Novel Insights into *RPGR* Exon ORF15: Could G-Quadruplex Folding Lead to Challenging Sequencing?

Luigi Donato[1,2,*], Concetta Scimone[1,2], Simona Alibrandi[1], Rosalia D'Angelo[1,2] and Antonina Sidoti[1,2]

[1]*Department of Biomedical and Dental Sciences and Morphofunctional Imaging, Division of Medical Biotechnologies and Preventive Medicine, University of Messina, Messina, Italy*

[2]*Department of Cutting-Edge Medicine and Therapies, Biomolecular Strategies and Neuroscience, Section of Applied Neuroscience, Molecular Genetics and Predictive Medicine, I.E.ME.S.T., Palermo, Italy*

Abstract: Hereditary retinal dystrophies (HRDs) represent a wide group of chronic and hereditary disorders affecting the retina, which constitute an important source of disability. Among inherited retinal dystrophies, retinitis pigmentosa (RP) represents the most genetically and clinically heterogeneous group. X-linked forms (OMIM 26800), the most severe subtypes of this disease, account for about 15% of RP cases. *RPGR*, one of the most X-linked RP involved genes, involved in ciliogenesis, microtubule organization and regulation of transport in primary cilia, presents a splicing variant, called exon ORF15, which represents a mutational hot spot in a huge number of patients. The most challenge peculiarity of exon ORF15 is its repetitive nature, especially of guanine (G)-rich sequences, that makes it very difficult to screen. Thus, we investigate the possible molecular causes that determine such difficulties by an in-silico approach, evaluating the possibility that, due to its nature, exon ORF15 could show a G-quadruplex structure. All the three algorithms exploited confirmed the possibility that several G-quadruplex could be folded in *RPGR* exon ORF15, providing new insights towards a better sequencing approach to RPGR diagnostic screening.

Keywords: Retinitis pigmentosa, RPGR, G-quadruplex, Bioinformatics, Retina.

INTRODUCTION

Hereditary retinal dystrophies (HRDs) represent a wide group of chronic and hereditary disorders affecting the retina, which constitute an important source of disability [1]. Among them, retinitis pigmentosa (RP) represents the most genetically and clinically heterogeneous group, with an overall incidence of 1 in 4000 in the general population [2]. X-linked forms (OMIM 26800), the most severe subtypes of this disease, account for about 15% of RP cases, and they are generally characterized by early onset with nyctalopia and constriction of visual fields, frequently progressing to complete blindness by the third or fourth decade [3]. Nowadays, two genes are known as main causes of X-linked RP: *RPGR* (OMIM 312619; Xp21.1) and *RP2* (OMIM 312600; Xp11.3) [4]. *RP2* encodes a protein of 350 amino acids, a GTPase-activating protein (GAP) for tubulin in concert with tubulin-specific chaperone C, involved in trafficking between the Golgi and the ciliary membrane. Variants in protein N-terminus alter normal targeting to the plasma membrane [5]. *RPGR* (RP3) codes for a protein with a series of six RCC1-like domains (RLDs), characteristic of the highly conserved guanine nucleotide exchange factors. RPGR is involved in ciliogenesis, probably by regulating actin stress filaments and cell contractility, and plays an important role in photoreceptor integrity, microtubule organization and regulation of transport in primary cilia [6]. Several splice variants were discovered in different human tissues in humans, and mutations in exon ORF15 were already identified in a mutant dog characterized by distinguishable X-linked progressive retinal atrophy [7]. The most interesting alternatively spliced *RPGR* transcript presents a 1.7-kb 3' terminal exon (ORF15) resulting from the retention of 1554 nucleotides of intron 15. Such mRNA is characterized by a purine-rich repetitive region, and codes for 567 C-terminal amino acids which shows many glycine and glutamic acid residues. Exon ORF15 represents a mutational hot spot in a huge number of patients affected by X-linked RP, probably accounting for about 20% of all RP cases [8]. The most challenge peculiarity of exon ORF15 is its repetitive nature, especially of guanine (G)-rich sequences, that, together with in-frame duplications and deletions diffused in the general population, makes it very difficult to screen [9]. Even if several strategies were developed to avoid this problem, such the use of nested primers from a single PCR product to sequence both strands of the repetitive stretch [10], obtained results are frequently not suitable for diagnostics. Thus, we investigate the possible molecular causes that

*Address correspondence to this author at the Department of Biomedical and Dental Sciences and Morphofunctional Imaging, Division of Medical Biotechnologies and Preventive Medicine, University of Messina, via C. Valeria 1, 98125 Messina, Italy; Tel: +39 0902213372; Fax: +39 090692449;
E-mail: ldonato@unime.it

E-ISSN: 2309-6136/19

determine such still present difficulties, evaluating the possibility that, due to its nature, exon ORF15 could show a G-Quadruplex structure [11]. So, in this work, we illustrate how G-quadruplexes can be arise on *RPGR* exon ORF15, presenting novel bioinformatic methods used to predict and identify G-quadruplex structures.

MATERIALS AND METHODS

Typically, a classic nucleic acid sequence presenting four sets of at least three guanines, disjointed by short tracts with other bases, can theoretically form an intramolecular G-quadruplex. Therefore, the possibility to fold into these motifs can be predicted from primary sequence. Interestingly, while an intermolecular G-quadruplex contains all the guanine series on both sense and antisense strands, in an intramolecular G-quadruplex they occur on the same strand of DNA [12]. Recently, several algorithms have been developed to predict the probable formation of G-quadruplexes directly from DNA sequence, including QGRS Mapper [13], G4-Predictor [14] and the newly G4 Hunter [15]. These algorithms exploit two main different approaches: 1) seeking sequences containing four runs of three guanines in close proximity; 2) considering supplementary factors which can influence G4 folding, e. g. the nature and extension of loops between guanine series. Main goal of these algorithms is the prediction of supposed G-quadruplex sequences (PQSs) that a target sequence be expected to contain at random and, thus, whether these sequences are under- or over-represented. Nevertheless, there are several problems to be faced, as biased base dyad frequencies (the nonrandom likelihood that any G will be followed by a T, A, C, or G) or inconstant genome composition (the global percentage of guanines may not be distributed equally through a genome). Additionally, even if the consensus sequence for PQSs has conventionally been $G_3\ N_{1-7}\ G_3\ N_{1-7}\ G_3\ N_{1-7}\ G_3$, it is now known that exists other PQSs with different structure in a genome, such as non-guanine knots and motifs with wider loops. Short loops represent the key factor in G-quadruplex stability, and considered algorithms include this, together with other factors, into a sliding score for G-quadruplex stability and propensity. Furthermore, such algorithms permit a user-defined loop length, resulting able to analyze large loops, *e.g*. N = 30, that have been already assessed to support G-quadruplex formation *in vitro*. Regarding this, it was recently seen that two instead of three guanine quartets can be sufficient, giving rise to $G_2\ N_x$ quadruplexes [16]. In our hypothesis, the latter could be the case of *RPGR* exon ORF15 (Figure **1**).

QGRS MAPPER

QGRS Mapper is a web-based tool that permits to seek putative G-quadruplexes starting from a wide range of inputs. It is possible to enter a FASTA or raw nucleotide sequence or search and analyze gene sequences by Gene ID, Gene name or symbol, accession number or GI number for an NCBI nucleotide sequence entry. Settings are very feasible, ranging from the possibility to change the maximum length of QGRS that will be searched for (default maximum length = 30) and the minimum sized G-group (default = 2) to the option to specify that the loops in the QGRS fall within a given size range and that one or more loops of the QGRS present a given string. After input selection, QGRS Mapper will start the analysis of QGRS within the query sequence. QGRS are assigned scores (G-scores) related to their likelihood to form G-quadruplexes. The scoring method is based on several parameters: 1) shorter loops are more common than longer loops; 2) G-quadruplexes generally contain loops approximately with the same size; 3) the greater the number of guanine tetrads, the more stable the quadruplex. Higher scoring sequences represent better candidates for G-quadruplexes. The computed G-scores depends on the chosen maximum QGRS length: using the default maximum QGRS length of 30 the highest possible G-score is 105. Moreover, the computed G-score is used to eliminate overlapping QGRS (it is well known that two QGRS overlap if their positions in the nucleotide sequence overlap). Overlaps are removed by selecting the higher scoring QGRS.

