Analysis of G-Quadruplex Formation in mRNA Transcripts of Phospholemman/*FXYD1*

by

Hansraj DHAYAN

Submitted in partial fulfilment of the requirements of the University of Hertfordshire for the degree of Master of Science by Research

Title of sponsoring department: School of Life Sciences
Name of company/institution: University of Hertfordshire
Date: 27th November 2013

UNIVERSITY OF HERTFORDSHIRE

Research Degrees Board

Name of Candidate: **HANSRAJ DHAYAN**

Award of the degree of: **MSc by Research in Molecular Biology, Biophysics and Biochemistry**

DECLARATION REGARDING FINAL SUBMISSION

My submission for examination was temporarily bound.

I confirm that the contents of my final, approved submission are identical with the version submitted for examination, except where amendments have been made to meet the requirements of the examiners.

Signed:

Date: 10th June 2014

ABSTRACT

G-quadruplexes are higher-order nucleic acid structures formed by tetrads of guanine bases (G-tetrads) through non-canonical base interactions. Two G-tetrads are stabilised by a potassium-ion sandwiched between the tetrads. It has emerged from recent studies that G-quadruplexes occur widely throughout the human genome and have significant biological roles. In this study the *FXYD1* pre-mRNA encoding the protein Phospholemman (PLM) is investigated. PLM is highly expressed in cardiomyocytes and forms a third subunit of the Na^+/K^+ pump (NKA). PLM is a major phosphorylation target and thus regulates NKA activity. *FXYD1* pre-mRNA was investigated for its ability to form G-quadruplexes. By computational analysis, it was found that *FXYD1* can fold into G-quadruplex and multiple sequence alignment of ortholog *FXYD1* sequences indicated that G-quadruplex-forming potential is conserved in evolution, hinting at a potential regulatory mechanism of *FXYD1* expression. Comparative analysis confirmed that *FXYD1*-009, a variant of *FXYD1*, is a product of alternative splicing of *FXYD1*'s pre-mRNA. G-quadruplex formation in human and bovine *FXYD1*-derived oligonucleotides was detected experimentally by non-denaturing poly acrylamide gel electrophoresis that showed an increased mobility rate of G-quadruplexes in contrast to controls. Further analysis by fluorescence emission spectroscopy confirmed G-quadruplex formation in the human and bovine *FXYD1*-oligonucleotides that was triggered by the presence of K^+ ions. The results provided clear evidence of G-quadruplex formation *in vitro* and together with evolutionary conservation point to potential role in regulating expression of *FXYD1* possibly through alternative splicing and thus regulate indirectly the

activity of Na$^+$/K$^+$-ATPase. Further *in-vivo* works should address whether alternative splicing of *FXYD1* to *FXYD1*-009 is associated with G-quadruplex formation.

TABLE OF CONTENTS

List of Tables

List of Figures

List of abbreviations

Acronym	Definition
PLM	Phospholemman
GQS	G-quadruplex forming sequences
+VE	Positive control
-VE_A	Negative control A
-VE_B	Negative control B
MFE	Minimum Free Energy
QGRS	Quadruplex forming G-Rich Sequences
MSA	Multiple Sequence Alignment
TrisOAc	Tris Acetate buffer
KCl	Potassium Chloride
KOAc	Potassium acetate
NKA	Sodium Potassium pump/ Na^+/K^+-ATPase
PAGE	Poly Acrylamide Gel Electrophoresis
DNA	Deoxyribonucleic Acid
RNA	Ribonucleic Acid
mRNA	messenger Ribonucleic acid
UTR	Untranslated region
ETDA	Ethylenediaminetetraacetic acid

1. INTRODUCTION

1.1 G-quadruplex overview

If there is something in particular that intrigues scientists about Guanine rich (G-rich) nucleic acid sequences, it is their ability to form higher order secondary structures called G-quadruplexes (Phong Lan Thao, Mergny, & Alberti, 2011; Stegle, Payet, Mergny, MacKay, & Huppert, 2009; Tluckova et al., 2013; Yuan et al., 2013; Zhang, Liu, Zheng, Hao, & Tan, 2013). G-rich nucleic acid sequences form G-quadruplexes when in the presence of cations, which help stabilizing the structures. G-tetrads, being the tetrahedral arrangement of four guanines residues linked via hydrogen bonds, are the building blocks of G-quadruplexes (Lech, Heddi, & Anh Tuan, 2013; Wu & Brosh, 2010). The role of cations is to stabilize the G-tetrads by sitting in their core (Figure 1A). Stacks of G-tetrads are called G-quadruplexes and a minimum of two G-tetrads are required to form a stable G-quadruplex. Structural polymorphisms in G-quadruplexes have been previously reported and some of the variants are shown in Figure 1B (Musetti, Krapcho, Palumbo, & Sissi, 2013; Palacky, Vorlickova, Kejnovska, & Mojzes, 2013; Xu, Xu, Shang, Feng, & Zhou, 2012).
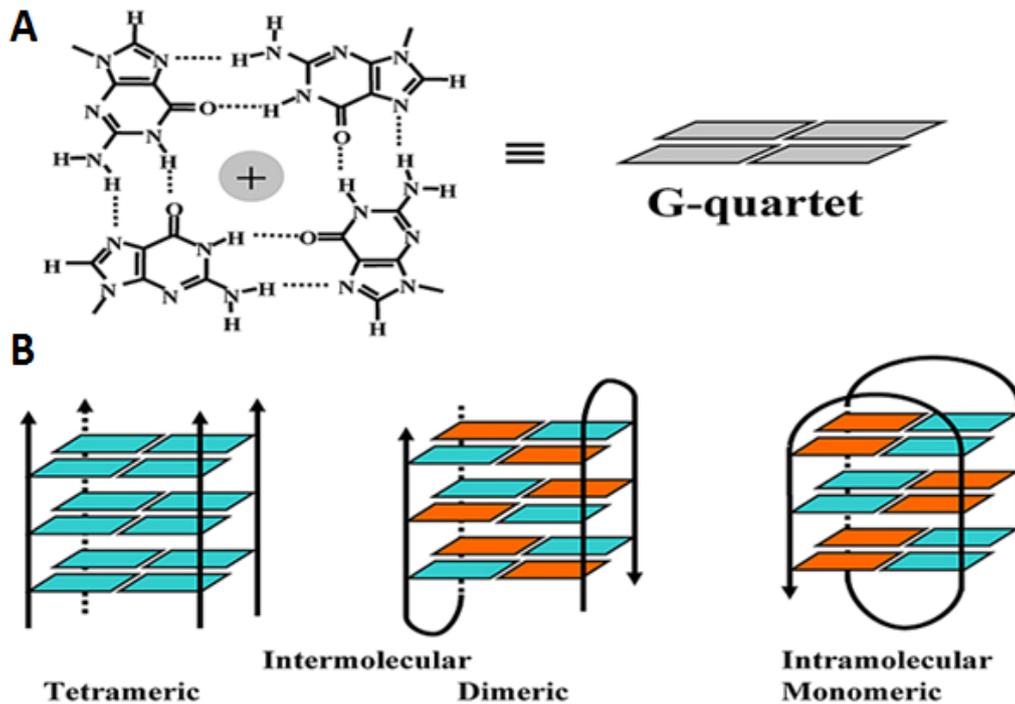
Figure 1: (A) Schematic representation of a G-quartet. Four guanines are linked together via hydrogen bonds (dotted lines) and the quartet is stabilised by a cation, K^+. **(B)** G-quartets stack to form G-quadruplexes. Three different types of G-quadruplex are shown here. Intermolecular tetrameric G-quadruplex involves 4 separate strands of nucleic acid with the participation of one guanine residue from each strand in the G-quartets. Dimeric G-quadruplex involves participation of two separate strands with two guanines (*Anti* and *Syn*) from each strand participating in the G-quartets. Intramolecular monomeric G-quadruplex involves only one strand. *Anti*-guanines are coloured cyan and *Syn* guanines are coloured orange. Adapted from Moon & Jarstfer (2007).

1.2 G-quadruplex folding motif and prediction tools

In order for any nucleic acid to fold into G-quadruplex, the sequence of the latter should be rich in guanine residues and the arrangement of guanines within the nucleic acid should comply with particular motifs. Many algorithms have been developed to identify nucleic acid sequences rich in guanines with the appropriate motifs, and allow easy prediction of *intra*molecular G-quadruplex formation. Algorithms such as Quadruplex forming G-Rich sequences (QGRS) mapper (Kikin O, D'Antonio & Bagga, 2006), Quadfinder (Scaria, Hariharan, Arora &Maiti, 2006), QuadPredict (Wong, Stegle, Rodgers & Huppert, 2010), G-Rich sequence Database (GRSD), G-Rich Sequences UTR DataBase (GRS UTRdb), non-B DNA Motif Search Tool (nBMST), Quadbase and others are readily available on the internet (Kostadinov, Malhotra, Viotti, Shine, D'Antonio & Bagga, 2006). The most common folding motif was devised by Kikin *et al*. (2006) and is as follows: $G_xN_{y1}G_xN_{y2}G_xN_{y3}G_x$. G stands for guanine and N stands for any other nucleotide residue, subscripts denote the number of occurrences of these nucleotides. According to the folding rule, x should be at least two as a minimum of two quartets is required to stack on top of each other to form a G-quadruplex. N is representative of the other bases involved in the loops of the G-quadruplex, N can be any base including guanine. Y1, Y2 & Y3 is the number of the different residues that participate in the three different loops, and can vary.
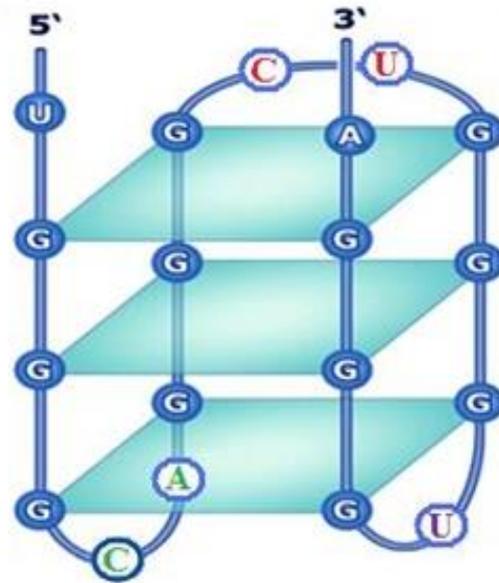
Figure 2: Intramolecular G-quadruplex formed by a RNA molecule with the sequence: 5'-UGGGCAGGGCUGGGUGGGA-3'. This particular intramolecular RNA (5'-UGGGCAGGGCUGGGUGGGA-3') G-quadruplex corresponds to the motif $G_3N_2G_3N_2G_xN_1G_x$. Note that the first base 5'-U and last base A-3' did not participate in the G-quadruplex structure. (Adapted and edited from GRS UTRdb Database, 2007)

Lorenz *et al.* (2011) stated that most of the putative G-quadruplex forming sequences in RNA are more likely to form secondary structures based on conventional base pairing rather than G-quadruplexes. The Vienna RNA package developed by Lorenz *et al.* (2013) provides a suitable platform for detecting secondary structures in sequences based on thermodynamic parameters and properties; it also allows users to predict the formation of G-quadruplexes alongside other possible competing secondary structures. The three main types of computational structural predictions are based on (i) Zuker & Stiegler's (1981) Minimum Free Enegery (MFE) algorithm, which will predict a single structure for a particular sequence based on its MFE requirement (ii) McCaskill's (1990) Partition Function algorithm, providing

statistical insights about the base parining probabilties in RNA emsembles allowing the prediction of more than one secondary structures within the same species of RNA (iii) Suboptimal Folding algorithm (Wuchty, Fontana, Hofacker, & Schuster, 1999) that computes structures within a given range of optimal energy, hence allowing users to screen for competing secondary structures with respect to G-quadruplex in RNA molecules. All of the three prediction methods are implemented in the Vienna RNA package, mostly independent of each other such as predicting MFE structure of a particular sequence or sometimes combined when for instance predicting structures in a particular sequence over a range of optimal energy. The webserver of the Vienna RNA package provides a suitable platform for users to predict structures in desired RNA sequences and is available at http://rna.tbi.univie.ac.at/.

1.3 Existence and significance of G-quadruplexes

There have been reports in the past about the existence of G-quadruplexes occurring *in-vitro* (Yuan, Tian, Chen, Yan, Xing, Zhang, Zhai, Xu, Wang, Weng, Yuan, Feng & Zhou, 2013; Biffi, Tannahill, McCafferty & Balasubramanian, 2013; Xu, Suzuki, Ito & Komiyama, 2010). *In-sillico* analysis of the human genome has revealed many potential sequences that can fold into G-quadruplex, with quite a large fraction falling into gene promoter regions of DNA and UTR, exon, intron and exon-intron boundary regions of pre-mRNAs (Beaudoin *et al.*, 2010; Johnson *et al.,* 2010; Onyshchenko *et al.,* 2009). The biological significance of these G-quadruplexes has been discussed in literatures. Controversies have revolved around G-quadruplex as being a potential down-regulator of gene expression. These structures may have a specific role like the hairpin-stem loops that form within palindromic sequences and aid terminating translation in prokaryotes (Wilson *et al.*, 1995). Many roles have been associated to G-quadruplexes. Some of the proposed functions associated with G-quadruplex formation include: G-quadruplexes can up-regulate genes by keeping promoter or upstream regions of genes in a more open structure, therefore enabling easy access for transcriptional factors to bind (Du, Zhao, & Li, 2008). With the recent advances in molecular techniques and latest technological assets, G-quadruplexes' existence within cells and elucidation of the roles of some G-quadruplexes have been characterised. The formation of a G-quadruplex structure within a promoter region has been reported to sterically hinder access to negative regulators and enhance gene expression following the work led by Gu, Lin, Xu, Yu, Du, Zhang, Yuan & Gao in 2012. Their work led to the proposition that the formation of a G-quadruplex in the rat relaxin-1 (*RLN1*) gene promoter restricts access to the transcriptional activator STAT3. STAT3 is known to negatively regulate the expression of

relaxin-1 and Gu *et al.,* (2012) hypothesized that G-quadruplex formation in the RLN1 promoter region led to enhanced expression of relaxin-1. Down regulation of genes has also been reported to be associated with G-quadruplexes, for example in case of the oncogene *c-myc* (Ou *et al.*, 2007). Ou and colleagues (2007) reported that the stabilisation of a G-quadruplex within the *c-myc* gene promoter lead to its down regulation. G-quadruplexes are largely unexploited in the cancer therapeutics field. Reports have confirmed the fact that telomeric ends of *Homo sapiens* chromosomes are guanine rich and have the potential to fold into G-quadruplexes (Zhu, Xiao & Liang, 2013; Long, Parks, Bagshaw & Stone, 2013). The survival of cancer cells depends on the enzymatic action of telomerase on telomeric ends of chromosomes (Shay, Zhou, Hiyama & Wright, 2001). Telomerase is known to elongate ends of telomeres and helping cancer cells to survive. Stabilized G-quadruplexes in telomeres will inhibit telomerase and eventually stops telomeric elongation that will prove difficult for the cancer cells to survive (Li, Xiang, Zhang & Tang, 2012). It was the report published by Siddiqui-Jain, Grand, Bearss & Hurley in 2002 that drew major attention to considering G-quadruplexes as potential target for anti-canceral drugs. The former group successfully stabilised a G-quadruplex entity upstream the promoter of the pro-oncogene *c-myc*, using the ligand porphyrin TMPyP4. The stable G-quadruplex suppressed the expression of *c-myc* significantly, and their work was the first direct evidence of ligand mediated G-quadruplex stabilisation in the *c-myc* promoter region.

1.4 Prevalence of G-quadruplex in RNA

RNA G-quadruplexes have been reported in the past and the high occurrence of G-quadruplex in UTR regions of RNA has lead to hypothesizing on their role as translational regulators (Huppert *et al.*, 2008, Bugaut & Balasubramanian, 2012). Huppert, Bugaut, Kumari & Balasubramanian (2008), proposed that G-quadruplex in 5'-UTRs of RNA can down regulate translation by caging the 5'-cap end or by disrupting small ribosome subunits (Figure 3**A**). Alternatively, Huppert *et al.* (2008) proposed that G-quadruplexes in the 3'-UTR region of template DNA can effectively allow mRNA processing, by supporting the cleavage of pre-mRNA at the poly adenylation site (Figure 3**B**).
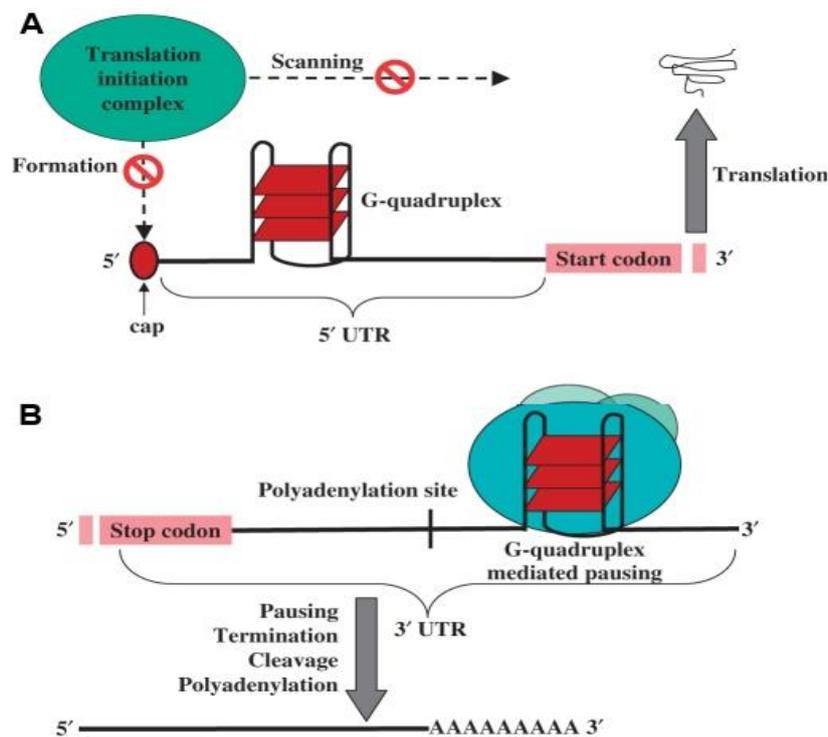
Figure 3: Proposed roles of G-quadruplexes associated with UTR regions of RNA. **(A)** G-quadruplex formation within the 5'-UTR region of an mRNA molecule. Cap-dependent initiation of translation is compromised in this instance, by the presence of the G-quadruplex that restricts the initiation complex to scan along the mRNA for the start codon.

Translation is prevented in this instance. **(B)** Formation of a G-quadruplex in the 3' region of the template DNA strand, just after the polyadenylation site. The presence of the G-quadruplex pauses RNA polymerase complex and allows effective termination of transcription. Adapted and Edited from Huppert *et al.*, (2008)

As previously stated, some G-quadruplexes and their *in-vivo* roles have been characterised in the past. Kumari, Bugaut, Huppert & Balasubramanian (2007) reported that G-quadruplex within the 5'-UTR of the *NRAS* oncogene reduces expression of the latter. Another group of researchers proposed that G-quadruplexes in RNA leads to alternative splicing. Marcel *et al*. (2011) reported that the formation of a G-quadruplex in the pre-mRNA of tumour suppressor protein, P53, leads to alternative splicing. Eventually this has an impact on the type of P53 that is formed. The usual form of p53 is FSP53, which is a fully processed mRNA, while P5312 is the alternative form that is derived from a partially unspliced pre-mRNA. The P5312 form retains its intron two, which is not spliced. The finding from Marcel & colleagues' work led to the suggestion that G-quadruplex formation in intron three of the pre-mRNA has an impact on the splicing frequency of intron two. The more G-quadruplex that was stabilized in intron three, the more FSP53 was made. Another group of researchers have also demonstrated that G-quadruplex formation led to alternative splicing patterns in hTERT intron 6, which caused down regulation of the activity of telomerase in A549 carcinoma cells (Gomez *et al., 2004*). Bugaut *et al*.(2012) reported that a significantly large number of clinically important genes have been analysed and shown to have sequences that can form G-quadruplexes, especially post transcriptional. Previous reports supported the fact that conformational changes within mRNA molecules have the potential of regulating protein formation (Gray & Hentze, 1994; Van der velden & Thomas, 1999). Van der velden

et al. (1999) reported that the 5'-UTR of most mRNA is an important site where ribosomes will bind to initiate protein synthesis and any structural changes, G-quadruplexes in this instance, will affect this process. Many of the genes proposed by Bugaut *et al*.,(2012) fall into the oncogene family and the study and elucidation of G-quadruplexes in these genes is of clinical importance.

1.5 *FXYD1/phospholemman*

One clinically important gene, highly expressed in cardiomyocytes is the *FXYD1* gene. *FXYD1* codes for the protein phospholemman (PLM) and is part of the FXYD family, which are involved mainly in regulating the Na$^+$/K$^+$-ATPase in different tissues (Teriete, Franzin, Choi, & Marassi, 2007; Cheung, 2010). *FXYD1* is located on chromosome 19 in *Homo sapiens* (Figure 4).

A.

B.

Figure 4: Genomic location of *FXYD1* in Chromosome 19 of Homo sapiens and the structure of the *FXYD1* gene. **A.** Chromosome 19 of *Homo sapiens* showing the genomic location (red rectangle) of the *FXYD1* gene on the q arm of chromosome 19 (Adapted and edited from Ensembl 2013). **B.** Gene structure of the *H. sapiens FXYD1* gene located in the region chr19: 35,138,789-35,143,055. The *FXYD1* gene is represented by the green line and green rectangles. The coding regions are represented by the red rectangles from the red line,

which translate to phospholemman. Introns are represented by the solid horizontal black lines at the bottom, while exons are located between the introns boundaries, red vertical lines at the top. The coding exons are exons 2 to 8. (Adapted and edited from NCBI 2014).

1.5.1 *The phospholemman protein*

Phospholemman (PLM) is 72 residues long and a single-span transmembrane protein. Characterised by Larry Jones in 1985, PLM is an important phosphorylation target of protein kinase A/C (PKA/PKC) (Crambert, Füzesi, Garty, Karlish & Geering, 2002). PLM is part of the Na^+/K^+-ATPase (NKA) ion pump and contributes to the proper functioning of NKA (Fuller *et al*, 2004; Silverman *et al*, 2005). PLM is therefore considered as a key physiological regulator of cardiomyocytes and poses as a potential target site for cardiac therapeutics (Shattock, 2009). The 72-residue single-span transmembrane protein forms alpha helical tetramers *in vitro* (Beevers & Kukol, 2006) and *in vivo* (Bossuyt, Despa, Martin, & Bers, 2006; Song, Pallikkuth, Bossuyt, Bers, & Robia, 2011).

1.5.2 PLM primary, secondary and tertiary structure

A
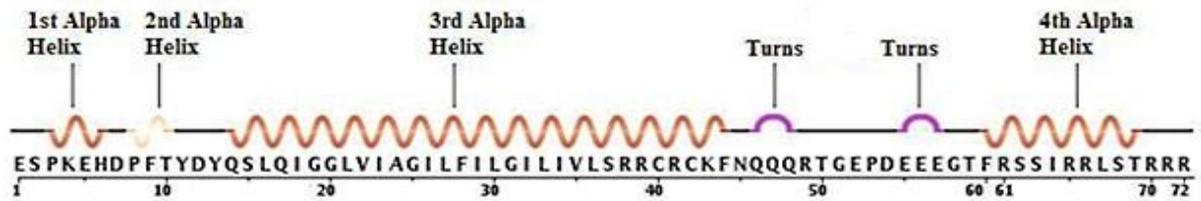
B
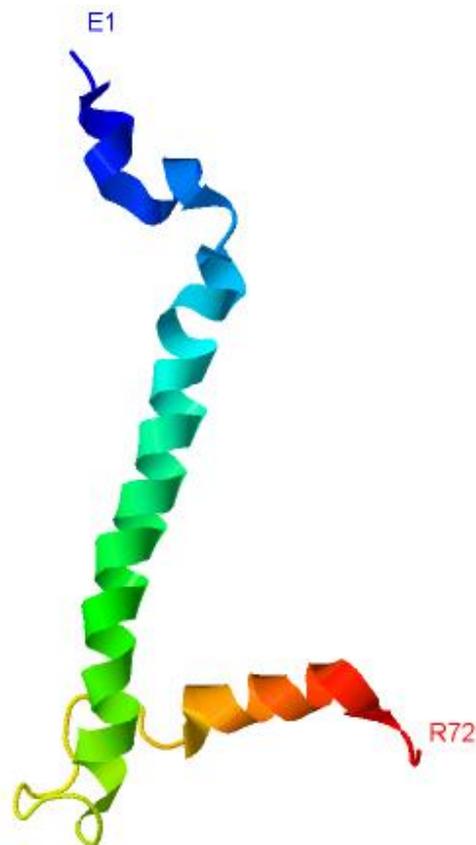
Figure 5: (**A**) Primary and tertiary structure of PLM showing the 72 amino acid residues. (**B**) The cartoon 3-D structure of the PLM monomer obtained by NMR spectroscopy in detergent micelles. The polypeptide chain is made up of one long transmembrane alpha helix and

three shorter helices that are connected by turns. Adapted from RCSB Protein Data Bank (2013).

