GRSDB: a database of quadruplex forming G-rich sequences in alternatively processed mammalian pre-mRNA sequences

Rumen Kostadinov, Nishtha Malhotra, Manuel Viotti, Robert Shine, Lawrence D'Antonio and Paramjeet Bagga*

Bioinformatics Group, School of Theoretical and Applied Science, Ramapo College of New Jersey, Mahwah, NJ 07430, USA

ABSTRACT

Guanine-rich nucleic acids are known to form highly stable G-quadruplex structures, also known as G-quartets. Recently, there has been a tremendous amount of interest in studying G-quadruplexes owing to the realization of their biological importance. G-rich sequences (GRSs) capable of forming G-quadruplexes are found in the vicinity of polyadenylation regions and are involved in regulating 3′ end processing of mammalian pre-mRNAs. G-rich motifs are also known to play an important role in alternative, tissue-specific splicing by interacting with hnRNP H protein subfamily. Whether quadruplex structure directly plays a role in regulating RNA processing events requires further investigation. To date there has not been a comprehensive effort to study G-quadruplexes near RNA processing sites. We have applied a computational approach to map putative Quadruplex forming GRSs within the transcribed regions of a large number of alternatively processed human and mouse gene sequences that were obtained as fully annotated entries from GenBank and RefSeq. We have used the computed data to build the GRSDB database that provides a unique avenue for studying G-quadruplexes in the context of RNA processing sites. GRSDB website offers visual comparison of G-quadruplex distribution patterns among all the alternative RNA products of a gene with the help of dynamic graphics. At present, GRSDB contains data from 1310 human and mouse genes, of which 1188 are alternatively processed. It has a total of 379 223 predicted G-quadruplexes, of which 54 252 are near RNA processing sites. GRSDB is a good resource for researchers interested in investigating the functional relevance of G-quadruplexes, especially in the context of alternative RNA processing. It can be accessed at http://bioinformatics.ramapo.edu/grsdb/.

INTRODUCTION

Guanine-rich nucleic acids are known to form higher order structures. Their ability to form highly stable quadruplex structures was discovered more than four decades ago (1). The G-quadruplex structure, also known as a G-quartet, is composed of stacked G-tetrads, which are square co-planar arrays of four guanine bases each. Cyclic Hoogsteen hydrogen bonding between the four guanines within each tetrad renders a high level of stability to the quadruplex (Figure 1). Although structures with three or more G-tetrads are considered to be more stable, many nucleotide sequences are known to form quadruplexes with two G-tetrads (2,3). G-quadruplexes may be formed by repeated folding of a single nucleic acid molecule (unimolecular G-quadruplex) or by interaction of two or four strands. The former is more likely to be encountered in physiological conditions (4,5). (The present work focuses only on the unimolecular quadruplexes.) Formation of G-quadruplexes *in vivo* is facilitated by proteins (6). Some proteins are also implicated in resolving the G-quadruplex structure (7,8).

G-quadruplex sequence motifs have been reported in telomeric, promoter and other regions of mammalian genomes. Formation of a G-quadruplex in the promoter region has been associated with transcription regulation of the c-myc oncogene and is being considered as a potential target for therapeutic purposes (9,10). Owing to the realization of their biological importance, recently, there has been a tremendous amount of interest in studying G-quadruplexes. This is evident from a surge in the published literature. [for reviews see (8,11)].

Although initially most of the studies focused on G-quadruplexes in the DNA, lately there have been many efforts

*To whom correspondence should be addressed. Tel: +1 201 684 7722; Fax: +1 201 684 7637; Email: pbagga@ramapo.edu
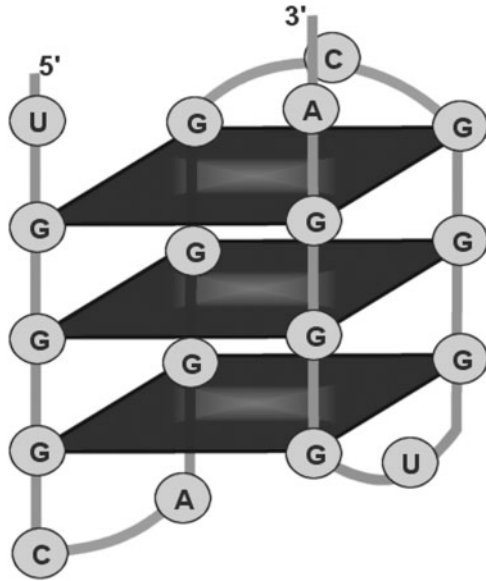
Figure 1. QGRS: 5′-GGGCAGGGCAGGUGGGA-3′. Predicted intramolecular G-quadruplex formed by a GRS.

