RESEARCH ARTICLE / ARAŞTIRMA MAKALESİ
Journal of Advanced Research in Health Sciences /
Sağlık Bilimlerinde İleri Araştırmalar Dergisi 2023
DOI: 10.26650/JARHS2023-1349345

THE FREQUENCY AND FITNESS OF m6A-ASSOCIATED VARIANTS COULD BE MODULATED BY THE THERMODYNAMIC STABILITY OF AN OVERLAPPING G-QUADRUPLEX

m6A İLE İLİŞKİLİ VARYANTLARIN SIKLIĞI, ÖRTÜŞTÜĞÜ G-KUADRUPLEKS YAPISININ TERMODİNAMİK KARARLILIĞI İLE DEĞİŞEBİLİR

Çağrı GÜLEÇ[1]

[1] Istanbul University, Istanbul Faculty of Medicine, Department of Medical Genetics, Istanbul, Turkiye

ORCID ID: C.G. 0000-0002-1256-9574

Citation/Atıf: Gulec C. The frequency and fitness of m6A-associated variants could be modulated by the thermodynamic stability of an overlapping G-quadruplex. Journal of Advanced Research in Health Sciences 2023;6(3):219-228. https://doi.org/10.26650/JARHS2023-1349345

ABSTRACT

Objective: Post-transcriptional modifications like m6A (N6-methyladenosine) and secondary structures like G-quadruplex (G4) are formations that play a vital role in RNA processing. Their synergy also has functional consequences. Since m6A is known to be enzymatically created in the DRACH-motif, and that genetic variants can create a novel DRACH-motif or abolish a pre-existing DRACH-motif, we can hypothesize that variants which affect the gene product level through modulating m6A-G4 colocalization, may also consequently affect fitness and change the allele frequency. To test this hypothesis, the rare and common variants in selected human genes were investigated to determine their effect on DRACH-G4 colocalization.

Material and Methods: Genomic sequences and variant information were retrieved from the GRCh37/hg19 and Biomart-Ensembl databases. Experimentally determined G4 sequences were obtained from two different studies.

Results: Common variants leading to the formation of a novel DRACH-motif were found to be significantly higher inside the G4 structure than outside. In contrast, rare variants with the same feature were higher outside the G4-structure and had uneven distribution alongside the pre-mRNA. The uneven distribution of the DRACH-creating rare variants was observed to correlate with their effect on thermodynamic stability of the overlapping G4.

Conclusion: Selective DRACH-G4 colocalization suggests that m6A is evolutionally favorable when overlapping with G4. The thermodynamic stability could lead to uneven distribution of DRACH-G4 colocalization, favorable in 3-prime-side, but not in 5-prime-side. We can conclude that the fitness, and consequently frequency of a DRACH-creating variant is prone to become higher or lower depending in its position and effect on the overlapping-G4 stability.

Keywords: RNA modification, N6-methyladenosine, G-quadruplex, allele frequency

ÖZ

Amaç: M6A (N6-metiladenozin) gibi post-transtranskripsiyonel modifikasyonlar ve G-kuadrupleks (G4) gibi ikincil yapılar, RNA işlenmesinde önemli rol oynayan oluşumlardır. Bu iki oluşumun birlikteliğinin de işlevsel sonuçları vardır. M6A oluşumunun DRACH motifi üzerinde enzimatik olarak meydana geldiği, genetik varyantların yeni DRACH motifi oluşturabildiği veya var olan bir DRACH motifini ortadan kaldırabildiği dikkate alındığında, bu tür varyantların, mRNA üzerinde m6A-G4 örtüşme durumunu değiştirerek gen ürün düzeyini etkileyebileceğini, bunun da nesiller boyunca ilgili varyantın alel sıklığını değiştireceğini varsayabiliriz. Bu hipotezi test etmek için seçilmiş hastalık ilişkili genlerdeki nadir ve sık varyantlar DRACH-G4 örtüşmesi yönünden incelendi.

Gereç ve Yöntemler: Genomik diziler ve varyant bilgileri sırasıyla GRCh37/hg19 ve Biomart-Ensembl veritabanlarından çekildi. Deneysel olarak saptanmış G4 dizileri iki farklı çalışmadan elde edildi.

Bulgular: Yeni bir DRACH motifi oluşumuna yol açan yaygın varyantlar, G4 yapısı içinde yüksek bulundu. Aynı özelliğe sahip nadir varyantlar ise G4 yapısı dışında yüksek bulunurken, pre-mRNA üzerinde eşit dağılım göstermedikleri belirlendi. Yeni bir DRACH motifi oluşumuna yol açan nadir varyantların eşit olmayan dağılımı, örtüştüğü G4 yapısının termodinamik kararlılığı üzerindeki etkisi ile ilişkili bulundu.

Sonuç: Beklenenden sık gözlenen DRACH-G4 örtüşmesi, m6A modifikasyonunun G4 ile örtüştüğü durumların evrimsel bir avantaj sağlıyor olabileceğini düşündürmektedir. Nadir varyantlara bağlı ortaya çıkan DRACH-G4 örtüşmelerinin pre-mRNA'da eşit dağılım göstermemesi ise, m6A'nın G4 termodinamik kararlılığını değiştirmesi ve bu değişikliğin pre-mRNA'nın 5' kısmına göre 3' kısmında daha fazla tolere ediliyor olmasına bağlı görünmektedir. Sonuç olarak, DRACH motifi oluşturan varyantların seçilim baskısı ve bunun sonucunda biçimlenen alel sıklığı, bu varyantın pre-mRNA üzerindeki konumuna ve örtüştüğü G4 oluşumunun kararlılığı üzerindeki etkisine göre değişiklik göstermektedir.

Anahtar kelimeler: RNA modifikasyonu, N6-metiladenozin, G-kuadrupleks, alel sıklığı

Corresponding Author/Sorumlu Yazar: Çağrı GÜLEÇ **E-mail:** cagri@istanbul.edu.tr

Submitted/Başvuru: 25.08.2023 • **Revision Requested/Revizyon Talebi:** 25.08.2023 • **Last Revision Received/Son Revizyon:** 01.09.2023 • **Accepted/Kabul:** 01.09.2023 • **Published Online/Online Yayın:** 13.10.2023

INTRODUCTION

The genetic variants in protein-coding genes display their impact on the phenotype mostly through quantitative and qualitative fluctuation in protein products. In this respect, post-transcriptional RNA modifications play a significant role, especially in a quantitative manner. To date, more than 150 chemical modifications have been identified in RNAs (1, 2). Though most of these modifications are found in non-coding RNAs (rRNA, tRNA, snRNA and long-non-coding RNAs), it has been shown that coding RNAs are also subject to modifications (3). N6-methyladenosine (m6A) is one of the most abundant chemical modifications found in mRNAs (4, 5). The function of m6A was elucidated long after its discovery (6, 7). We now know that m6A plays a significant role in processing, nuclear export, translational regulation, and decay of mRNAs (8, 9). In line with these functions, m6A modification displays an unequal distribution throughout mRNA (4, 5, 10).