G4-PREDICTOR

G4-predictor online tool is able to predict the possible G-quadruplex folding sequence simultaneously in both sense and antisense strands. It also produces the output of the start and end positions of each potential G-quadruplex making motif and export total number of putative G-quadruplex motif in the queried sequence. G4 predictor tool is capable to perform analysis of very large sequences on any genome size, starting from manually inserted or locally selected sequences. Putative G-quadruplex forming sequences prediction is based on pattern matching of [G]{Y1} [X]{Y2}[G]{Y1} [X]{Y2}[G]{Y1} [X]{Y2}[G]{Y1} motif, where G represents the Guanine nucleotide and X represented any nucleotide. The length of guanine tracts (Y1) ranges from 2 to 7 in number and length of loop (Y2) varies between 1 and 7 nucleotides. The stability of predicted G-quadruplex structure will

AAGCATTTTCAGATGAGGAAGTAGGTAATGACACAGGCCAGGTGGGACCTCAGGCTGACACTGATGGAGA
GGGTTTACAAAAAGAGGTATATAGACATGAAAATAATAATGGTGTTGATCAACTTGATGCTAAGGAAATAG
AAAAGGAAAGTGATGGAGGACACAGTCAGAAGGAATCAGAAGCAGAAGAAATAGACAGTGAGAAGGAA
ACTAAACTGGCAGAAATAGCAGGTATGAAGGATTTAAGAGAAAGGGAAAAGAGTACAAAAAAGATGAGT
CCTTTCTTTGGCAACTTACCAGATAGAGGTATGAATACTGAGAGTGAAGAAAATAAAGATTTTGTTAAGAA
AAGGGAAAGTTGCAAGCAAGATGTGATCTTTGACAGTGAAAGAGAATCAGTAGAAAAGCCAGACAGTTAC
ATGGAAGGTGCAAGTGAGAGTCAACAGGGTATAGCTGATGGATTCCAGCAGCCTGAGGCAATAGAATTTA
GTAGTGGAGAGAAAGAGGATGATGAAGTGGAAACTGACCAAAACATACGGTATGGCAGGAAATTGATTG
AACAAGGAAATGAAAAAGAGACTAAACCCATAATATCCAAATCCATGGCAAAGTATGATTTTAAATGTGAT
CGCTTGTCAGAGATCCCAGAGGAGAAGGAAGGAGCAGAGGATTCAAAAGGAAATGGAATAGAGGAGCAA
GAGGTAGAAGCAAATGAGGAAAATGTGAAGGTGCATGGAGGAAGAAAGGAGAAAACAGAGATCCTATCA
GATGACCTTACAGACAAAGCAGAGGTGAGTGAAGGCAAGGCAAAATCAGTGGGAGAAGCAGAGGATGG
GCCTGAAGGTAGAGGGGATGGAACCTGTGAGGAAGGTAGTTCAGGAGCAGAACACTGGCAAGATGAGGA
GAGGGAGAAGGGGGAGAAAGACAAGGGTAGAGGAGAAATGGAGAGGCCAGGAGAGGGAGAGAAGGA
ACTAGCAGAGAAGGAAGAATGGAAGAAGAGGGATGGGGAAGAGCAGGAGCAAAAGGAGAGGGAGCAG
GGCCATCAGAAGGAAAGAAACCAAGAGATGGAGGAGGGAGGGGAGGAGGAGCATGGAGAAGGAGAAG
AAGAGGAGGGAGACAGAGAAGAGGAAGAAGAGAAGGAGGGAGAAGGGAAAGAGGAAGGAGAAGGGG
AAGAAGTGGAGGGAGAACGTGAAAAGGAGGAAGGAGAGAGGAAAAAGGAGGAAAGAGCGGGGAAGG
AGGAGAAAGGAGAGGAAGAAGGAGACCAAGGAGAGGGGGAAGAGGAGGAAACAGAGGGGAGAGGGG
AGGAAAAAGAGGAGGGAGGGGAAGTAGAGGGAGGGGAAGTAGAGGAGGGGAAAGGAGAGAGGGAAG
AGGAAGAGGAGGAGGGTGAGGGGGAAGAGGAGGAAGGGGAGGGGGAAGAGGAGGAAGGGGAGGGG
GAAGAGGAGGAAGGAGAAGGGAAAGGGGAGGAAGAAGGGGAAGAAGGAGAAGGGGAGGAAGAAGGG
GAGGAAGGAGAAGGGGAGGGGGAAGAGGAGGAAGGAGAAGGGGAGGGAGAAGAGGAAGGAGAAGG
GGAGGGAGAAGAGGAGGAAGGAGAAGGGGAGGGAGAAGAGGAAGGAGAAGGGGAGGGAGAAGAGG
AGGAAGGAGAAGGGAAAGGGGAGGAGGAAGGAGAGGAAGGAGAAGGGGAGGGGGAAGAGGAGGAA
GGAGAAGGGGAAGGGGAGGATGGAGAAGGGGAGGGGGAAGAGGAGGAAGGAGAATGGGAGGGGGA
AGAGGAGGAAGGAGAAGGGGAGGGGGAAGAGGAAGGAGAAGGGGAAGGGGAGGAAGGAGAAGGGG
AGGGGGAAGAGGAGGAAGGAGAAGGGGAGGGGGAAGAGGAGGAAGGGGAAGAAGAAGGGGAGGAA
GAAGGAGAGGGAGAGGAAGAAGGGGAGGGAGAAGGGGAGGAAGAAGAGGAAGGGGAAGTGGAAGG
GGAGGTGGAAGGGGAGGAAGGAGAGGGGGAAGGAGAGGAAGAGGAAGGAGAGGAGGAAGGAGAAG
AAAGGGAAAAGGAGGGGGAAGGAGAAGAAAACAGGAGGAACAGAGAAGAGGAGGAGGAAGAAGAGG
GGAAGTATCAGGAGACAGGCGAAGAAGAGAATGAAAGGCAGGATGGAGAGGAGTACAAAAAAGTGAGC
AAAATAAAAGGATCTGTGAAATATGGCAAACATAAAACATATCAAAAAAAGTCAGTTACTAACACACAGGG
AAATGGGAAAGAGCAGAGGTCCAAAATGCCAGTCCAGTCAAAACGACTTTTAAAAAACGGGCCATCAGGT
TCCAAAAAGTTCTGGAATAATGTATTACCACATTACTTGGAATTGAAGTAACAAACCTTAAATGTGACCCGA
TTATGGCCAGTCAGACAATTTAAATGCCTTGCATATAACGGGCACTCATTACGTGTTATTAAATTGATTTTAT
GTCAATTATTTTATGTGTAGTAAAAAAAAAAGCAACTGATGCAGCTGTGTTAAGGAGCCAAAGACAATAGG
AGGCACTGGTAAATTTTGGCCTCTCTCAAACTAAAATTTTCGTGTATTTCCCCCCCAAATTATAAAAACATAA
CTAGAAAATATTAAAAGGTCATATCAGATTATTAACATTATATATTCATTAAAGGCAGCTTTAGGAAACAGG
AATATACTACAAGAGTGTTTTGTTTGTGTATACAAATCATTCCATTTTTAAATGGCACAGATGCTTAAGGGCT
ATAAAAACTTCTAATTTCTTATAAATATGTTAGCACTTTTTTTAAGTTAGTGATTACAGTTTACCTACTGTATA
GAATAATTTTCTAATAATGGATGGTATTCTAAAACTCAATTGAGGCATTCACATTTTAAAGAAAGTATTGTCT
TTCAC