to study G-quadruplex forming RNA (12–16). In fact, G-rich sequences capable of forming G-quadruplexes in the RNA have been implicated in a variety of important biological activities, such as mRNA turnover (6), Fragile X Mental Retardation Protein (FMRP) binding (14), translation initiation (15) as well as repression (16).

We have previously shown that a conserved auxiliary G-rich sequence (GRS) found near the polyadenylation regions can mediate efficient 3′ end processing of mammalian pre-mRNAs (17,18) by interacting with DSEF1/hnRNP H/H′ protein (19). However, hnRNP F has been shown to be a negative regulator of 3′ end processing (20). Regulated polyadenylation is an important component of differential gene expression. More than 50% of human and 32% of mouse genes are known to have alternative polyadenylation (21). An interplay among GRS-binding proteins, hnRNP H/H′ and F, helps in regulating alternative polyadenylation of immunoglobulin pre-mRNA (20) which, combined with alternative splicing, plays an important role in mouse B lymphocyte development (22).

In addition to differential gene expression, alternative splicing affects disease processes (23) and is a major source of protein diversity. More than two-thirds of human genes are thought to undergo alternative splicing (24). Members of the hnRNP H protein subfamily, that bind G-rich motifs, are known to be involved in alternative, tissue-specific, regulated splicing events (25–27). GRS motifs that are present near splice sites act as splicing regulators by interacting with hnRNP H (28). For example, binding of hnRNP H and F to G-rich tracts near 5′ splice site favors production of alternative pro-apoptotic Bcl-x_s product (29). The regulatory G-rich motifs may be capable of forming quadruplex structures. Whether quadruplex structure directly plays a role in regulating RNA processing events requires investigation.

The majority of the mammalian poly(A) region GRS sequences that we had surveyed in our previous studies (18,19) are capable of forming unimolecular G-quadruplexes. Our preliminary analysis of ~100 alternatively

processed human transcripts has also revealed the presence of quadruplex forming sequences near alternative splice sites (30). However, a more detailed investigation into the distribution of G-quadruplex sequences near RNA processing sites requires a systematic large-scale analysis of mammalian genes. Although, there have been two recent surveys of quadruplexes in the human genome (31,32), to date there has not been a comprehensive effort to study G-quadruplexes near RNA processing sites.

We have used a computational approach (30) to map putative G-quadruplex forming sequences within the transcribed regions of a large number of alternatively processed human and mouse genes. The fully annotated genomic nucleotide sequences are obtained from NCBI-based GenBank and RefSeq for computational analysis. Based on our analysis of alternatively spliced and alternatively polyadenylated human and mouse genes, we have built the GRSDB database. GRSDB provides a unique avenue for studying G-quadruplex forming sequences in the context of RNA processing sites. In addition to providing data on composition and locations of mapped quadruplexes relative to the processing sites in the pre-mRNA sequence, GRSDB offers simultaneous visual comparison of G-quadruplex distribution patterns among all the alternative RNA products of individual genes with the help of dynamically generated graphics.

Researchers interested in investigating the functional relevance of G-quadruplex structure, in particular its role in regulating the gene expression by alternative processing, will find GRSDB to be of great value. It allows a comprehensive large-scale analysis as well as detailed studies in individual genes. GRSDB is also a good resource for performing large-scale analysis of G-quadruplex sequence composition, including study of loops, in the transcribed regions.