The transmembrane domain of PLM was shown to form tetramers in lipid bilayers (Beevers & Kukol, 2006). Using site specific infrared dichroism combined with molecular modelling (reviewed in Kukol, 2005) an atomic model of the tetramer was obtained that revealed the potential to interact with NKA, which was proposed to lead to a subsequent dissociation of the tetramer (Beevers & Kukol 2007). Further *in vivo* studies have shown that a tetramer exists *in vivo* and that there is a delicate balance between monomer and tetramer, which also depends on the phosphorylation of PLM (Song *et al.*, 2011). X-ray crystallography studies of the sodium-pump (NKA) in other tissues and species have shown that monomeric *FXYD1* (PLM) homologs, such as FXYD2 in porcine renal tissue (Morth *et al.*, 2007) and *FXYD1*0 in the shark rectal gland (Shinoda, Ogawa, Cornelius, & Toyoshima, 2009) act as a third subunit of NKA. NKA exchanges three Na^+ ions against two K^+ ions that are pumped back into the cell and ensures the resting electrical membrane potential of cells is maintained. When not phosphorylated, PLM reduces the NKA pump's affinity for intracellular Na^+. This will cause an overload of intracellular Na^+ and create an ionic imbalance, eventually causing accumulation of Ca^{2+} ions. Contrary to when PLM is phosphorylated, this intracellular accumulation of Na^+ ions is reduced as affinity of the NKA pump for sodium ions is restored. Protein kinase A activation reduces K_M of NKA for Na^+, while protein kinase C activation increases v_{max} (Han, Bossuyt, Despa, Tucker, & Bers, 2006). The transmembrane domain of PLM on its own is responsible for changes in the sodium affinity (Lifshitz, Lindzen, Garty, & Karlish, 2006). As previously stated, an imbalance of Na^+ will lead to accumulation of Ca^{2+}, which is reported to lead to arrhythmia (Parham,

Mehdirad, Biermann, & Fredman, 2006; Thandroyen *et al*., 1991). Any factors that cause an increase in intracellular Na$^+$ ions will cause a build-up of Ca^{2+} inside cells. Previous papers have reported that G-quadruplex formation is positively correlated with the concentration of cations, especially K$^+$ ions (Kan *et al*., 2006; Samatanga *et al*., 2013), which have been proposed to be the best stabilizers of G-quartets, eventually G-quadruplexes, when compared to other cations such as Na$^+$, Ca^{2+}, Li$^+$ etc (Sun *et al*., 2013; Nguyen Thuan, Haselsberger, Michel-Beyerle, & Anh Tuan, 2011).

1.6 Techniques for G-quadruplex detection

Detection of G-quadruplex ensembles within nucleic acid species made use of biophysical, biochemical and molecular assays as well as bioinformatics-based predictions. As previously stated, the prediction of G-quadruplex in nucleic acid sequences can be done by computational techniques (Kikin *et al.,* 2006; Lorenz *et al.,* 2011). Biophysical assays exploit the different physical properties of G-quadruplexes compared to normal DNA/RNA. Such assays include circular dichroism spectroscopy (Paramasivan, Rujan, & Bolton, 2007; Randazzo, Spada, & da Silva, 2013) and light absorption (UV/VIS) spectroscopy (Goncalves, Ladame, Balasubramanian, & Sanders, 2006; Rubis *et al.*, 2009) that investigated the interactions of different ligand with G-quadruplex forming sequences. UV melting (Liu *et al.*, 2012; Mergny & Lacroix, 2009) experiments were aimed at measuring the folding and unfolding of G-quadruplexes under different cations concentration over a range of temperatures. Nuclear Magnetic Resonance (NRM) spectroscopy (Adrian, Heddi, & Anh Tuan, 2012; da Silva, 2007), can be used to detect the presence of G-quadruplexes due to characteristic resonances in the 1-dimensional spectrum. Upon the formation of G-quadruplexes, the imino guanine protons become trapped within the G-quadruplex entity and cannot be exchanged with the H_2O present in the buffer. This signal can be detected within the chemical shift range of 10-12 ppm, by a proton 1-D NMR spectrum. 2-D NMR techniques have been used to determine the three-dimensional structure of an anti-parallel intramolecular G-quadruplex (PDB-ID: 2KM3, fig. 6) derived from human telomeric ends (Lim, Alberti, Guedin, Lacroix, Riou, Royle, Mergny & Phan, 2009). The 2KM3 sequence was used as positive control in this work.
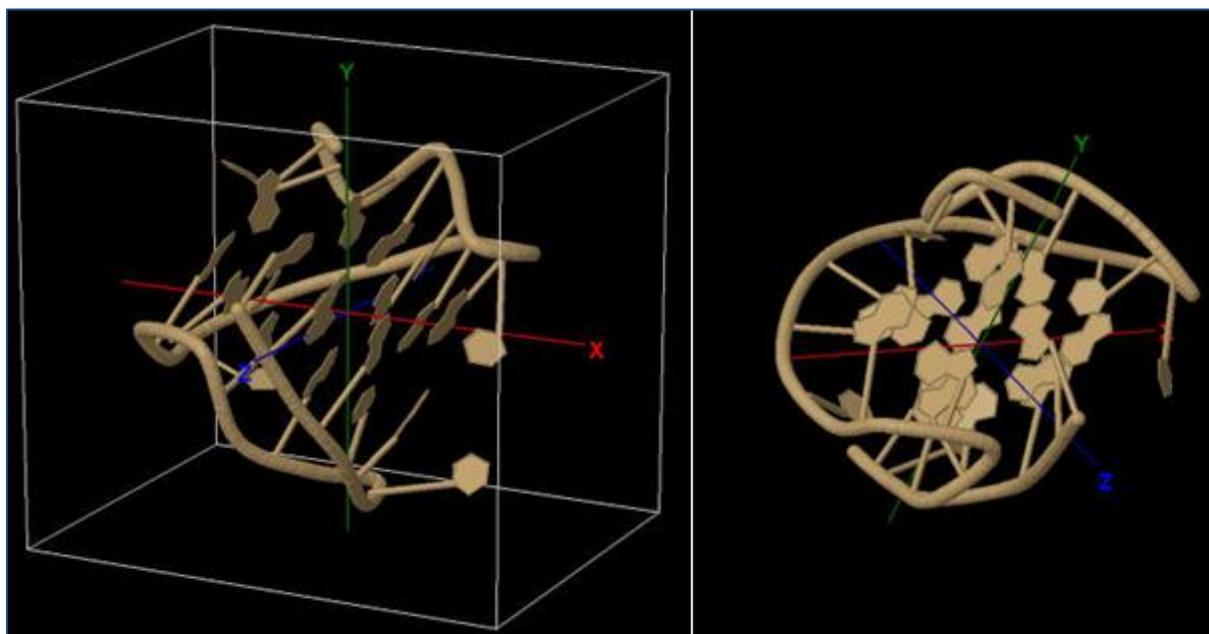
Figure 6: Cartoon representations of 3D structure of an anti-parallel intramolecular G-quadruplex formed from DNA viewed in two orientations (left and right part) (adapted and edited from RCSB PDB, 2013).

Other techniques used included surface plasmon resonance (Redman, 2007), isothermal titration calorimetry (Musetti *et al.*, 2013), mass Spectrometry (G. Yuan, Zhang, Zhou, & Li, 2011) and others. One of the most widely employed techniques used in the detection of G-quadruplex is fluorescence spectroscopy (Hong *et al.*, 2008; Tseng *et al.*, 2013; Vummidi, Alzeer, & Luedtke, 2013). The most commonly used fluorescence technique is based on the Förster resonance energy transfer (FRET) technique. A donor and an acceptor fluorophore are attached on either the 5' or the 3' ends of nucleic acids. In the G-quadruplex the 5' and 3' ends of the nucleic acid come into close proximity that allows FRET to occur. In one of the few *in vivo* studies, Xu *et al* (2010) investigated whether G-quadruplex can be formed *in vivo* by Telomeric Repeat-containing RNA (TERRA). A modified TERRA oligonucleotide containing

a pyrene monomer on each end was used and G-quadruplex formation will bring the monomers close together to form a pyrene dimer that emits light at wavelength 480 nm. Xu *et al* have found that TERRA can form G-quadruplex *in vivo*. Another approach utilises intrinsic fluorescence of nucleic acids, which has the advantage that it does not require labelling. G-quadruplexes are known to have increased fluorescence intensities. Nguyen Thuan *et al.* (2011) reported increased intrinsic fluorescence emission of previously characterised G-quadruplex structures.

Biochemical and molecular techniques include assays such as Polymerase Chain Reaction (PCR) stop assay (Ou *et al.*, 2007; Yan *et al.*, 2010), nuclease assays (Zhou *et al.*, 2013), Gel electrophoresis (Lin *et al.*, 2010; Moon & Jarstfer, 2010; Viglasky, Bauer, Tluckova, & Javorsky, 2010), antibody engineering (Biffi *et al.*, 2013) etc. The PCR stop assay gives information about ligand that can stabilize G-quadruplexes. PCR products are screened and any disturbance of the enzymatic activity of polymerase in guanine rich regions are attributed to stabilized G-quadruplexes by the ligand in that specific region. Nuclease assays enables detection by using restriction endonucleases to cut nucleic acid at specific sites. G-quadruplexes can restrict endonucleases and running the products on gels will generate a distinct band in nucleic acids that formed G-quadruplex, while nucleic acid that did not form G-quadruplex will produce more bands. Antibodies that selectively bind G-quadruplexes have been engineered and allowed easy detection of G-quadruplexes. The method is however very expensive. The basic principle resembles that of Enzyme Linked Immunosorbent Assay (ELISA). The engineered antibody will bind the G-quadruplex DNA, and usually the antibody is conjugated with a molecule that will allow visual detection. In early 2013, Biffi *et al.*, have reported the development of a specific antibody that has high

selectivity for DNA G-Quadruplexes. This labelled antibody allowed the visual detection of DNA G-quadruplexes inside human cancer cells. Gel electrophoresis is by far the easiest way to detect G-quadruplex formation within nucleic acid. Cheap and reliable, this simple method exploits the electrophoretic migration properties of compact vs. linear species in gels. G-quadruplex species have been reported to migrate faster on Poly Acrylamide Gel (PAGE) than non-G-quadruplex species. PAGE is preferred to other gels mainly because the nucleic acid sequences used for G-quadruplex assays are relatively short and PAGE gives better resolution.

1.7 Aim

The aim of this work was to investigate whether or not *FXYD1* pre-mRNA can form G-quadruplexes. This work took into account the ability of the *FXYD1* gene to form G-quadruplex and various techniques used to detect G-quadruplex formation. The initial stages involved *in-sillico* analysis of *FXYD1* pre-mRNA and ortholog sequences using QGRS mapper, Quadbase and the Vienna RNA Package. Later stages involved the detection of G-quadruplexes in synthetic oligonucleotides by native PAGE and intrinsic fluorescence spectroscopy.

2. MATERIALS & METHODS

2.1 MATERIALS

2.1.1 *Software, databases, web-servers*

Algorithms and software used for G-quadruplex prediction:

1. G-quadruplex online prediction algorithm; QGRS mapper (http://bioinformatics.ramapo.edu/QGRS/analyze.php) (Kikin *et al.*, 2006)&Quadbase (http://quadbase.igib.res.in/proquad/quad_input.jsp) (Yadav *et al.,* 2008)

2. Vienna RNA Package version 2.1.2 (Lorrenz *et al.*, 2011)

FXYD1 **pre-mRNA sequences and control sequence database:**

1. *FXYD1* pre-mRNA sequence accession numbers for *H. sapiens* (ENST00000351325), *M. musculus*(ENSMUSG00000036570), *C. familiaris*(ENSCAFT00000011368), *P. troglodytes*(ENSPTRT00000020057), *B. taurus*(ENSBTAG00000017816), *R. norvegicus*(ENSRNOG00000021079), *M. domestica*(ENSMODT00000033163), *F. catus*(ENSFCAG00000008890), *O. garnettii*(ENSOGAG00000014401), *E. caballus*(ENSECAG00000014815), *A. melanoleuca*(ENSAMEG00000000212), *P. abelii*(ENSPPYG00000009851), *O. cuniculus*(ENSOCUG00000022123), *G. gorilla*(ENSGGOT00000026217), *S. scrofa*(ENSSSCT00000027321), *O. aries*(ENSOARG00000004709), T. truncates(ENSTTRG00000001446)

2. *FXDY1* variant pre-mRNA sequence*: FXYD1*-009 (ENST00000589121)

3. Positive control DNA sequence, PDB ID: 2KM3, sequence from RSCB PDB

Web servers for sequence conversion, genome comparison and sequence alignment:

1. DNA<>RNA converting tool (http://www.attotron.com/cybertory/analysis/trans.htm)

2. DNA/Protein sequence randomizer software (http://www.cellbiol.com/python.html)

3. Multiple Sequence Alignment of orthologous *FXYD1* sequences using the MAFFT web based alignment tool Version 7 available at (http://mafft.cbrc.jp/alignment/server/)

4. Pre-mRNA comparison of *FXYD1* and variant-009 using the 1000 genomes transcript comparison available at (http://browser.1000genomes.org/Homo_sapiens/Gene/TranscriptComparison?db= core;g=ENSG00000266964;r=19:35629712- 35634013;t=ENST00000589121;t1=ENST00000589121;time=1396457246372.372)

2.1.2 Sample preparation

Oligonucleotides used for laboratory analysis:

1. Oligonucleotides purchased from EurogentecLtd.(Southampton, UK) and used without further modification;

   - Positive (+VE) control DNA (**AGG-GCT-AGG-GCT-AGG-GCT-AGG-G**)purified by Reverse-phase cartridge purification (RP-Cartridge)

   - Negative control_A (-VE_A) DNA (**CGT-GGG-GAG-ATT-GGG-GAG-CGC-A**) purified by RP-Cartridge

   - Negative control_B (-VE_B) DNA (**GGT-GTG-CGT-GTG-CGA-GCG-AGA-GAG-**

AGU-GG) purified by RP-Cartridge

- *H. sapiensFXYD1* (Human_PLM) RNA (**GGG-AGA-CUG-CGG-GUA-UUC-UGG-GGA-GAG-GG**) purified by Reversed Phase High Performance Liquid Chromatography (RP-HPLC)

- *B. Taurus FXYD1* (Bovine_PLM) RNA (**GGG-CGC-GGG-GGG-UCG-GGG-AUC-GGG**) purified by RP-HPLC

Solutions used for preparing G-quadruplex samples:

1. 10 ml of 1M Potassium Chloride (KCl) solution

2. 20 ml of RNAase free H_2O

3. 500 ml of 1M Tris-Acetate Buffer (TrisOAc) pH 7.5

4. 100 ml of 1 M Potassium Acetate (KOAc)

NOTE: All solutions were autoclaved and kept at room temperature prior to use.

2.1.3 NATIVE PAGE

Solutions for preparing Native PAGE and staining:

1. 100 ml 40% acrylamide solution

2. 10 x TBE Buffer solution

3. 100 ml of 0.05M &1M KCl/KOAc solution, sterile distilled water

4. Ammonium persulfate (APS) at 10% (w/v) in water

5. N,N,N',N'-tetramethylethylenediamine (TEMED)

6. Mini gel stop mix; 1 x TBE + 20% (w/v) sucrose + 10 % (w/v) Ficoll + 10mM EDTA and

0.25% (w/v) bromophenol blue

7. 1 x TBE gel running buffer

8. SYBR Green IS32717& SYBR Green II Nucleic Acid Stain S9430

2.1.4 Fluorescence and UV-vis spectroscopy

Equipments used for fluorescence spectroscopy:

1. Fluor cuvette Type C quartz glass with 10 mm light path

2. Perkin Elmer LS 55 fluorimeter

3. UV/VIS CARY 100 dual-beam spectrophotometer (Varian Inc.)

4. Quinine solution at 24 ppm

2.1.5 Data processing

Software used to process raw data from Native PAGE and Fluorescence spectroscopy:

1. Gene Tool Syngene (Copyright © 2009-2011 Syngene, A Division of Synoptics Ltd)

2. PerkinElmer UV WinLab Data Processor and Viewer Version1.00.00

3. Microsoft®Excel®2010 Version 14.0.7109.5000

2.2 METHODS

2.2.1 *In-sillico* analysis

G-quadruplex prediction using QGRS mapper and Quadbase

The raw FASTA pre-mRNA sequences of the *FXYD1* orthologs were analysed online using

QGRS mapper (Figure 7) and Quadbase prediction software (Figure 8).

Figure 7: *H. sapiens' FXYD1* pre-mRNA sequence (4286 bp) in the QGRS mapper analyzer

box.

The parameters were left at the defaults, with maximum length of potential of Quadruplex forming sequences set at 30 bases. The minimum G-group was set at 2, which is the minimum number of G-tetrads and finally the loop length was set between the range of 0 – 36 bases. Clicking on the "Analyze" button in the bottom right corner initiates screening of the sequence and search for putative G-quadruplex forming sequences (GQS). All the other orthologous *FXYD1* pre-mRNA sequences were analysed using the same settings as *H. sapiens*.

```
AAAGUGCUCAGCCCCCGGGGCACAGCAGGACGUUUGGGGGCCUUCUUUCAGCAGGGGACAGCCCGAUUGGGGUGAGCGUCCCCCACUCCUUCCCUCC
AGGCCUCACCCCUGGUCUGGCUGGGCCGCCUAUUUUGGGAGCAGGAGUGGCCAGCCCGAGGCUUCCCAGGCAGGCCAACCCAAGAGGGAGGGAGUGU
GGUUGAGGCAGUGGGUUCUGCAGGGUGGGAUGUGGGGUGACUCCUCCCUGCCCUGCUGGUGCGUGUGCACCCUGGCAGGGUGUGGAGUUGGGACACAC
ACGUGUGUAGGGCUGGUUGCGUCACUGCGUGGGGGCACCGGAGGCCCAGAGGAGGAGUAUUGGAUGCCUGACGGUGUUUACACCCCACGUCCUGCUC
CAACCAGCAGUUUGGGGAGAGGUUGUUGUUCAUGUCCAUUCCGGCCCCACUGUGUGUGUGUGUGUGUGUGUGUGUGUGUGUGUGUGUAUCCCU
GCCCCAGCAUGUGUUUUCUAUCUCUAAGGCCCACUGGGCUGGGCCUCAUGUCACUUGCCUGACAUCCGAUUGUGAAAGAUGUCACCCAGAGGCGGGC
AGAGGGGCUGUCUUUUCCUUUUCUCGUUGCUGCCCAGGGAGGAGACGGGGUGACCUUUCCCACAGGGGCAGCCUGUGGCGAUGUGGCAGCUGGGCCU
CACCCCGGCAGGGCUGUGCGUGACCCCCUGAGUGGGGGAAGGCAGGCUGUUGCCAUGGUGGCCUGAGCGAGCAGAAUUCCUCCAGGGUGAAGUGGGA
GAUAUUUAUACCCGGGGUCAGGCCCGCGAGCGGGCGGGCGGGAGAGGGCAGGGAGCUGGGAUUUCGCGGGGCACAGUGAGGCCGGGCAUGUAGGCAGGU
GGGACUUGGGCGUGCCCUGCUGUCUCCUGCUCUGUGUUUGUGUGAGGCAGCGCCUCCUCUGCCCUGCCAGGGUAGGUCUGGGAAUCGGGGGCCUGCU
GCGGGAGGUGGAGGCCCAAGGGAGGCCCCCGGGGACUGUGUGUCUCACCCCCGUCCCUGCUACGUUGUGUUGUUGUGUGAUCCCAUCGUGGAGGUU
GUUUUGGUGACACUGUGUCCCCACGAAGCUGGGGAUACCCGUUUCUCUAGCUUGGAGCCACCAAGAUAGAGGACAAACACUUCUGUGAUUCAGUCCC
CAGACUGUCUCUGACUUAAUCCCUUGGGUUCAAGCCCUAUGUGGGGAGAGCAAGGGCACACACUGCCUAAUCCGUGGUGUCCCCCCCAGGACAAUGGC
GUCUCUUUGGCCACAUCUUGGUUUUCUGUGUGGGUCUCCUCACCAUGGCCAAGGCAGGUGAGUGCAGGGGAGGCUGCCCGCUACCCACCUCAGCCCCA
GGGGUGGCGGUGGGGACCGAAGAACCAAGUUUGGAGACCCCAACCUAGACUAAGUCGGCUGGGGGUACCAAGAAGUUUGGGGGUCUCCACGUGGGGUCC
AGUCACAGGCUGGUAUUUGGGGGAGGGGAGAGGAAGCCCCAGAUCAGGCAAAGAUGGGGUGGGGAUGGGGCUGAAUCCCCGAUGGGAUAACUGGGUCA
CAGACAGCCUGCCGUGAGUCAGGGAGCUGGGGCAGUUAGGUGCCACCUGCCCCAUCUGGGACAGUGCAGAGGGGGCAGCUGGGACCCAGAGAGUGUG
GGCAGCCUGCCCAGACACCCUCAGACUCUAAGCCCAGCAAGGCAGAGCCUCCAGUGGUCUCCUCAUGCCCCUCCCUGCCAGGACCCCAGGAAGCAUU
CAACCCCUGAUUUCUCUCUCUUUCCAGAAAGUCCAAAGGAACACGACCCGUUCACUUACGGUGAGCGGGGGGGUCUAAUUUUGGAGUCCUGGGGGAGAG
CCUGGCUUUGCUGGUCCUUUGAUUUCCCCCUGCCCCUCCCCCAGAGUCCCAGUAUUGAUAUCUCUGUCAUUCUUCCUUCCUCUAUUUUGUCCUUCCUC
UCUGAUUCCACCUGUCUGCAUCUUUUCCUGUCUGUGUCUAUCUGUGUCACUGUCUAUGUGAUACCUCUCUGGUUCUCUUUCUCUUGCCUGCGUCUGU
CUCAGCAUCUCGUGGCCCAUCCUCUGCUUCUUUCCUGUCUUCUCUCCCCCCUGUCCUCCUCCUCCCUGUCCCCUCCCUCCCUUUCCUAUACACCCCUU
UCCUCUCCCUGGUACCCCACUUUCCUCCUCCCAUAUCUGCUCCCCCUUAAUUUAUCUUACUUCCCCCUUCUGCCUGCUGGUCCUUUCUCCCUGUUCC
CUCCUUCCCAAUUUACCCCUCUCCUAUUCUCCCUCCUGUCUUCCCUGCCCUCACCUUCCCUGCUCUGCUGCUCACAGACUACCAGUCCCUGCAGAUC
GGAGGCCUCGUCAUCGCCGGGAUCCUCUUCAUCCUGGGCAUCCUCAUCGUGCUGAGUGAGUGCCCCUAGCUCCCGCCCUCUACCCCGCCUCUCCCUG
GCCCCACCUCUCUCUGGCCCCGCCUCUCCCUGGCCCCGCCUCUCCCUUAGCCCCCCUCUCCCUGGGCCCCGCCUUCUCUCUGGGUCCCUUCUCCCUGGCC
CCCGCCCCCGCCCCAACCCCUCCCCAGGCCCUUGCCCCGGCCCUACCCUGCCUUGGUUCCCCGGCCCCCGGUCUCGCUCUAGCCCGCCCCGUCCCCCAAG
CCCCGCCCCUCGCGAGGGCGAGCUGGAGCUACAGCGCCGCCUUGGCGCCCGGCGGGAGGGAGCCUCAGCUUCUCCUACCUCUCCACGCCCACAGGCAG
AAGAUGCCGGUGCAAGUUCAACCAGCAGCAGAGGUAAGACGCCCCUCCCCGCCCUCCUUCGCCCGCUCCUGCUCUGGAGGGCGCCGCGGGUGAGGCG
GGGAGUACCCCUGACCCGCAGCCCGAUCCCCGUCAGCGACUAUGUAUUAAGCACCUACUAUGUGCCAUGGCCCAAGCCUGGCCCUGGGACCAAGCGA
GGAAAAAAACCUCCCGCCCUUCCUGGCCGAGCUCCCAGCCUAGUGGAGGCGGUGGCCGUGGGUUCCAACAGCCCCACAGAUAGAAAAAAUCACAAAGCG
UGAUAACACAAAGUGCAGGAAAGAAGAAACGGCGGUGAAAAUGAGAUCAUCUCACACGCGGCCAGUUUAGCUUGUGAGUCUUUGGUUCCUAGCUCUUUGA
UUCCUCUUCGAAUAAAAUGUUAAAGCAUGGACAAUGUAUGAAUAUGUUAGAACAAUUAUAGAUAAUUUAUCAUAAGUAGUAGCUAAUAUUUACUGGGUG
UGUACCACGUGUUAGAUACGGUUUCACUUCCUCUGGGAGGGAGGUGCUGUUAUUAACCCCAUUUGACAGAUGAGGAAACUAAGGCACAGGGAGGUAA
AGUCACUUUGUUCAAGAUCACUCAAGUGGAAGAUGGGGGGUUCUGGGUUUCCAACCCAGGCCAUCUCAUGGCAGUCUGCCAAGUCCCCAUGACUAUC
CCUCCCCCACCAACUUCACAUCCCUGCCCCCAAAUCCGCGGAGGUACUCACUGUUAACCAGCUUAGAAGCCCCCUGCCAGCACAUAAGCUGCUCCUG
GGUGCUCCUCAUUUCUGGCGGACCCCGAGCCUGCUCUUCGUCCAUAUCUGGGCCUAGUUACACCAAUCUGGGAAAGGAGGCUUGUACUGGGGGGGUUC
CUAGAAGGGCAGCCUCUCCCCUCUUCCAUCCCGAAAUCCCUCUGCCUCUGGUCUGUCUUCCCAGGACUGGGGAACCCGAUGAAGAGGAGGGGAACUUUCCGCA
GCUCCAUCCGCCGUGAGUCUGGGGGAGACUGCGGGUAUUCUGGGGGAGAGGGCUGGUUCCAAGGACCGCUUUUCCCGGCCCUCCCUGGCUGCGUAGAGG
GAAGGGCUGGAUCUGAAAGCGGAGGGCGGGGAGUUGCCCCGCCGCGGGCCCCACCUGCCCAGGAGCUGGGGAUGCCUCUCCAGAAUGACCCCGAUC
UCCGUGUUCCCCCCAGGUCUGUCCACCCGCAGCCGGUAGAAACACCUGGAGCGAUGGAAUCCGGCCAGGUGCUGCAGCUCUGACACGGCGGUGGGAG
GGAAGGAGGGAGGAAGGAAAGGCGGGAGAGGGAGGGGGCCAAGUGCCAGAGUUGAAGGGCGGCGAGGGGUGGGGCUGGACGUCCCCCCUCGCCUCUC
ACCCUUUUCACCCUCACAGGACUCCCCUGGCACCUGACAUCUCCCACGCUCCACCUGCGCGCCCACCGCCCCCCUCCGCCGCCCCUUCCCCAGCCCUG
CCCCCGCAGACUCCCCCUGCCGCCAAGACUUCCAAUAAAACGUGCGUUCCUCUCGACAGCACUUUGUCGGUCUCGGUCCCUCAGCGCGAAACGCCAG
CGCCACUGGGCCCCAGCA
```

[Submit]   [Reset]

Figure 8: H. sapiens' FASTA *FXYD1* pre-mRNA (4286 bp) in the Pattern finder search box of Quadbase that screens nucleic sequences for patterns that can form G-quadruplex.