Figure 1: *RPGR* ORF15 sequence. Figure represents the ORF15 retained exon of *RPGR*.

depend upon length of the internal loop, the number of tandem repeats of the motif sequence and the number of guanine tracts. The default maximum length of putative G-quadruplex forming sequence is 49 bases. The G4 Prediction Score, based on the efficiently calculated cG and cC score by considering few base pairs upstream and downstream of putative G4 motifs, has been validated as robust and reliable score, reducing the false positive. The cG score calculation is based on the following equation and applied for each predicted substring (s) that has the length of (n), with i=1: $cGs = \sum\left(Gs \quad i*10*i\right)$. A value of 10 is assigned to each G, a value of 20 to each paired GG, a value of 30 to each triplet GGG, and so on. The cC score calculation is similar, with the cytosine nucleotide used instead of guanine nucleotide. The cG/cC score is based on the ratio of both cG and cC scores. The putative G-quadruplex motifs with higher cG/cC score have more probability to readily fold into G-quadruplex structure.

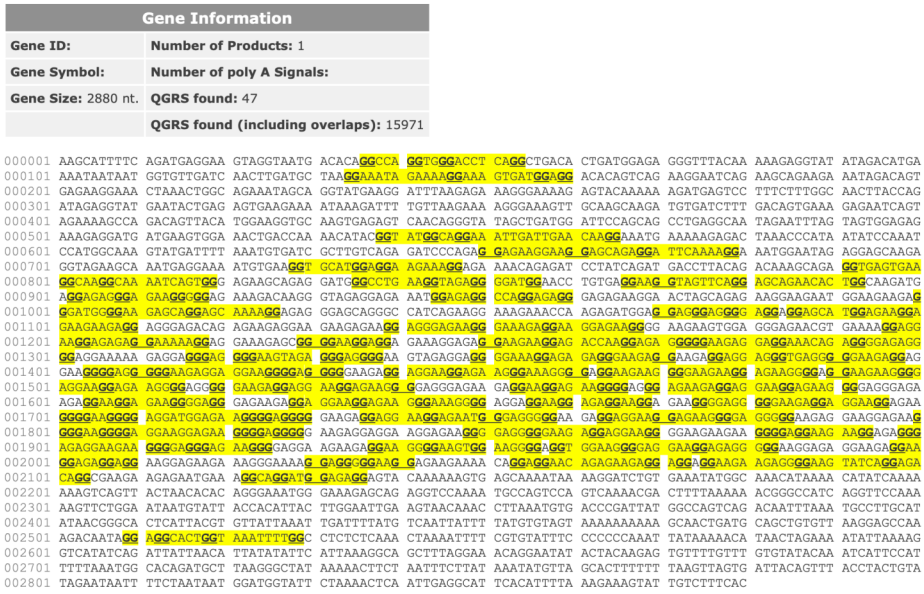
G4-HUNTER

G4Hunter represents the most recent developed algorithm, which tries to reduce the number of false positives and negatives detected by other tools, as well as providing a quantitative analysis that would allow correlation of a given quadruplex “strength” metric with other genomic or functional parameters. It takes into account G-skewness and G-richness of an input sequence and outputs a quadruplex propensity score.

Skewness reflects G/C asymmetry between the complementary strands, while richness deals with the fraction of Gs in the sequence. In order evaluate G skewness and G richness, each position in a sequence is given a score between −4 and 4. The score is positive for G, negative for C and 0 for A and T (neutral). To account for G-richness (or C-richness, meaning G-richness on the complementary strand), a single G is given a score of 1, in a GG sequence each G is given a score of 2, and so one, until in a sequence of 4 or more Gs each G is given a score of 4. The Cs are computed similarly but with negative values. This scoring scheme also enables simultaneous scoring of the complementary strand. For a given sequence, the G4Hunter score (G4Hscore) represents the arithmetic mean of this "sequence" of numbers. By construction, the G4Hscore is centered on 0 for random sequences, independently of GC content.

RESULTS

RPGR exon ORF15 analysis for G-quadruplex structures by QGRS mapper predicted and interesting distribution of QGRS, especially between 550 and 2100 nts (Figures **2-3**). In details, 15971 QGRS including overlaps and 47 QGRS with no overlaps throughout the whole fragment were detected. Curiously, three of QGRS without overlaps (position=1404, length=30, sequence=GGGGAGGGGGAAGAGGAGGAAGGGGA GGGG; position=1701, length=30; sequence=GGGGAAGGGGAGGATGGAGAAGGGGA GGGG; position=1800, length=30, sequence=GGGGAAGGGGAGGAAGGAGAAGGGGA GGGG) reached the highest score of 53 (Table **1**). G4-predictors, instead, has correctly predicted only 21 of 75 possible G-quadruplex folding sequences from sense strand (5' to 3') and all the three possible G-quadruplex folding sequences from antisense strand (3'