METHODS

Quadruplex forming GRS

The basic unit of study in GRSDB is the putative G-quadruplex that we have called QGRS (**Q**uadruplex forming **GRS**). These sequences follow the motif $G_xN_aG_xN_bG_xN_cG_x$. Here G_x refers to the group of guanines (which we will refer to as a G-group) that form a complex of x stacked G-tetrads. In the individual gene entries stored in GRSDB, x is generally 2, 3 or 4. The intervening arbitrary bases, N_a, N_b and N_c, are called gaps or loops.

Two sequences are said to be overlapping if their positions in the nucleotide sequence do overlap. The default action of GRSDB is to display non-overlapping sequences, but the user can display all QGRS.

Structure of GRSDB

GRSDB is a relational database built using MySQL. The GRSDB website can be accessed at http://bioinformatics.ramapo.edu/grsdb/. This database primarily stores information about putative G-quadruplex sequences (QGRS) for genes that are alternatively processed (either alternatively spliced or alternatively polyadenylated). GRSDB is structured to facilitate queries about alternatively processed genes and to display information on the G-quadruplex sequences contained in the transcribed regions of the gene and their locations relative to

RNA processing sites. Table 1 shows the types of objects found in the database.

GRSDB is populated using an auxiliary program, QGRS-Mapper, that is based on previously published methods (30) and was developed using BioPerl. Once appropriate genes have been identified, this program links to GenBank or RefSeq, downloads the corresponding genomic nucleotide sequence entry of the gene, and parses the entry for product, intron, exon, poly(A) and related information. The program then processes the nucleotide sequence to find all QGRS and map their location within the gene and their distance from relevant RNA processing sites.

QGRS scoring

A scoring method is applied to each QGRS. The computed score, called a G-score (30), is formulated to reward sequences with smaller, more even gaps between the G-groups in addition to larger G-group size, thereby favoring the arrangement that

is more likely to form a unimolecular complex. This choice of scoring system is in agreement with the existing literature on loop structures in G-quadruplexes (31–35). In particular, the data gathered in this research points to loop sizes tending to be small and preferentially equal or nearly equal.

Interfaces

The data flow for GRSDB is summarized in Table 2. After the gene information is downloaded from NCBI, parsed, processed for QGRS, and scored, it is then uploaded into GRSDB. At this point the database is ready for user queries. There are three different interfaces provided for viewing database contents: the gene view, the data view and the graphical view.

Table 1. Structure of GRSDB

Data type	Comments
Gene	Includes gene name, gene location, whether or not gene is alternatively spliced or alternatively polyadenylated
Product	Includes RNA product name, and numbers of introns, exons and poly(A) signals
Intron/exon map	Maps start and stop positions of introns and exons
Poly(A) map	Maps start and stop position of poly(A) signal
QGRS	Includes actual sequences, their position and G-scores

Table 2. Data flow of GRSDB

Data sources	GenBank RefSeq
Data filtering	Selection for mouse and human genomic entries that exhibit alternative processing information
Parsing	Gene entries are parsed for gene information, product information, exon/intron/Poly(A) information
QGRS detection	QGRS-Mapper analyzes genes for occurrences of QGRS
QGRS scoring	Scoring system based on loop size is applied to all QGRS
Data interfaces	Gene/data/graphical views of QGRS

Gene Information			
NCBI Accession number	AF258674	Organism	Homo sapiens
Gene name	MUCDHL	Chromosome number	11
GI	10334768	Alternatively spliced	YES
Locus	AF258674	Alternatively polyadenylated	YES
Total number of products	2	Date of entry	2005-04-20
Gene range	995 - 10285	Comment	Viotti

The gene has 2 known alternatively spliced RNA products

Product 1 : MUCDHL-ALT			
Number of introns	13	Total # of QGRS	73
Number of exons	14	QGRS mapped near RNA processing sites	34
Number of PolyA signals	2	PolyA signal # 1 position	9598
PolyA signal # 2 position	10264		

Map non-overlapping QGRS in this product

Data View   Graphic View   NOTE: Version 1.4.2_05 or higher version of Java needs to be enabled for Graphic View. Preferred browser is IE.