In addition to post-transcriptional modifications, secondary structures have a key role in translational efficiency of mRNAs. One of these secondary structures is the G-quadruplex (G4) which is spontaneously formed in guanine-rich (G-rich) sequences in DNA or RNA. This structure is formed via Hoogsten base pairing of adjacent guanines in G-rich sequences (11). Although Hoogsten hydrogen binds between G bases are essential for G4 structure, there are many additional factors that could affect G4 folding in a G-rich sequence, like the length of the sequence, the presence of an alternative Watson-Crick pair-based stable structure and free metal cations (12-15). Bioinformatic studies which are based on the acknowledged consensus sequence and experimental genome-wide studies using G4 specific probes revealed a non-random distribution of G4 structure throughout the genome and transcriptome (16). These studies demonstrated that the G4 sequences are enriched at the telomeres, promoter regions and replication origins in genomic DNA, and UTRs (Untranslated Region) in mRNA. Though G4 structures are formed both in DNA and RNA, it has been shown that G4 structures in RNA are more stable and have less topological diversity than G4 structures in DNA (17). Searching the canonical consensus sequence (5'-$G_3N_{1-7}G_3N_{1-7}G_3N_{1-7}G_3$-3' where N is A, C, G, T, or U) from which a typical G4 structure is known to form, putative G4 structures were found to be present in 5'UTR of more than 9.000, and in 3'UTR of more than 8.000 human genes (15, 18, 19, 20-22). Moreover, it showed that more than 1.600 human genes have G4 structures in their ORFs (Open Reading Frame) (23). G4 structures in RNAs are supposed to play a role in stability, splicing, and translation of RNAs through binding specific proteins like eIF4G, LARK, SLIRP, AFF3, AFF4, eIF4A and hnRNP A2 (24-30).

Despite the enormous difference in their formation and structure, m6A modification and the G4 structure seem to share common functional consequences in RNA processing. Both are separately shown to modulate splicing, nuclear export, translation, and decay of the RNAs. Although mutual or synergic activities of the m6A and G4 are not well understood, recent studies revealed that m6A modification can modulate G4 structure formation and vice versa. The m6A modification was found to modulate the G4 structure through affecting its stability as shown in R-loops, while the G4 structure was demonstrated to modulate m6A modification through facilitating the adenine N6-methylation in target motif as shown in viral RNAs (31, 32). Their synergic activity is supported by overlapping m6A-G4 in eukaryotic mRNAs (33, 34).

MATERIAL and METHODS

Aim: Considering previous studies, this study aimed to investigate whether the functional consequences of overlapping m6A-G4 have selective pressure on the variants that lead to colocalization of m6A and G4. To address this question, defined single nucleotide variants in selected disease-associated human genes were investigated to determine their effect on the colocalization of m6A and G4.

Reference sequence data: Human genome sequence (GRCh37/hg19) was downloaded from NCBI (https://www.ncbi.nlm.nih.gov/assembly/GCF_000001405.25) and recorded as a sql database using Sqlite3 library of Python3.7. The name list and exonic coordinates of disease associated human genes (28,250 transcripts from 3,306 MIM genes) and genomic coordinates, alleles, and allele frequencies of the variants within these regions were fetched from Ensembl (Biomart; http://grch37.ensembl.org/biomart/martview/c4f6ef5ffc88cf1d9a00b63173228cda) and UCSC Genome Browser (https://genome.ucsc.edu/cgi-bin/hgTables).

Real G4 data: Genomic coordinates of experimentally identified G4 regions were yielded from two GEO datasets (GSE77282 and GSE181373) (37, 38). The neighboring gene name, SNP (Single Nucleotide Polymorphism) record name, variant alleles, and allele frequency of the variants within these coordinates were fetched from the UCSC Genome Browser and recorded as text files.

Artificial variant sequences: Considering their genomic position, strand, alleles and allele frequency, each single nucleotide variant and their flanking reference sequence was retrieved from a hg19-sql database, and separately stored as 'wild-type sequence pieces'. By replacing the reference base with a variant base, a 'variant sequence piece' was produced for each corresponding 'wild-type sequence piece.' Each 'variant sequence piece was tagged as 'common' or 'rare' regarding the minor allele frequency (MAF) ('common' for alleles with MAF ≥0.01, and 'rare' for alleles with MAF <0.01) of the corresponding variant. Variants with no recorded MAF value in the database were evaluated as rare variants.

Searching and counting the putative G4 structures and m6A motifs: Both wild-type and matched variant sequence pieces were evaluated with 're' module and 'Pandas' library of the Python3.7, for the presence and count of a putative G4-forming sequence ($G_{1-3}[N_{1-7}G_{1-3}]_3$), for putative G4 structures and consensus DRACH motif ([A/G/U][A/G]AC[A/C/U]), and for putative

m6A modification sites. The effect of the variant on the m6A (DRACH) motif number was evaluated according to putative DRACH motif Number (n) in the presence of a reference allele. The resulting effect was assessed as inert (change from 'n' to 'n'), augmentative (from 'n' to 'n+1') or reductive (from 'n' to 'n-1').

Thermodynamic stability: Minimum free energy (MFE) value of each sequence piece was calculated with 'seqfold' library of the Python3.7.

Statistical analysis: The Chi-square, t-test, Z-test, and correlation test were performed with the 'statsmodels' library of the Python3.7.

RESULTS

The number of common variants creating the DRACH (m6A) motif inside the G4 structure are higher than the rare variants. Since, both m6A and G4 found in pre-mRNA are known to affect splicing, localization, translational efficiency, and decay of the RNAs, variants in a non-spliced sequence of the selected MIM genes were evaluated in terms of their effect on a m6A motif Number within putative G4 structures. While most of the variants (94,6%, 50,933 variants) were found not to relate to the m6A motif (n→n;inert variants), 75% (2,009 variants) of the remaining variants created a novel m6A motif (n→n+1;augmentative variants), and 25% (851 variants) abolished a preexisting m6A motif (n→n-1;reductive) inside a putative G4-structure (Supplementary File 1). The research found that the number of rare variants which create a novel m6A motif or increase the number of preexisting m6A motifs inside a G4 structure were statistically lower (p=3.02e-6) than that of the common variants (Table 1).