Gene Information	
Gene ID:	**Number of Products:** 1
Gene Symbol:	**Number of poly A Signals:**
Gene Size: 2880 nt.	**QGRS found:** 47
	QGRS found (including overlaps): 15971

```
000001 AAGCATTTTC AGATGAGGAA GTAGGTAATG ACACAGGCCA GGTGGGACCT CAGGCTGACA CTGATGGAGA GGGTTTACAA AAAGAGGTAT ATAGACATGA
000101 AAATAATAAT GGTGTTGATC AACTTGATGC TAAGGAAATA GAAAAGGAAA GTGATGGAGG ACACAGTCAG AAGGAATCAG AAGCAGAAGA AATAGACAGT
000201 GAGAAGGAAA CTAAACTGGC AGAAATAGCA GGTATGAAGG ATTTAAGAGA AAGGGAAAAG AGTACAAAAA AGATGAGTCC TTTCTTTGGC AACTTACCAG
000301 ATAGAGGTAT GAATACTGAG AGTGAAGAAA ATAAAGATTT TGTTAAGAAA AGGGAAAGTT GCAAGCAAGA TGTGATCTTT GACAGTGAAA GAGAATCAGT
000401 AGAAAAGCCA GACAGTTACA TGGAAGGTGC AAGTGAGAGT CAACAGGGTA TAGCTGATGG ATTCCAGCAG CCTGAGGCAA TAGAATTTAG TAGTGGAGAG
000501 AAAGAGGATG ATGAAGTGGA AACTGACCAA AACATACGGT ATGGCAGGAA ATTGATTGAA CAAGGAAATG AAAAAGAGAC TAAACCCATA ATATCCAAAT
000601 CCATGGCAAA GTATGATTTT AAATGTGATC GCTTGTCAGA GATCCCAGAG GAGAAGGAAG GAGCAGAGGA TTCAAAAGGA AATGGAATAG AGGAGCAAGA
000701 GGTAGAAGCA AATGAGGAAA ATGTGAAGGT GCATGGAGGA AGAAAGGAGA AAACAGAGAT CCTATCAGAT GACCTTACAG ACAAAGCAGA GGTGAGTGAA
000801 GGCAAGGCAA AATCAGTGGG AGAAGCAGAG GATGGGCCTG AAGGTAGAGG GGATGGAACC TGTGAGGAAG GTAGTTCAGG AGCAGAACAC TGGCAAGATG
000901 AGGAGAGGGA GAAGGGGGAG AAAGACAAGG GTAGAGGAGA AATGGAGAGG CCAGGAGAGG GAGAGAAGGA ACTAGCAGAG AAGGAAGAAT GGAAGAAGAG
001001 GGATGGGGAA GAGCAGGAGC AAAAGGAGAG GGAGCAGGGC CATCAGAAGG AAAGAAACCA AGAGATGGAG GAGGGAGGGG AGGAGGAGCA TGGAGAAGGA
001101 GAAGAAGAGG AGGGAGACAG AGAAGAGGAA GAAGAGAAGG AGGGAGAAGG GAAAGAGGAA GGAGAAGGGG AAGAAGTGGA GGGAGAACGT GAAAAGGAGG
001201 AAGGAGAGAG GAAAAAGGAG GAAAGAGCGG GGAAGGAGGA GAAAGGAGAG GAAGAAGGAG ACCAAGGAGA GGGGGAAGAG GAGGAAACAG AGGGGAGAGG
001301 GGAGGAAAAA GAGGAGGGAG GGGAAGTAGA GGGAGGGGAA GTAGAGGAGG GGAAAGGAGA GAGGGAAGAG GAAGAGGAGG AGGGTGAGGG GGAAGAGGAG
001401 GAAGGGGAGG GGGAAGAGGA GGAAGGGGAG GGGGAAGAGG AGGAAGGAGA AGGGAAAGGG GAGGAAGAAG GGGAAGAAGG AGAAGGGAGG GAAGAAGGGG
001501 AGGAAGGAGA AGGGGAGGGG GAAGAGGAGG AAGGGAGAAG GGAGGGAGAA GAGGGAAGGAG AAGGGGAGGG AGAAGAGGAG GAAGGAGAAG GGGAGGGAGA
001601 AGAGGAAGGA GAAGGGGAGG GAGAAGAGGA GGAAGGAGAA GGGAAAGGGG AGGAGGAAGG AGAGGAAGGA GAAGGGGAGG GGGAAGAGGA GGAAGGAGAA
001701 GGGGAAGGGG AGGATGGAGA AGGGGAGGGG GAAGAGGAGG AAGGAGAATG GGAGGGGGAA GAGGAGGAAG GAGAAGGGGA GGGGGAAGAG GAAGGAGAAG
001801 GGGAAGGGGA GGAAGGAGAA GGGGAGGGGG AAGAGGAGGA AGGAGAAGGG GAGCGGGAAG AGGAGGAAGG GGAAGAAGAA GGGGAGGAAG AAGGAGAGGG
001901 AGAGGAAGAA GGGGAGGGAG AAGGGGAGGA AGAAGAGGAA GGGGAAGTGG AAGCGGAGGT GGAAGGGGAG GAAGGAGAGG GGGAAGGAGA GGAAGAGGAA
002001 GGAGAGGAGG AAGGAGAAGA AAGGGAAAAG GAGGGGGAAG GAGAAGAAAA CAGGAGGAAC AGAGAAGAGG AGGAGGAAGA AGAGGGGAAG TATCAGGAGA
002101 CAGGCGAAGA AGAGAATGAA AGGCAGGATG GAGAGGAGTA CAAAAAAGTG AGCAAAATAA AAGGATCTGT GAAATATGGC AAACATAAAA CATATCAAAA
002201 AAAGTCAGTT ACTAACACAC AGGGAAATGG GAAAGAGCAG AGGTCCAAAA TGCCAGTCCA GTCAAAACGA CTTTTAAAAA ACGGGCCATC AGGTTCCAAA
002301 AAGTTCTGGA ATAATGTATT ACCACATTAC TTGGAATTGA AGTAACAAAC CTTAAATGTG ACCCGATTAT GGCCAGTCAG ACAATTTAAA TGCCTTGCAT
002401 ATAACGGGCA CTCATTACGT GTTATTAAAT TGATTTTATG TCAATTATTT TATGTGTAGT AAAAAAAAAA GCAACTGATG CAGCTGTGTT AAGGAGCCAA
002501 AGACAATAGG AGGCACTGGT AAATTTTGGC CTCTCTCAAA CTAAAATTTT CGTGTATTTC CCCCCCAAAT TATAAAAACA TAACTAGAAA ATATTAAAAG
002601 GTCATATCAG ATTATTAACA TTATATATTC ATTAAAGGCA GCTTTAGGAA ACAGGAATAT ACTACAAGAG TGTTTTGTTT GTGTATACAA ATCATTCCAT
002701 TTTTAAATGG CACAGATGCT TAAGGGCTAT AAAAACTTCT AATTTCTTAT AAATATGTTA GCACTTTTTT TAAGTTAGTG ATTACAGTTT ACCTACTGTA
002801 TAGAATAATT TTCTAATAAT GGATGGTATT CTAAAACTCA ATTGAGGCAT TCACATTTTA AAGAAAGTAT TGTCTTTCAC
```

Figure 2: *RPGR* Exon ORF15 QGRS distribution by QGRS mapper. This figure represents the whole analyzed fragment, with highlighted (in yellow background) sites for QGRS.

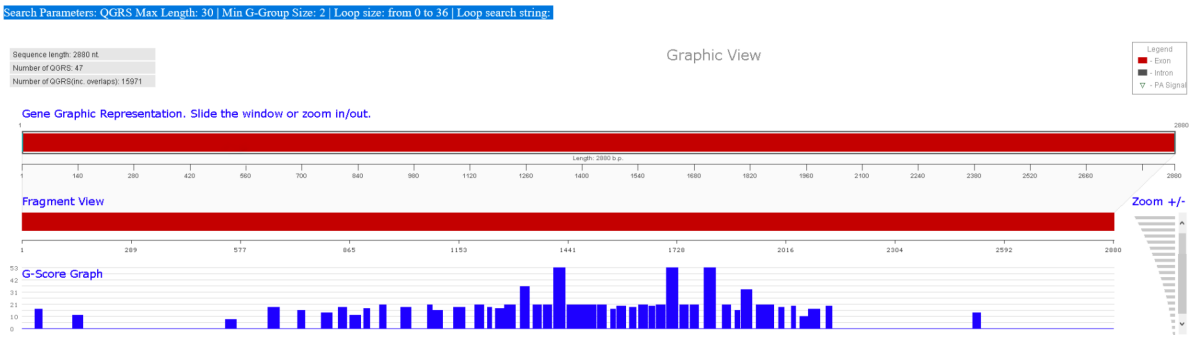


Figure 3: *RPGR* Exon ORF15 Graphics View by QGRS mapper. A graphic representation of the selected fragment showing the location and G-score for the non-overlapping QGRS in that fragment. The nucleotide sequence is visible at maximum zoom levels (not showed).