Figure 2. MUCDHL—*Gene View*. The first table in this view gives basic information about the gene being analyzed, including the alternatively spliced or polydenylated status of the gene. The second table provides information about the first alternatively spliced RNA product of the gene, including numbers of exons/introns/poly(A) signals. Also displayed are the total number of QGRS in the product and near RNA processing sites (i.e. within 120 nt of a site). Located below each product table are buttons allowing the user to view the QGRS information for that product in two ways; the data view and the graphic view, shown in Figures 3 and 4. Not shown are the second product table and the button placed at the bottom of the page allowing the user to analyze all RNA products simultaneously. To view the entire screen shot, see the Supplementary Data.

Database users are given a variety of options in formulating a query, including searching for genes that are alternatively spliced or alternatively polyadenylated. Once a query has been entered, a table is displayed of all genes satisfying the query. Information for individual genes is displayed in a table as shown in Figure 2 for the particular gene MUCDHL. This is what we call the gene view. One can see that MUCDHL is both alternatively spliced and alternatively polyadenylated.

At this point the user can choose to analyze one of the products or all products simultaneously. There are two types of analysis possible, the data view or graphical view. Figure 3 represents the data view analysis for Product 1 of MUCDHL, showing all non-overlapping QGRS. The table shows the location of each QGRS, its distance from the nearest splice site, and its G-score.

Alternatively, the user can select the graphical view of any product (or again, all products together), which is shown in Figure 4. A visual model of the product is displayed, showing the location of exons, introns and untranscribed regions. Further, the location of each QGRS is indicated by a vertical

Map of non-overlapping QGRS:				
QGRS in Exons				
Exon number	QGRS Position	QGRS Distance from 3' or 5'ss	QGRS Sequence	G-Score
1	1000	3:5	GGGGCTCGGGGG	0.83333
2	1999	3:95	GGCAGTGGGCTGGACCAGCCGG	0.77778
2	2072	3:168	GGCGTCTGGCTGGGGCGAAGG	0.94444
2	2170	5:82	GGGGTCTTGGGCCCTGCTGTGG	0.50000
2	2230	5:22	GGGGACCATGGCCCAGG	0.72222
3	2420	3:79	GGAGGGCCAGGAGG	0.88889
5	5103	5:9	GGTGGAGGAGG	1.00000
7	5516	3:49	GGCTGGACCGGCCCCTGG	0.83333
7	5564	5:14	GGCTGCTGGTGCGGGTGAGCAGG	0.88889
8	5843	5:0	GGTAATGGGGAGG	0.77778
10	6857	5:1	GGGTNGGGCGGNGGGTCAGG	1.00000
11	7043	3:47	GGCTGTGGCTGCGGCCGG	0.88889
11	7095	5:98	GGCACCGTGGCGCGTGGCGCTGG	0.88889
11	7165	5:28	GGATCCAGGCTCAGGACCCGG	0.94444
12	7332	3:49	GGATGGAGGGAGAGG	0.94444
12	7396	5:0	GGTGCGGGGCAGGGCAGG	0.94444
14	9941	3:69	GGGGAGGCTGAGG	0.77778
14	9973	3:101	GGCGACGCACTTGGGAGCCGGTGG	0.50000
14	10086	5:199	GGACTGAGGCGCGCTGGGGGTGG	0.88889