To clarify whether this situation is limited to G4-included regions, the variants which have a flanking region that is not included in any putative G4 (G4-not-included regions) were analyzed. Contrary to G4-included regions, the number of rare augmentative variants, which created a novel m6A motif or increased the number of preexisting m6A motifs inside G4-not-included regions, were found statistically higher (p=6.98e-30) than that of common variants (Table 1).

Since both the m6A modification and G4 structures participated in the translational rate, localization and stability of mRNA, variants were evaluated in the spliced RNA sequence, as well. While the difference between rare and common variants in G4-not-included regions remained statistically significant (p<0.001), there was no statistical significance in the distribution of variant counts inside G4 (Table 2).

Rare variants creating the DRACH (m6A) motif inside G4 avoid being close to 3'-side and prefer to be near 5'-side of pre-mRNAs.
Independent of their frequency, both inert and reductive variants displayed an equal distribution throughout pre-mRNAs. The distribution of augmentative rare variants displayed a statistically significant shift from 3'-end to 5'-end of the pre-mRNAs (Figure 1).

Table 1: Count of variants which decrease (n ->n-1), increase (n ->n+1) or do not change (n ->n) the m6A motif number (n) inside or outside the G4 structure in pre-mRNA. Rare variants (MAF<0.01) increasing the m6A motif number were significantly higher outside the G4 structure, while common variants (MAF ≥0.01) increasing the m6A motif number were significantly higher inside the G4 structure

m^{6}A motif location	m^{6}A motif number		MAF value	Variant count		Chi-square
	Reference allele	Variant allele		Observed	Expected	
Inside G4-motif	n	n-1	<0.01	509	529.913	0.82
			≥0.01	763	742.087	0.58
	n	n	<0.01	31,814	31,661.47	0.73
			≥0.01	44,186	44,338.53	0.52
	n	n+1	<0.01	1,174	1,305.619	13.26*
			≥0.01	1,960	1,828.381	9.47*
Outside G4-motif	n	n-1	<0.01	994	919.87	5.97
			≥0.01	7,913	7,987.12	0.68
	n	n	<0.01	23,410	23,836.62	7.63
			≥0.01	207,396	206,969.37	0.87
	n	n+1	<0.01	1,516	1,163.5	106.79**
			≥0.01	9,750	10,102.49	12.29**

* p=3.02e-6, Total chi-square:25.41, df=2, ** p=6.98e-30, Total chi-square: 134.26, df=2

Table 2: Count of variants which decrease (n ->n-1), increase (n ->n+1) or does not change (n ->n) the m6A motif number (n) inside or outside the G4 structure in mRNA. Rare variants (MAF<0.01) increasing the m6A motif number were significantly higher outside the G4 structure, while common variants (MAF≥0.01) were lower

m^{6}A motif location	m^{6}A motif number		MAF value	Variant count		Chi-square
	Reference allele	Variant allele		Observed	Expected	
Inside G4-motif	n	n-1	<0.01	0	1.09	0.128
			≥0.01	13	11.9	1.395
	n	n	<0.01	71	68.2	7.995
			≥0.01	737	739.79	86.728
	n	n+1	<0.01	1	2.73	0.320
			≥0.01	31	29.29	3.434
Outside G4-motif	n	n-1	<0.01	695	655.42	2.39
			≥0.01	5,443	5,482.57	0.28
	n	n	<0.01	12,698	12,857.69	1.98
			≥0.01	107,714	107,554.3	0.23
	n	n+1	<0.01	756	635.88	22.69*
			≥0.01	5,199	5,319.11	2.71*

* p<0.0001, Total chi-square: 30.299, df=2

Because the total length and exon-intron contents of each transcript are different, relative position rather than absolute position of the variants is more appropriate for the comparison of their position-dependent effect in pre-mRNAs. To investigate the position-dependent effect of the variants on m6A-G4 colocalization in more detail, their relative positions were clustered as positional quartiles. Distribution of the rare variant ratio through positional quartiles revealed that the ratio of inert variants displayed an equal distribution throughout pre-mRNAs. Except for the second quartile, the ratio of reductive variants displayed an equal distribution. In the second quartile, the ratio of reductive rare variants was found to be significantly lower than expected. The ratio of augmentative variants displayed unequal distribution in all quartiles. In the first quartile, the ratio of augmentative rare variants was higher than expected, howerver in the last quartile it was lower than expected (Figure 2).

To evaluate whether there is a difference between the distribution of variant positions throughout the G4 structure, the relative position of variants in the G4 sequence were compared. There was no difference found between the over-G4 distribution of the variants (Data not shown.).

Colocalization of m6A and G4 seems to increase the thermodynamic stability of the G4 structure.

Because of the crucial involvement of thermodynamic stability in G4-structure formation, minimum free energy (MFE) values of the putative G4-structures were compared considering the variants and m6A motif numbers.

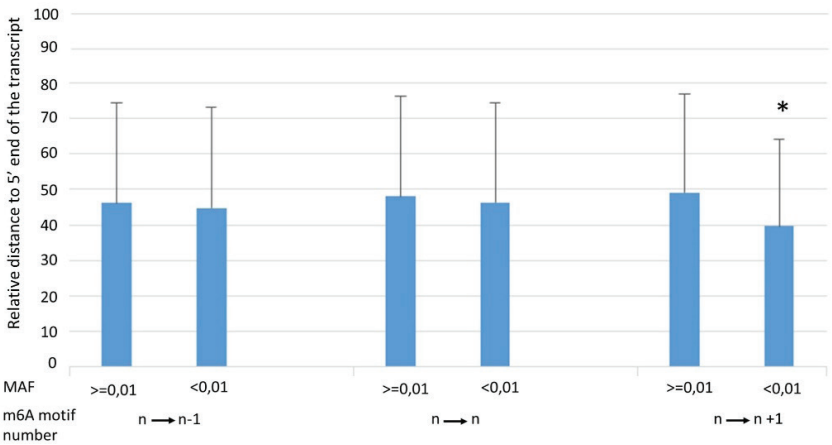

Figure 1: Comparison of mean distribution of common (MAF≥0.01) and rare (MAF<0.01) variants regarding their effect on m6A-motif number inside the putative G4 motifs. Distribution of the rare variants that increase m6A number have closer location to 5'-side. *p<0.001