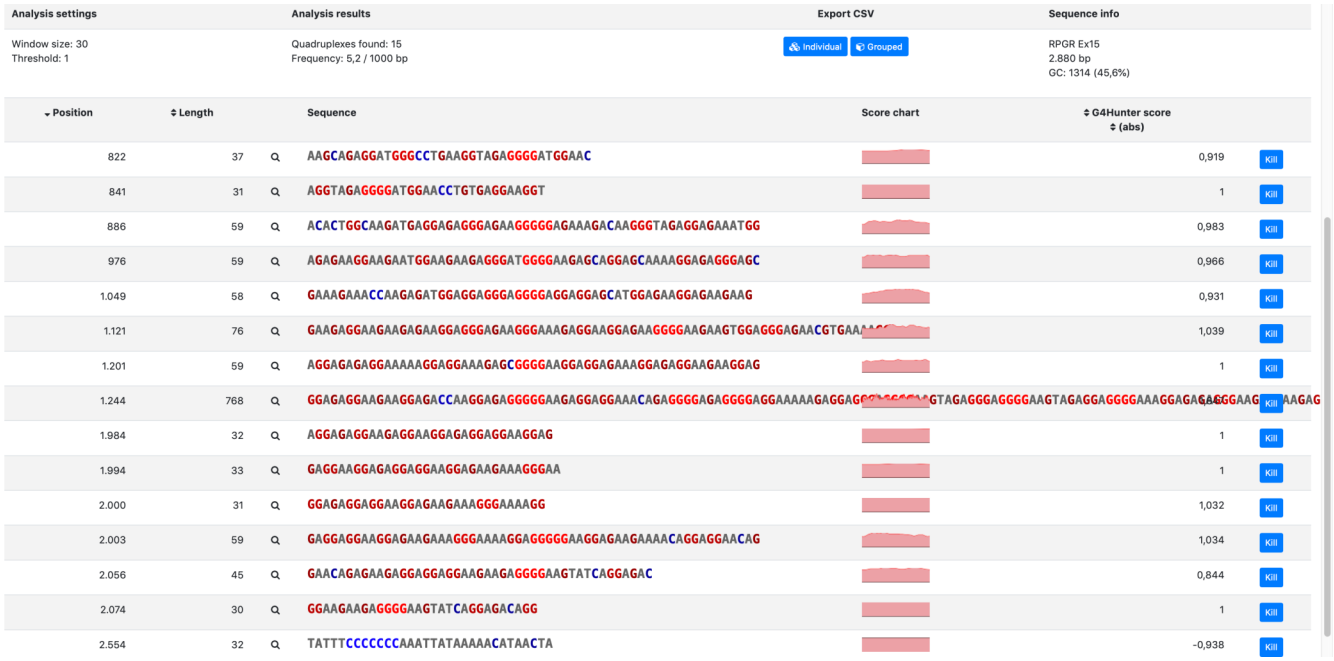


Figure 4: G4-Hunter main result page analysis. The sequence browser component showed the nucleotide sequences of the results and a cut-out of bases which fits the screen/browser window. The sequences corresponding to the analysis parameters are marked by colors (G in red—the longer the G-track, the brighter the intensity, C in blue). Position in the sequence, length, sequence, score chart and G4Hunter score are shown.

to 5') (Table **2**). Among them, two putative G-quadruplex motifs ((length=44, position=476, sequence=GGCAATAGAATTTAGTAGTGGAGAGAAA GAGGATGATGAAGTGG; length=27, position=892, sequence=GGCAAGATGAGGAGAGGGAGAAGGGG G) highlighted the best score (based on cG/cC ratio) of 17 and 16, respectively. Finally, G4-Hunter prediction showed 15 putative aggregated G-quadruplex, with one (position=1244, length=768) reaching the highest score of 1.65 (Figure **4** and Table **3**).

DISCUSSION

DNA and RNA are biopolymers essential for all life forms, and play a fundamental role in in encoding, transmitting and expressing genetic information. Nucleic acid sequences can fold into a huge number of structural motifs, such as hairpins, pseudoknots, duplexes, triplexes and G-quadruplexes, in order to reach the functional structural conformation needed for their specific biological roles related to cellular environments [18]. One of the most interestingly aspect regards guanine (G)-rich sequences, which can self-associate into stacks of G-quartets to form composite structural motifs called G-quadruplexes [19]. Such complexes are of growing interest in biology, mainly due to their peculiar and different molecular structures, that include parallel and antiparallel topologies. During last years, several studies highlighted that G-quadruplexes play critical regulatory roles in biological processes, such as transcription, translation and DNA replication, depicting new fundamental mechanisms involved in gene expression regulation and genome stability [20]. Moreover, G-quadruplexes have been used multiple times to develop molecular tools binding to different classes of targets [21]. Among them, it is interesting the possibility to use G-quadruplex-containing aptamers as therapeutic and diagnostic agents due to their thermodynamical and chemical stability, resistance to various serum nucleases, low immunogenicity and good cellular uptake [22]. Furthermore, specific G-quadruplex-containing aptamers have also been found to recognize small molecules, being capable to act as biosensors [23]. This study will propose to analyze the possible G-quadruplex folding of the *RPGR* gene, one of the retinitis pigmentosa-associated genes that localize to the connecting cilium, the typical structure of photoreceptor that mediate vesicular trafficking between inner and outer segment of photoreceptors. Mutation analysis of *RPGR*-XLRP patients highlighted that mutations in exon ORF15 are frequently associated with clinically heterogeneous disease, ranging from RP to cone-rod degeneration (especially in variants toward the 3′ end of exon ORF15) and macular degeneration. This scenario is challenging, because the RPGR-ORF15 isoform expression can be detected in both rods and cones, and personalized

Table 1: *RPGR* Exon ORF15 Data View. The Table Lists All QGRS Mapped to the Product, Including Information about the Position of the QGRS, its Distance from 3' and 5' Splice Sites, the Actual Sequence (Underlining the G-Groups) and its G-Score

Position	Length	QGRS	G-Score
36	19	GGCCAGGTGGGACCTCAGG	17
134	27	GGAAATAGAAAAGGAAAGTGATGGAGG	12
538	28	GGTATGGCAGGAAATTGATTGAACAAGG	8
650	30	GGAGAAGGAAGGAGCAGAGGATTCAAAAGG	19
728	20	GGTGCATGGAGGAAGAAAGG	16
791	29	GGTGAGTGAAGGCAAGGCAAAATCAGTGG	14
835	22	GGCCTGAAGGTAGAGGGGATGG	19
866	28	GGAAGGTAGTTCAGGAGCAGAACACTGG	12
902	17	GGAGAGGGAGAAGGGGG	18
944	17	GGAGAGGCCAGGAGAGG	21
1000	27	GGGATGGGGAAGAGCAGGAGCAAAAGG	19
1070	14	GGAGGGAGGGGAGG	21
1085	26	GGAGCATGGAGAAGGAGAAGAAGAGG	16
1139	30	GGAGGGAGAAGGGAAAGAGGAAGGAGAAGG	19
1196	23	GGAGGAAGGAGAGAGGAAAAAGG	21
1229	11	GGGGAAGGAGG	19
1250	24	GGAAGAAGGAGACCAAGGAGAGGG	18
1274	29	GGAAGAGGAGGAAACAGAGGGGAGAGGGG	21
1316	23	GGGAGGGGAAGTAGAGGGAGGGG	37
1349	23	GGGGAAAGGAGAGAGGGAAGAGG	21
1376	23	GGAGGAGGGTGAGGGGGAAGAGG	21
1404	30	GGGGAGGGGGAAGAGGAGGAAGGGGAGGGG	53
1439	23	GGAGGAAGGAGAAGGGAAAGGGG	21
1463	26	GGAAGAAGGGGAAGAAGGAGAAGGGG	21
1490	26	GGAAGAAGGGGAGGAAGGAGAAGGGG	21
1519	23	GGGAAGAGGAGGAAGGAGAAGGG	21
1553	14	GGAAGGAGAAGGGG	17
1569	24	GGAGAAGAGGAGGAAGGAGAAGGG	20
1604	17	GGAAGGAGAAGGGGAGG	19
1628	23	GGAGGAAGGAGAAGGGAAAGGGG	21
1655	15	GGAAGGAGAGGAAGG	20
1674	23	GGGGAGGGGGAAGAGGAGGAAGG	21
1701	30	GGGGAAGGGGAGGATGGAGAAGGGGAGGGG	53
1736	23	GGAGGAAGGAGAATGGGAGGGGG	21
1763	23	GGAGGAAGGAGAAGGGGAGGGGG	21
1800	30	GGGGAAGGGGAGGAAGGAGAAGGGGAGGGG	53
1848	23	GGGGAGGGGGAAGAGGAGGAAGG	21
1881	14	GGGGAGGAAGAAGG	16
1898	28	GGGAGAGGAAGAAGGGGAGGGAGAAGGG	34
1937	20	GGAAGGGGAAGTGGAAGGGG	21