Figure 3. MUCDHL—*Data View*. A table of the mapping data for the non-overlapping QGRS in the product (refer to the text for explanation of non-overlapping) is shown. The table provides the location of QGRS in exons, the distance from 3′ and 5′ splice sites, the actual sequence, and its G-score (refer to the text for explanation). Not shown are the tables for QGRS mapping data in introns and in poly(A) regions. To view the entire screen shot, see Supplementary Data.
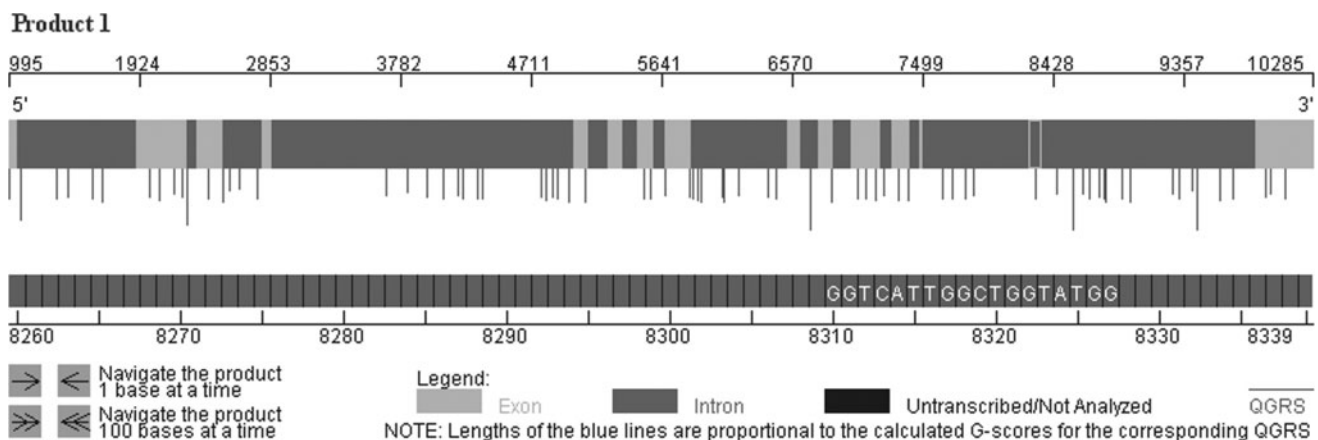

Figure 4. MUCDHL gene—*Graphic View*. A visual representation for RNA Product 1 of this gene. The upper graph shows the location of exons/introns in the product, along with a scale to locate their positions. The QGRS in the product are indicated by the vertical bars, whose length is proportional to the G-scores of the QGRS. The lower graph represents a zoom-in of the RNA product 1 displaying the QGRS at position 8310. The arrows at the bottom left are used to navigate the RNA product with the interactive-zoom tool. It is possible to visually compare both the alternatively spliced products, for example, to identify differential association of QGRS with alternative sites. See Supplementary Data for product comparison.

line. The length of the line is proportional to the G-score for the sequence.

CONCLUSIONS

GRSDB provides curated information on composition and distribution of putative QGRSs in the transcribed regions of alternatively processed human and mouse genes. The data are based on the analysis of fully annotated GenBank/RefSeq human and mouse genomic nucleotide entries that exhibit alternative processing information. Although the NCBI databases contain a large number of mRNA sequence records, at present the number of genomic entries that will provide information needed for our studies is limited.

At present, our database contains information obtained from 1310 human and mouse genes, of which 1188 are alternatively processed. A total of 30 584 introns and 33 816 exons were analyzed, containing a total of 3231 RNA products. These products taken together contain a total of 379 223 putative G-quadruplexes, of which 54 252 are near RNA processing sites [within 120 nt of a splice site or a poly(A) signal]. Note that while GRSDB currently contains data only on human and mouse genes, our computational tools and the database are designed to include other organisms as well.

GRSDB is continuously being updated with new data entries. The database is structured to facilitate a wide variety of queries and to map G-quadruplex sequences relative to the RNA processing sites in both data and graphic formats. The user friendly interface allows comparisons of all the alternative RNA products of individual genes on the same screen. GRSDB is a good resource for researchers interested in investigating the functional relevance of G-quadruplexes, especially in the context of alternative RNA processing.

We are using the database to conduct detailed bioinformatics studies on the distribution patterns of QGRS near RNA processing sites. In particular, we are investigating whether there is a correlation between the distribution pattern of QGRS and alternative processing. Our group is also studying the loop composition of these sequences.

SUPPLEMENTARY DATA

Supplementary Data are available at NAR online.