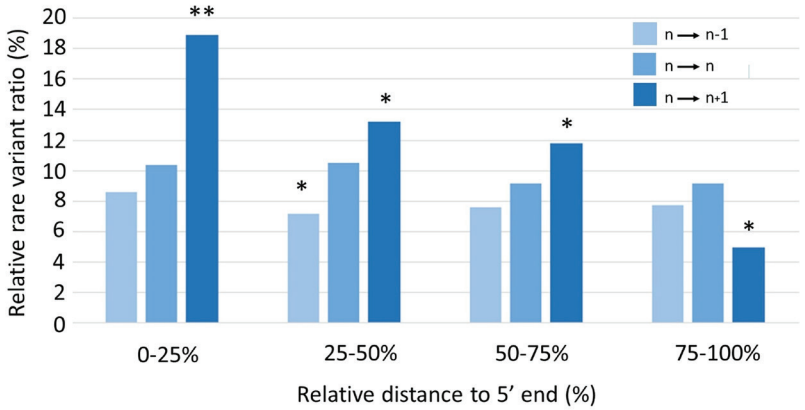

Figure 2: Distribution of rare variant/common variant ratio throughout relative position quartiles. The rare variants that increase m6A number display a gradually decreasing pattern from 5'- to 3'- side. *p<0.001, **p<0.0001

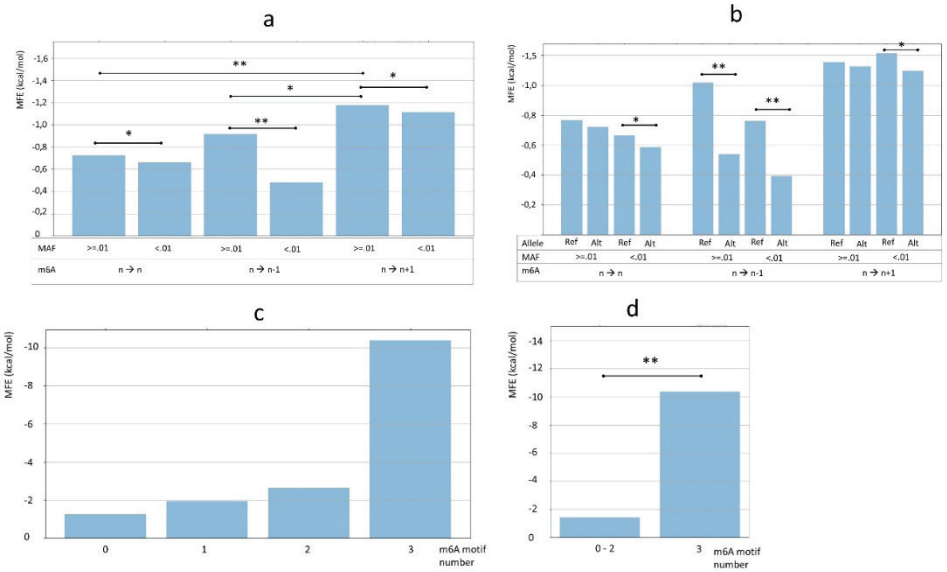

Figure 3: Distribution of G4-MFE values throughout relative position quartiles (a, b) and the effect of m6A number on G4-MFE value (c, d). The effect of m6A on MFE value depends on frequency (a), alleles (b) and m6A number (c), especially three m6As (d). *p<0.001, **p<0.0001

Rare variants were shown to have lower MFE values, especially in reductive variants (Figure 3a). When reference and variant bases were considered, it was observed that the alternative variant had a lower MFE than the reference allele in each group (Figure 3b). While reductive rare variants were found to have the lowest MFE, augmentative common variants were found to have the highest MFE value (Figure 3a). To evaluate whether this result was due to the presence of the m6A motif, the MFE values of G4 sequences were compared to the m6A motif number. The MFE values were observed to be inversely related to the m6A motif count (Figure 3c). Putative G4 structures with three m6A motifs were found to have the lowest MFE values (Figure 3d).

Overlapping m6A-G4 could modulate the G4 stability in a position-dependent manner.

Due to position-dependent colocalization of m6A-G4 and m6A-dependent stability of G4, we can hypothesize that G4-stability could also be modulated in a position-dependent manner. Additionally, because of the relationship between overlapping m6A-G4 and allele frequency, it is possible that G4-stability could depend on allele frequency of the variants. To test this theory, the MFE values of putative G4 structures were compared while considering the relative position and minor allele frequency of the variants that created the m6A motif.

Distribution of the MFE values over positional quartiles showed that MFE values were likely to decrease from the 5'-end to the

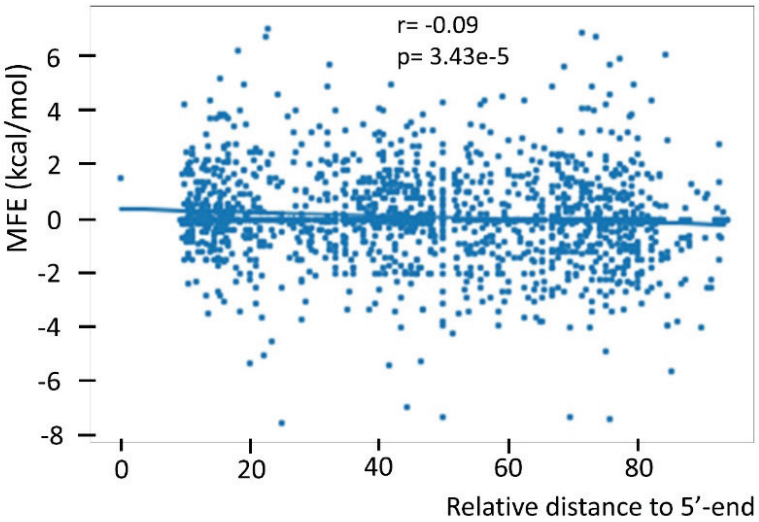

Figure 4: Correlation between G4-MFE values led by m6A-creating variants and their relative position. Decreasing MFE value suggests an increased stability of G4 from 5'- to 3'-end.