The parameters were those of the default settings, which was between two and five guanines for the G-tetrads. The loop sizes were set between 1 and 7, 1 being the minimum and 7the maximum integer available.

Prior to analysis by QGRS mapper and Quadbase the raw FASTA sequences were converted to RNA using the DNA<>RNA converter tool. The sequence of a +VE control DNA previously known to form a G-quadruplex was obtained from the RSCB Protein Database (NDB-ID: 2KM3) and was also analysed in QGRS mapper and Quadbase. A DNA/Protein randomiser tool was used to shuffle the sequence of the +VE control and generate possible sequences of a –VE control, -VE_A, which were of the same length as the +VE control and had the same base composition as the +VE control. Following analysis by QGRS mapper and Quadbase, a second –VE control, -VE_B, was also generated using the DNA/Protein randomiser based on the *H. sapiens'* highest scoring sequence generated by QGRS mapper.

G-quadruplex and secondary structures prediction using the Vienna RNA package

The *Vienna RNA package* was used to predict G-quadruplex and other secondary structures that are likely to compete against G-quadruplex formation. The +VE control, -VE_A, -VE_B, Human_PLM and Bovine_PLM were analysed using the RNAfold, RNAsubopt and RNAeval algorithms from the package. RNAplot option was used to produce graphical display of the proposed structures by RNAfold and RNAsubopt for all sequences.

The following command lines were used in the Command Prompt (Microsoft©) for the Human_PLM sequence:

```
1) C:\Users> rnafold –g < Human_PLM.txt
```

```
2) C:\Users> rnaeval -g < RNA_struct.txt

3) C:\Users> rnasubopt -e3 < Human_PLM.txt

4) C:\Users> rnaplot -o ps <RNA_struct.txt
```

The first commands predict the minimum free energy (MFE) structure of the sequence contained in the text file taking into account G-quadruplex formation (-g option) The second command calculates the energies of given secondary structures, taking into account G-quadruplex formation. The third command determines other secondary structures within 3 kcal/mol above the MFE structure. The same command lines were executed for the +VE, -VE_A, -VE_B and Bovine_PLM sequences. The last command line produces graphical display of secondary structures predicted by RNAfold and RNAsubopt in post script format (-o ps option).

Multiple sequence alignment by MAFFT web server version 7.0

The MAFFT web server was used to align all orthologous *FXYD1* pre-mRNA sequences. All parameters were the default settings. The slow iterative refinement method was used.

Pre-mRNA comparison of *FXYD1* and *FXYD1*-009

The mRNA and pre-mRNA sequences of *H. sapiens' FXYD1* and *FXYD1*-009 were compared against each other to look for alternative splicing. Using the 1000 Genomes Transcript Comparison option, mutations were screened in potential G-quadruplex forming sequences from the pre-mRNA sequences of *FXYD1* and variant 009.

2.2.2 G-quadruplex preparation

G-quadruplex was induced by incubating the oligonucleotides in K^+ containing and K^+ free buffers (as controls). The G-quadruplex folding buffer contained K^+ (mixture of KCl and KOAc) at 0.1 or 0.05 M and 0.02 M TrisOAc pH 7.5. Controls were prepared in K^+ free buffer that contained 0.02 M TrisOAc pH 7.5 only. The samples were prepared in sterile microfuge tubes. The mixtures were heated at 90^oC for 10 minutes to disrupt any intramolecular interactions. After heating, the –VE_A and -VE_B samples in K^+ containing buffer and all control samples were cooled to 4^oC by keeping the tubes on ice, to disfavour formation of G-quadruplex. The +VE, Human_PLM and Bovine_PLM samples in K^+ containing buffer were allowed to cool down to 25^oC over 2.5 hours by removing the heating block from the heating source. Once the samples reach the annealing temperature, the tubes were then stored at 0^oC to preserve the G-quadruplex structures for later use.

All plastic wares were heated at 230^oC, including pipette tips, to inactivate any RNAase.

2.2.3 Native PAGE preparation

30 % polyacrylamidegels were used to run the samples. Samples incubated in K^+ containing buffer were ran on separate gels from samples incubated in K^+ free buffer, to keep experimental conditions constant. Gels prepared for K^+ containing samples was made by adding 9.375 ml of 40% acrylamide solution + 1.250 ml of 10 x TBE supplemented with KCl & KOAc to match the concentration of K^+ of the folding buffers, e.g. for samples incubated in 0.1 M K^+, 10 x TBE+0.1 M K^+ mixture was used for preparation of the gel. This was followed

by the addition of 1.875 ml of sterile distilled water and 150 µl of 10% APS. This mixture was degassed under vacuum to remove any molecular oxygen that would inhibit the polymerisation process. Degassing was followed by the addition of 15 µl of TEMED. Gels used for K^+ free samples were made in the same way as previously described for K^+ containing samples, except that the 10 x TBE was used without K^+.

2.2.4 Detection of G-quadruplex by Native PAGE

Samples for electrophoresis were thawed at room temperature and 8 µl of mini gel stop mix (1 x TBE + 20% (w/v) sucrose + 10 % (w/v) Ficoll + 10mM EDTA and 0.25% (w/v) bromophenol blue) was added to each tube. The final oligonucleotide concentration of each species was 3µM. After thoroughly mixing the samples with the dye, 10 µg of each sample was loaded onto the gels. The gels were run in different tanks and the buffer used for non-G-quadruplex gels was 1 x TBE buffer, while G-quadruplex gels were ran using 1 X TBE containing either KCl and KOAc at a final concentration of 0.1M or 0.05 M. The buffers were pre-chilled at 4^oC to minimize overheating of the tanks. Electrophoresis was performed at 140V and the run time was on average 3-4 hours. Following electrophoresis each gel was removed and cut at the upper right hand corner to track orientation. The gels were stained using SYBR Green I RNA stain S9430 and SYBR Green I nucleic S32717 exposed at 254 nm for 15.5s.

The ratio of the distance migrated by each samples relative to the distance migrated by the tracking dye, R_f value, was calculated using the software Gene Tool Syngene (Copyright 2009-2011 Syngene, A Division of Synoptics Ltd). Student 2-tailed-t-test was carried out for the samples under different incubation conditions.

2.2.5 Detection of G-quadruplex by fluorescence spectroscopy

G-quadruplex induced and uninduced samples were prepared at a final oligonucleotide concentration of 1.5 µM for RNA species and 5.0 µM for DNA species. The reason behind the choice of these concentrations was that these are the minimum detectable concentrations for either RNA or DNA by the Perkin Elmer LS 55 fluorimeter. Buffers for non G-quadruplex samples was 0.02 M TrisOAC only and that of G-quadruplex samples was 0.02 M TrisOAC + 0.1 M K^+. Samples prepared overnight were allowed to thaw and attain room temperature, 20^oC, before readings were taken. Emission spectra were recorded over the wavelength range of 300-500 nm using a Perkin Elmer LS 55 in a Type C Fluor micro cuvette with a 10 mm light pathway. Samples were excited at a wavelength of 260 nm and both excitation and emission slit widths were set at 5 nm. The scan rate was 150 nm/min. Emission spectra of buffers were also recorded. UV-VIS spectra of each sample were recorded using a UV-VIS CARY 100 dual-beam spectrophotometer between the range of 200-400 nm with the appropriate buffer placed into the second beam.

The fluorescent compound quinine was used to test the fluorimeter by recording the emission spectra in the presence of either 0.02 M TrisOAc or 0.1 M K^+ and 0.02 M TriOAc. The spectra were recorded by exciting Quinine at a final concentration of 0.6 ppm at

wavelength of 250 and 350 nm independently over the range 335-485 and 355-505 nm respectively. The scan speed was 150 nm/min and both excitation and emission slits were set at 5nm each.

The data generated by the fluorimeter were processed with PerkinElmer UV WinLab Data Processor and Viewer Version 1.00.00 into graphical display. The original spectra were processed using Microsoft Excel 2010 Version 14.0.7109.5000 to obtain smooth curves. Trendline with moving average of 30 data points per period was produced for each emission spectrum.

3. RESULTS

3.1 *In-sillico* analysis

3.1.1 QGRS mapper and Quadbase findings

Analysis of orthologous *FXYD1* pre-mRNA sequences by QGRS mapper revealed several G-Quadruplex forming Sequences (GQS) for most organisms. The whole pre-mRNA sequence *of H. sapiens FXYD1* contains 41 GQS as seen in the FASTA sequence below:

AAAGUGCUCAGCCCCCGGGGCACAGCAGGACGUUUGGGGGCCUUCUUUCAGCAGGGGACAGC
CCGAUUGGGgugagcgucccccacuccuucccuccaggccucaccccuggucuggcugggcc
gccuauuuugggagcaggaguggccagcccgaggcuucccaggcaggccaacccaagagggagggagugugguugaggcaguggguucugcaggguggggaugugggugacuccucccugcccugcuggugcgugugcaccccuggcaggugugguuggagcccagaggaggaguauuggaugccugacggug
uuuacaccccacguccugcuccaaccagcaguuuggggagaaggcaccggaggcccagaggaggaguauuggaugccugacgguguuuacaccccacguccugcuccaaccagcaguuuggggagagguuguuguucauguccauuccggccccacuguguguguguguguguguguguguguguguguguguguguauccugccccagcaugug
uuuucuaucucuaaggcccacugggcugggccucaugucacuugccugacauccgauugu
gaaagaugucacccagaggcgggcagaggggcugucuuuuccuuuucucguugcugcccagg
gaggagacggggugaccuuucccacaggggcagccuguggcgauguggcagcugggccucac
cccggcagggcugugcgugacccccugaguggggggaaggcaggcuguugccaugguggccug
agcgagcagaauuccuccaggugaaguggggagauauuuauacccggggucaggccgcgagc
gggcgggcggagagggcagggagcugggauuucgcggggcacagugaggccgggcauguagg
cagguggggacuugggcgugcccugcugucuccugcucuguguuugugugagugcagcgccucc
ucugcccugccagggguaggucugggaaucggggggccugcugcgggagguggaggcccaaggg
aggcccccggggacugugugucucaccccugucccugcuacguuguguuguuguguugaucc
caucguggagguuguuuuuggugacacuguguccccacgaagcuggggauacccguuucucua
gcuuggagccaccaagauagaggacaaacacuucugugauucagucccccagacugucucuga
cuuaauccccuugggguucaagcccuaugugggagagcaagggcacacacugccuaauccgugg
uguccccccagGACAAUGGCGUCUCUUGGCCACAUCUUGGUUUUCUGUGUGGGUCUCCUCA
CCAUGGCCAAGGCAGgugagugcaggggaggcugcccgcuacccaccucagccccagggguggcgguggggaccgaagaaccaaguuggagaccccaaccuagacuaagucggcugggguaccaagaaguuugggggucuccacguggggguccagucacaggcugguauuuggggagggagaggaagccccagaucaggcaaagauggggugggaugggggcugaauccccgaugggauaacugggucacagacagccugccgugagucagggagcuggggcaguuaggugccaccugccccaucgggacagugcagaggggcagcugggacccagagaguguggggcagccugcccagacaccucaga
cucuaagcccagcaaggcagagccuccaguggucuccucaugcccucccugccaggaccccaggaagcauucaaccccugauuucucucucuuuccagAAAGUCCAAAGGAACACGACCCGUU
CACUUACGgugagcgggggggucuaauuuugaguccuggggggagagccuggcuuugcugguccuuugauucccccucgcccuccccagagucccaguauugauaucucugucauucuccuuccc

ucuauuuuguccuuccucucugauuccaccugucugcaucuuuuccugucuguugucuaucug
ugucacugucuaugugauaccucucugguucucuuucucuugccugcgucugucucagcauc
ucguggcccauccucugcuucuucccgucuucucuccccccuguccuccuccucccuguccc
cucccucccuuuccuauacacccccuuuccucucccugguaccccacuuuccuccucccauau
cugcucccccuuaauuaucuuacuuccccccuucugccugcugguccuuucucccuguuccc
uccuuccaauuuaccccucuccuauucuccuccugucuucccugcccucaccuucccugc
ucugcugcucacagACUACCAGUCCCUGCAGAUCGGAGGCCUCGUCAUCGCCGGGAUCCUCU
UCAUCCUGGGCAUCCUCAUCGUGCUGAgugagugcccccuagucccgcccucuaccccgccu
cucccuggcccccaccucucucuggccccgccucucccuggccccgccucucccuagccccc
ucucccuggccccgcuucucccuggucccgccccucccuggccccgccccgccccaacccccu
cccaggccuugccccgccuaccugccuugguucccoggccccccggucucgccucuagcccc
gccccguccccaagccccgccccucgcgagggcgagcuggagcuacagcgccgcuuggcgc
ccgccgggagggagccucagcuucuccuaccucuccacgcccacagGCAGAAGAUGCCGGUG
CAAGUUCAACCAGCAGCAGAGguaagacgcccccuccccgcccuccuucgcccgcuccugcuc
uggaggGCGCCGCGGGUGAGGCGGGGaguaccccugacccgcagcccgauccccgucagcga
cuauguauuaagcaccuacuaugugccauggcccaagccuggcccugggaccaagcgaggaa
aaaaccucccgcccuuccuggccgagcucccagccuaguggaggcgguggccguggguucca
acagcccccacagauagaaaaaucacaaagcgugauaacacaaagugcaggaaagaagaaacg
gcggugaaaugagaucaucucacacgcggcccaguuuagcuuagagucuuguuccuagcucu
uugauuccucuucgaauaaaauguuaaagcauggacaauguaugaauauguuagaacaauua
uagauauuaucauaaguaguagcuaauauuuacuggguguguuaccacugucuuagauacgguu
ucacuuccucugggagggaggugcuguuuauuaaccccauuugacagaugaggaaacuaaggc
acagggagguaaagucacuuuguucaagaucacucaaguggaagaugggggguucuggguuu
ccaacccaggccaucucauggcagucugccaaguccccaugacuauccucccccaccaacu
ucacaucccugcccccaaauccgcggaguacucacuguuaaccagcuuagaagcccccugc
cagcacauaagcugcuccugggugcuccucauuucggcggacccgagccugcucuucuguc
cauaucgggccuaguuacaccaaucggaaaggaggcuuguacuggggggguuccuagaag
ggcagccucuccccuuuccaucccgaaauccucugccucugucuucccagGACUGGGGAA
CCCGAUGAAGAGGAGGGAACUUUCCGCAGCUCCAUCCGCCgugagucug<mark>gggagacugcggg</mark>
<mark>uauucuggggagaggg</mark>cugguuccaaggaccgcuuuucccggcccucccuggcugcguagag
ggaagggcuggaucugaaagcggagggcggggaguugccccgccgcgggccccaccugccca
ggagcuggggaugccucuccagaaugacccccgaucuccguguuccccccagGUCUGUCCAC
CCGCAGGCGGUAGAAACACCUGGAGCGAUGGAAUCCGGCCAGgugcugcagcucugacacgg
cgguggggagggaaggaggaggaaggaaaggcgggagagggagggggccaagugccagaguu
gaagGGCGGCGAGGGGugGGGcuggacguccccccugccucucacccuuuucacccucaca
gGACUCCCCUGGCACCUGACAUCUCCCACGCUCCACCUGCGCGCCCACCGCCCCCUCCGCCG
CCCCUUCCCCAGCCCUGCCCCCGCAGACUCCCCCUGCCGCCAAGACUUCCAAUAAAACGUGC
GUUCCUCUCGACAGCACUUUGUCGGUCUCGGUCCCUCAGCGCGAAACGCCAGCGCCACUGGG
CCCCAGCA

Key: UTR region
        Intronic sequence
        Exonic translated sequence

In the above FASTA sequence, alternate exons are in uppercase and introns are in lowercase

blue characters respectively. Purple uppercase characters represent UTR regions of the gene

and black uppercase characters represent translated region of the gene. Lowercase blue characters represent intron sequences of the gene. The predicted GQSs by QGRS mapper have been mapped and are underlined in the gene and have different G-scores as shown in Table 1. The most stable G-quadruplex in the gene is located in the intron between exon 6 and 7, highlighted yellow. Each GQS has different G-scores; influenced by several factors such as loop sizes, number of guanine residues taking part in G-quartet formation

The 41 GQSs obtained upon analysis of the pre-mRNA of *H. sapiens' FXYD1* sequence and their respective G-scores are listed in Table 1.

Table 1: List of all 41 GQSs and the G-scores for *H. sapiens' FXYD1* pre-mRNA predicted by QGRS mapper (Kikin *et al.*, 2006).

Length	GQS	G-Score
29	GGGAGACUGCGGGUAUUCUGGGGAGAGGG	39
24	GGGCGGAGAGGGCAGGGAGCUGGG	38
24	GGGUUCUGCAGGGUGGGAUGUGGG	36
30	GGGGUGGGAUGGGGCUGAAUCCCCGAUGGG	31
30	GGGAGGAAGGAAAGGCGGGAGAGGGAGGGG	31
11	GGAGGUGGAGG	21
11	GGUGGCGGUGG	21
11	GGAGGCGGUGG	21
14	GGUGGGAGGGAAGG	21
18	GGCAGGGUGUGGAGUUGG	20

13	GGGGGAAGGCAGG	20
25	GGCAUGUAGGCAGGUGGGACUUGGG	20
19	GGGUAGGUCUGGGAAUCGG	20
10	GGGGGAGGGG	20
16	GGGCGGCGAGGGGUGG	20
20	GGGAGGGAGUGUGGUUGAGG	19
28	GGCAGCCUGUGGCGAUGUGGCAGCUGGG	19
11	GGAGGGCGGGG	19
30	GGCACCGGAGGCCCAGAGGAGGAGUAUUGG	18
21	GGCGCCGCGGGUGAGGCGGGG	18
19	GGAAGAUGGGGGGUUCUGG	18
22	GGAGCGAUGGAAUCCGGCCAGG	18
14	GGCGGGCAGAGGGG	17
13	GGAGGAGACGGGG	17
30	GGACAAUGGCGUCUCUUGGCCACAUCUUGG	17
27	GGCCAAGGCAGGUGAGUGCAGGGGAGG	16
21	GGAAACUAAGGCACAGGGAGG	16
29	GGCCCUCCCUGGCUGCGUAGAGGGAAGGG	16
22	GGCACAGCAGGACGUUUGGGGG	15
25	GGGAGCAGGAGUGGCCAGCCCGAGG	15
16	GGGAGGCCCCCCGGGG	15
19	GGGAGCUGGGGCAGUUAGG	15
23	GGGGGAGAGCCUGGCUUUGCUGG	15

24	GGAAAGGAGGCUUGUACUGGGGGG	15
24	GGCCUCACCCCUGGUCUGGCUGGG	14
19	GGGGUCAGGCCGCGAGCGG	13
18	GGGGGUCUCCACGUGGGG	12
25	GGACAGUGCAGAGGGGGCAGCUGGG	12
25	GGUUUCACUUCCUCUGGGAGGGAGG	10
28	GGACUGGGGAACCCGAUGAAGAGGAGGG	10
25	GGCCCCACCUGCCCAGGAGCUGGGG	8

The GQS listed in Table 1 are sorted in the order of highest to lowest G-scoring. The underlined guanines are those taking part in G-tetrad formation to form G-quadruplexes. The highest scoring GQS from *H. sapiens* pre-mRNA is 29 bases long and has a G-score of 39. The G-quadruplex structure formed by the latter is comprised of 3 G-tetrads. The guanines are connected by loops of length 7, 6 and 4 bases in length.

The highest scoring GQS from each ortholog are listed in Table 2 alongside the controls used in this work.

Table 2: The highest scoring predicted GQS from *FXYD1* orthologs and their location within the gene. The analysis was performed with QGRS mapper (Kikin *et al.*, 2006).

Organism	Sequence of highest scoring GQS	G-score	Genomic location[1]
Homo sapiens	GGGAGACUGCGGGUAUUCUGGGGAGAGGG	39	Intronic (6:7)
Mus musculus	GGGAGGAAGGAGGGAGAGGGGUUUGGAGGG	38	Intronic (7:8)
Canis lupus	GGGGGCGAAGGGUGGGCUGGGAUGGCCGGG	42	3'-UTR
Pan troglodytes	GGGAGACUGCGGGUAUUUGGGGAGAGGG	39	Intronic (6:7)
	GGGUUGAAGGGCGGCGAGGGGUGGGG	39	Intronic (7:8)
Bos taurus	GGGCGCGGGGGGUCGGGGAUCGGG	42	Intronic (6:7)
Rattus norvegicus	GGGCAGGUGAGGCUGGG	21	Intronic (1:2)
	GGAUGGAAGGUAGG	21	Intronic (2:3)
	GGCGGUGGGGG	21	Intronic (5:6)
	GGCACGGGGAGGUAAGG	21	Intronic (5:6)
	GGGAGGAAGGAGGG	21	Intronic (7:8)
	GGCGGGUUGGAGGG	21	Intronic (7:8)
Felis catus	GGGAGACUUUGGGGGUUUGGGGGUGAGGG	40	Intronic (5:6)
Otolemur garnettii	GGGCGCAGGGUGGGGUGGGUGAGGCGGG	40	Intronic (4:5)
Tursiops truncatus	GGGAGUUAGGGGGUGCUGGGCUGGG	38	Intronic (2:3)
Equus caballus	GGGAGUUGGGGAGUGGGGUUUGGG	42	Intronic (3:4)
Ailuropoda	GGGAGACUUCGGGUGUUUGGGGGUGAGGG	40	Intronic (5:6)

melanoleuca			
	GGGAGACUGCGGGUAUUUUGGGGAGAGGG	39	Intronic (5:6)
Pongo abelii	GGGUUGAAGGGCGGCGAGGGGUGGGG	39	Intronic (6:7)
Oryctolagus cuniculus	GGGAGAGUGGGUGGGGGUCCUGGG	40	Intronic (5:6)
Gorilla gorilla gorilla	GGUGGCGGUGG	21	Intronic (1:2)
Sus scrofa	GGGGGUGGGGGUGGGGGUGGGGG	83	Intronic (2:3)
Ovis aries	GGGCUG*GGG*CAAAGGGGGAGGG	41	Intronic (1:2)
Monodelphis domestica	GGGGGUGGGGAGGAGGGAUGGG	40	5'-UTR
	GGGAGAUGGGGGGGGGUAGGUGGG	40	Intronic (2:3)
positive control[2]	AGGGCTAGGGCTAGGGCTAGGG	42	N/A
negative control_A[3]	CGTGGGGAGATTGGGGAGCGCA	0	N/A
negative control_B	GGTGTGCGTGTGCGAGCGAGAGAGAGTGG	0	N/A

[1]The genomic location specifies the intron between the numbered exons

[2]The G-quadruplex structure of this DNA sequence was determined by nuclear magnetic resonance (Protein databank-ID: 2KM3)

[3] All controls were DNA. The negative control_A is a randomised sequence with the same base composition as the positive control. The negative control_B has the same base composition as the *Homo sapiens* GQS.

In Table 2, the sequences that had highest scores within the whole pre-mRNA of respective organism are listed. The G-scores obtained from QGRS mapper for most organisms are comparable to that of the +VE control, with the exception of *R. norvegicus* and *G. gorilla*. The –VE controls have G-score of 0 as they cannot fold into G-quadruplex. Underlined are the guanine residues participating in the G-quartets. The genomic location of the GQSs is also listed in Table 2, with the majority of them being intronic. For instance the *M. musculus'* highest GQS is Intronic (7:8), which is indicative of the intron located between exon 7 & 8. *R. norvegicus* has 6 GQSs with G-scores of 21 each and are at different locations in the gene. *M. domestica* has its highest GQS occurring in the 5' UTR region while *C. Lupus* has its highest scoring GQS located in its 3'-UTR. *P. troglodyte, P. abelii* and *M. domestica* have 2 GQSs with highest G-score from different locations. *S. scrofa* possesses a GQS that has a score of 83, indicative of a very stable G-quadruplex. Quadbase does not have a scoring system unlike QGRS mapper; however the putative sequences predicted by Quadbase correlated with the highest scorers from QGRS mapper.

The QGS listed in Table 2 have been mapped for respective organisms (Appendix I).

With the exception of *F. catus, O. garnettii, T. truncatus, A. melanoleuca, O. cuniculus, G. gorilla* and *O. aries*, which lack UTR regions, every other orthologs that possess UTR regions in their *FXYD1* gene have GQS located in their UTR regions. However, given the low scores, it does not seem likely that these UTR GQS form stable G-quadruplexes when compared to the +VE control's G-score. The UTR GQSs from each ortholog are shown in Table 3.

Table 3: GQS located in UTR regions from the orthologs, revealed by QGRS mapper and Quadbase.