ACKNOWLEDGEMENTS

REFERENCES

1. Gellert,M., Lipsett,M.N. and Davies,D.R. (1962) Helix formation by guanylic acid. *Proc. Natl Acad. Sci. USA*, **48**, 2013–2018.
2. Zharudnaya,M.I., Kolomiets,I.M., Potyahaylo,A.L. and Hovorun,D.M. (2003) Downstream elements of mammalian pre-mRNA polyadenylation signals: primary, secondary and higher-order structures. *Nucleic Acids Res.*, **31**, 1375–1386.
3. Kankia,B.I., Barrany,G. and Musier-Forsyth,K. (2005) Unfolding of DNA quadruplexes induced by HIV-1 nucleocapsid protein. *Nucleic Acids Res.*, **33**, 4395–4403.
4. Schaffitzel,C., Berer,I., Postberg,J., Hanes,J., Lipps,H.J. and Plückthun,A. (2001) *In vitro* generated antibodies specific for telomeric guanine-quadruplex DNA react with *Stylonychia lemnae* macronuclei. *Proc. Natl Acad. Sci. USA*, **98**, 8572–8577.
5. Halder,K. and Chowdhury,S. (2005) Kinetic resolution of bimolecular hybridization versus intramolecular folding in nucleic acids by surface plasmon resonance: application to G-quadruplex/duplex competition in human c-myc promoter. *Nucleic Acids Res.*, **33**, 4466–4474.
6. Paeschke,K., Simonsson,T., Postberg,J., Rhodes,D. and Lipps,H.J. (2005) Telomere end-binding proteins control the formation of G-quadruplex DNA structures *in vivo*. *Nature Struct. Mol. Biol.*, **12**, 847–854.
7. Zaug,A.J., Podell,E.R. and Cech,T.R. (2005) Human POT1 disrupts telomeric G-quadruplexes allowing telomerase extension *in vitro*. *Proc. Natl Acad. Sci. USA*, **102**, 10864–10869.
8. Simonsson,T. (2001) G-quadruplex DNA structures—variations on a theme. *Biol. Chem.*, **382**, 621–628.
9. Simonsson,T., Pecinka,P. and Kubista,M. (1998) DNA tetraplex formation in the control region of c-myc. *Nucleic Acids Res.*, **26**, 1167–1172.
10. Phan,A.T., Modi,Y.S. and Patel,D.J. (2004) Propeller-type parallel-stranded G-quadruplexes in the human c-myc promoter. *J. Am. Chem. Soc.*, **126**, 8710–8716.
11. Davis,J.T. (2004) G-quartets 40 years later: from 5′-GMP to molecular biology and supramolecular chemistry. *Angew. Chem. Int. Ed. Engl.*, **43**, 668–698.
12. Liu,H., Matsugami,A., Katahira,M. and Uesugi,S. (2002) A dimeric RNA quadruplex architecture comprised of two G:G(:A):G:G(:A) hexads, G:G:G:G tetrads and UUUU loops. *J. Mol. Biol.*, **322**, 955–970.
13. Bashkirov,V.I., Scherthan,H., Solinger,J.A., Buerstedde,J.M. and Heyer,W.D. (1997) A mouse cytoplasmic exoribonuclease (mXRN1p) with preference for G4 tetraplex substrates. *J. Cell. Biol.*, **136**, 761–773.
14. Darnell,J.C., Jensen,K.B., Jin,P., Brown,V., Warren,S.T. and Darnell,R.B. (2001) Fragile X mental retardation protein targets G quartet mRNAs important for neuronal function. *Cell*, **107**, 489–499.
15. Bonnal,S., Schaeffer,C., Creancier,L., Clamens,S., Moine,H., Prats,A.-C. and Vagner,S. (2003) A single internal ribosome entry site containing a G quartet RNA structure drives fibroblast growth factor 2 gene expression at four alternative translation initiation codons. *J. Biol. Chem.*, **278**, 39330–39336.
16. Oliver,A., Bogdarina,I., Schroeder,E., Taylor,I.A. and Kneale,G.G. (2000) Preferential binding of fd gene5 protein to tetraplex nucleic acid structures. *J. Mol. Biol.*, **301**, 575–584.
17. Bagga,P.S., Ford,L.P., Chen,F. and Wilusz,J. (1995) The G-rich auxiliary downstream element has distinct sequence and position requirements and mediates efficient 3′ end pre-mRNA processing through a *trans*-acting factor. *Nucleic Acids Res.*, **23**, 1625–1631.
18. Bagga,P.S., Arhin,G.K. and Wilusz,J. (1998) DSEF-1 is a member of the hnRNP H family of RNA binding proteins and stimulates pre-mRNA cleavage and polyadenylation. *Nucleic Acids Res.*, **26**, 5343–5350.
19. Arhin,G.K., Boots,M., Bagga,P.S., Milcarek,C. and Wilusz,J. (2002) Downstream sequence elements with different affinities for the hnRNP H/H′ protein influence the processing efficiency of mammalian polyadenylation signals. *Nucleic Acids Res.*, **30**, 1842–1850.
20. Veraldi,K.L., Arhin,G.K., Martincic,K., Chung-Ganster,L.-H., Wilusz,J. and Milcarek,C. (2001) hnRNP F influences binding of a 64 kDa subunit of cleavage stimulation factor to mRNA precursors in mouse B cells. *Mol. Cell. Biol.*, **21**, 1228–1238.
21. Tian,B., Hu,J. and Lutz,C. (2005) A large-scale analysis pf mRNA polyadenylation of human and mouse genes. *Nucleic Acids Res.*, **33**, 201–212.
22. Bruce,S.R., Dingle,R.W.C. and Peterson,M.L. (2003) B-cell and plasma-cell splicing differences: a potential role in regulated immunoglobulin RNA processing. *RNA*, **9**, 1264–1273.