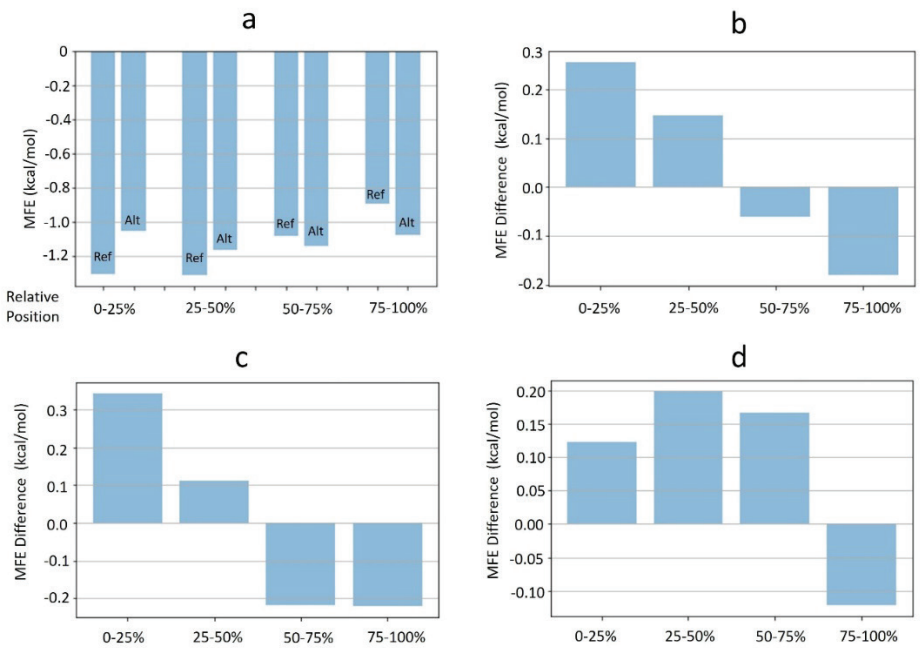

Figure 5: Position-dependent decrease in MFE values shows difference between reference (Ref) and alternative (Alt) allele of the variants (a). The position dependency of MFE values is clearer when MFE differences between Ref and Alt alleles are considered (b). MFE differences of common variants (c) and rare variants (d) have different distribution patterns.

3'-end of pre-mRNAs (Figure 4). To see whether this position-dependent decrease depended on variant alleles, the MFE values were evaluated in terms of alleles. The effect of variant alleles on the MFE values was found to depend on their relative position (Figure 5a). The difference between MFE values of the variant allele and reference allele showed a dependence on relative position (Figure 5b). The MFE-difference (Variant allele MFE minus reference allele MFE) value had a decreasing pattern throughout pre-mRNA. When the distribution of the MFE-difference was reevaluated considering the frequency of variant alleles, it revealed that the decrease in the MFE-difference started closer to the 5'-end with common alleles (Figure 5c), and closer to the 3'-end with rare variants (Figure 5d).

Real Data Analysis

To evaluate the validity of the findings, experimental

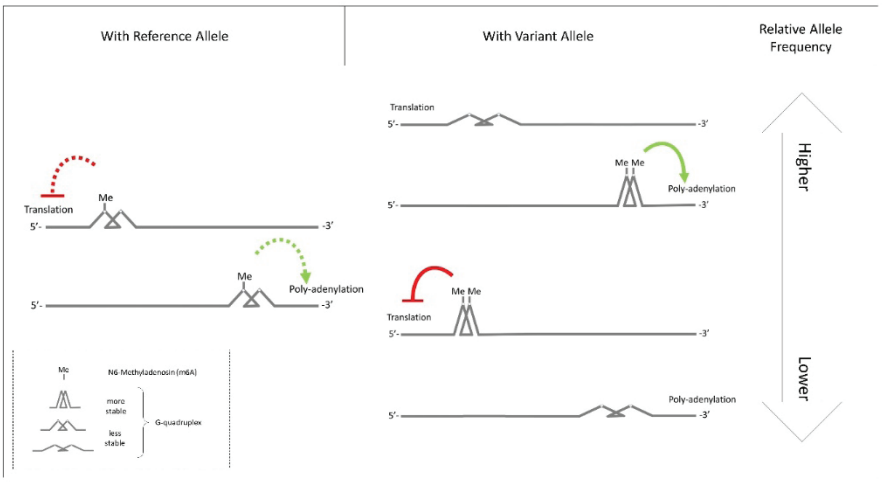

Figure 6: A hypothetical model to explain the position-dependent effect of m6A-G4 colocalization. A variant that creates m6A motif overlapping with G4 is prone to have higher allele frequency in 3'-side, because G4 supports poly-adenylation. In 5'-side, however, it is prone to have lower allele frequency to allow translation.

G-quadruplex data were analyzed. The first set of data included 13,423 genomic coordinates of G-quadruplex regions yielded from the HeLa human transcriptome study dataset, GSE77282 (35). The second set was retrieved from the GSE181373 dataset that obtained single-cell mapping of DNA G-quadruplex structures in three human cancer cell lines (K562, MCF7 and U2OS), and included 223,696 genomic regions (36). All the SNPs and their flanking sequences located inside these genomic coordinates were gotten from the USCS Browser (https://genome.ucsc.edu/).

The first dataset included 13,360 unique genomic regions in 6,613 genes. A total of 15,169 known variants were found within these regions. When the regions with extreme length (median length: 11,657 bp, min length: 10 bp, maximum length: 110,841 bp) were filtered out, 823 regions covering at least one variant were yielded (Supplementary File-2). Of these regions, only 110 (13.36%) were found to have classical G4-quadruplex motifs ($G_{1-3}[N_{1-7}G_{1-3}]_3$). The G4-quadruplex motif-included sequences that were evaluated in terms of DRACH (m6A) motif number with variant alleles. Approximately 10% of them were found to have a changed DRACH motif number; From 0 to 1:21 variants, from 1 to 2:13 variants, from 2 to 3:2 variants, from 1 to 0: 15 variants and from 2 to 1:4 variants.

The second dataset included 223,696 genomic regions of which 198,959 were unique. Of these experimentally detected quadruplex regions, which had a 2,939.46 bp length in mean (median 1,766 bp), 1,5487 (43.5%) were found to have at least one classical G4-quadruplex motif. Though 34,482 variants (35,603 alleles) were identified to be covered by these regions, only 406 (1.14%) were located inside a classical G4-quadruplex motif. While 81 variants were found to remove a pre-existing classical G4-quadruplex motif, 117 were observed to create a novel classical G4-quadruplex motif. Of the remaining 299 variants, which did not affect the G4-motif, 288 did not change the m6A-motif number within the G4-quadruplex motif. Only six variants were found to increase the m6A-motif number within the G4-quadruplex motif, while five variants decreased the m6A-motif number.