Organism	UTR GQS	UTR QGRS G-Score
H. sapiens	GGCACAGCAGGACGUUUGGGGG*	15
	GGAGCGAUGGAAUCCGGCCAGG**	18
M. musculus	GGGUGGAGCAUCCAGUUCUGGGCCAGGG*	10
	GGUGCACAGCUGGACAUUUGGGGG*	13
	GGAGGGAAAGAGAGCAGGGCAGAGG*	13
C. lupus	GGCGGCGCAGGACCAGCUCUGGAACAGGGG*	18
	GGCACAGCCGGACGUUUGGGGG*	15
	GGCGGUAGAGACACCUGGCGCGAUGG**	11
	GGGCUAGGCUGGGGGGCGGGGGG**	35
	GGGGGCGAAGGGUGGGCUGGGAUGGCCGGG**	42
P. troglodytes	GGCACAGCAGGACGUUUGGGGG*	15
	GGAGCGAUGGAAUCCGGCCAGG**	18
B. taurus	GGCAGCGCAGCCAGCUCUGGGCCAGGGGG*	6
	GGCCCCGGGGCACAGCCGGACGUUUGGG*	20
	GGCCUUCUUUCGGCAGGGG*	19
	GGCGGUAGAGACACCUGGCGCGAUGGG**	11
	GGCUGGGGGAGGGAGGAUAGAGG**	21
	GGGCAAAGGGCUGGGUAGCGGG**	40
R. norvegicus	GGCGGUAGAACCUCCACCUGGCUCCAGG**	8
Felis catus	N/A	N/A

Otolemur garnettii	N/A	N/A
Tursiops truncatus	N/A	N/A
Equus caballus	<u>GG</u>CCCCUG<u>GG</u>CACAGCC<u>GG</u>ACGUUG<u>GGG</u>*	20
Ailuropoda melanoleuca	N/A	N/A
Pongo abelii	<u>GG</u>AGU<u>GG</u>CCAGCCCGA<u>GG</u>CUUCCCA<u>GG</u>*	15
	<u>GGG</u>AG<u>GG</u>AGUGU<u>GG</u>UUGA<u>GG</u>*	19
	<u>GGG</u>UUCUGCA<u>GGG</u>U<u>GGG</u>AUGU<u>GGG</u>*	36
	<u>GG</u>CAG<u>GG</u>UGU<u>GG</u>AGUUU<u>GG</u>*	19
	<u>GG</u>CACCGGA<u>GG</u>CCCAGA<u>GG</u>AGGAGUACU<u>GG</u>*	18
	<u>GGG</u>ACGACGGUGGU<u>GGG</u>CGGGGGCG<u>GGGG</u>**	34
Oryctolagus cuniculus	N/A	N/A
Gorilla gorilla gorilla	N/A	N/A
Sus scrofa	<u>GGGG</u>A<u>GGGG</u>U<u>GGGG</u>U<u>GGGG</u>*	63
	<u>GGG</u>AG<u>GGG</u>ACACCGCUGA<u>GGG</u>C<u>GG</u>*	13
	<u>GGG</u>CCAG<u>GGGG</u>UCCAGCC<u>GG</u>CCGUUU<u>GGG</u>*	21
Ovis aries	N/A	N/A
M. domestica	<u>GGG</u>UG<u>GGG</u>AGGA<u>GGG</u>AU<u>GGG</u>*	40

* represents GQS from 5'UTR regions

** represents GQS from 3'UTR regions

Most orthologs have more than one GQS in their UTR regions that can fold into a G-quadruplex, but are relatively unstable in comparison to the GQS from Table 2. It is seen here that 5 organisms, namely (i) *C. lupus* (ii) *B. taurus* (iii) *M. domestica* (iv) *S. scrofa* & (v) *P. abelii* have GQS of G-scores comparable to the positive control in their UTR regions, indicative of the formation of highly stable G-quadruplexes.

Analysis of fully processed *FXYD1* human and ortholog mRNA did not contain high scoring GQS in comparison to the +VE control. The results for the mRNA of *H. sapiens* are shown in Table 4.

Table 4: GQS predicted by QGRS mapper for *H. sapiens'* fully processed mRNA

Length	GQS	G-Score
26	GGCAGCUGGGCCUCACCCCGGCAGGG	15
13	GGGGGAAGGCAGG	20
30	GGACAAUGGCGUCUCUUGGCCACAUCUUGG	17
28	GGACUGGGGAACCCGAUGAAGAGGAGGG	10
22	GGAGCGAUGGAAUCCGGCCAGG	18

Five potential GQS were predicted, but they have relatively low scores and the highest scoring one has a G-score of 20, which is about half of the score obtained for the +VE control's G-quadruplex. Also the GQSs have only 2 quartets (underlined guanines), making the G-quadruplexes less stable.

3.1.2 Stability calculations of secondary/tertiary structures

The calculations executed by RNAfold, RNAeval and RNAsubopt on the controls, Human_PLM and Bovine_PLM sequences confirmed the potential of the +VE control, Human_PLM and Bovine_PLM sequences to form G-quadruplexes. The calculated and proposed structure based Minimum Free Energy (MFE) calculation is listed in Table 5.

Table 5: Analysis of the oligonucleotide sequences considered for laboratory work by the Vienna RNA Package. The proposed dot bracket notation of the MFE structure generated by RNAfold and other secondary structures by RNAeval and RNAsubopt are shown.

Name	Dot bracket annotation[1]	Free Energy (kcal/mol)	Diversity of MFE structure[2]	Frequency of MFE structure
+VE	.+++...+++...+++...+++	-12.65	0.00	1.001
	....((((.........))))...	-3.80		
	..((((.........))))....	-2.80		
	...(((.........))).....	-2.10		
	....((..(((....))).))..	-2.00		
-VE_A	.(((..............))).	-0.30	4.10	0.259
	....	0.00		
	..((.(............).))	0.30		
	(((..............))))..	0.30		

	Structure[1]	MFE	Diversity[2]	Frequency
	(.(.((....)).).).......	0.30		
-VE_B	.((.(((....))).))..............	-2.70	4.68	0.385
	.((..((....))..))..............	-2.20		
	......(((......)))..........	-1.80		
	....(((.........)))..........	-1.60		
	.....((.........))............	-1.30		
Human	......((.(((.....))).))......	-4.30	1.69	0.476
_PLM	......((.((((...)))).))......	-4.10		
	+++.......+++......+++....+++	-3.51	0.01*	0.504**
	...(..((.(((.....))).))..)...	-2.90		
	...(..((.((((...)))).))..)...	-2.70		
Bovine	+++...+++.....+++....+++	-8.37	0.00	0.125
_PLM	...(.(((....))).).......	-3.20		
	((.(.(((....))).)..))...	-2.50		
	...(.((......)).).......	-2.30		
	.(((.(...).)))..........	-2.20		

[1]The symbols '(' and ')' represent canonical base pairs, '+' represents guanine bases taking part in G-tetrad formation, '.' represents unpaired bases.

[2] = diversity of the proposed structure, the average distance separating bases involved in pairing of the structure

* diversity of the 3rd structure with respect to its MFE for Human_PLM sequence

** frequency of Human_PLM's 3rd structure after its MFE structure

The data presented in Table 5 lists the MFE structures and other secondary structures according to their free energies. The results indicate that the +VE control forms a highly stable G-quadruplex, with a structural diversity of d = 0.00. The intramolecular G-quadruplex formed is the minimum energy requiring structure for the +VE control at -12.65 kcal/mol. For the Human_PLM sequence, G-quadruplex was the 3rd energy favourable entity, giving -3.51 kcal/mol. The highly stable G-quadruplex formed by Bovine_PLM, with diversity of d = 0.0, was its MFE structure, at -8.37 kcal/mol. The structures proposed for –VE_A and -VE_B are quite unstable, with high free energies. The frequencies of the MFE structures vary for the different species. The G-quadruplex for the +VE control is expected to be the only structure present with a frequency of 1.00. The MFE structure proposed for the remaining species will be in equilibrium with other structures as indicated by frequencies < 0.5. All G-quadruplex entities have d = 0.00, which indicate highly stable G-quadruplexes from the +VE, Human_PLM and Bovine_PLM sequences.

The proposed MFE structure for the +VE, -VE_A, -VE_B and Bovine_PLM sequences was obtained by RNAplot and are shown below.

A.
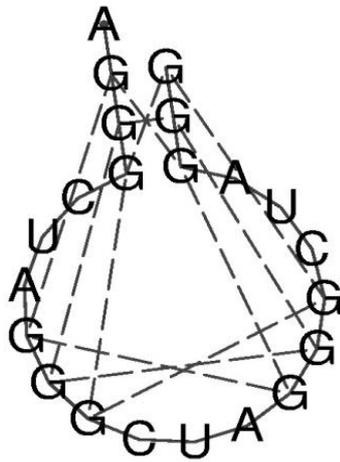
B.
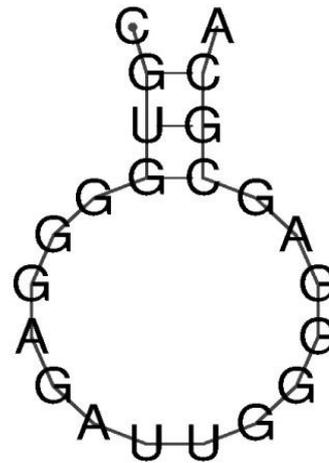
C.
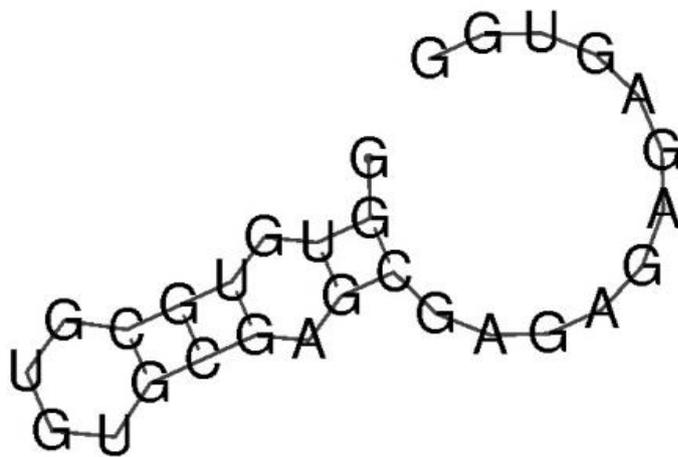
D.
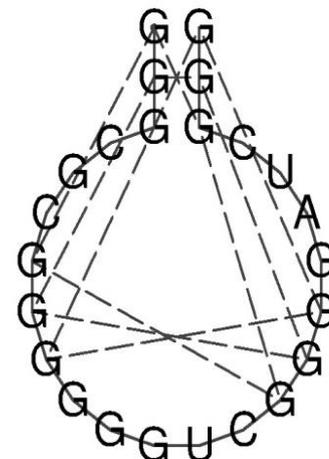
Figure 9: MFE structures generated by RNAplot as calculated by RNAfold and RNAsubopt. **A.** +VE control's intramolecular G-quadruplex at -12.00 kcal/mol. **B.** –VE_A MFE structure at -0.30 kcal/mol. **C.** Proposed –VE_B MFE ensemble at -2.70 kcal/mol. **D.** Intramolecular G-quadruplex entity formed by the Bovine_PLM sequence at – 8.37 kcal/mol.

Graphical plots of the 3 lowest energy state structures for Human_PLM sequence were produced by RNAplot and are shown below.
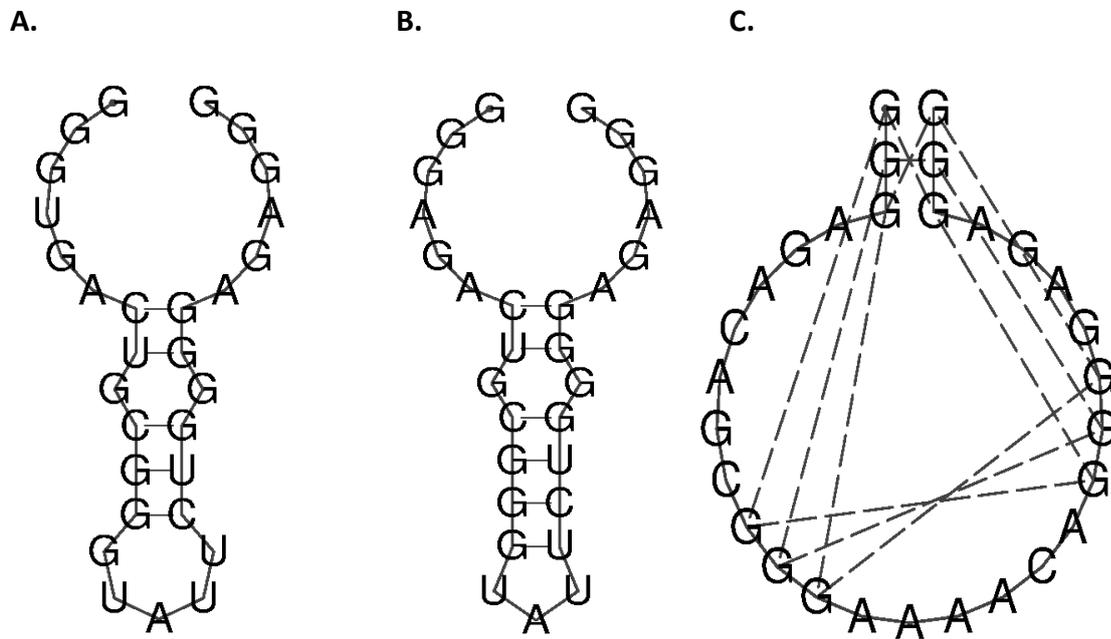
Figure 10: Graphical plot by RNAplot for the proposed secondary and G-quadruplex structures for Human_PLM sequence predicted by RNAfold and RNAsubopt. **(A)** MFE structure at -4.30 kcal/mol **(B)** second lowest energy state structure at -4.10 kcal/mol **(C)** G-quadruplex structure of Human_PLM at -3.51 kcal/mol.

In comparison to the G-quadruplex structure proposed for Human_PLM, the MFE and structure at -4.10 kcal/mol are relatively broader and longer in size. The G-quadruplex is compacter.

These MFE values were used to estimate the relative amount of secondary structure in equilibrium with G-quadruplex structure for Human_PLM, assuming that the MFE values correspond approximately to the free enthalpy of folding ΔG.

$$\Delta G = -RTlnK$$

,where ΔG is the Gibbs free energy, R is the gas constant and T is temperature(R = 1.987 x 10^{-3} kcal K^{-1} mol^{-1}, T = 298 K) and K is the equilibrium constant.

$$K1 = \frac{[G4\ emsemble]}{[Non - folded]}$$

$$K2 = \frac{[2°\ emsemble]}{[Non - folded]}$$

$$\frac{K1}{K2} = \frac{[G4\ ensemble]}{[2°\ emsemble]}$$

$$K = e\left(\frac{-\Delta G}{RT}\right)$$

$$K1 = e\left(\frac{-\left(-3.51\frac{kcal}{mol}\right)}{1.987E - 3\frac{kcal}{kmol}\ x\ 298\ K}\right) \equiv 375.33$$

$$K2 = e\left(\frac{-\left(-4.30\frac{kcal}{mol}\right)}{1.987E - 3\frac{kcal}{kmol}\ x\ 298\ K}\right) \equiv 1425.06$$

$$\frac{375.33}{1425.06} = \frac{G4 \; ensemble}{2° \; emsemble}$$

$$2° \; emsemble = (3.80)G4 \; emsemble \; with \; \Delta G - 4.30 \; kcal/mol$$

Or $\quad\quad 2° \; emsemble = (2.71)G4 \; emsemble \; with \; \Delta G - 4.10 \; kcal/mol$

The calculation revealed that the other two lower energy secondary structures exist at about a fourfold higher concentration than G-quadruplex for Human_PLM. Or in other words, the concentration of the G-quadruplex species takes approximately 25% of the concentration of all molecular species. The two secondary structures are likely to compete against G-quadruplex formation in Human_PLM.

3.1.3 Multiple Sequence Alignment of GQS from Table 2 against orthologous *FXYD1* pre-mRNA sequences

The MSA carried out using the MAFFT server shows G-rich regions that are conserved across the genome of the orthologs, having the ability to fold into G-quadruplexes. *H. sapiens'* GQS from Table 2 was aligned alongside the pre-mRNA sequence of the remaining orthologs (Figure 11).
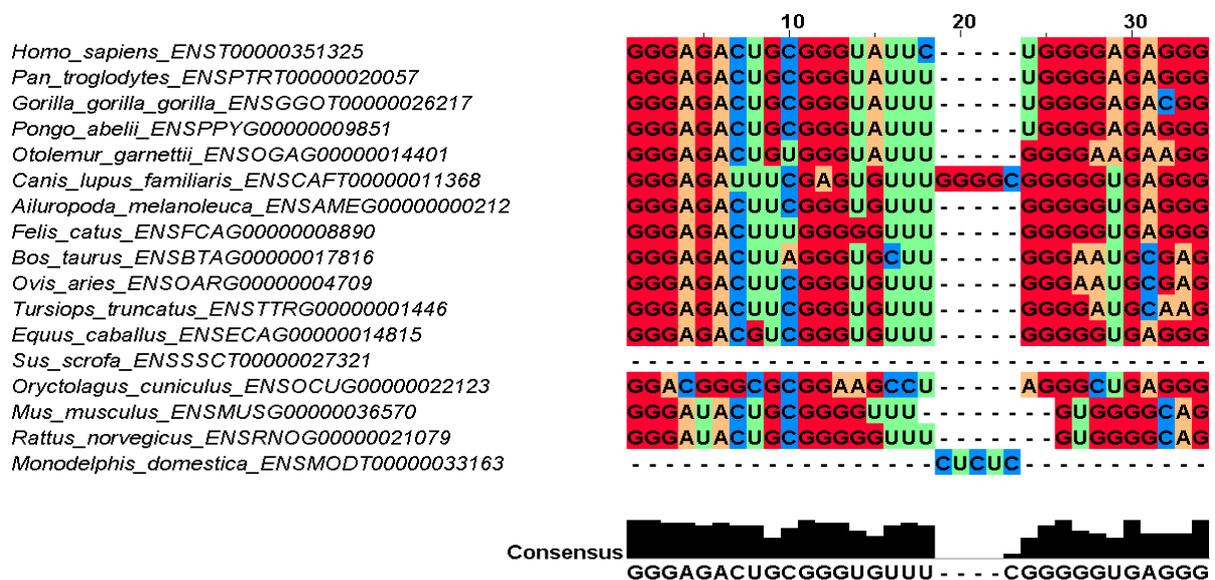
Figure 11: Highest scoring GQS for *H. sapiens* aligned with the remaining orthologs by MAFFT Version 7. Accession numbers are listed next to respective organism's name. Alignment of the *H. sapiens* GQS from Table 2 has indicated conserved sequences within most orthologs that can form G-quadruplexes. The consensus sequence was GGGAGACUGCGGGUGUUUCGGGGGUGAGGG and can form a G-quadruplex with a G-score of 40.

The conserved sequences from each ortholog from Figure 11 have been mapped (Appendix I) onto the *FXYD1* gene of the respective organism and are listed in Table 6.

50

Table 6: Genomic location and G-scores of the conserved sequences with respect to the highest scoring GQS of *H. sapiens*

Organism	Conserved sequence after aligning *H. sapiens* highest scoring GQS	Genomic location	G-score
P. troglodytes	GGGAGACUGCGGGUAUUUUGGGGAGAGGG	Intronic (6:7)	39
G. gorilla	GGGAGACUGCGGGUAUUUUGGGGAGACGG	Intronic (5:6)	20
P. abelii	GGGAGACUGCGGGUAUUUUGGGGAGAGGG	Intronic (6:7)	39
O. garnettii	GGGAGACUGUGGGUAUUUGGGGAAGAAGG	Intronic (5:6)	20
C. lupus	GGGAGAUUUCGAGUGUUUGGGGCGGGGGUGAGGG	Intronic (6:7)	20
A. melanoleuca	GGGAGACUUCGGGUGUUUGGGGGUGAGGG	Intronic (5:6)	40
F. catus	GGGAGACUUUGGGGGUUUGGGGGUGAGGG	Intronic (5:6)	40
B. Taurus	GGGAGACUUAGGGUGCUUGGGAAUGCGAG	Intronic (6:7)	0
O. aries	GGGAGACUUCGGGUGUUUGGGAAUGCGAG	Intronic (5:6)	0
T. truncates	GGGAGACUUCGGGUGUUUGGGGAUGCAAG	Intronic (5:6)	14
E. caballus	GGGAGACGUCGGGUGUUUGGGGGUGAGGG	Intronic (5:6)	40
S. scrofa	N/A	N/A	N/A
O. cuniculus	GGACGGGCGCGGAAGCCUAGGGCUGAGGG	Intronic (6:7)	19
M. musculus	GGGAUACUGCGGGGUUUGUGGGGCAG	Intronic (6:7)	14
R. norvegicus	GGGAUACUGCGGGGGUUUGUGGGGCAG	Intronic (6:7)	16
M. domestica	N/A	N/A	N/A

From Table 6, the conserved sequences for almost every ortholog have the potential to form G-quadruplex and all the sequences are located in introns. The conserved sequences from *B. taurus* and *O. aries* are the only sequences that do not fold in G-quadruplex.

The consensus sequences obtained after aligning each GQS from Table 2 with the pre-mRNA of other orthologs are listed in Table 7.

Table 7: Consensus sequence obtained after aligning each sequence from Table 2 by MAFFT Version 7.0

GQS aligned with pre-mRNA of other orthologs	Consensus Sequence	G-score
H. sapiens	GGGAGACUGCGGGUGUUUCGGGGGUGAGGG	40
P. troglodytes 1st	GGGAGACUGCGGGUGUUUCGGGGGUGAGGG	40
P. troglodytes 2nd	GGGUUGGAGGGCGGCGAGGGGUGGGG	39
G. gorilla	GGUGGCAAGGGU+G	19
P. abelii 1st	GGGAGACUGCGGGUGUUUCGGGGGUGAGGG	40
P. abelii 2nd	GGGUUGGAGGGCGGCGAGGGGUGGGG	39
O. garnettii	GGGCGCUGUGGGGGUGAGGC+GG	19
C. lupus	GGAGGAAGGCGGGAGAGGCA+GGGGCCAAGUGCCAGGGUUGGA	20[1], 14[2]
A. melanoleuca	GGGAGACUGCGGGUGUUUCGGGGGUGAGGG	40
F. catus	GGGAGACUGCGGGUGUUUCGGGGGUGAGGG	40
B. taurus	GGGCGCGGGGGGUUGGAGGGAGGG	37
O. aries	AGGUCAGGCAAAGGUGGGGGG	20
T. truncatus	GGGAG+UGGGAGGGGGGAGGGCCUGGG	41
E. caballus	GGGAG+UGGGAGGGGGGAGGGCCUGGG	41

S. scrofa	--------------------GGGG	0
O. cuniculus	ACUGGAAGAUGGAGGGUUCUGGG	18
M. musculus	ACUAGGCUGGGGGAGGGAGGGAGGGGGGGG	42
R. norvegicus 1st	GCAGGUGGG+CCUUGGG	17
R. norvegicus 2nd	GGAUGGAGGCCGGC	20
R. norvegicus 3rd	GGCGCUGUGGGGG	0
R. norvegicus 4th	GGCAC+GGGAGGUGAAG	18
R. norvegicus 5th	ACUAGGCUGGGGGA	0
R. norvegicus 6th	GGGGGGGAGGA	20
M. domestica 1st	CCUGGC+GGGUGUGGGGUUUGG	20
M. domestica 2nd	GCAAGGGU+GGGGGAAACCCUGCAAGAGAA	0

[1] The G-score obtained for the consensus sequence after aligning the highest scoring GQS of *C. lupus* is 20 after substituting the + with the base G.

[2] The consensus sequence after aligning *C. lupus'* highest scoring GQS has a G-score of 14 after substituting the + with the base A.

Besides *S. scrofa*, the consensus sequences obtained after aligning the highest scoring QGS from Table 2 for every ortholog, have the ability to form G-quadruplexes as seen in Table 6. Two out of the six GQSs from *R. norvegicus* gave consensus sequences that cannot form G-quadruplexes and one out of the two highest scoring GQS from *M. domestica* gave a consensus sequence that does not fold into a G-quadruplex.

3.1.4 Alternative splicing

The comparison between *H. sapiens'* *FXYD1* and the variant *FXYD1*-009 mRNA and pre-mRNA sequences suggests that alternative splicing takes place.