23. Faustino,N.A. and Cooper,T.A. (2003) Pre-mRNA splicing and human disease. *Genes Dev.*, **17**, 419–437.

24. Johnson,J.M., Castle,J., Garrett-Engele,P., Kan,Z., Loerch,P.M., Armour,C.D., Santos,R., Schadt,E.E., Stoughton,R. and Shoemaker,D.D. (2003) Genome-wide survey of human alternative pre-mRNA splicing with exon junction microarrays. *Science*, **302**, 2141–2144.

25. Min,H., Chan,R.C. and Black,D.L. (1995) The generally expressed hnRNP F is involved in a neural-specific pre-mRNA splicing event. *Genes Dev.*, **9**, 2659–2671.

26. Chou,M.-Y., Rooke,N., Turck,C.W. and Black,D.L. (1999) hnRNP H is a component of a splicing enhancer complex that activates a c-src alternative exon in neuronal cells. *Mol. Cell. Biol.*, **19**, 69–77.

27. Caputi,M. and Zahler,A.M. (2002) SR proteins and hnRNP H regulate the splicing of the HIV-1 tev-specific exon 6-D. *EMBO J.*, **21**, 845–855.

28. Han,K., Yeo,G., An,P., Burge,C. and Grabowski,P. (2005) A combinatorial code for splicing silencing: UAGG and GGG motifs. *PLoS Biol.*, **3**, 0843–0860.

29. Garnea,D., Revil,T., Fisette,J.-F. and Chabot,B. (2005) Heterogeneous nuclear ribonucleoprotein F/H proteins modulate the alternative splicing of the apoptotic mediator Bcl-x. *J. Biol. Chem.*, **24**, 22641–22650.

30. D'Antonio,L. and Bagga,P.S. (2004) Computational methods for predicting intramolecular G-quadruplexes in nucleotide sequences. *Proceedings of the IEEE Computational Systems Bioinformatics Conference (CSB'04)*, August 16–19, Stanford, CA, 561–562.

31. Huppert,J.L. and Balasubramanian,S. (2005) Prevalence of quadruplexes in the human genome. *Nucleic Acids Res.*, **33**, 2908–2916.

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

33. Risitano,A. and Fox,K. (2004) Influence of loop size on the stability of intramolecular DNA quadruplexes. *Nucleic Acids Res.*, **32**, 2598–2606.

34. Črnugelj,M., Šket,P. and Plavec,J. (2003) Small change in a G-rich sequence, a dramatic change in topology: new dimeric G-quadruplex folding motif with unique loop orientations. *J. Am. Chem. Soc.*, **125**, 7866–7871.

35. Hazel,P., Huppert,J., Balasubramanian,S. and Neidle,S. (2004) Loop-length-dependent folding of G-quadruplexes. *J. Am. Chem. Soc.*, **126**, 16405–16415.