1958	26	**GGTGGAAGGGGAGGAAGGAGAGGGGG**	21
1997	14	**GGAAGGAGAGGAGG**	19
2030	12	**GGAGGGGGAAGG**	20
2053	21	**GGAGGAACAGAGAAGAGGAGG**	11
2075	30	**GGAAGAAGAGGGGAAGTATCAGGAGACAGG**	17
2122	15	**GGCAGGATGGAGAGG**	20
2509	21	**GGAGGCACTGGTAAATTTTGG**	14

Table 2: G4-Predictor Results. Table Shows Details of G-Quadruplex Predicted by G4-Predictor. The First Block Highlights Sense Motifs, while Antisense Motifs are Shown at the Bottom. The Last Column Represents the cG/cC Ratio, Indicative of G4-Predictor Ranking Score

#	Length (bp)	Start Position	End Position	PG4 Motifs (Sense)	cG Score	cC Score	cG/cC
1	26	17	42	GGAAGTAGGTAATGACACAGGCCAGG	100	40	2,50
2	30	44	73	GGGACCTCAGGCTGACACTGATGGAGAGGG	130	60	2,17
3	27	134	160	GGAAATAGAAAAGGAAAGTGATGGAGG	110	0	N/A
4	35	206	240	GGAAACTAAACTGGCAGAAATAGCAGGTATGAAGG	110	40	2,75
5	39	422	460	GGAAGGTGCAAGTGAGAGTCAACAGGGTATAGCTGATGG	160	40	4
6	44	476	519	GGCAATAGAATTTAGTAGTGGAGAGAAAGAGGATGATGAAGTGG	170	10	17
7	28	538	565	GGTATGGCAGGAAATTGATTGAACAAGG	100	20	5
8	20	650	669	GGAGAAGGAAGGAGCAGAGG	110	10	11
9	25	678	702	GGAAATGGAATAGAGGAGCAAGAGG	110	10	11
10	24	716	739	GGAAAATGTGAAGGTGCATGGAGG	110	10	11
11	30	791	820	GGTGAGTGAAGGCAAGGCAAAATCAGTGGG	130	30	4,33
12	23	830	852	GGATGGGCCTGAAGGTAGAGGGG	130	20	6,5
13	26	855	880	GGAACCTGTGAGGAAGGTAGTTCAGG	110	30	3,67
14	27	892	918	GGCAAGATGAGGAGAGGGAGAAGGGGG	160	10	16
15	22	929	950	GGGTAGAGGAGAAATGGAGAGG	120	0	N/A
16	31	954	984	GGAGAGGGAGAGAAGGAACTAGCAGAGAAGG	150	20	7,5
17	18	991	1008	GGAAGAAGAGGGATGGGG	110	0	N/A
18	24	1016	1039	GGAGCAAAAGGAGAGGGAGCAGGG	130	20	6,5
19	27	1049	1075	GGAAAGAAACCAAGAGATGGAGGAGGG	120	20	6
20	10	1077	1086	GGGGAGGAGG	80	0	N/A
21	23	1092	1114	GGAGAAGGAGAAGAAGAGGAGGG	130	0	N/A
22	25	1127	1151	GGAAGAAGAGAAGGAGGGAGAAGGG	140	0	N/A
23	14	1157	1170	GGAAGGAGAAGGGG	90	0	N/A
24	23	1178	1200	GGAGGGAGAACGTGAAAAGGAGG	120	10	12
25	19	1203	1221	GGAGAGAGGAAAAAGGAGG	100	0	N/A
26	11	1229	1239	GGGGAAGGAGG	80	0	N/A
27	23	1245	1267	GGAGAGGAAGAAGGAGACCAAGG	110	20	5,5
28	14	1271	1284	GGGGGAAGAGGAGG	100	0	N/A
29	11	1292	1302	GGGGAGAGGGG	90	0	N/A
30	20	1304	1323	GGAAAAAGAGGAGGGAGGGG	120	0	N/A
31	17	1331	1347	GGGAGGGGAAGTAGAGG	110	0	N/A
32	17	1349	1365	GGGGAAAGGAGAGAGGG	110	0	N/A

33	15	1370	1384	GGAAGAGGAGGAGGG	100	0	N/A
34	14	1388	1401	GGGGGAAGAGGAGG	100	0	N/A
35	10	1404	1413	GGGGAGGGGG	90	0	N/A
36	11	1418	1428	GGAGGAAGGGG	80	0	N/A
37	14	1430	1443	GGGGGAAGAGGAGG	100	0	N/A
38	16	1446	1461	GGAGAAGGGAAAGGGG	100	0	N/A
39	18	1463	1480	GGAAGAAGGGGAAGAAGG	100	0	N/A
40	16	1485	1500	GGGGAGGAAGAAGGGG	110	0	N/A
41	14	1502	1515	GGAAGGAGAAGGGG	90	0	N/A
42	14	1517	1530	GGGGGAAGAGGAGG	100	0	N/A
43	14	1533	1546	GGAGAAGGGGAGGG	100	0	N/A
44	14	1553	1566	GGAAGGAGAAGGGG	90	0	N/A
45	18	1568	1585	GGGAGAAGAGGAGGAAGG	110	0	N/A
46	16	1590	1605	GGGGAGGGAGAAGAGG	110	0	N/A
47	14	1608	1621	GGAGAAGGGGAGGG	100	0	N/A
48	16	1628	1643	GGAGGAAGGAGAAGGG	100	0	N/A
49	10	1647	1656	GGGGAGGAGG	80	0	N/A
50	19	1659	1677	GGAGAGGAAGGAGAAGGGG	120	0	N/A
51	14	1679	1692	GGGGGAAGAGGAGG	100	0	N/A
52	16	1695	1710	GGAGAAGGGGAAGGGG	110	0	N/A
53	14	1712	1725	GGATGGAGAAGGGG	90	0	N/A
54	14	1727	1740	GGGGGAAGAGGAGG	100	0	N/A
55	16	1743	1758	GGAGAATGGGAGGGGG	110	0	N/A
56	17	1763	1779	GGAGGAAGGAGAAGGGG	110	0	N/A
57	15	1781	1795	GGGGGAAGAGGAAGG	100	0	N/A
58	10	1800	1809	GGGGAAGGGG	80	0	N/A
59	14	1811	1824	GGAAGGAGAAGGGG	90	0	N/A
60	14	1826	1839	GGGGGAAGAGGAGG	100	0	N/A
61	16	1842	1857	GGAGAAGGGGAGGGGG	120	0	N/A
62	11	1862	1872	GGAGGAAGGGG	80	0	N/A
63	14	1881	1894	GGGGAGGAAGAAGG	90	0	N/A
64	17	1898	1914	GGGAGAGGAAGAAGGGG	110	0	N/A
65	14	1916	1929	GGGAGAAGGGGAGG	100	0	N/A
66	14	1937	1950	GGAAGGGGAAGTGG	90	0	N/A
67	10	1953	1962	GGGGAGGTGG	80	0	N/A
68	11	1965	1975	GGGGAGGAAGG	80	0	N/A
69	14	1979	1992	GGGGGAAGGAGAGG	100	0	N/A
70	14	1997	2010	GGAAGGAGAGGAGG	90	0	N/A
71	25	2013	2037	GGAGAAGAAAGGGAAAAGGAGGGGG	140	0	N/A
72	31	2040	2070	GGAGAAGAAAACAGGAGGAACAGAGAAGAGG	130	20	6,5
73	16	2072	2087	GGAGGAAGAAGAGGGG	100	0	N/A
74	32	2096	2127	GGAGACAGGCGAAGAAGAGAATGAAAGGCAGG	140	30	4,67
75	27	2493	2519	GGAGCCAAAGACAATAGGAGGCACTGG	100	50	2
#	Length (bp)	Start Position	End Position	PG4 Motifs (Antisense)	cG Score	cC Score	cG/cC