- *FXYD1* mRNA FASTA sequence:

```
guggcagcugggccucaccccggcagggcugugcgugaccccccugagugggggaaggcag
gcuguugccaugguggccugagcgagcagaauuccuccaggGACAAUGGCGUCUCUUGGC
CACAUCUUGGUUUUUCUGUGUGGGUCUCCUCACCAUGGCCAAGGCAGaaaguccaaaggaa
cacgacccguucacuuacgACUACCAGUCCCUGCAGAUCGGAGGCCUCGUCAUCGCCGGG
AUCCUCUUCAUCCUGGGCAUCCUCAUCGUGCUGAgcagaagaugccggugcaaguucaac
cagcagcagagGACUGGGGAACCCGAUGAAGAGGAGGGAACUUUCCGCAGCUCCAUCCGC
CgucuguccacccgcaggcgguagaaacaccuggagcgauggaauccggccagGACUCCC
CUGGCACCUGACAUCUCCCACGCUCCACCUGCGCGCCCACCGCCCCCUCCGCCGCCCCUU
CCCCAGCCCUGCCCCCGCAGACUCCCCCUGCCGCCAAGACUUCCAAUAAAACGUGCGUUC
CUCUCGA
```

FXYD1 amino acid sequence:

```
MASLGHILVFCVGLLTMAKAESPKEHDPFTYDYQSLQIGGLVIAGILFILGILIVLSRRCRCKFNQQQRT
GEPDEEEGTFRSSIRRLSTRRR
```

- *FXYD1*- 009 mRNA FASTA sequence:

```
uuuucugugugggucuccucaccauggccaaggcagAAAGUCCAAAGGAACACGACCCGU
UCACUUACGacuaccagucccugcagaucggaggccucgucaucgccgggauccucuuca
uccugggcauccucaucgugcugaCCCCGCCCCUCGCGAGGGCGAGCUGGAGCUACAGCG
CCGCUUGGCGCCCGCCGGGAGGGAGCCUCAGCUUCUCCUACCUCUCCACGCCCACAGGCA
GAAGAUGCCGGUGCAAGUUCAACCAGCAGCAGAGgacuggggaacccgaugaagaggagg
gaacuuuccgcagcuccauccgccGUCUGUCCACCCGCAGGCGGUAGAAACACCUGGAGC
GAUGGAAUCCGGCCAGgacuccccuggcaccugacaucucccacgcuccaccugcgcgcc
caccgcccccuccgccgccccuuccccagcccugccccgcagacuccccccugccgccaa
gacuuccaauaaaacgugcguuccucucgaca
```

FXYD1-009 amino acid sequence:

```
XFCVGLLTMAKAESPKEHDPFTYDYQSLQIGGLVIAGILFILGILIVLTPPLARASWSYSAAWRPPGGSL
SFSYLSTPTGRRCRCKFNQQQRTGEPDEEEGTFRSSIRRLSTRRR
```

In the mRNA sequences (*FXYD1* & *FXYD1*-009), alternate exons are shown in lower and uppercase. *FXYD1* and *FXYD1*-009 have similar mRNA stretches, highlighted in blue and green. Likewise the amino acids derived from the blue highlighted stretch of bases are highlighted blue, and green highlighted bases code for amino acids highlighted green. The UTR regions are not highlighted. The underlined bases in *FXYD1*-009 mRNA are located in exon 4 of the transcript, which also occur in the intronic region between exons 4 and 5 of *FXYD1* pre-mRNA as shown on the following page.

The FASTA sequence below is that of exon 4-intron4-exon5 of *H. sapiens*' pre-*FXYD1* mRNA. Exon 4 is highlighted pink and exon 5 is highlighted blue. The intronic bases that are bold and underlined are the extra bases that occur in *FXYD1*-009's exon 4, and the strikethrough ones are bases that are spliced.

ACUACCAGUCCCUGCAGAUCGGAGGCCUCGUCAUCGCCGGGAUCCUCUUCAUCCUGGGCAUC
CUCAUCGUGCUGA gu gagugccccuagcucccgcccucuacccgccucucccuggccccac
cucucucuggccccgccucucccuggccccgccucucccuagccccccucucccuggccccg
cuucucccuggucccgccccucccuggccccgccccgccccaacccccucccaggccuugccc
cgccuacccugccuuggguucccccggcccccgggucucgccucuagccccgccccguccccca a
g ccccgccccucgcgagggcgagcuggagcuacagcgccgcuuggcgcccgccgggagggag
ccucagcuucuccuaccucuccacgcccac ag GCAGAAGAUGCCGGUGCAAGUUCAACCAGC
AGCAGAG GUA

The consensus di-nucleotide bases to which sub units U1 and U2 from spliceosome complexes bind at the 5' and 3' ends of introns in order to establish splicing sites have been highlighted green and red or yellow respectively. It has been reported that the highly conserved and invariant di-nucleotide bases GU is the 5' binding site of subunit U1 and AG is the 3' binding site for sub unit U2 (Rogers & Wall, 1980; Shapiro & Senapathy, 1987). The binding of spliceosome sub units U1 at 5'-GU and U2 at 3'-AG will result in a fully spliced

intron 4 from the *FXYD1* pre-mRNA, producing *FXYD1*. Alternatively, the binding of the sub

units at 5'-`GU` and 3'-`AG` will result in a partially spliced intron 4 of the *FXYD1*-pre mRNA,

producing *FXYD1*-009. Hence *FXYD1*-009 is a consequence of alternative splicing of the

intron 4 of *FXYD1* pre-mRNA.

A comparison of the pre-mRNA of *FXYD1* and *FXYD1*-009 was performed by 1000 Genomes

Transcript comparison option. *H. sapiens'* highest scoring GQS,

GGGAGACUGCGGGUAUUCUGGGGGAGAGGG, was compared in the two pre-mRNA transcripts

for mutations. The result is shown in Figure 12.

Figure 12: Part of the pre-mRNA comparison of *FXYD1* (intron6) and *FXYD1*-009 (intron5)

that maps the highest scoring GQS of *H. sapiens*. The highest scoring GQS of *H. sapiens'* pre-

mRNA is highlighted blue.

The GQS is present in both *FXYD1* and *FXYD1*-009. However, two variations within the

sequence of *FXYD1*-009 are present. Base A in the second loop that was a A/G variant and

base G from the fourth quartet, which was a G/A variant as seen in Figure 12. The G/A

variant is more likely to affect the G-quadruplex structure than the A/G variant as the A/G is

a loop base instead of G/A from a quartet. The sequence of the variant-009 after G → A

substitution is GGGAGACUGCGGGUAUUCUGGGGGAGAGAG.

Analysis of the variant sequence by QGRS mapper revealed a GQS (<u>GG</u>AGACUGC<u>GG</u>GUAUUCU<u>GGGG</u>) with a G-score of 14, which is very low in stability. Further analysis by the Vienna RNA Package did not predict G-quadruplex formation by the variant sequence from *FXYD1*-009 pre mRNA. The results obtained after analysis by RNAfold and RNAsubopt are shown in Table 8.

Table 8: Dot bracket annotations of the MFE and secondary structures of the variant *FXYD1*-009 sequence (GGGAGACUGCGGGUAUUCUGGGGAGAGAG) by RNAfold and RNAsubopt.

Dot bracket annotation	Free energy (kcal/mol)	Frequency of MFE	Diversity of MFE
......((.(((.....))).))......	-4.30	0.520	2.00
......((.((((...))))).))......	-4.10		
((((.(((...))).))))..........	-1.70		
...............((((....))))..	-1.50		
((((..(....)...))))..........	-1.40		
.(((((((...))).))))..........	-1.30		

The data presented in Table 8 indicate that RNAfold did not predict G-quadruplex formation for the variant sequence, even though the –g option was used in the command lines. GQS that were predicted to form low stability G-quadruplex with low G-scores from the *H. sapiens*' pre-mRNA were also predicted not to form G-quadruplex by the Vienna RNA Package Data not shown). The MFE structure has free energy of -4.30 kcal/mol and structural diversity of 2.00, indicative of high instability.

The data obtained from the comparative analysis suggest the likelihood of the intronic G-quadruplex between exon 6 and 7 of *H. sapiens FXYD1* to play a major role in the splicing of the intron occurring between exons 4 and 5, impacting on the formation of *FXYD1*-009

3.2 G-quadruplex detection by Native PAGE

Comparison of R_f values of samples in the presence and absence of K^+ supports conformational changes in the +VE, Human_PLM and Bovine_PLM sequences (Figure 13). 30% Native PAGE gels were run at 140 mV with oligonucleotides of final concentration 3 µM. The results obtained after exposure in the presence of SYBR Green I RNA stain S9430 and SYBR Green I nucleic S32717 are shown in Figure 13A-C.
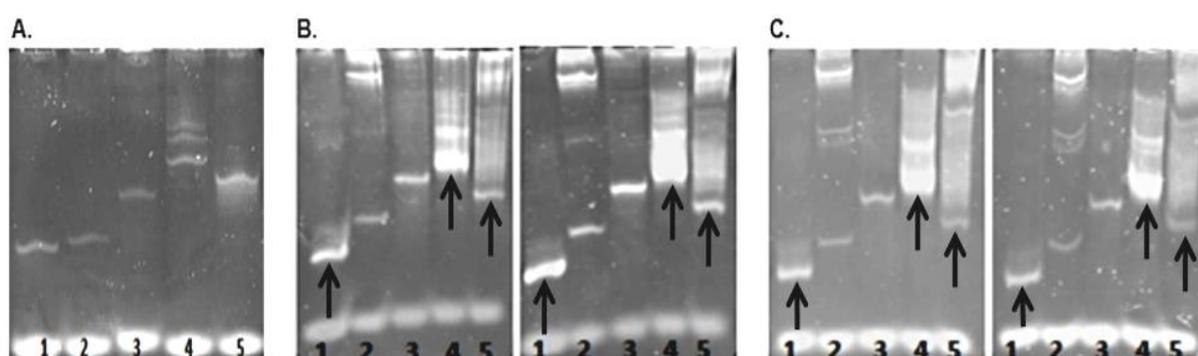
Figure 13: Native 30% PAGE of GQS oligos (Table 5) in absence and presence of K+. Lanes**1:** +VE**; 2:**-VE_A**; 3:**-VE_B**; 4:**Human_PLM**; 5:**Bovine_PLM**. A.**30% PAGE loaded with samples incubated in 0.02 M TrisOAc buffer solution only. **B.** 30% PAGE loaded with samples incubated in 0.02 M TrisOAc and 0.05 M K^+ buffer solution on two separate gels, duplicates. **C.** Two separate 30% PAGE, duplicates, loaded with samples incubated in 0.02 M TrisOAc and 0.10 M K^+ buffer solution. Arrows in **B** & **C** point putative G-quadruplexes. The control samples in lanes 1, 2 and 3 under K^+ free conditions and samples in lanes 2 & 3 in the presence of K^+ acted as markers.

Under K+ free condition, the +VE and -VE_A controls migrated at the same rate on the gel. -VE_B migrated less than other controls but faster than Human_PLM and Bovine_PLM. Human_PLM was the slowest migrating sample. In the presence of 0.05 M K^+, the +VE

control was the fastest migrating sample followed by the Bovine_PLM that migrated faster than -VE_B. In the order of fastest to slowest migrating sample: +VE > -VE_A > Bovine_PLM > -VE_B > Human_PLM is observed for samples in 0.05 M K^+. Samples incubated in 0.1 M K^+ migrated with a similar trend as samples incubated in 0.05 M K^+. The tracking dye is at the bottom of the gels in A-C. The DNA species under K^+ free conditions (Fig 13A) produced a distinct single band as well as Bovine_PLM, except for Human_PLM that produces 3 bands. In K^+ containing buffer, the +VE and –VE_B controls produced single bands on the gels (Fig 13B & 13C), while –VE_A produced several bands and Bovine_PLM and Human_PLM produced smears.

The ratio of the distance migrated by each sample with respect to that of the tracking dye (R_f value) on the gel is represented graphically for 5 separate experiments.
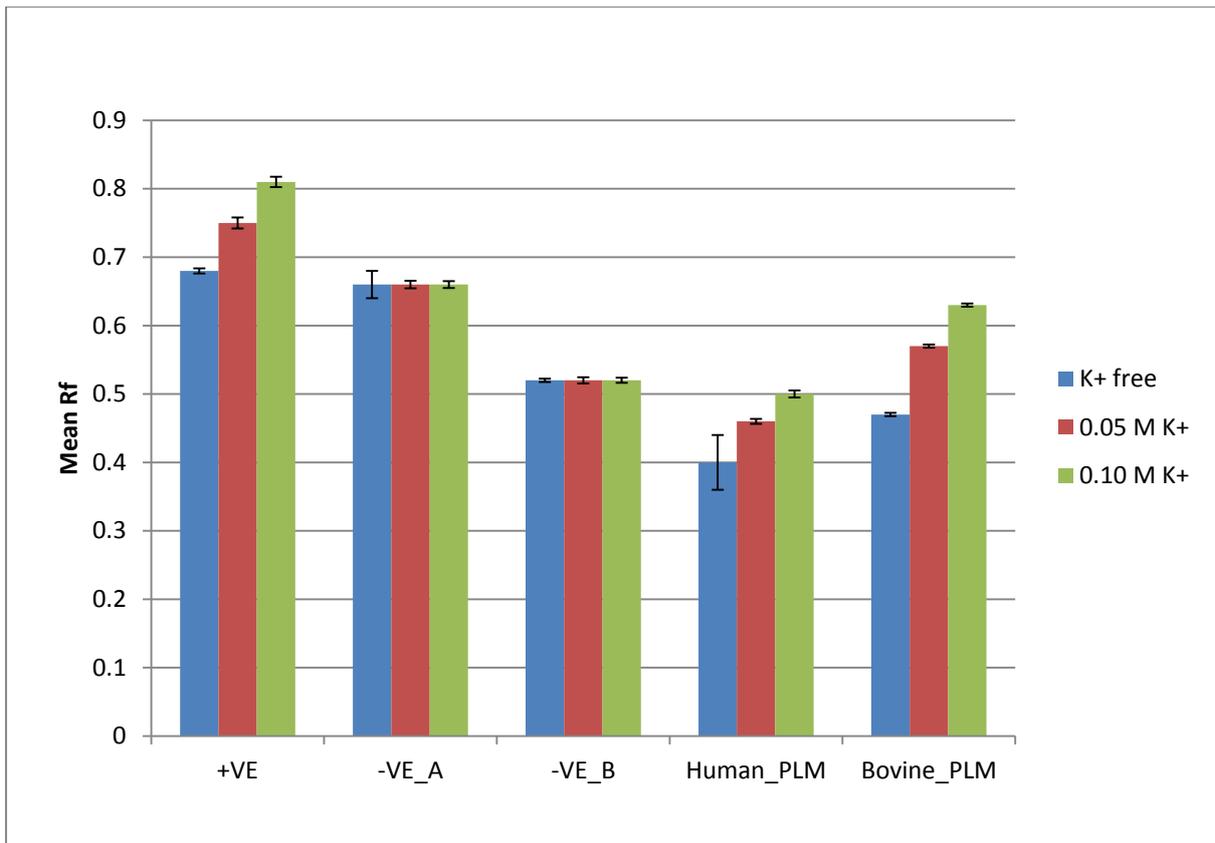
Figure 14: Comparison of the relative migration distance, R_f, obtained from native PAGE experiments for samples treated in K^+ free and K^+ containing buffer. The error bars show the standard error (n=5 experiments).

The +VE control, Human_PLM and Bovine_PLM samples migrated significantly faster when in the presence of K^+ than when in K^+ free buffer. At higher K^+ concentration the +VE, Human_PLM and Bovine_PLM samples migrated even faster. −VE_A and −VE_B samples in K^+ buffer migrated by the same rate in comparison to their respective counterparts in K^+ free buffer.

Table 9: Student 2-tailed-t-test of R_f values for samples in the presence of K^+ containing buffer against samples in K^+ free buffer(n=5)

Sample	+VE	-VE_A	-VE_B	Human_PLM	Bovine_PLM
0p-value for 0.05 M K+	4.18E-05	0.74044	0.45537	3.34E-06	1.97E-07
p-value for 0.10 M K+	4.23E-07	0.72447	0.34053	3.09E-07	5.33E-11

Table 10: Student 2-tailed-t-test of R_f values for samples incubated in different concentration of K^+ containing buffer(n=5)

Sample	+VE	-VE_A	-VE_B	Human_PLM	Bovine_PLM
p-value	0.00153	0.60751	0.67254	0.00032	3.83E-07

At a confidence level of 5%, the difference in migration for the +VE Control, Human_PLM and Bovine_PLM in the presence of K^+ compared to their respective K^+ free incubated counterparts are statistically significant as they support p values < 5% as seen in Table 9. –VE_A and –VE_B have p values > 5%, ruling out the fact that the differences in migration of these samples in the presence or absence of K^+ are significant. This significant difference in migration supports G-quadruplex formation in the +VE control, Human_PLM and Bovine_PLM samples. Comparison between similar samples at different K^+ concentrations (Table 10) supported p values < 5% for the +VE control, Human_PLM and Bovine_PLM suggesting G-quadruplex formation depends on availability of K^+ ions.

The slow migrating band of –VE_A can be interpreted as intermolecular G-quadruplex formation between four strands of oligonucleotides; this is also supported by the fluorescence data shown in Figure 15.

3.3 Detection of G-quadruplexes by Fluorescence spectroscopy

The recorded spectrum of the controls, Human_PLM and Bovine_PLM sequences in the presence and absence of K^+ ions are shown below.
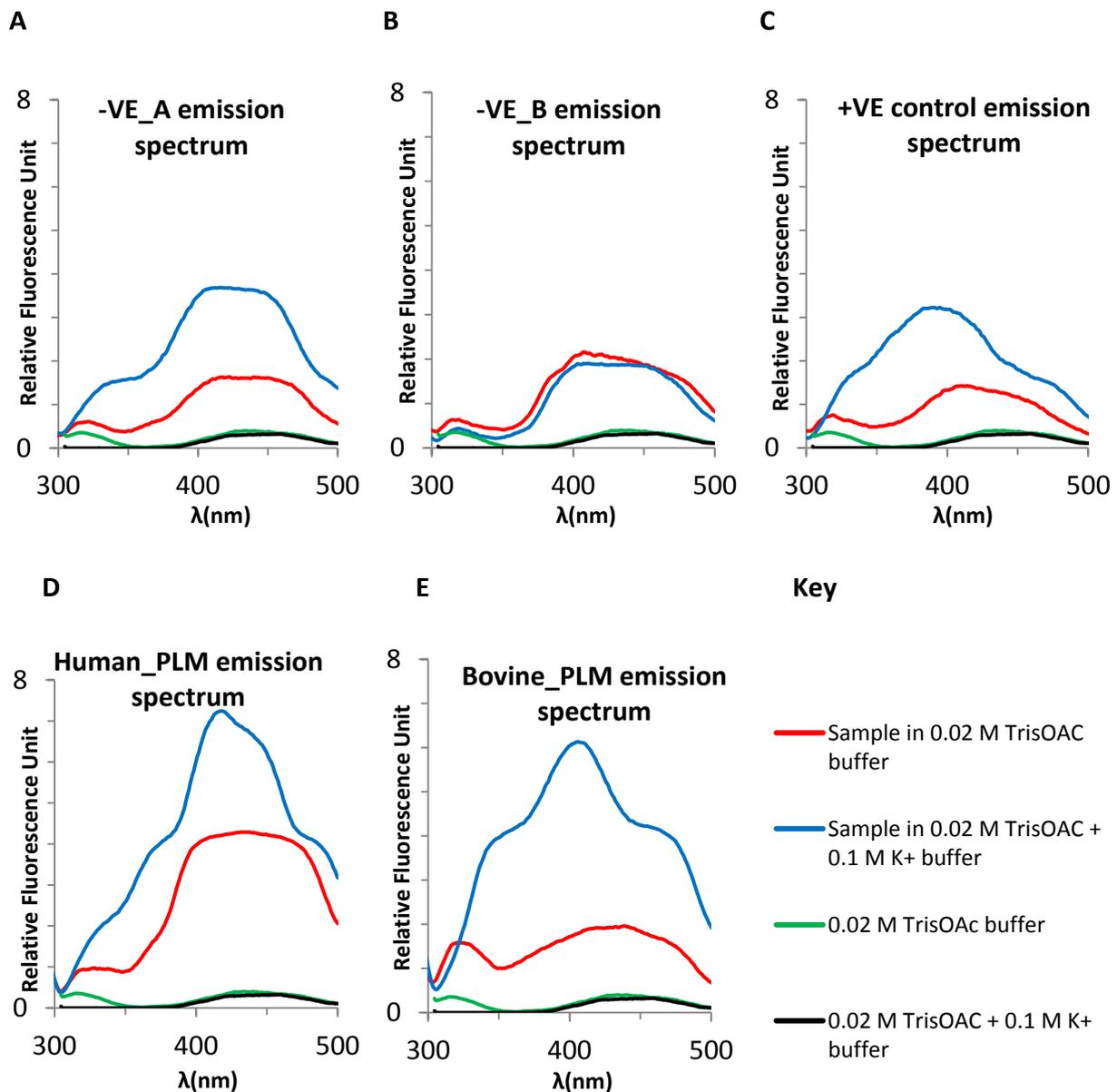
Figure 15: Emission spectra of samples recorded over the range 300-500 nm, excited at 260 nm. The oligonucleotide concentrations were 5.0 µM for positive and negative control DNA samples 1.5 µM for RNA samples.

The fluorescence intensity of –VE_A shows an unexpected increase upon addition of 0.1 M K$^+$. This indicates the possibility that –VE_A can form G-quadruplex structures. As detailed in the discussion, the formation of an *inter*molecular G-quadruplex formed from four strands of DNA is possible with –VE_A. The fluorescence spectra of -VE_B (Fig. 15B) in the presence or absence of K$^+$ do not show significant differences. This is indicative that –VE_B cannot fold into a G-quadruplex as it was predicted by the *in-sillico* studies. The emission spectra of the +VE control shows an increased fluorescence intensity in the presence of K$^+$ (Fig. 15C). This increase is most likely caused by intramolecular anti-paralell G-quadruplex structures formed by the +VE control. The increase of the peak maximum of the +VE control in K$^+$ buffer is ≈ 2.0 relative intensity units. The emission spectrum of Human_PLM over the range 300-500 nm (Fig. 15D) shows an increase of the fluorescence emission intensity upon addition of K$^+$ containing buffer. Notably, in the absence of K$^+$, the Human_PLM RNA has a higher fluorescence emission intensity than other samples with a peak maximum of ≈ 4.0, albeit there is a clear difference in K$^+$ buffer conditions with a maximum of ≈ 7.5. Possibly canonical base-pairing due to secondary structure formation contributes to the higher fluorescence intensity in the absence of K$^+$. The Bovine_PLM RNA sample in 0.1 M K$^+$ buffer shows a clear increase of fluorescence from 2.0 in K$^+$-free conditions to ≈6.1 in K$^+$ buffer. The spectra of the buffers were also recorded and had a relatively low fluorescence emission compared to the oligonucleotides.

In order to assess potential fluorescence quenching effects of KCl on the fluorescence, emission spectra of quinine were measured with and without potassium ions in the buffer and at two different excitation wavelengths (250 and 350nm).
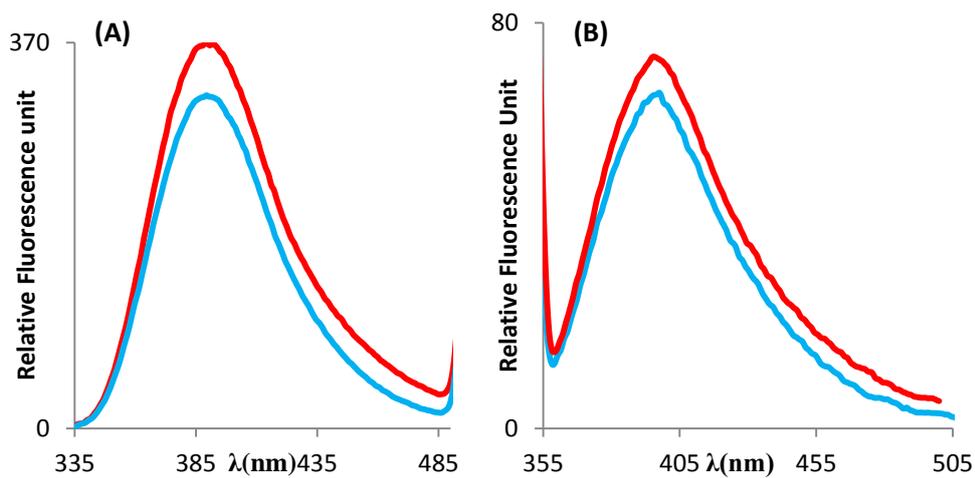
Figure 16: Emission spectra of Quinine at 0.6 ppm in the presence of 0.02 M TrisOAc buffer only (Red curve) or 0.02 M TrisOAC + 0.1 M K$^+$ buffer(Blue curve). **(A)** λ_{ex} = 250 nm, **(B)** λ_{ex} = 350 nm.

The spectra shown in figure 16A and B indicate a small reduction of fluorescence due to the presence of 0.1 M KCl. It should be noted that this reduction effect is opposite to the fluorescence enhancement seen in the G-quadruplex fluorescence experiments.

4. DISCUSSION

4.1 Computational sequence Analysis

Computational analysis by QGRS mapper and Quadbase on *FXYD1* and orthologous pre-mRNA has revealed sequences that can fold into G-quadruplex. The G-scores generated by QGRS mapper for *FXYD1* and ortholog sequence GQS shown in Table 2 are comparable to that of the +VE control, indicative of a stable G-quadruplex, except for *R. norvegicus* and *G. gorilla* with G-scores of 21. Looking back at Kikin's folding motif, it can be seen that all the ortholog *FXYD1* GQS, excluding those of *R. norvegicus*, *G. gorilla* and the –VE controls, have three successive guanines. This indicates the existence of three stacks of G-quartets in the G-quadruplex of these ortholog GQSs as seen in the +VE control DNA. *G. gorilla's* highest scoring GQS was <u>GG</u>U<u>GG</u>C<u>GG</u>U<u>GG</u>, with a G-score of 21 and as per Kikin's folding motif this particular sequence has only two stacks of G-quartets. Similarly, *R. norvergicus* has 6 GQS of G-score 21 and each GQS has only 2 stacks of G-quartets participating in G-quadruplex formation as seen in Table 2, hence accounting for low stability G-quadruplex from *R. norvegicus*. Stability of G-quadruplexes is enhanced by more G-quartets (Kikin *et al*., 2006), while loop size has a smaller effect. GQSs having at least three guanine tetrads and loops of equal length connecting them, will be highly stable and have high G-scores (Kikin *et al*., 2006). The GQS obtained for *S. scrofa*, <u>GGGGG</u>U<u>GGGGG</u>U<u>GGGGG</u>U<u>GGGGG</u>, has a G-score of 83, which makes its G-quadruplex twice as stable as that of the +VE control. *S. scrofa* has 5 G-quartets that stack on top of each other to form a G-quadruplex that has loops of equal length of 1 base each. This makes the G-quadruplex from *S. scrofa* highly stable. On the

other hand, both –VE_A and –VE_B have G-scores of 0 and this means that these sequences were not predicted to form any intramolecular G-quadruplexes.