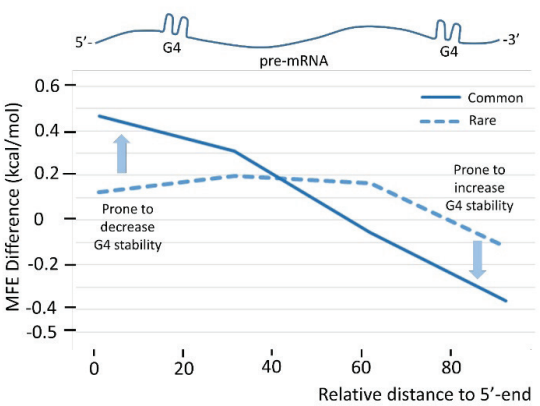

Figure 7: The position-dependent effect of m6A-G4 colocalization can be explained by MFE difference. The variants that create m6A motif overlapping with G4 may have better fitness due to decreased G4 stability near 5'-side and increased G4 stability near 3'-side.

When the allele frequency was considered, no significant difference was found in the distribution of the m6A motif counts (Supplementary Table-1). On the other hand, the statistically significant relationships observed in the data yielded from the presumed G4-motifs that we were not able to evaluate in the real data, because the count of rare variants were not sufficient (Supplementary File-2).

DISCUSSION

Post-transcriptional modifications like m6A, and secondary structures like G-quadruplex, are principal actors in RNA processing. While the m6A modification is controlled by specific enzymes, the formation of G4 structures relies on physico-chemical conditions and thermodynamic rules. Both lead to similar consequences in terms of their effects on RNA processing, but with different manners. The dependence on a consensus sequence is another common feature of the m6A and G4. While the m6A modification targets the adenine in the third position of the DRACH ([A/G/U][A/G]AC[A/C/U]) motif, the G4 structure is formed from the sequence with $G_{1-3}[N_{1-7}G_{1-3}]_3$ motif (15, 18, 19, 37).

Besides the separate roles of the m6A modification and G4 structure, their colocalization has been shown to have functional importance. The crosstalk between the m6A modification and G4 structures seems bidirectional. For instance, the m6A modification is shown to affect the stability of the R-loop, and a G4 structure formed by DNA: RNA hybrid strands (31, 32). Mutually, the G4 structures are shown to modulate m6A modification in some viral genomes like HIV, Zika, Hepatitis B, and SV40 (34). If we suppose that a variant within the G4-forming sequence can create or abolish the m6A motif (DRACH), this variant could be under selective pressure depending on the functional effect of the resulting overlapping m6a-G4 status. In this respect, we can expect a correlation between the frequency and colocalization ability of the variants. In this study, the impact of the m6A-G4 colocalization on the variant frequency was investigated. For this purpose, the single nucleotide variants in selected disease-associated human genes were evaluated for their allele frequency and m6A-G4 colocalization ability.

The outstanding result of this study is that the variants creating the m6A motif inside a G4 structure are prone to have a higher frequency. In contrast, such variants have lower frequency if located outside of the G4 structure. The Colocalization-dependent higher allele frequency of such variants suggest that overlapping m6A-G4 could have a protective role.

To understand how a variant creating m6A motif inside a G4 structure may play a protective role and gain a favorable feature, distribution of the colocalization-leading variants throughout pre-mRNAs was evaluated, since both the m6A modification and G4 structure display their functional effects in a position-dependent manner (38).

The results showed that the position of the colocalization-leading variants has importance for the underlying mechanism. Previous studies, showed that the m6A motif distribution throughout RNA molecules was not equal. Many studies observed that the m6A residues were enriched in 5'UTRs, around stop codons and in 3' UTRs adjacent to stop codons in mammalian mRNAs (4, 5, 39). Similarly, the G4 structures were also reported to be overrepresented in 5'- and 3'-UTRs (20, 40).

Findings of this study suggest that the functional consequence of the m6A-G4 colocalization may have selective pressure on the colocalization-leading variants in a position-dependent manner. While the G4 structures hosting a m6A motif created by a common variant showed equal distribution throughout the pre-mRNA, G4 structures overlapping with m6A motifs created by a rare variant are likely to avoid the 3'-side of the pre-mRNAs. This result could mean that the m6A-G4 colocalization can be tolerated when found near the 3'-side, but not in the central region and especially near the 5'-side of the pre-mRNA. Then, it can be deduced that any variant that creates a novel m6A motif inside a G4 structure could become favorable, and display an increased population frequency, if found near the 3'-side. On the other hand, the same type of variant would not bepreferable if found in the transcript body, and undesirable if found near the 5'-side of the pre-mRNA. The latter is expected to display a decreasing allele frequency. The favorability of a m6A-G4 colocalization in the 3'-side is due to the supported function of the pre-existing G4 structure. For instance, as represented in Fig 6, the G4 structure supported by the m6A in 5'UTR may reduce the translational efficiency of mRNA, since G4 structures in 5'UTR are known to affect cap-dependent and cap-independent translation (41). Similarly, the G4 structure supported by an overlapping m6A in the 3'-side may enable the transcript to produce alternative products, through modulating alternative splicing and alternative poly-adenilation of pre-mRNA (21, 41). It is also possible that the m6A can modulate miRNA binding through stabilizing the G4 structure after splicing (40).

There could be consequences for the m6A-G4 colocalization. The creation of novel docking sites for specific proteins must be considered in this respect. Both m6A and G4 structures are known to be recognized by specific proteins (24, 43, 44). Therefore, colocalization of the m6A modification and G4 structure may lead to competitive or cooperative interaction between these proteins. Evaluating these possibilities requires experimental methods dealing *in vitro* RNA-protein interactions. Another consequence of the m6A-G4 colocalization is the changing thermodynamic properties of the structure. Thermodynamic stability is crucial for the formation of G4 structures (45). In contrast, m6A formation is regulated by specific enzymes rather than the thermodynamic status of the flanking sequence (46). However, m6A modification can affect the thermodynamic stability of RNA, and marginally reduce the stability of an A: U base pairing (47). Similarly, in recent studies, the colocalization of m6A and G4 was shown to alter the stability of the DNA: RNA hybrid quadruplexes, known as R-loop (31, 32). Some studies reported that the m6A promoted G4 folding, while others demonstrated that the m6A downregulated the G4 formation. These contradictory findings are still discussed. The G4 was also disputed to modulate the m6A modification in viral RNAs (32). The overlapping m6A and G4 in 3'UTR of viral RNAs revealed that the folded G4 structures may guide the enzymatic adenine methylation in the DRACH motifs (34). Moreover, enrichment of the overlapping m6A and G4 in viral RNAs are reported to be critical for the impact of the m6A on viral fitness; as shown in HIV-1 (48). In eukaryotes, the overlapping m6A and G4 are shown in the 3'UTRs of mRNAs, as well (33, 49).