1	37	315	351	GGGGGGGAAATACACGAAAATTTTAGTTTGAGAGAGG	140	20	7
2	45	487	531	GGCATTTAAATTGTCTGACTGGCCATAATCGGGTCACATTTAAGG	110	80	1,375
3	45	584	628	GGAACCTGATGGCCCGTTTTTTAAAAGTCGTTTTGACTGGACTGG	130	80	1,625

Table 3: G4-Hunter Aggregated Sequences output. Table Shows Results of G4-Hunter Predictive Analysis. POSITION in Sequence, LENGTH of the Longest Continuous Sequence with G4Hunter Scores above Threshold, its SCORE and the Part of the Sequence. The SUB_SCORE shows Scores for each Window Position Inside the Concatenated Sequence

ID	POSITION	LENGTH	SCORE	SEQUENCE	SUB_SCORE
1	822	37	0.918	AAGCAGAGGATGGGCCTGAAGGTAGAGGGGATGGAAC	1.0666667,1.0666667,1.0666667,1.1,1.2,1.2,1.1666666,1.1
2	841	31	1.0	AGGTAGAGGGGATGGAACCTGTGAGGAAGGT	1.0333333,1.0333333
3	886	59	0.983	ACACTGGCAAGATGAGGAGAGGGAGAAGGGGGAGAAAGACAAGGGTAGAGGAGAAATGG	1.0,1.1333333,1.3,1.3,1.3666667,1.3666667,1.3,1.2333333,1.3,1.3,1.2666667,1.2333333,1.2333333,1.3333334,1.4,1.5,1.4333333,1.3666667,1.4,1.3666667,1.4333333,1.4,1.3,1.2333333,1.2333333,1.2,1.2,1.2,1.1333333,1.0666667
4	976	59	0.96	AGAGAAGGAAGAATGGAAGAAGAGGGATGGGGAAGAGCAGGAGCAAAAGGAGAGGGAGC	1.0,1.1333333,1.2333333,1.2333333,1.2,1.2333333,1.2333333,1.2,1.1,1.1,1.1666666,1.2,1.2,1.2333333,1.2,1.1333333,1.0666667,1.0666667,1.0666667,1.1,1.1666666,1.1666666,1.1666666,1.1666666,1.1666666,1.1666666,1.1666666,1.1666666,1.2,1.0333333
5	1049	58	0.931	GAAAGAAACCAAGAGATGGAGGAGGGAGGGGAGGAGGAGCATGGAGAAGGAGAAGAAG	1.0,1.0666667,1.0666667,1.1333333,1.2,1.1666666,1.2333333,1.3,1.3,1.4,1.4333333,1.4333333,1.4333333,1.4666667,1.5333333,1.5,1.5333333,1.5333333,1.4666667,1.4666667,1.5333333,1.4666667,1.4333333,1.4333333,1.3333334,1.2666667,1.1666666,1.1666666,1.0666667
6	1121	76	1.04	GAAGAGGAAGAAGAGAAGGAGGGAGAAGGGAAAGAGGAAGGAGAAGGGGAAGAAGTGGAGGGAGAACGTGAAAAGG	1.0666667,1.0333333,1.0333333,1.0333333,1.0333333,1.0333333,1.0333333,1.0333333,1.0333333,1.0333333,1.0666667,1.1333333,1.1333333,1.1333333,1.1333333,1.1,1.2333333,1.3666667,1.4333333,1.5,1.5,1.4,1.3333334,1.2333333,1.2333333,1.2333333,1.2333333,1.3,1.2666667,1.1666666,1.1666666,1.2666667,1.3666667,1.3666667,1.3666667,1.3666667,1.3,1.2,1.2333333,1.2333333,1.2,1.1333333,1.1333333,1.1,1.1,1.1666666,1.1
7	1201	59	1.0	AGGAGAGAGGAAAAAGGAGGAAAGAGCGGGGAAGGAGGAGAAAGGAGAGGAAGAAGGAG	1.0333333,1.1666666,1.1,1.0333333,1.1,1.1333333,1.1333333,1.1666666,1.2333333,1.1666666,1.1333333,1.1333333,1.1333333,1.1333333,1.2,1.2666667,1.2,1.1666666,1.1666666,1.1666666,1.1666666,1.1666666,1.1666666,1.2,1.1666666,1.1666666,1.2,1.3,1.1666666,1.0666667
8	1244	768	1.65	GGAGAGGAAGAAGGAGACCAAGGAGAGGGGGAAGAGGAGGAAACAGAGGGGAGAGGGGAGGAAAA…...	1.0666667,1.1333333,1.0666667,1.0666667,1.0666667,1.0666667, …
9	1984	32	1.0	AGGAGAGGAAGAGGAAGGAGAGGAGGAAGGAG	1.0333333,1.0333333,1.0
10	1994	33	1.0	GAGGAAGGAGAGGAGGAAGGAGAAGAAAGGGAA	1.0,1.0666667,1.0666667,1.0
11	2000	31	1.03	GGAGAGGAGGAAGGAGAAGAAAGGGAAAAGG	1.0,1.0
12	2003	59	1.03	GAGGAGGAAGGAGAAGAAAGGGAAAAGGAGGGGGAAGGAGAAGAAAACAGGAGGAACAG	1.0666667,1.1666666,1.3,1.3666667,1.4333333,1.4333333,1.3666667,1.3666667,1.4333333,1.4333333,1.4,1.3333334,1.3333334,1.3333334,1.3333334,1.3333334,1.3,1.3,1.2666667,1.2666667,1.2333333,1.2,1.1,1.1666666,1.2333333,1.2333333,1.2333333,1.1333333,1.0666667,1.1
13	2056	45	0.844	GAACAGAGAAGAGGAGGAGGAAGAAGAGGGGAAGTATCAGGAGAC	1.0,1.0666667,1.0666667,1.0666667,1.1333333,1.1333333,1.1,1.1,1.0333333,1.0333333,1.1,1.1333333,1.1333333,1.1,1.0333333,1.0

14	2074	30	1.0	GGAAGAAGAGGGGAAGTATCAGGA GACAGG	1.0
15	2554	32	0.937	TATTTCCCCCCCAAATTATAAAAACA TAACTA	-1.0,-1.0,-1.0

genotype-phenotype correlation in RPGR-ORF15 patients might depend on spatially restricted effect of the exon ORF15 mutations on the retina [24]. Though both photoreceptors may be affected by ORF15 mutations, wide spatial distribution of the impaired photoreceptors was observed, and the degree of effect of such variants on rods versus cones was variable yet bilaterally consistent in patients [25]. So, as just described, the challenging screening of *RPGR* ORF15 is fundamental to realize a personalized diagnosis of the correct inheritable dystrophy. However, due to high G richness of ORF15, it is very difficult to sequence this exon, even if several scientists tried to develop procedures *ad hoc*. The real problem probably relies on the possible creation of G-quadruplex structures within the *RPGR* ORF15, as shown by output of the realized in-silico analyses. All three algorithms used to predict G-quadruplex showed a potential arising of these structures at different intervals within the query sequence, especially between 500 and 2000 nucleotides, region with the greatest number of Gs. Such hypothesis could explain why sequencing *RPGR* exon ORF15 with Sanger method could not provide clear electropherograms, especially in previously described sequence, and even using nested primers. Moreover, as already said, the case of ORF15 could represent one example in which two instead of three guanine tracks can generate $G_2 N_x$ quadruplexes.