Additionally, there were putative GQSs within the untranslated (UTR) regions of all the orthologs from the *FXYD1*pre-mRNA that have the potential to fold into G-quadruplexes but the G-scores for the majority of these GQS as seen in Table 3 do not compare well with the +VE control, resulting in G-quadruplexes that have low stability. As previously reported, the high occurrence of G-quadruplex in UTR regions leads to hypothesizing on their role as translational regulators (Huppert *et al*., 2008, Bugaut *et al.,* 2012), the *FXYD1* gene in this instance. G-quadruplexes of low stability could support a rapid folding and unfolding of G-quadruplex ensembles and thus support the conformational heterogeneity within the UTR regions. Instead of inhibiting translation, this could support translation of the *FXYD1* gene. The 5'-UTR region contains the ribosomal binding site and low stability G-quadruplexes at that site can ensure that translation is not perturbed, as it would have been if highly stable G-quadruplexes or secondary structures are formed within that region. Alternatively, under stress conditions such as cell growth, mitosis etc., where cap-dependent translation is compromised at the 5'UTR, G-quadruplex formation can assist initiation of translation of the *FXYD1* gene via cap-independent translation (Bugaut *et al*., 2012).

4.2 Stability calculations of secondary/tertiary structures

The +VE and Bovine_PLM GQS were both predicted to form highly stable G-quadruplexes by minimum free energy calculations using the Vienna RNA package. Bovine_PLM's G-quadruplex is the minimum free energy (MFE) structure, which is the most stable structure with a free energy of -8.37 kcal/mol in comparison to other secondary structures predicted. The +VE control was predicted to form a very stable G-quadruplex (-12.65 kcal/mol) with a frequency of 1.00 in the structural ensemble. This indicates that the +VE control was a suitable positive control for further studies. Two lower energy state secondary structures were predicted to compete against G-quadruplex formation for Human_PLM. The equilibrium constant of the G-quadruplex formed by the Human_PLM with respect to the two competing structures indicates a significant proportion of G-quadruplex structure present, which may increase upon increasing potassium concentration. Note that the energy model of the Vienna RNA package for G-quadruplex structures did not take the potassium ion concentration into account (Lorenz *et al*., 2012). The data obtained from minimum free energy calculations suggest that Human_PLM will form a mixture of secondary and G-quadruplex structures.

4.3 Evolutionary conservation of G-rich sequences in *FXYD1* pre-mRNA

The evolutionary trait of G-rich sequences in the *FXYD1* gene was confirmed by the MSA experiment. The alignment of the *H. sapiens* GQS from Table 2, was found conserved among all orthologs except in *M. Domestica* & *S. scrofa* and the consensus sequence obtained has the ability to fold into G-quadruplex. Consensus sequences obtained from Table 7, with the exception of *S. scrofa* can form G-quadruplexes, indicating that G-rich sequences among the orthologs are conserved. The existence of evolutionary conserved GQS based on a pairwise alignment of two sequences has been proposed as a method of validation and emphasis of their functional significance (Menendez, Frees & Bagga, 2012). The presence of G-rich sequences in orthologs points to an evolutionary conservation of that feature, which supports the hypothesis that G-quadruplex formation is a control mechanism of *FXYD1* pre-mRNA processing.

4.4 Alternative splicing

G-quadruplexes have been reported to regulate gene expression *in-vivo* at the translational level via alternative splicing (Gomez *et al.*, 2004; Marcel *et al.*, 2011). The comparative analysis suggests that the G-quadruplex formed in intron 6 of the *H. sapiens*' pre-mRNA could be affecting the splicing pattern of intron 4 in the *FXYD1* pre-mRNA. The analysis suggests that the presence of a G-quadruplex in intron 6 is promoting the full splicing of intron 4. On the other hand the absence or presence of a G-quadruplex of low stability is causing partial splicing of intron 4, which will lead to the production of the variant *FXYD1-*

009. As reviewed by Clancy in 2008, the consensus 5'-GU and 3'-AG are the binding sites for spliceosome sub units, which determine splicing points in introns. The downstream G-quadruplex in intron 6 ensures that the sub unit U2 from the spliceosome complex binds to the most 3'-AG in intron 4 and ensures the latter is fully spliced. The absence or a lowly stable G-quadruplex in intron 6 causes the sub unit U2 to bind to an alternate 3'-AG, rather than the most 3'-AG, resulting in a longer transcript, *FXYD1*-009.

4.5 Laboratory experimental results support G-quadruplex formation

G-quadruplex formation in the +VE, Human_PLM and Bovine_PLM GQS were successfully detected by 30% native PAGE. Under K^+ free condition (Figure 13 A), the +VE control and -VE_A samples migrated almost a similar distance on the gels as they are both 22 bases long. The -VE_B, which is 29 bases long migrates slower than the +VE and -VE_A control sample under K^+ free condition. Under similar conditions, the 29 bases long Human_PLM and 24 bases long Bovine_PLM samples migrated slower than the control samples. The Bovine_PLM sample was expected to migrate faster than the -VE_B sample. Under the non-denaturing conditions used here, samples not only migrate according to their size but also according to their shape. RNA under normal physiological conditions form loops that makes RNA behave like longer molecules on gels in comparison to DNA molecules of same size (Rio, Ares, Hannon & Nilsen, 2010). In lane 3 from Figure 13 A, Human_PLM produced three distinct bands that moved at different rates on the K^+ free gel. It is proposed that the three bands are due to linear RNA and the two secondary structures predicted by minimum free energy calculations. Addition of K^+ (Figures 13 B & C) altered the migration properties of the +VE, Human_PLM and Bovine_PLM. The +VE control migrated fastest, while under K^+ free

condition it has the same mobility as -VE_A. Bovine_PLM also migrated faster than -VE_B, which would seem opposite under K^+ free conditions. Comparing the R_f of similar species under K^+ free and K^+ conditions from Figure 14 supports the fact that the -VE controls did not change structures and rather stayed in their linear conformations. *Intra*-molecular G-quadruplexes are compact in shape and confer high mobility rates in gels in comparison to linear species and *inter*-molecular G-quadruplexes (Williamson, Raghuraman & Cech, 1989; Bryan&Baumann, 2011). The dependence of R_f on the K^+ concentration strongly supports that intramolecular G-quadruplex formed by the +VE control, Human_PLM and Bovine_PLM sequences. The -VE_A sample (lanes 2 in Figure 13) showed in addition to the expected fast migrating band a slow migrating band at high molecular mass in presence of potassium. This can be attributed to the formation of *inter*molecular G-quadruplexes, which is possible in -VE_A. The two stretches of four consecutive guanines in the sequence of –VE_A (CGT**GGGG**AGATT**GGGG**AGCGCA) can participate in the formation of intermolecular G-quadruplexes (Figure 17). At a final oligonucleotide concentration of 3µM, Moon *et al.*(2007) reported that intermolecular G-quadruplex formation is favoured.
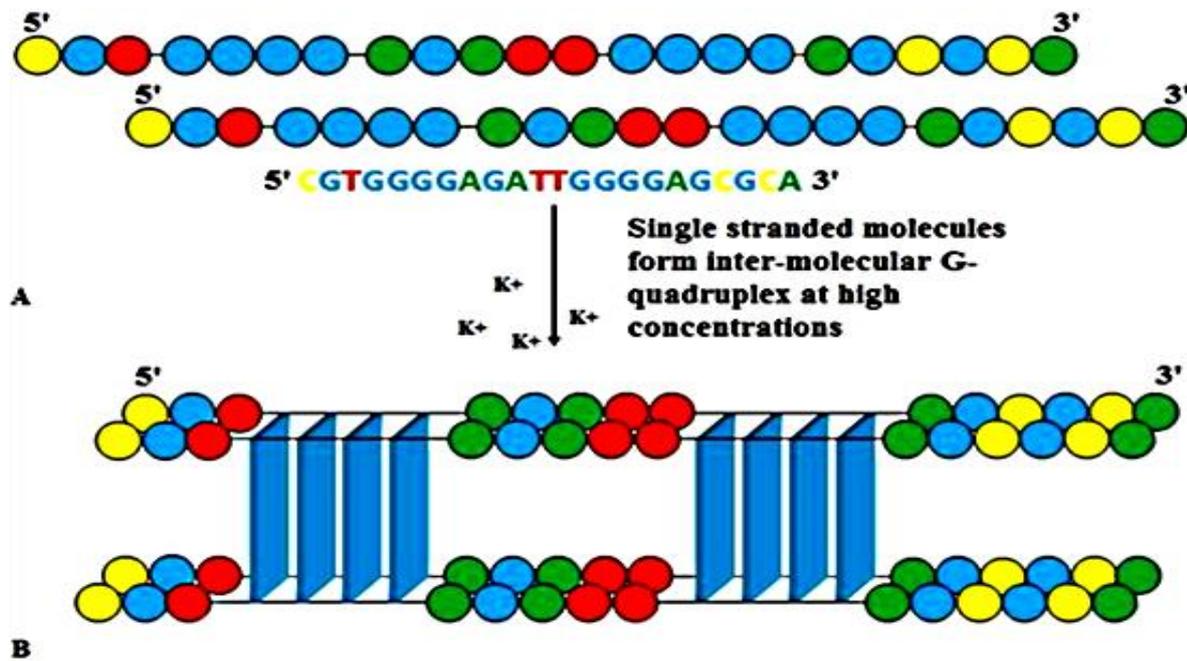
Figure 17: schematic illustration of the *inter*molecular G-quadruplex formed by −VE_A. **A.** single −VE_A species present in high amount, 3µM for the Native PAGE experiments, associate to form intermolecular G-quadruplex in the presence of K⁺. Circles represent unpaired bases and are colour coded according to −VE_A's sequence. **B.** Two sets of four consecutive guanines from four separate strands of −VE_A arrange into G-quartets (blue rectangles) to form a tetrameric parallel intermolecular G-quadruplex.

The proposed intermolecular G-quadruplex by -VE_A makes it difficult for the ensemble to move along the gel, resulting in slow migrating bands. In the K⁺ containing gel, the smeared bands of Human_PLM in lanes 4 (Figure 13 B & C) may be explained by the formation of other secondary structures due to Watson-Crick base pairing. The intramolecular G-quadruplex formed by Human_PLM is the fastest migrating structure in comparison to the other structures, as the G-quadruplex is compacter than the other structures proposed in Table 5. Similarly, the smear pattern by Bovine_PLM in the presence of potassium could be due to the formation of secondary structures predicted in Table 5.

Further confirmation of G-quadruplex formation was achieved by exploiting the intrinsic fluorescent properties of nucleic acid. The emission intensities of the +VE, Human_PLM and Bovine_PLM samples in the presence of 0.1 M K^+ was significantly higher compared to K^+ free buffer (Figures 15 A-E). This was due to the formation of G-quadruplexes in these species. G-quadruplex entities have been reported to have increased intrinsic fluorescence emission in contrast to non-G-quadruplex complexes due to the stacking of G-tetrads (Nguyen Thuan *et al.*, 2011; Kwok, Sherlock, & Bevilacqua, 2013). The higher fluorescence intensity of Human_PLM in K^+ free buffer could be due to the presence of other secondary structures as computed by the Vienna RNA package. Nonetheless, the fluorescence intensity in the presence of K^+ was clearly increased, which confirms the formation of G-quadruplexes. The -VE_A sequence showed higher fluorescence intensity in K^+ containing buffer compared to K^+ free buffer most likely due to the formation of *inter*molecular G-quadruplex (Figure 17), as was seen earlier in the native PAGE experiment. The presence of eight potential tetrads supports the high fluorescence intensity of the intermolecular G-quadruplex of -VE_A. Measuring the fluorescence emission spectrum of quinine in the same buffers, indicated that the K^+ containing buffer had a weak quenching effect on fluorescence. Hence, RNA samples in K^+ containing buffer were expected to show slightly less fluorescence than their respective counterparts incubated in K^+ free buffer, if they would assume the same structure. This was observed for the -VE_B control sample, which in the presence of K^+ had slightly reduced fluorescence intensity (Figure 15B).

In conclusion, using a computational scan of the *FXYD1* pre-mRNA potential G-quadruplex forming sequences (GQS) were identified in *Homo sapiens*, *Bos taurus* and other orthologs. Through energy calculations it was established that the G-quadruplex was either the most stable structure or existent in a significant proportion next to secondary structures. The stability of these G-quadruplex structures is likely higher *in vivo* considering the intracellular K^+ concentration of 120-150 mM. Using native PAGE and fluorescence emission spectroscopy the theoretical calculations were confirmed and the existence of G-quadruplex structures established. Multiple sequence alignment of ortholog GQS indicated that the G-quadruplex forming potential may be conserved in evolution, rendering it possible that it may occur *in vivo* as a mechanism to control phospholemman expression levels and ultimately the activity of the cardiac sodium-potassium ATPase.

4.6 Limitations and further work

Overall, the Native PAGE experiments were challenging due to the low molecular mass samples and electrophoresis in the presence of ionic species, which caused heating of the gel due to increased conduction. A high percentage acrylamide gel was used, as lower percentage gels would cause the +VE control's G-quadruplex to migrate faster than the tracking dye. This was observed in 20% and 25% acrylamide gels (data not shown). The heat generated during electrophoresis, mostly in the K^+ containing buffers may interfere with the electrophoresis and affect the migration of the samples. Often the heat caused the glass plate used to encase the gels to break and the gels were discarded. Heat also caused the voltage of the power supply to fluctuate, which also affected the process of electrophoresis.

The heat issue was addressed by using buffers pre-chilled at 4^0C, which required a longer running time for the gels.

Apart from technical challenges and limitations, a more fundamental limitation is the relevance of the results obtained on short oligonucleotides for the longer pre-mRNA transcript *in vitro* and ultimately the existence of G-quadruplexes of *FXYD1* pre-mRNA *in vivo*. Once the existence of G-quadruplex structures in vivo has been established, the functional consequences on phospholemman expression need to be investigated. Therefore, the present study provides the basis for extensive further work in this area.

Further work should investigate the formation of G-quadruplex structures in longer oligonucleotides using gel electrophoresis, NMR, intrinsic fluorescence and fluorescence - resonance energy transfer (FRET). The formation of G-quadruplex should be investigated *in-vivo* as previously described (Xu *et al*., 2010). Single-molecule FRET can also be used to establish the dynamics and stability of the G-quadruplex, as for example in the work by (Okumus & Ha, 2010; Ying, Green, Li, Klenerman, & Balasubramanian, 2003). A modified construct of the Human_PLM sequence containing an acceptor molecule at one of its end can be used, alongside a complementary strand that will be covalently linked to a glass surface and also modified to contain a donor molecule. Hybridisation of the Human_PLM oligo to the complementary oligo, followed by the formation of a G-quadruplex will allow energy exchange between the donor and acceptor molecule and this can be detected by using Total Internal Reflection Microscopy (TIRM).

Alternative splicing of intron 4 of *FXYD1* has been linked to *FXYD1*-009 formation. G-quadruplexes have been reported in the past to influence splicing (Marcel *et al.*, 2011, Gomez *et al.*, 2004). Could G-quadruplex formation be the influential factor behind variant 009? Further work, similar to Marcel *et al.*, (2011), should address the consequences of G-quadruplex structure on the splicing of pre-mRNA. This can be addressed with constructs using the reporter gene Green Fluorescence Protein (GFP). A suitable construct would include encode an *FXYD1*-GFP fusion protein, while a stop codon is included in the particular intron under investigation. Alternatively the expression levels of mature mRNA species could be measured with quantitative PCR techniques.

5. REFERENCES

Adrian, M., Heddi, B., & Anh Tuan, P. (2012). NMR spectroscopy of G-quadruplexes. *Methods, 57*(1), 11-24.

Beaudoin, J. D., & Perreault, J. P. (2010). 5'-UTR G-quadruplex structures acting as translational repressors. *Nucleic Acids Research, 38*(20), 7022-7036.

Beevers, A. J., & Kukol, A. (2006). Secondary structure, orientation, and oligomerization of phospholemman, a cardiac transmembrane protein. *Protein Science, 15*(5), 1127-1132.

Beevers, A. J., & Kukol, A. (2007). Phospholemman transmembrane structure reveals potential interactions with Na+/K+-ATPase. *Journal of Biological Chemistry, 282*(45), 32742-32748.

Biffi, G., Tannahill, D., McCafferty, J., & Balasubramanian, S. (2013). Quantitative visualization of DNA G-quadruplex structures in human cells. *Nature Chemistry, 5*(3), 182-186.

Bossuyt, J., Despa, S., Martin, J. L., & Bers, D. M. (2006). Phospholemman phosphorylation alters its fluorescence resonance energy transfer with the Na/K-ATPase pump. *Journal of Biological Chemistry, 281*(43), 32765-32773.

Bryan, T. M., & Baumann, P. (2011). G-Quadruplexes: From Guanine Gels to Chemotherapeutics. *Molecular Biotechnology, 49*(2), 198-208.

Bugaut, A., & Balasubramanian, S. (2012). 5 '-UTR RNA G-quadruplexes: translation regulation and targeting. *Nucleic Acids Research, 40*(11), 4727-4741.

Chang, T.-C., & Chang, C.-C. (2010). Detection of G-quadruplexes in cells and investigation of G-quadruplex structure of d(T2AG3)4 in K+ solution by a carbazole derivative: BMVC. *Methods in molecular biology (Clifton, N.J.), 608*, 183-206.

Clancy, S. (2008) RNA splicing: introns, exons and spliceosome. *Nature Education* 1(1):31

Crambert, G., Fuzesi, M., Garty, H., Karlish, S., & Geering, K. (2002). Phospholemman (*FXYD1*) associates with Na,K-ATPase and regulates its transport properties. *Proceedings of the National Academy of Sciences of the United States of America, 99*(17), 11476-11481.

da Silva, M. W. (2007). NMR methods for studying quadruplex nucleic acids. *Methods, 43*(4), 264-277.

Daehnel, K., Harris, R., Maddera, L., & Silverman, P. (2005). Fluorescence assays for F-pill and their application. *Microbiology-Sgm, 151*, 3541-3548.

Du, Z., Zhao, Y. Q., & Li, N. (2008). Genome-wide analysis reveals regulatory role of G4 DNA in gene transcription. Genome Research, 18(2), 233-241.

Emsembl.(2012).Chromosome 19:35,629,728-35,634,013.Retrieved November 38, 2012, fromhttp://www.ensembl.org/Homo_sapiens/Location/View?g=ENSG00000221857; r=19:35629728-35634013

Ensembl.(2012). *FXYD1*. Retrieved Novmber 26 2012, from http://www.ensembl.org/Multi/Search/Results?q=*FXYD1*;y=0;site=ensembl_all;x=0; page=1;fall_species=1#

Fuller, W., Howie, J., McLatchie, L. M., Weber, R. J., Hastie, C. J., Burness, K., . . . Shattock, M. J. (2009). *FXYD1* phosphorylation in vitro and in adult rat cardiac myocytes: threonine 69 is a novel substrate for protein kinase C. *American Journal of Physiology-Cell Physiology, 296*(6), C1346-C1355.

Gomez, D., Lemarteleur, T., Lacroix, L., Mailliet, P., Mergny, J. L., & Riou, J. F. (2004). Telomerase downregulation induced by the G-quadruplex ligand 12459 in A549 cells is mediated by hTERT RNA alternative splicing. *Nucleic Acids Research, 32*(1), 371-379.

Goncalves, D. P. N., Ladame, S., Balasubramanian, S., & Sanders, J. K. M. (2006). Synthesis and G-quadruplex binding studies of new 4-N-methylpyridinium porphyrins. *Organic & Biomolecular Chemistry, 4*(17), 3337-3342.

Gray, N.K., & Hentze, M. W. (1994). Iron regulatory protein prevents binding of the 43S translation pre-initiation complex to ferritin and eALAS mRNAs.EMBO J, 13, 3882–3891

Gu, H.-P., Lin, S., Xu, M., Yu, H.-Y., Du, X.-J., Zhang, Y.-Y., . . . Gao, W. (2012). Up-Regulating Relaxin Expression by G-Quadruplex Interactive Ligand to Achieve Antifibrotic Action. *Endocrinology, 153*(8), 3692-3700.

Han, F., Bossuyt, J., Despa, S., Tucker, A. L., & Bers, D. M. (2006). Phospholemman phosphorylation mediates the protein kinase C-dependent effects on Na+/K+ pump function in cardiac myocytes. *Circulation Research, 99*(12), 1376-1383.

Hong, Y., Haeussler, M., Lam, J. W. Y., Li, Z., Sin, K. K., Dong, Y., . . . Tang, B. Z. (2008). Label-free fluorescent probing of G-quadruplex formation and real-time monitoring of DNA folding by a quaternized tetraphenylethene salt with aggregation-induced emission characteristics. *Chemistry-a European Journal, 14*(21), 6428-6437.

Huppert, J. L., Bugaut, A., Kumari, S., & Balasubramanian, S. (2008). G-quadruplexes: the beginning and end of UTRs. *Nucleic Acids Research, 36*(19).

Johnson, J. E., Cao, K., Ryvkin, P., Wang, L.-S., & Johnson, F. B. (2010). Altered gene expression in the Werner and Bloom syndromes is associated with sequences having G-quadruplex forming potential. *Nucleic Acids Research, 38*(4).

Kan, Z. Y., Yao, Y. A., Wang, P., Li, X. H., Hao, Y. H., & Tan, Z. (2006). Molecular crowding induces telomere G-quadruplex formation under salt-dericient conditions and enhances its competition with duplex formation. *Angewandte Chemie-International Edition, 45*(10), 1629-1632.

Kikin, O., D'Antonio, L., & Bagga, P. S. (2006). QGRS Mapper: a web-based server for predicting G-quadruplexes in nucleotide sequences. *Nucleic Acids Research, 34*, W676-W682.

Kostadinov, R., Malhotra, N., Viotti, M., Shine, R., D'Antonio, L., & Bagga, P. (2006). GRSDB: a database of quadruplex forming G-rich sequences in alternatively processed mammalian pre-mRNA sequences. *Nucleic Acids Research, 34*, D119-D124.

Kumari, S., Bugaut, A., Huppert, J. L., & Balasubramanian, S. (2007). An RNA G-quadruplex in the 5 ' UTR of the NRAS proto-oncogene modulates translation. *Nature Chemical Biology, 3*(4), 218-221.

Kwok, C. K., Sherlock, M. E., & Bevilacqua, P. C. (2013). Effect of Loop Sequence and Loop Length on the Intrinsic Fluorescence of G-Quadruplexes. *Biochemistry, 52*(18), 3019-3021.

Lech, C. J., Heddi, B., & Anh Tuan, P. (2013). Guanine base stacking in G-quadruplex nucleic acids. *Nucleic Acids Research, 41*(3), 2034-2046.

Li, Q., Xiang, J.-F., Zhang, H., & Tang, Y.-L. (2012). Searching Drug-Like Anti-cancer Compound(s) Based on G-Quadruplex Ligands. *Current Pharmaceutical Design, 18*(14), 1973-1983.

Lim, K. W., Alberti, P., Guedin, A., Lacroix, L., Riou, J.-F., Royle, N. J., . . . Phan, A. T. (2009). Sequence variant (CTAGGG)(n) in the human telomere favors a G-quadruplex structure containing a G center dot C center dot G center dot C tetrad. *Nucleic Acids Research, 37*(18), 6239-6248.

Lin, J., Yan, Y. Y., Ou, T. M., Tan, J. H., Huang, S. L., Li, D., . . . Gu, L. Q. (2010). Effective Detection and Separation Method for G-Quadruplex DNA Based on Its Specific Precipitation with Mg2+. *Biomacromolecules, 11*(12), 3384-3389.

Liu, W., Zhu, H., Zheng, B., Cheng, S., Fu, Y., Li, W., . . . Liang, H. (2012). Kinetics and mechanism of G-quadruplex formation and conformational switch in a G-quadruplex of PS2.M induced by Pb2+. *Nucleic Acids Research, 40*(9), 4229-4236.

Long, X., Parks, J. W., Bagshaw, C. R., & Stone, M. D. (2013). Mechanical unfolding of human telomere G-quadruplex DNA probed by integrated fluorescence and magnetic tweezers spectroscopy. Nucleic Acids Research, 41(4), 2746-2755.

Lorenz, R., Bernhart, S. H., Siederdissen, C. H. Z., Tafer, H., Flamm, C., Stadler, P. F., & Hofacker, I. L. (2011). ViennaRNA Package 2.0. *Algorithms for Molecular Biology, 6*.

Marcel, V., Tran, P. L. T., Sagne, C., Martel-Planche, G., Vaslin, L., Teulade-Fichou, M.-P., . . . Van Dyck, E. (2011). G-quadruplex structures in TP53 intron 3: role in alternative splicing and in production of p53 mRNA isoforms. *Carcinogenesis, 32*(3).

McCaskill, J. S. (1990). The equilibrium partition-function and base pair binding probabilities for rna secondary structure. *Biopolymers, 29*(6-7), 1105-1119.

Menendez, C., Frees, S., & Bagga, P. S. (2012). QGRS-H Predictor: a web server for predicting homologous quadruplex forming G-rich sequence motifs in nucleotide sequences. *Nucleic Acids Research, 40*(W1), W96-W103.

Mergny, J.-L., & Lacroix, L. (2009). UV Melting of G-Quadruplexes. *Current protocols in nucleic acid chemistry / edited by Serge L. Beaucage ... [et al.], Chapter 17*, Unit 17.11-Unit 17.11.

Moon, I. K., & Jarstfer, M. B. (2007). The human telomere and its relationship to human disease, therapy, and tissue engineering. *Frontiers in Bioscience, 12*. doi: 10.2741/2412

Moon, I. K., & Jarstfer, M. B. (2010). Preparation of G-quartet structures and detection by native gel electrophoresis. *Methods in molecular biology (Clifton, N.J.), 608*, 51-63.