Based on previous studies that suggested the synergy between m6A and G4, we can deduce that significance in the distribution of variants leading to the m6A-G4 colocalization may result from the changed stability of the G4 structures. To assess this possibility, the MFE values of G4 structures were calculated. DRACH (m6A) motifs were observed to cause a decreased MFE, which means increased thermodynamic stability of the G4 structure. The number of m6A motifs inside the G4 structure also seems crucial for the degree of stability.

The position-dependency of overlapping m6A-G4 may explain the mechanism responsible for difference in allele frequency among the variants leading to m6A-G4 colocalization. Based on the findings of this study, we can suppose that a m6A-G4 colocalization-leading variant will prone to have higher allele frequency, if decreases G4 stability when located near 5'-side and increases it when located near 5'-side of the pre-mRNA (Figure 7).

The preliminary results of this study suggest that m6A overlapping with G4 structure may have functional consequences with an unknown mechanism. At least, it seems likely that this mechanism needs position-dependent stability of the G4 structure.

Since all findings of the study need to be validated with real data, sequences yielded from genomic coordinates of experimentally identified quadruplex regions in two GEO datasets were also investigated for the covered variants and putative m6A motif numbers. However, because the counts of the experimentally detected quadruplex regions which had classical motif ($G_{1-3}[N_{1-7}G_{1-3}]_3$) were not sufficient for further statistical analysis, findings of the study could not be supported or disproved by the real data. Choosing the classical G4-quadruplex motif was a limitation of recent study, so the theoretical findings of this study were restricted to the classical G4-quadruplex motif, which seems not to be common in the real quadruplex pools. Therefore, different subtypes of the G4-quadruplex motifs, together with the $G_{1-3}[N_{1-7}G_{1-3}]_3$ motifs, should be considered in further studies to manage the synergy of the overlapping G4-m6A motifs and allele frequency of the m6A-related variants.

In summary, we conclude that the fitness, and consequently the frequency of a variant creating the m6A motif is prone to become higher or lower depending on whether it is located inside or outside the classical G4 structure. Furthermore, the frequency of these variants may depend on both their position and their effect on the thermodynamic stability of the overlapping G4 structure. If located near the 5'-side it destabilizes the G4 structure or if located near the 3'-side it stabilizes the G4, a variant creating m6A motif is prone to have higher fitness and frequency.

Study limitations: The recent study included only the selected MIM genes, and the known SNPs reported in these genes. Findings revealed from the available experimental data were insufficient to evaluate theoretical findings of the study.

CONCLUSION

The starting point of this study was a suspicion as to whether there is a functional crosstalk between chemical modifications (like m6A) and secondary structures (like G4-quadruplex) in RNA. In case of such a crosstalk, the genetic variations at the chemically modified base are expected to have functional consequences, which create selective pressure on the variant allele. Though the lack of experimental validation and supports, the findings of this study suggest that the fluctuation in allele frequency of the human SNP's could be a consequence of the crosstalk between the m6A and G4-quadruplex. In addition to the potential contribution to population genetics and evolutionary genetics, such a crosstalk between the m6A and G4-quadruplex also has the potential to help us understand the polygenic nature of the complex disorders and the modifying genetic factors in the single-gene disorders.

Ethics Committee Approval: The author declared that an ethics committee approval is not required for this study.

Peer Review: Externally peer-reviewed.

Conflict of Interest: The authors have no conflict of interest to declare.

Financial Disclosure: The authors declared that this study has received no financial support.

REFERENCES

1- Motorin Y, Helm M. RNA nucleotide methylation. Wiley Interdiscip Rev RNA 2011;2(5):611-31.
2- Machnicka MA, Milanowska K, Oglou OO, Purta E, Kurkowska M, Olchowik A, et al. MODOMICS: a database of RNA modification pathways--2013 update. Nucleic Acids Res 2013;41(Database issue):D262-7.
3- Desrosiers R, Friderici K, Rottman F. Identification of methylated nucleosides in messenger RNA from Novikoff hepatoma cells. Proc Natl Acad Sci U S A 1974;71(10):3971-5.
4- Dominissini D, Moshitch-Moshkovitz S, Schwartz S, Salmon-Divon M, Ungar L, Osenberg S, et al. Topology of the human and mouse m6A RNA methylomes revealed by m6A-seq. Nature 2012;485(7397):201-6.
5- Meyer KD, Saletore Y, Zumbo P, Elemento O, Mason CE, Jaffrey SR. Comprehensive analysis of mRNA methylation reveals enrichment in 30 UTRs and near stop codons. Cell 2012;149(7); 1635-46.
6- Roundtree IA, Evans ME, Pan T, He C. Dynamic RNA modifications in gene expression regulation. Cell 2017;169(7):1187-200.
7- Zhao BS, Roundtree IA, He C. Post-transcriptional gene regulation by mRNA modifications. Nat Rev Mol Cell Biol 2017;18(1):31-42.
8- Shi H, Wang X, Lu Z, Zhao BS, Ma H, Hsu PJ, et al. YTHDF3 facilitates translation and decay of N6-methyladenosine-modified RNA. Cell Res 2017;27(3):315-28.
9- Roignant JY, Soller M. m6A in mRNA: an ancient mechanism for fine-tuning gene expression. Trends Genet 2017;33(6):380-90.
10- Gilbert WV, Bell TA, Schaening C. Messenger RNA modifications: form, distribution, and function. Science 2016;352(6292):1408-12.
11- Bochman ML, Paeschke K, Zakian VA. DNA secondary structures: stability and function of G-quadruplex structures. Nat Rev Genet 2012;13(11):770-80.