CONCLUSIONS

We analyze the challenging case of *RPGR* exon ORF15, whose mutations could determine the onset of different forms of inherited retinal dystrophies, as X-linked retinitis pigmentosa. An in-silico approach was exploited, using three different algorithms to predict the possibility that this exon could form several G-quadruplex, altering its sequencing. Since no predictive algorithm is able to obtain certain results, users must define their parameters appropriately and balance the chances of false positives against false negatives. Therefore, it is critical to confirm in silico predictions by *in vitro* approaches, such as cell imaging or omics ones, and/or *in vivo* methods. Deciphering the complex picture involving G-quadruplex, suggesting their connection with gene regulation and disease development, could clear underlying biochemical mechanism and the molecular basis of pathologies, facilitating the rational design and development of G-quadruplex-related tools for various biological applications.

REFERENCES

[1] Broadgate S, Yu J, Downes SM, Halford S. Unravelling the genetics of inherited retinal dystrophies: Past, present and future. Prog Retin Eye Res. 2017; 59: 53-96. https://doi.org/10.1016/j.preteyeres.2017.03.003

[2] Zhang Q. Retinitis Pigmentosa: Progress and Perspective. Asia Pac J Ophthalmol (Phila). 2016; 5(4): 265-71. https://doi.org/10.1097/APO.0000000000000227

[3] Tsang SH, Sharma T. X-linked Retinitis Pigmentosa. Adv Exp Med Biol. 2018; 1085: 31-5. https://doi.org/10.1007/978-3-319-95046-4_8

[4] Lyraki R, Megaw R, Hurd T. Disease mechanisms of X-linked retinitis pigmentosa due to RP2 and RPGR mutations. Biochem Soc Trans. 2016; 44(5): 1235-44. https://doi.org/10.1042/BST20160148

[5] Jayasundera T, Branham KE, Othman M, Rhoades WR, Karoukis AJ, Khanna H, *et al*. RP2 phenotype and pathogenetic correlations in X-linked retinitis pigmentosa. Arch Ophthalmol. 2010; 128(7): 915-23. https://doi.org/10.1001/archophthalmol.2010.122

[6] Parmeggiani F, Barbaro V, Migliorati A, Raffa P, Nespeca P, De Nadai K, *et al*. Novel variants of RPGR in X-linked retinitis pigmentosa families and genotype-phenotype correlation. European journal of ophthalmology. 2017; 27(2): 240-8. https://doi.org/10.5301/ejo.5000879

[7] Charng J, Cideciyan AV, Jacobson SG, Sumaroka A, Schwartz SB, Swider M, *et al*. Variegated yet non-random rod and cone photoreceptor disease patterns in RPGR-ORF15-associated retinal degeneration. Human molecular genetics. 2016; 25(24): 5444-59. https://doi.org/10.1093/hmg/ddw361

[8] Raghupathy RK, Gautier P, Soares DC, Wright AF, Shu X. Evolutionary Characterization of the Retinitis Pigmentosa GTPase Regulator Gene. Investigative ophthalmology & visual science. 2015; 56(11): 6255-64. https://doi.org/10.1167/iovs.15-17726

[9] Li J, Tang J, Feng Y, Xu M, Chen R, Zou X, *et al*. Improved Diagnosis of Inherited Retinal Dystrophies by High-Fidelity PCR of ORF15 followed by Next-Generation Sequencing. J Mol Diagn. 2016; 18(6): 817-24. https://doi.org/10.1016/j.jmoldx.2016.06.007

[10] Oldak M, Ruszkowska E, Siwiec S, Pollak A, Stawinski P, Szulborski K, *et al*. ORF15 exon of the RPGR gene in retinitis pigmentosa - technically difficult, diagnostically important. Klin Oczna. 2016; 118(2): 139-43.

[11] Harkness RWt, Mittermaier AK. G-quadruplex dynamics. Biochim Biophys Acta Proteins Proteom. 2017; 1865(11 Pt B): 1544-54. https://doi.org/10.1016/j.bbapap.2017.06.012

[12] Zhang S, Wu Y, Zhang W. G-quadruplex structures and their interaction diversity with ligands. ChemMedChem. 2014; 9(5): 899-911. https://doi.org/10.1002/cmdc.201300566

[13] Kikin O, D'Antonio L, Bagga PS. QGRS Mapper: a web-based server for predicting G-quadruplexes in nucleotide sequences. Nucleic Acids Res. 2006; 34(Web Server issue):

W676-82.
https://doi.org/10.1093/nar/gkl253

[14] Mishra SK, Tawani A, Mishra A, Kumar A. G4IPDB: A database for G-quadruplex structure forming nucleic acid interacting proteins. Sci Rep. 2016; 6: 38144.
https://doi.org/10.1038/srep38144

[15] Brazda V, Kolomaznik J, Lysek J, Bartas M, Fojta M, Stastny J, *et al*. G4Hunter web application: a web server for G-quadruplex prediction. Bioinformatics. 2019; 35(18): 3493-5.
https://doi.org/10.1093/bioinformatics/btz087

[16] Qin M, Chen Z, Luo Q, Wen Y, Zhang N, Jiang H, *et al*. Two-quartet G-quadruplexes formed by DNA sequences containing four contiguous GG runs. J Phys Chem B. 2015; 119(9): 3706-13.
https://doi.org/10.1021/jp512914t

[17] Sripathi SR HW, Um JY1, Moser T, Dehnbostel S, Kindt K, Goldman J, Frost MC, Jahng WJ. Nitric oxide leads to cytoskeletal reorganization in the retinal pigment epithelium under oxidative stress. Adv Biosci Biotechnol. 2012; 3: 1167-78.
https://doi.org/10.4236/abb.2012.38143

[18] Mondal S, Bhat J, Jana J, Mukherjee M, Chatterjee S. Reverse Watson-Crick G-G base pair in G-quadruplex formation. Mol Biosyst. 2016; 12(1): 18-22.
https://doi.org/10.1039/C5MB00611B

[19] Duarte AR, Cadoni E, Ressurreicao AS, Moreira R, Paulo A. Design of Modular G-quadruplex Ligands. ChemMedChem. 2018; 13(9): 869-93.
https://doi.org/10.1002/cmdc.201700747

[20] Marchand A, Gabelica V. Folding and misfolding pathways of G-quadruplex DNA. Nucleic Acids Res. 2016; 44(22): 10999-1012.
https://doi.org/10.1093/nar/gkw970

[21] Narayan SP, Choi CH, Hao L, Calabrese CM, Auyeung E, Zhang C, *et al*. The Sequence-Specific Cellular Uptake of Spherical Nucleic Acid Nanoparticle Conjugates. Small. 2015; 11(33): 4173-82.
https://doi.org/10.1002/smll.201500027

[22] Wang SK, Wu Y, Ou TM. RNA G-Quadruplex: The New Potential Targets for Therapy. Curr Top Med Chem. 2015; 15(19): 1947-56.
https://doi.org/10.2174/1568026615666150515145733

[23] Zhang XF, Xu HM, Han L, Li NB, Luo HQ. A Thioflavin T-induced G-Quadruplex Fluorescent Biosensor for Target DNA Detection. Anal Sci. 2018; 34(2): 149-53.
https://doi.org/10.2116/analsci.34.149

[24] Wright RN, Hong DH, Perkins B. RpgrORF15 connects to the usher protein network through direct interactions with multiple whirlin isoforms. Investigative ophthalmology & visual science. 2012; 53(3): 1519-29.
https://doi.org/10.1167/iovs.11-8845

[25] Genini S, Zangerl B, Slavik J, Acland GM, Beltran WA, Aguirre GD. Transcriptional profile analysis of RPGRORF15 frameshift mutation identifies novel genes associated with retinal degeneration. Investigative ophthalmology & visual science. 2010; 51(11): 6038-50.
https://doi.org/10.1167/iovs.10-5443

Received on 17-11-2019 Accepted on 18-12-2019 Published on 27-12-2019

DOI: https://doi.org/10.12974/2309-6136.2019.07.1