Morth, J. P., Pedersen, B. P., Toustrup-Jensen, M. S., Sorensen, T. L. M., Petersen, J., Andersen, J. P., . . . Nissen, P. (2007). Crystal structure of the sodium-potassium pump. *Nature, 450*(7172), 1043-U1046.

Musetti, C., Krapcho, A. P., Palumbo, M., & Sissi, C. (2013). Effect of G-Quadruplex Polymorphism on the Recognition of Telomeric DNA by a Metal Complex. *Plos One, 8*(3).

National Center for Biotechnology Information.(2014). Homo sapiens chromosome 19, GRCh38 Primary Assembly. Retrieved April 14 2014 from http://www.ncbi.nlm.nih.gov/projects/sviewer

Nguyen Thuan, D., Haselsberger, R., Michel-Beyerle, M.-E., & Anh Tuan, P. (2011). Following G-quadruplex formation by its intrinsic fluorescence. *Febs Letters, 585*(24), 3969-3977.

Okumus, B., & Ha, T. (2010). Real-time observation of G-quadruplex dynamics using single-molecule FRET microscopy. *Methods in molecular biology (Clifton, N.J.), 608*, 81-96.

Onyshchenko, M. I., Gaynutdinov, T. I., Englund, E. A., Appella, D. H., Neumann, R. D., & Panyutin, I. G. (2009). Stabilization of G-quadruplex in the BCL2 promoter region in double-stranded DNA by invading short PNAs. *Nucleic Acids Research, 37*(22).

Ou, T.-M., Lu, Y.-J., Zhang, C., Huang, Z.-S., Wang, X.-D., Tan, J.-H., . . . Gu, L.-Q. (2007). Stabilization of G-quadruplex DNA and down-regulation of oncogene c-myc by quindoline derivatives. *Journal of Medicinal Chemistry, 50*(7), 1465-1474.

Palacky, J., Vorlickova, M., Kejnovska, I., & Mojzes, P. (2013). Polymorphism of human telomeric quadruplex structure controlled by DNA concentration: a Raman study. *Nucleic Acids Research, 41*(2), 1005-1016.

Pandey, S., Agarwala, P., & Maiti, S. (2013). Effect of Loops and G-Quartets on the Stability of RNA G-Quadruplexes. *Journal of Physical Chemistry B, 117*(23), 6896-6905.

Paramasivan, S., Rujan, I., & Bolton, P. H. (2007). Circular dichroism of quadruplex DNAs: Applications to structure, cation effects and ligand binding. *Methods, 43*(4), 324-331.

Parham, W. A., Mehdirad, A. A., Biermann, K. M., & Fredman, C. S. (2006). Hyperkalemia revisited. *Texas Heart Institute Journal, 33*(1), 40-47.

Phong Lan Thao, T., Mergny, J.-L., & Alberti, P. (2011). Stability of telomeric G-quadruplexes. *Nucleic Acids Research, 39*(8), 3282-3294.

Presti, C. F., Jones, L. R., & Lindemann, J. P. (1985). Isoproterenol-induced phosphorylation of a 15-kilodalton sarcolemmal protein in intact myocardium. *Journal of Biological Chemistry, 260*(6), 3860-3867.

Randazzo, A., Spada, G. P., & da Silva, M. W. (2013). Circular Dichroism of Quadruplex Structures. *Quadruplex Nucleic Acids, 330*, 67-86.

Redman, J. E. (2007). Surface plasmon resonance for probing quadruplex folding and interactions with proteins and small molecules. *Methods, 43*(4), 302-312.

Rio, D. C., Ares, M., Jr., Hannon, G. J., & Nilsen, T. W. (2010). Nondenaturing agarose gel electrophoresis of RNA. *Cold Spring Harbor protocols, 2010*(6), pdb.prot5445-pdb.prot5445.

Rogers, J., & Wall, R. (1980). A mechanism for rna splicing. *Proceedings of the National Academy of Sciences of the United States of America-Biological Sciences, 77*(4), 1877-1879

RSCB PDB. (2012). Structure of an intramolecular G-quadruplex containing a G.C.G.C tetrad formed by human telomeric variant CTAGGG repeats. Retrieved November 28, 2012, from http://www.rcsb.org/pdb/explore/explore.do?structureId=2km3

RSCB PDB. (2012). Structure of the Na,K-ATPase regulatory protein *FXYD1* in micelles. Retrieved November 30, 2012, from http://www.rcsb.org/pdb/explore/explore.do?structureId=2JO1

Rubis, B., Kaczmarek, M., Szymanowska, N., Galezowska, E., Czyrski, A., Juskowiak, B., . . . Rybczynska, M. (2009). The biological activity of G-quadruplex DNA binding papaverine-derived ligand in breast cancer cells. *Investigational New Drugs, 27*(4), 289-296.

Samatanga, B., Dominguez, C., Jelesarov, I., & Allain, F.H.-T. (2013). The high kinetic stability of a G-quadruplex limits hnRNP F qRRM3 binding to G-tract RNA. *Nucleic Acids Research*, 41(4), 2505-2516

Scaria, V., Hariharan, M., Arora, A., & Maiti, S. (2006). Quadfinder: server for identification and analysis of quadruplex-forming motifs in nucleotide sequences. *Nucleic Acids Research, 34*, W683-W685.

Shay, J. W., Zou, Y., Hiyama, E., & Wright, W. E. (2001). Telomerase and cancer. *Human Molecular Genetics, 10*(7), 677-685.

Shapiro, M. B., & Senapathy, P. (1987). Rna splice junctions of different classes of eukaryotes - sequence statistics and functional implications in gene-expression. *Nucleic Acids Research, 15*(17), 7155-7174

Shinoda, T., Ogawa, H., Cornelius, F., & Toyoshima, C. (2009). Crystal structure of the sodium-potassium pump at 2.4 angstrom resolution. *Nature, 459*(7245), 446-U167.

Siddiqui-Jain, A., Grand, C. L., Bearss, D. J., & Hurley, L. H. (2002). Direct evidence for a G-quadruplex in a promoter region and its targeting with a small molecule to repress c-MYC transcription. Proceedings of the National Academy of Sciences of the United States of America, 99(18), 11593-11598.

Song, Q., Pallikkuth, S., Bossuyt, J., Bers, D. M., & Robia, S. L. (2011). Phosphomimetic Mutations Enhance Oligomerization of Phospholemman and Modulate Its Interaction with the Na/K-ATPase. *Journal of Biological Chemistry, 286*(11), 9120-9126.

Stegle, O., Payet, L., Mergny, J.-L., MacKay, D. J. C., & Huppert, J. L. (2009). Predicting and understanding the stability of G-quadruplexes. *Bioinformatics, 25*(12), I374-I382.

Sun, H., Li, X., Li, Y., Fan, L., & Kraatz, H.-B. (2013). A novel colorimetric potassium sensor based on the substitution of lead from G-quadruplex. *Analyst, 138*(3), 856-862.

Teriete, P., Franzin, C. M., Choi, J., & Marassi, F. M. (2007). Structure of the Na,K-ATPase regulatory protein *FXYD1* in Micelles. *Biochemistry, 46*(23), 6774-6783.

Thandroyen, F. T., Morris, A. C., Hagler, H. K., Ziman, B., Pai, L., Willerson, J. T., & Buja, L. M. (1991). Intracellular calcium transients and arrhythmia in isolated heart-cells. *Circulation Research, 69*(3), 810-819.

Tluckova, K., Marusic, M., Tothova, P., Bauer, L., Sket, P., Plavec, J., & Viglasky, V. (2013). Human papillomavirus g-quadruplexes. *Biochemistry, 52*(41), 7207-7216.

Tseng, T.-Y., Chien, C.-H., Chu, J.-F., Huang, W.-C., Lin, M.-Y., Chang, C.-C., & Chang, T.-C. (2013). Fluorescent probe for visualizing guanine-quadruplex DNA by fluorescence lifetime imaging microscopy. *Journal of biomedical optics, 18*(10), 101309-101309.

van der Velden, A. W., & Thomas, A. A. M. (1999). The role of the 5 ' untranslated region of an mRNA in translation regulation during development. *International Journal of Biochemistry & Cell Biology, 31*(1), 87-106.

Viglasky, V., Bauer, L., Tluckova, K., & Javorsky, P. (2010). Evaluation of human telomeric g-quadruplexes: the influence of overhanging sequences on quadruplex stability and folding. *Journal of nucleic acids, 2010*.

Vummidi, B. R., Alzeer, J., & Luedtke, N. W. (2013). Fluorescent Probes for G-Quadruplex Structures. *Chembiochem, 14*(5), 540-558.

Williamson, J. R., Raghuraman, M. K., & Cech, T. R. (1989). Mono-valent cation induced structure of telomeric dna - the g-quartet model. *Cell, 59*(5), 871-880.

Wilson, K. S., & Vonhippel, P. H. (1995). Transcription termination at intrinsic terminators - the role of the rna hairpin. *Proceedings of the National Academy of Sciences of the United States of America, 92*(19).

Wong, H. M., Payet, L., & Huppert, J. L. (2009). Function and targeting of G-quadruplexes. *Current Opinion in Molecular Therapeutics, 11*(2), 146-155.

Wong, H. M., Stegle, O., Rodgers, S., & Huppert, J. L. (2010). A toolbox for predicting g-quadruplex formation and stability. *Journal of nucleic acids, 2010*.

Wong, W. C., Zhuang, J. Y., Ng, S. L. L., New, L. L. L., Hiew, S., Guo, J. J., . . . Li, T. H. (2010). Conformational organizations of G-quadruplexes composed of d(G(4)T(n))(3)G(4). *Bioorganic & Medicinal Chemistry Letters, 20*(15), 4689-4692.

Wu, Y., & Brosh, R. M., Jr. (2010). G-quadruplex nucleic acids and human disease. *Febs Journal, 277*(17), 3470-3488.

Wuchty, S., Fontana, W., Hofacker, I. L., & Schuster, P. (1999). Complete suboptimal folding of RNA and the stability of secondary structures. *Biopolymers, 49*(2), 145-165.

Xu, L., Xu, Z., Shang, Y., Feng, S., & Zhou, X. (2012). Structural polymorphism of human telomere G-quadruplex induced by a pyridyl carboxamide molecule. *Bioorganic & Medicinal Chemistry Letters, 22*(8), 2988-2992.

Xu, Y., Suzuki, Y., Ito, K., & Komiyama, M. (2010). Telomeric repeat-containing RNA structure in living cells. *Proceedings of the National Academy of Sciences of the United States of America, 107*(33), 14579-14584.

Yan, Y.-Y., Lin, J., Ou, T.-M., Tan, J.-H., Li, D., Gu, L.-Q., & Huang, Z.-S. (2010). Selective recognition of oncogene promoter G-quadruplexes by Mg2+. *Biochemical and Biophysical Research Communications, 402*(4), 614-618.

Ying, L. M., Green, J. J., Li, H. T., Klenerman, D., & Balasubramanian, S. (2003). Studies on the structure and dynamics of the human telomeric G quadruplex by single-molecule fluorescence resonance energy transfer. *Proceedings of the National Academy of Sciences of the United States of America, 100*(25), 14629-14634.

Yuan, G., Zhang, Q., Zhou, J., & Li, H. (2011). Mass Spectrometry of G-quadruplex DNA: formation, recognition, property, conversion, and conformation. *Mass Spectrometry Reviews, 30*(6), 1121-1142.

Yuan, L., Tian, T., Chen, Y., Yan, S., Xing, X., Zhang, Z., . . . Zhou, X. (2013). Existence of G-quadruplex structures in promoter region of oncogenes confirmed by G-quadruplex DNA cross-linking strategy. *Scientific Reports, 3.*

Zhang, C., Liu, H.-h., Zheng, K.-w., Hao, Y.-h., & Tan, Z. (2013). DNA G-quadruplex formation in response to remote downstream transcription activity: long-range sensing and signal transducing in DNA double helix. *Nucleic Acids Research, 41*(14), 7144-7152.

Zhou, Z., Zhu, J., Zhang, L., Du, Y., Dong, S., & Wang, E. (2013). G-quadruplex-Based Fluorescent Assay of S1 Nuclease Activity and K+. *Analytical Chemistry, 85*(4), 2431-2435.

Zhu, H., Xiao, S., & Liang, H. (2013). Structural Dynamics of Human Telomeric G-Quadruplex Loops Studied by Molecular Dynamics Simulations. Plos One, 8(8).

Zuker, M., & Stiegler, P. (1981). Optimal computer folding of large rna sequences using thermodynamics and auxiliary information. *Nucleic Acids Research, 9*(1), 133-148.

APPENDIX I

GQS mapping from Table 2 and conserved sequence from Table 6 in the *FXYD1* pre-mRNA of each ortholog

Alternate exon sequences are represented in uppercase characters, where purple characters are UTR sequence bases and black characters represent translated sequence. Intron sequences are denoted by lowercase blue characters. The highest scoring GQS from Table 2 are underlined, while conserved sequences from Table 6 are highlighted yellow to map their position in the gene for each organism.

```
Key:    UTR region
        Intronic sequence
        Exonic translated sequence
```

Mus musculus

GGGUGGAGCAUCCAGUUCUGGGCCAGGGGGUCCAAAGUGCUUAGCUCCUAGGGUGCACAGCU
GGACAUUUGGGGGUCUUCUGUCAACAGGGGACAGCGUGAAUGGGgugagcgucccccagccc
ucccuccgggcccccucagcuccccuagcugggaggccuauuuugggaacaagaguggccagc
cuguggcuucucaggcaggccugacccaagagggaggagagagugugggggacagggguugca
caggggcgggagaguuagagacuccucccuuuuucagcggccacugcgagaccccuggcagg
gggugaggcucagauacucauuuguauaggucuguuucugucucuguuuggggggcacagaa
ggcccagagcgagagaaugugcuuaugucuaaacccugcgcucucuaaucaacaguuugggg
agaggguguuguuugcuccuguuccagcuacacccacucugugugugugugguaccuguac
auaaaugugucugugcccguaugugugucccugaaacaacaucugacuucucucaggcaugg
gccgccugucacucacuggccuaaagucuuguugugaaagaugucacccagagguggacaaa
gagagggauguucccccuuuucucacagcuucaagaaaaggagauggggguggccuguaggga
uguggcuccuggcugggccucacccccagcagcaguguuauacaggacccccugagucuuuggggg
gggagcuguugccauggugggccgugugcagcaaauuccucccggguggaagugggagauauu
uauacccaggguuagggagagagcgggcaggcggccgagggcaggagagcugggacggccug
gguacagagagaccacugguugaggguguguagggcaggugggggcugggcaugccugcugu
augucgccuaguguucccaccuaugucuagaggcagcuugcuucccuuacaagggguagguuu
ugagacucugggaccagcuaacggauggaguagccuguggggagccacccccaucccccccagg
acucugccucuccccaucugucuugcugcuguguaugcugugguucuccuggucccuauuauu
gaguuguuuuagggacaugguuuuugggugaagcaaggagccauucaacuugaauggcugac
acuuaaguccuuaggccgguccuugacuuaaccccacggguucagAUCCUCUUUAGGAGGGAA
AGAGAGCAGGGCAGAGGACAUUUCUUGACCCUGGCUGACUCCCUAGGGCAAUGGCAUCUCCC

GGCCACAUCCUGGCUCUGUGUGUGUGUCUCCUCUCCAUGGCCAGUGCAGgugaguccaaagg
aggugcucagcaucucagccauguggguggcagagguagggaaaagcccccaagaaaaccc
agugagaccccaaacuagacacuauguaacaggaaaucaggggucuccuaggcaacagcg
gggguggauggaagguggauucccggggugacggaaacaugaacaaucugcaagauaucagg
gugccggaccagugcuggagguuccaggacgcagagcaggcuggcagccugccaguggcuca
guagcaccuaaacucccaguccaagaagaucaguggugccccugaagggcuccuuauacuu
uccccuguuacagaccccaggagcuacucuaacgcugcucuuuauuucucuuucuagAAGCU
CCACAGGAACCGGAUCCAUUCACCUACGgugaggggaagcuacguggggguuuggagagagg
gccugcguguuucuauucccucucucugucucucugucacugucucugucucugucugucuc
ucucucucauacacacacaccacauccuccacucucuccugcucuaguugucuuagacccc
uugccuguuuguuugaccccccccccgucucauuugaacugguuuaucaguaucaguaccu
caaccucuacgacccccacuuucagcagcauccuccucuccauccuuucccucuuccuacca
ccuccucccugccuuccuuccucuaaccuguccucuccuccccucucacaucuguccuuuccc
cuucucucuccuccuccuugagcugcauacuauacucuuccccgcuggucacccuccuccgg
ggcuguccacagAUUACCACACCCUGCGGAUCGGCGGCCUCACUAUCGCUGGGAUCCUCUUC
AUCUUGGGCAUCCUUAUCAUCCUUAgugagugucugcaccugucuucuccauccgccucca
gccuucccuccccaaaccccacucccagcaacacaugcagccugugcgcuauucacgcccaa
cacaggcucagucuccaaccacuucucuuaggucccacccugacuccaaccucucaccgcac
agaccccaccccaccucaaaccacgcccacuccacacccagaccagaccacgccccacaccc
uacccuucuucagccgcgccucaggccccaccccugcggccuccccaguuuggaggagagc
cuagggaugcgggagcgagagggaagcucagccuauacccucccacuccacagGCAAGAGAU
GUCGAUGCAAAUUCAACCAACAGCAGAGgugaguggucccucugggccucccucgcuccuuc
cgcaguggagaggcgguuggggcgaggcagcaagugcacccauccugcagugaacauguauu
aagcgcuuagugugugugguuaaacccuaagacaggcuccugggccggagcgauggucaagcucc
cucucugccuggccuagcgcccagccaguggggggcgcugggcucccaccagccccgcagaug
gaaacucaagcguggugaggcgaagcugggcagaagcagccgcgcugaaaugagaucaccuc
acagggcggcccaguuuagcuccagucccgauccucgcgcaggauuccucucgaaauaaacc
uuuaaagcgcagaaaacguaggcaugccuucugcgugcuaagaugaucacagaugccuccau
gccgacccucuggcaugcuugagugcgcacuacgcgccaggugcccgauuccuucucuauua
uguacuccuuaccccacccggaugaggugcgcuccaucaucaaucccauuuugcauaugagg
aaacugaggcacggggcgguaaggugacuuagggucacucacuggacccaagucccaagucu
cccaagacgcugugaccgucccucucccaacaaagucaaguaccuuccccauuaucccggag
gccuucacccgugacaggcugggagcaccuccugccgcacccccgaaacagcagccgggcgcu
cuuguuucugacggaccgcguucauaucaaguccacgguggggggucgggaaauaaggccugc
auuagggggcuucucggaagcggcugccucuccugguccauccgaauuccucuaucuguuccc
uuuuuagAACUGGGGAACCCGACGAAGAGGAGGGAACUUUCGCAGCUCCAUCCGCCgugag
uucg<mark>gggauacugcgggguuuguggggcag</mark>cguuuuucaagagcccccucuucuuggacuca
gaauggguuuggcggagaguuagcccugguugcagacucccccuuaccccagaacucucugc
gucccucucagGUCUGUCAUCCCGCAGGCGGUAGAACCUCCACCUGACUCCAGGAAACUCAG
CCAGguucugcaguugaugggaggaaggagggagagggguuuggagggagggaggcaagggg
cccagugccaggguggauaguggcgagagcuugccacuuacucuuuucacccccgcagAGC
CCCCUUAGCACCUGACACCUCUCCCCACCCAGAUGCUCGCCUGUGACCACCCCCAGCGUUCC
CUGCAUCAGCCCUGCCUUUCGGACACCCCUUGCUGCUCAGACCUCUAAUAAAACUCGGUUUU
CCUUCUUG

Canis lupus familiaris

GGGAGGGGAGACCGCUGAGGGCGGCGCAGGACCAGCUCUGGAACAGGGGGUCCAAAGUGCUC
ACCCCCGGCACAGCCGGACGUUUGGGGGCCUUCUUUCAGCAGGGGACAGCCUGACUGGGgug
agcgucccugcccucccccccaggccucaccccuggccuggccacgccaccuauuuuggga
gcaggaguggccagcccguggcuucccaggcaggccagacccaagaggaagggagugugguu
gggacggccgugggguuccccaggagccgagugugggugcgacuccccuccccugcugcug
cgugugcaggcuggggugggguuuugagcacugauguguguaggucggguggcgucacuuug
gggggcacuggaggcucagaugaagaguaucguuuuagugacugugucuccaccccaugucu
ugcucaaccagagguuugggggaguaggcguugucuccauuccagccccauugcguaugugu
gugugucugugugucuguguguucugugugugucccuggcccuggcaucugucuuccgucucu
cagccccgcugcucugggccucaugccacucgccuggugguguccgccugugagagauguca
cccagaggcaggcggagggggauguuuauccuuguucucaacauuuccagaagagaaggggu
ggccuuuccccaggggcagccuguagcgauguggcagccgggccucaccccggcagaguug
ugcgugaccccgaguggggaaggaaggcguugccaugguggccugugugaggcaaauuc
cuccagggugaagugggagauauuuauacccggggucaggagagagccggccagcggccgag
gccaggggagccgggauauuacacggacacagcgaggcccugugugugggcagguggcccuu
gggcgugucuugggcugcagugucgcucaccccucacggcgugcgucugucuccccaccugg
uggggugagccauccugcucuucccacgggguagguguggggaaccuggggcccgcugucagg
aggcccaggccuugggcgaccccgguggggcugugugugucucaccccugcccucgcugugug
ucguaccacccccacgguggguuuguuuuggggacaccgguccacaugaagcaguggccaag
cuggaugauggggugcuuuuggaauuaagccccuggacucugauuuaaccccuuggguucaag
ccccaaugguggagggagacgaggaacacauuucuugacccuugcugccucccagGACGAUGG
CACCUCUCCACCACAUCUUGGUUCUCUGUGUGGGUUUCCUCACCAUGGCCACCGCAGgugag
ucuagggcggguagcccacaacccaccucagccccaaggggaggccagggaggggaaagcccu
ucaagagaaccaacuuggagacuccaaccuccacccagucaguuggggguaccaggaaacugg
ggacuucgcgugggcccaaacaccagcuaguauuuggggcaggggagcggaugcaaggucag
gcaaaggcggaugguggggauggaggccucuaaccucaaaugacagaaucauugacaacaggg
gauaagacaggaagcgggggcaguuaggagcccccugccccacccagcacagggcagaugag
gcggcuaaggcccagggaggguggggcagccugcucagggcccagcgguccgcagacucucag
cccacgggaguaaagcccccagaggggucucuuuacgcccuucccugugcagagucccaggaa
gcagccaaccucugacuuccuucucuuuccagAAGCGCCACAGGAACACGACCCGUUCACCU
ACGgugagggaggggagaggcaucuacuggggagccgggagggcgcauggccuaggugugccc
guccuguccucauuaccucucuucuccuuguccaccugucucucgggcucgggugguuuguuauc
ucggucacucuccggucucucgcgucucucagagaucuggcucucugacucccuguuuugau
cugcccuugucugugucucugugugucugucuacuuggucccucucugagucucuuucuccu
gucaccucugcguacuucauggcucauccucugcgugugccucccucucuccccucugcgguc
cccucacccucuccaccauccuccucuacuuccuuguuuccugccuccuccccgccugcccu
gcccaccuccucccucuccuucgcaccccgcuugaucucccucccugaucccgcucccacuu
ccuccuuucaccguucuccucucccucccgggguccuccucgcccccuccuccucccuucc
uccucccucccugcccucacccuccucccugccgcccacagACUACCAAUCCCUGCGGAUC
GGAGGCCUCAUCAUCGCCGGGAUCCUCUUCAUCCUCGGUAUCCUCAUCGUCCUGAgugagua
ccccagccccugccuccagccccgcgggugccguggugugcgugcccgccucgccgcggg
ucccgccccgccccgcccucgccccgccccgccgccuuagccccgcccacgucccgcccc
ugccgccucgccccaccccugccaccccagccccgcccccuccgccucgccccgcccacgccc
ccugccgccccgcccuagccccgcccuccgccccgcccuagccccgcccuccgccccgcccu
ccgccccgccccgccccuccgccucgccccgccccuacccccugcggcuccgccccucccccag
ccccaggcggcucaccagcgccgcuccgugccgcccggagcugaaucccggccuccuccua
ccccacccgcagGCAGAAGGUGCCGGUGCAAAUUCAACCAGCAGCAGAGguaagaggcccc
uccgggcucucacucaccuacuuucgcucuagaggggcgcugcggggugaggcagggcauc
caccagacccgcagccggagcccuuuccgcaaaaauguauuaagccccuacuaugugugcca

agccccaagccaagcccuggggaccagugaauaaauaaacugcccucccugccuggcugagc
ucccagcccagugggggcggcggccguggcuccaacagccccacaaauauaaaaucacaaa
gcgugauaagugcugugaaagagccacggcgaaaugagaucaccgcgcuuggcggcccaguu
uagcucagaguccuguuccuagcucuuugauucaucuuuaaauaaaaauuuaaggcaacgua
ugaacaccuucuccauguugaaacagucauggaugucauuuuaaauaauagcuaauauuuag
cgccacgcgccagauaccguuuuauugccuuaucugcuccccacaccccuaguaggagguag
gugcuguuauuaaccucauuuugcuggugaggaaaccggcgcacagggaggcgaagccacuu
gcccaagaucgcucaacgaggagaugaggguucugggguuuuucuaaccuaggcugucgccag
gccggccgcgcucaggcccccucggcgcccucccccaccaacuuccagucucuuccccugcg
cuccgggaggcacugggagcagcacccccuggcacacguuuaagccgcuccugggugcgccc
gauuuccuggcggccccccgagccugcuccucgcccaucuccagcccucagcgcaggggacug
ugaagugagggcugcccuggcggaggagggguccagggggcggcugccucucuccguuuuc
auccccaauuccuucugccagGACUGGGGAACCUGAUGAAGAGGAGGGAACUUUCCGCAGCU
CCAUCCGCCgugagucug<mark>gggagauuucgaguguuuggggcgggggugaggg</mark>cggguuccaa
gaaccccuuuccuggcccucccggggugcgaggugggaggggcuggcuuugaaacccgaggu
ccccgaguccccugcugcuucccgcuccgcaggagcuagggugcccccuccugacgcccgga
cucuccguguuccccucagGUCUGUCCACCCGCAGGCGGUAGAGACACCUGGCGCGAUGGA
ACCCAGCCAGGUCCUGCAGCUCGGGCUAGGCUGGGGGGCGGGGGGAGGAGAAGGGGGCGAAG
GGUGGGCUGGGAUGGCCGGGGGGCUCUCUUGCCUCUCACCUUUGUCACCCCCACAGGAUUCC
CCUGGCGCCUGAUGCCUCCCACCCACCACCUGUGCGCCCACCGCCACCUGGACUGCCCUCUU
CCCCAGCCCUGCCCCCGCAGGCUCCUCUCUGCCGCCCAGACUUCAAUAAAACGUGCUUUUCU
CUCUUGA