12- Rachwal PA, Brown T, Fox KR. Effect of G-tract length on the topology and stability of intramolecular DNA quadruplexes. Biochemistry 2007;46(11):3036-44.
13- Rachwal PA, Fox KR. Quadruplex melting. Methods 2007;43(4):291-301.
14- Rachwal PA, Brown T, Fox KR. Sequence effects of single base loops in intramolecular quadruplex DNA. FEBS Lett 2007;581(8):1657-60.
15- Mukundan VT, Phan AT. Bulges in G-quadruplexes: broadening the definition of G-quadruplex-forming sequences. J Am Chem Soc 2013;135(13):5017-28.
16- Lam EY, Beraldi D, Tannahill D, Balasubramanian S. G-quadruplex structures are stable and detectable in human genomic DNA. Nat Commun 2013;4:1796.
17- Bugaut A, Murat P, Balasubramanian S. An RNA hairpin to G-quadruplex conformational transition. J Am Chem Soc 2012;134(49):19953-6.
18- Huppert JL, Balasubramanian S. G-quadruplexes in promoters throughout the human genome. Nucleic Acids Res 2007;35(2):406-13.
19- Huppert JL. Hunting G-quadruplexes. Biochimie 2008;90(8):1140-8.
20- Kumari S, Bugaut A, Huppert JL, Balasubramanian S. An RNA G-quadruplex in the 5' UTR of the NRAS proto-oncogene modulates translation. Nat Chem Biol 2007;3(4):218-21.
21- Beaudoin JD, Perreault JP. 5'-UTR G-quadruplex structures acting as translational repressors. Nucleic Acids Res 2010;38(20):7022-36.
22- Huppert JL, Bugaut A, Kumari S, Balasubramanian S. G-quadruplexes: the beginning and end of UTRs. Nucleic Acids Res 2008;36(19):6260-8.
23- Thandapani P, Song J, Gandin V, Cai Y, Rouleau SG, Garant JM, et al. Aven recognition of RNA G-quadruplexes regulates translation of the mixed lineage leukemia protooncogenes. Elife 2015;4:e06234.
24- Brázda V, Hároníková L, Liao JC, Fojta M. DNA and RNA quadruplex-binding proteins. Int J Mol Sci 2014;15(10):17493-517.
25- Lyons SM, Kharel P, Akiyama Y, Ojha S, Dave D, Tsvetkov V, et al. eIF4G has intrinsic G-quadruplex binding activity that is required for tiRNA function. Nucleic Acids Res 2020;48(11):6223-33.
26- Niu K, Xiang L, Jin Y, Peng Y, Wu F, Tang W, et al. Identification of LARK as a novel and conserved G-quadruplex binding protein in invertebrates and vertebrates. Nucleic Acids Res 2019;47(14):7306-20.
27- Williams P, Li L, Dong X, Wang Y. Identification of SLIRP as a G Quadruplex-Binding Protein. J Am Chem Soc 2017;139(36):12426-9.
28- Serikawa T, Spanos C, von Hacht A, Budisa N, Rappsilber J, Kurreck J. Comprehensive identification of proteins binding to RNA G-quadruplex motifs in the 5' UTR of tumor-associated mRNAs. Biochimie 2018;144:169-84.
29- Lyonnais S, Tarrés-Solé A, Rubio-Cosials A, Cuppari A, Brito R, Jaumot J, et al. The human mitochondrial transcription factor A is a versatile G-quadruplex binding protein. Sci Rep 2017;7:43992.
30- Khateb S, Weisman-Shomer P, Hershco-Shani I, Ludwig AL, Fry M. The tetraplex (CGG)n destabilizing proteins hnRNP A2 and CBF-A enhance the in vivo translation of fragile X premutation mRNA. Nucleic Acids Res 2007;35(17):5775-88.
31- Yang X, Liu QL, Xu W, Zhang YC, Yang Y, Ju LF, et al. m6A promotes R-loop formation to facilitate transcription termination. Cell Res 2019;29(12):1035-8.
32- Abakir A, Giles TC, Cristini A, Foster JM, Dai N, Starczak M, et al. N6-methyladenosine regulates the stability of RNA:DNA hybrids in human cells. Nat Genet 2020;52(1):48-55.
33- Jara-Espejo M, Fleming AM, Burrows CJ. potential G-quadruplex forming sequences and N6-methyladenosine colocalize at human Pre-mRNA intron splice sites. ACS Chem Biol 2020;15(6):1292-130.
34- Fleming AM, Nguyen NLB, Burrows CJ. 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 2019;5(2):218-28.
35- Kwok CK, Marsico G, Sahakyan AB, Chambers VS, Balasubramanian S. rG4-seq reveals widespread formation of G-quadruplex structures in the human transcriptome. Nat Methods 2016;13(10):841-4.
36- Hui WWI, Simeone A, Zyner KG, Tannahill D, Balasubramanian S. Single-cell mapping of DNA G-quadruplex structures in human cancer cells. Sci Rep 2021;11(1):23641.
37- Csepany T, Lin A, Baldick CJ Jr, Beemon K. Sequence specificity of mRNA N6-adenosine methyltransferase. J Biol Chem 1990;265(33):20117-22.
38- Bushkin GG, Pincus D, Morgan JT, Richardson K, Lewis C, Chan SH, et al. m6A modification of a 3' UTR site reduces RME1 mRNA levels to promote meiosis. Nat Commun 2019;10(1):3414.
39- Batista PJ, Molinie B, Wang J, Qu K, Zhang J, Li L, et al. m(6)A RNA modification controls cell fate transition in mammalian embryonic stem cells. Cell Stem Cell 2014;15(6):707-19.
40- Rouleau S, Glouzon JS, Brumwell A, Bisaillon M, Perreault JP. 3' UTR G-quadruplexes regulate miRNA binding. RNA 2017;23(8):1172-9.
41- Jodoin R, Carrier JC, Rivard N, Bisaillon M, Perreault JP. 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 2019;47(19):10247-66.
42- Song J, Perreault JP, Topisirovic I, Richard S. RNA G-quadruplexes and their potential regulatory roles in translation. Translation (Austin) 2016;4(2):e1244031.
43- von Hacht A, Seifert O, Menger M, Schütze T, Arora A, Konthur Z, et al. Identification and characterization of RNA guanine-quadruplex binding proteins. Nucleic Acids Res 2014;42(10):6630-44.
44- Roundtree IA, He C. Nuclear m(6)A Reader YTHDC1 Regulates mRNA Splicing. Trends Genet 2016;32(6):320-1.
45- Xue Y, Liu JQ, Zheng KW, Kan ZY, Hao YH, Tan Z. Kinetic and thermodynamic control of G-quadruplex folding. Angew Chem Int Ed Engl 2011;50(35):8046-50.
46- Song Y, Xu Q, Wei Z, Zhen D, Su J, Chen K, Meng J. Predict Epitranscriptome Targets and Regulatory Functions of N6-Methyladenosine (m6A) Writers and Erasers. Evol Bioinform Online 2019;15:1176934319871290.
47- Kierzek E, Kierzek R. The thermodynamic stability of RNA duplexes and hairpins containing N6-alkyladenosines and 2-methylthio-N6-alkyladenosines. Nucleic Acids Res 2003;31(15):4472-80.
48- Kennedy EM, Bogerd HP, Kornepati AVR, Kang D, Ghoshal D, Marshall JB, et al. posttranscriptional m6A editing of HIV-1 mRNAs enhances viral gene expression. Cell Host Microbe 2017;22(6):830.
49- Jenjaroenpun P, Wongsurawat T, Wadley TD, Wassenaar TM, Liu J, Dai Q, et al. Decoding the epitranscriptional landscape from native RNA sequences. Nucleic Acids Res 2021;49(2):e7.