Pan troglodytes

AAAGUGCUCAGCCCCCGGGGCACAGCAGGACGUUUGGGGGCCUUCUUUCAGCAGGGGACAGC
CCGAUUGGGgugagcgucccccacuccuucccuccaggccucaccccuggucuggcuggggcc
gccuauuuugggagcaggaguggccagcccgaggcuucccaggcaggccaacccaagaggga
gggagugugguugaggcaguggguucugcaggguggggaugugggugacuccucccugcccug
cuggugcgugugcacccuggcagggguguggaguugggacacacacguguguagggcugguug
cgucacugcgugggggcaccggaggcccagaggaggaguacuggaugccugacgguguuuac
accccacguccugcuccaaccagaaguuugggggagagguuguuguucauguccauuccggcc
ccacugugugugugugugugugugugugugugugugugugugugugugugguaucccugccccagcaugguguu
uucuaucucuaaggcccacugggcugggccucaugucacuugccugacauccgauugugaaa
gaugucacccagaggcgggcagagggggcugucuuuuccuuuucucguugcugcccagggagg
agacggggguggccuuucccacaggggcagccuguggcgauguggcagcugggccucacccg
gcaggggcugugcgugaccccugaguggggggaaggcaggcuguugccaugguggccugagcg
agcagaauuccuccaggggugaagugggagauauuuauacccggggucaggccgcgagcgggc
gggcggagagggcagggagcugggauuucgcggggcacagugaggccgggcauguaggcagg
ugggacuugggcuugcccugcugucuccugcuccguguuugugugaggcagcgccuccucug
cccugccaggguaggucugggaaucgggggccugcugcgggagguggaggcccaagggaggc
cccccggggacugugugucucaccccccgucccugcuacguuguggguguugugugaucccauc
guggagguuguuuuggugacacugugucccacgaagcuggggauacccguuucucuagcuu
ggagccaccaagauagaggacgaacacuucgugauucaguccccagacugucucugacuua
aucccuuggguucaagcccaugugggagagcaagggcacacacugccuaauccgguggguguc
cccccagGACAAUGGCGUCUCUUGGCCACAUCUUGGUUUUCUGUGUGGGUCUCCUCACCAU
GGCCAAGGCAGgugagugcaggggaggcugcccgcuacccaccucagccccaggggguggcgg
uggggaccgaagaaccaaguuggagaccccaaccuagacuaagucggcugggguaccaagaa
guuuggggggucuccacgugggguccagucacaggcugguauuuggggggaggggagaggaagc
cccagaucaggcaaagaugggguggggaugggggcugaauccccgaugggauaacugggucaca
gacagccugccgugagucagggagcuggggcaguuaggugccaccugccccaucugggacag
ugcagagggggcagcugggacccagagagguguggggcagccugcccagacaccucagacucu
aagcccagcaaggcagagccuccaguggucuccucaugcccccucccugccaggaccccagga
agcauucaaccccugauuucucucucuuuccagAAAGUCCAAAGGAACACGACCCGUUCACU
UACGggagcggggggucuaaugaaunnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnn
nnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnn
nnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnn
nnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnuacauguggu
ggauucuauugaaucucugucauucuccuuuccucuauuuugucuuuccucucugauuccac
cugucugcaucuuuuccugucuguguucuaucuguguacacugucuaugugauaccucucuggu
ucucuuucucuugccugcgucugucucagcaucucguggcccauccucugcuucuucccauc
uucucucccccuguccuccuccucccuguccccuccucccuuuccuauacacccuuucc
ucucccugguaccccacuuuccuccucccauaucugcucccccuuaauuaucuuauuucccc
ccuucugccugcugguccuuucucccuguccccuccuucccaauuuaucccucuccuauucu
cccuccucucccuguccucaccuucccugcucugcugcucacagACUACCAGUCCCUGC
AGAUCGGAGGCCUCGUCAUCGCCGGGAUCCUCUUCAUCCUGGGCAUCCUCAUCGUGCUGAgu
gagugccccuagcccccgcccucuaccccgccucucccuggccccgccucucccuggccccg
ccucucccuggccccgccucucccuggccccgcuucucccuggucccgccccucccuggccc
cgcccgccccaaccccucccaggccuugccccgccuacccugccuugguuccccggccccc
ggucucgccucuagcccccgccccguccccaagccccgccccucgcgagggcgagcuggagc
uacagcgccgcuuggcgcccgccgggagggagccucagcuucuccuaccucuccaugcccac
agGCAGAAGAUGCCGGUGCAAGUUCAACCAGCAGCAGAGguaagacgcccccuccccgcccuc
cuucgcccgcuccugcucuggagggcgccgcgggugaggcggggaguaccccugacccgcag

cccgauccccaucagcgacuauguauuaagcaccuacuaugugccauggcccaagccuggcc
cugggaccaagcgaggaaaaaaccucccgcccuuccuggccgagcucccagccuaguggagg
cgguggccgugggguuccaacagccccacagauagaaaaaucacaaagcgugauaacacaaag
ugcaggaaagaagaaacggcggugaaaugagaucaucucacacgcggcccaguuuagcuuag
agucuuguuccuagcucuuugauuccucuucgaauaaaauguuaaagcauggacaauguaug
aauauguuagaacaauuauagauauuaucauaaguaguagcuaauauuuacuggguguguac
cacgugucagauacgguuucacuuccucugggagggaggugcuguuauuaaccccauuugac
agaugaggaaacuaaggcacagggagguaaagucacuuuguucaagaucacucaaguagaag
auggggggguucuggguuuccaacccaggccaucucauggcagucugccaaguccccaugacu
aucccuccccuaccaacuucacaucccugcccccaaauccgcggagguacucacuguuaacc
agcucagaagccccugccagcacagaagcugcuccggggugcuccucauuucuagcggacc
ccgagccugcucuucguccauaucgggccuaguuacaccaaucggggaaaggaggcuugua
cugggggguuccuagaagggcagccucuccccuuuccaucccgaaaucccucugccucugu
cuucccag**GACUGGGGAACCCGACGAAGAGGAGGGAACUUUCCGCAGCUCCAUCCGCC**guga
gucug<mark><u>gggagacugcggguauuuuggggagaggg</u></mark>cugguuccaaggaccccuuuuccuggcc
cucccuggcugcguagagggaagggcuggaucugaaagcggaggcgggggaguugccccgcc
gcgggccccaccugcccaggaacuggggaugccucuccagaaugaccccgaucuccguguu
cccccag**GUCUGUCCACCCGCAGGCGGUAG**AAACACCUGGAGCGAUGGAAUCCGGCCAGgu
gcugcagcucugacacggcgguggggggaggaaggagggaggaaggaaaggcgggagagggaggg
gggccaagugcca<u>gggguugaagggcggcgagggguggggg</u>cuggacgucccccucgccucuc
acccuuuucacccucacagGACUCCCCUGGCACCUGACAUCUCCCACGCUCCACCUGCGCGC
CCACCGCCCCCUCCGCCGCCCCUUCCCCAGCCCUGCCCCUGCAGACUCCCCCUGCCGCCCAG
ACUUCCAAUAAAACGUGCGUUCCUCUCGACAGCACUUUGUCGGUCUCGGUCCCUCAGCGCGA
AAGCCCAGCGCCACUGGGCCCCAGCA

Bos taurus

GGAGACCGCUGAGGGCAGCGCAGCCAGCUCUGGGCCAGGGGGUGCAAAGUGCUCGGCCCCGG
GGCACAGCCGGACGUUUGGGGGCCUUCUUUCGGCAGGGGACAGUCUGACUGGGgugagcguc
ccccgcccccuccccuccagggccucaccccugnnnnnnnnnnnnnnnnnnnnnnnnnnnnnn
nnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnn
nnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnn
nnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnn
nnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnn
nnnnnnnnnaaagaugucacccagaggcaggccaggggaguguuuucccuguucucacugcuu
ccagaagaggagaaggggguggccuuucccacgggggcagcuggugggcagugug gcagcccag
ccucaccccgggcaggguugugcaugacccccaagugggggggaggaaggcguugccaugu
ggccugugcgaggcaaauuccuccagggugaaguggg gagauauuuauacccagggucaggca
gagagcgggccagcggccgagggcaggagagccgggcuauuacggacacagcaaggccgugg
ugugugggcaggugg cccuugggcguuuccuggcugucucgucccuggcagugucugucucu
ccaccuggug gggugaggcaucugcucuccccugccaggguaggugcggggaaccuggggc
cuacagugagggaggccccggccugcgggagauccuuguggg cuguguguugccucaccccgu
ccuugcugcuggg ugacccucuuuuguguggguuguuuuggugacgcuguuuccugaugaagc
aggggcgccucuguuucucaggg ugaccaagccggaugauagaugcuuuuggaauuaaguc
ccuggacugucucugacuuaaccgcuugggu ucaggccccagugguggg agacugagggcac
gcauuuccugaccuuuggguauuuccccagGACAAUGGCAUCUCUCAGCCACAUCUUGGUUCUU
UGUGUGGGUCUCCUUGCCAUGGUCAACGCAGgugagucuggggg gaggcagcccacuaccacc
ucagccccaggggugguaagggu agaggaaaaauucugcaagagaaacaagcuggagccucca
gucugaacugagucucagugggg gucucaugugggguccagucacaggcugguaucugggggga
gggg agaggaugccccaggucgggcaaaggggg agggguggg auggaggccggcuucccgug a
augaugg aaucacagacaaccuggaaug agaugg agagcugaggcugucaggugccuccugc
cccauuccggacagugcagagaggcagccaagacccagagaggguggg cagccuacucagag
uccaacaguccccagacuccugccccgagacuguaaaacucccacaggg ucucgugaugccg
cucucuguccaggg ucccaggaagcauccaacccugauuucuucuucucuuuccagAAGCAC
CACAGGAACACGACCCAUUCACCUACGgugagggaggggcaucuuucucugggugcugggcu
ugguuagccugcccuaucucucucuguuucugucuccaucauucucuagccucucggugugu u
acaucuuuggcauucucauucuuuaucucuugauuuggcucucugacucuuucggucucgau
cuccccuuuguccccugucucucugagucucuguguucucuauuuguguccucucugucucucu
gauucauucuccuuuuaccucuuaguaucccacggccucuccucuguugccaccugccuuuc
uccucucccugccccccuuucccucuccaucauccuuccuuacucgauugcuuccucccucc
ucuugcccuguccaccuccuaccuuuucuicccugauccccuucucccacaucgcucccucc
uccucgcacugauguccuuuccuccucuuccccauccuuccucuuugccccuuccucccca
uucugccccuuuccuauccuuccuccuccucucccugcauucacccuccuccgugcauucac
ccgccucccugcugcccacagACUACCAAUCCCUGCGGAUCGGAGGCCUUAUAAUCGCCGGG
AUUCUCUUCAUCCUGGGCAUACUCAUCGUCUUAAgugaguccccuauccugccuucagucg
cuuccggugucugccccagccccgcccaaccucccgcuccgccuucaaccccgccccaucug
ugcccuuggccccgccuuuguuucuguccacccacuuuccucccugggccuugcaccuguu
cugccccugcccucagccacgucccaguuucugcccuuccccccauucccgccuccggccccg
cccgguccuuauccugccugcccucgggcucgccccgccucuauccccgccccggccccgcc
cccacuuauauccccgcuucuggccccgccccguuuccuauccccgccccggacacgcccc
cauuucuauccccgccccucggccccgccccugccccugccccacccuuauccucagacacgc
cagcccgggcaagcgagaucugugg cucgcgugcgcugugcggcuuagagggg aaucucag
cacacuccucuccacgcccccacccacagGCAGAAGAUGCCGGUGCAAAUUCAACCAGCAGC
AGAGguaagaagccucucuggacccucaucuccucccgcucuaaaggagcguuggggg ugag
accagaaguccacucaauccacagccagauccuucggcaaauauguauuaaacaccuacuac
gugcccagccccaagccaggccuuggcaccaagccacgaguuaacugcccucucugucugag

cuuccaaccaagugggggcgguggccaugggguuccaacagacccacaaauauaaaaccacaa
agcgugauaagugcugugaaagagucgcggcgaaaugagaucaccuuacuuggcggcccagu
uuagcucagaguccuguuccuagcucuuugauuuaucguuaaauaaaauuuuaaagcccaag
caacguaugaauaucuucuuccauuaagauacuuauagauauuauuauaaauaaagguagu
auuuuauugugucccauguacaagaugcuauuucauuaugughhuauuuaccccucacaccau
cucuauaaaguaggcgcuguuauuaaccucauuuuacagaugaggaaacugagacauggga
aguucacaugccccagaucauucaacuggaagaugagggcucuggguuucucaccuaggc
ugucuccaggcaaucugcccucaagucgcuuuggcuauauaccuccucuuccauauucuagu
uccuuccucaacacucugggaguacucgcguuaucggcuuggagcaccuccugccaaaca
cuuaaacugcuccuggguugcacuggauuuucuggcggcuccaaucugcuccuugcccaucu
guggguicucaguggauuggucugugaaaugggcgggcauuggggucuuccgaguguagcug
ucucucuccuuuuuccacccccaauucuugaugcaucuccuuucccagGACUGGGGAACCU
GAUGAAGAGGAGGGAACUUUCCGCAGUUCAAUCCGCCgugaguuug<mark>gggagacuuagggugc</mark>
<mark>uugggaaugcgag</mark>uugguuccaagaagcccuuugcuggcccucccu<u>gggcgcgggggucgg</u>
<u>ggaucgggg</u>cuugaucugcccugccgugcucccaucuuucagaggagcuggggugccccucu
ugacuuccggauucucugucuuccccccucagGUCUGUCCACCCGCCGGCGGUAGAGACACCU
GGCGCGAUGGGCCCCAGCCAGGUCCCGCAGCUCUAACUAGGCUGGGGGAGGGAGGAUAGAGG
AGGAGGAGAGGGCAAAGGGCUGGGUAGCGGGCCUGGAGGGCUCUCCUGCCCUUCACCCUUUU
CACCCCCACAGGAUUCCCCUGGCACUUGACGCUUCCACCCACCUUCUGCACGCCCACCACCA
CCUGGACUGCCCUCUUCCCUAGCCUGCCCCCACAGACUCUGACGCUCAGAUUUCCAAUAAA
ACGUGCUUUUCUCUCUUG

Rattus norvegicus

GAGAACUGGGACUGCCUGAGUACGGAGAGACCACUGGCUGAGguguguaggggGcaggugagg
cuggGcauguccuguuugucucacagcauuccaccuguguccggagauagcuugcuucccuu
ucaaggguagguuuggggacucugggaccaccugucuguuggggguagccguggggaagccc
cugaggagucugccuccccaccucugccucaauauucguaugcuguguagucuccugccc
ccauuauggacuuguuuuagacuuggauugggggugaagcagggagcuagucaaguugaaug
gcugacaguuaaguccuuaggccgcccuugacuuaaccccaugggguucaggcccuuuauag
aaggggaagagagcagagcagaggacauuuuuugacccuggaugacuccuuagGGCAAUGGC
AUCUCCCGGCCACAUCCUGAUUGUCUGUGUGUGUCUCCUCUCCAUGGCCAGUGCAGgugagu
ccaaaggagguacccagcaucucagccaggugggggcagaggcagggaaaagcccccaag
aagaccaagugggagacccaaacuagaucccaugugacaggagaucaggggucuccacaua
ggcucugaucauggcuggcaucuaggggagaggggaagaugccgagugagguaaaaguggg
aguggauggaaggguaggucucuggggugacagaaauaugaacaaccugcaaggugucaggguag
ccgggccagugccggagguuccagggugcagaauaggcuggaaaccugccgacggcucagua
gcacccaaacuccgggcucaagaagauaggcagagccccucuaccuuucccugccacagac
cccaggagcuacucuaauucggaucuuuauuuuucuuucuagAAGCUCCGCAGGAACCAGAU
CCAUUCACCUACGgugaggguuucgagagggggucuggguguucuauucucuccucuccucu
cucucucucucucucucucucucucucucucucucucucucucucucucacacacacaca
cacacacacacacacucaggcacacucuuccuccacucucuccugcucuaguugucugag
acucccuugccugucuauuugaccucucugucucacugggacugcacguuggggagauggg
ugccccauccccacggaguggccauccucgccauuccuucccccuuuccaccaccuccucc
cucccuccuuccucuacccuguccucuccucccucacacaucugcucuuuccccuucccuc
uccuccuccuggcccugcaaacuauacccuuccccauuggucaccuccuccggggcugccc
acagAUUACCACACCCUGCGGAUCGGCGGCCUCACUAUCGCUGGGAUCCUCUUCAUCUUGGG
CAUCCUUAUCAUCCUUAgugagugucugcaccugucuucuccaccccgccuccagccuuccc
uccccaagccccacucccagcaacacacacagccugugcacuauacacgcccaacacaggcu
cagucuccaaacacuucucccacaucccacccugacuccaaccucgcaacacacagacccca
ccccaccucaaaccacgccccucaacucccacccagaccacgcccccacauccuacccuuc
cuccagcccugccccaggucccacccccugcggccucccaguuuggaggagagccuaggga
ugcuggagcaagagggaaucucagccuauacccucccacuccacagGCAAAAGAUGCCGGUG
CAAAUUCAACCAACAGCAGAGgugaguggucccucuggaccucccucgcuccucccgcagug
gagaggcggugggggcgaggcggcaaaugcacccauucugcagucgaacacguauuaagcgc
uuagugugguguuaaacccuaagccaggcuccggggucggagcgaugguucaagcucccucucu
gccuggcucagcgcccagccagugggggcgcugggcucccaccagccccgcagauggaaacu
caagcguggugaggcgaagcugggcagaagcagccgcgcugaaaugagaucaccucacaggg
cggcccaguuuagcuccaguccggauccucgcgcaggauuccucucgaaauaaaccuuuaaa
gcacaggcaacguaggcaugccuucugcgugcuaagacgaucacagaugccuucccaccgac
ccucuagcaugguugagugcgcaccacgagccaggcgccguuuucuauuauguaucccuuac
cccacccggguggaggugcgcuccaucauccaucccguuugugcauaugaggaaacugaggca
cggggagguaaggugacuuagggucacucacggacccaagucccaaaucucccaagaccu
gugaccgucccucucccaacaaugucaagugccuucccaacacccggaggccuccaccg
cgcaggcugggagcacuccugccgcacccugaaacagcagcgggaagcuucuguuucugac
ggaccgcgucuaugucaaguccacgguggggggucuggaaauaaggccugcacuggggcuucu

cagaagcggcuaccucuccugguccauccgaauuccucugucuguuccuucccagAACUGGG
GAACCCGACGAAGAGGAGGGAACUUUCCGCAGCUCCAUCCGCCgugaguucggggauacugc
ggggguuugugggggcagcuguuuuuaagagcccccucuucuugggcuuagaauggguuuggcg
gagaguuaucccugguugcagacuccccugaccccagaacucucugcaucccucucagGUC
UGUCCACCCGCAGGCGGUAGAACCUCCACCUGGCUCCAGGAAACUCAGCCAGguccugcagu
ugaugggaggaaggagggagagagguuggcgggguuggagggaggcaaggggcaaagugccag
uguggggagcguggcgaggagagcuugccacucacucucuucaccccccgcagAGCCCCCUUGG
CACCUGACACCUCUCCCCACCCAGCUGCUCGCCUGUGACCACCUCCAGCGUCCCCUGCAUCA
GCCCUGCUUCUCGGACACCCCUUGCUGCUCAGACCUCCAAUAAAACUCGGUUUUUCCUUCUUG

93

Monodelphis domestica

NNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNGUUUU
UGUUCAGCGCUGACAGCCCUUGCCAAAGUGCAGCUGUGUUCCC<u>GGGGGUGGGGAGGAGGG</u>AU
<u>GGG</u>CUGGAGGGUUUCUGACUGCCUGGCUCCCCCUGACCAGUCACUGCAUCCCCCUCCGAGGG
GGUGCUUGGGCGUCCCUCUGGGAGUCAAAGGGAAGGAAUGGGUGCAGUUGCCAUGGCUACAC
AGGGAGGCAAAUUCCUCUGGGGGUGAUGCUGGGGAUAUUUAUACCCGUGAGUGCCAGGCCUGA
GGGGCAGGCAGGAUGCCAAGGGCAUUACAGUGACAGAGAGUGGACAGGAGCGGAGUGUGUGA
CCUUGUGUGUGUGCGCGCCUGCCUCAACUCUUCAUUCCAUGGGGAUCUAUCUGUGCGCCCCU
UUGGGGGGGCGCUCUCUCUCUCUCUCUCUCUCACACCCUCCGGGGGAGCUGCAGGUUCCCAU
GUGGAGgugaggcugggccccucaacucugccucccugaugcccauccuguguugcgggc
uguuucccaggcucucuguucuggcucugaaucccacccaccccaggagggcucucuuugu
aucucagggcgaccaagcggcccagauggcucuggccggauuaggucuugggcccggcacag
acuuaacccuuggggaucgggucagaccugguaguggggguuggugguggguagaaaaaugua
gagaaggaggaggacuugggggggaccuaacuggguccacuguaucccacagGACAAUGGAGC
CCCUUGCCCAACUCCUGCUCCUCUGCAUAGCCCUGAUCUCGGCAGCCAAGGCCGgugagucu
agagauuggagaguuuggccucuaauucccagauuuccagagcug<u>gggagauggggggggu</u>
<u>aggugg</u>ugcagcaguacaacuucuggccaaacucccagcucagcugccauagcguuaggca
aaacauuagacacuggccagugugggcugaccagaagacacgggcagccucaggcuucaagc
uuacacauaccugaaaauugucugugcacccagaaucuccgaagagaagccagccuccccca
ggugagaaccccuuguaacauccagucacaccaagccuaacagggacggggagcgaaucua
gggcaaaagacuagccacguggccaucccaggcagaagcagggcugcaaccaucagaaguau
caacagcucauuucuagaacacuauagucuugugaggccaccagggcagggauuaccagccu
ccuuuuaucccuggggaaacugaggcccagagaggaacaacaucuugcuuagguucccauag
ugagugugagauuggcagaccuuggauuagaacucaugccucccaacucccacugagacagg
aagggacaagggguuaugccccauccucuaaacuaucacuggggaggcugggaaccugaaag
guccaauggcuuuuguggggggacauagugacuuauaguagaaaucaggcuggagcuuucau
uccugaacaaggaugcccuuagcaucccccagucuaucucccuaaacucuaaauccuugucuu
uccagAAGCUCCGAAGGAAAGAGACCCAUUCACAUAUGgugagaucagacucacgugcaugu
cgguguuucaaugucuguccccaacccuuauccuggaucucugacucuguuucuguccuucc
cugccucugccucucccacucucuccuccauccuuccuuucucucucuccuucuccgucccc
ugacccuucauacccuguucucucccuuucucagACUACCAGUCCCUGAGGAUUGGAGGGC
UAGUUUUUGCUGGCAUCCUCUUCAUCCUUGGCAUCCUCAUCAUUCUCAguaagucuuccucc
cuacccaugggacacccuggggauggaggaacuggaguggggaggguauguggaggaaugau
uucucacacacacacacacacacacacacacacacacacacacacacacacacacgcaca
cacacaccugugaaggcauuugcccaccuugcagggacauuccuuugcauaccuguagucuc
cccagauagaacauaagcuccucggaggcagggccugcuuccuuuugucuuugcugucccau
ugccuggcacaugguagacacuuaaauacacacugaaugauugauugauuuccccagGCCG
AAGAUGUCGCUGCAAAUUCAACCAACAGCAGAAgugaguaaugcccucugugucccccucugc
ucuccccuugucucuguacccaucuuugucuccaugucucauccccaucucucucuccaucu
aucuuugacccuuuccauuuguauguuuaacuucauugcugacuuuuuaucuucuuccaccu
ccugauuguuucucuuugucucucacuaguucucauauucaccccccuucuucaucuucauuu
ccaacuaucuugucaucuccaccucuaauagucuuugucgggccaucucugccuuuccccuc
cugucucugaccccaauuucacucuuaacacuuggucucugacucuuuuuaaaauguucuuu
ucuuuauuaugaacucauaugacacaugacauuguacugcauacaaaggaaagaaaggaugu
auaugaaacuaugaacuuggugauguauuuuuuuacuuugucagccucuuucucuccaucu
cugcuucugcuucuucauuucucaccucuuuucuguauuuaauuguauuuucuucuguuucu
ccacuuuugauucccuuucucucucuucaugucucccuucuuuuuucucuguacauuugagc
aucuuugcuucagucucuuccauuucugccuucuccaucucuaauucuaucccucuguuuuca
ucucuaacucuuucuuuguaucuuuucaucuuucuccauccuggacacacacacucacacac
auauagucucucccucuuccuccuucccucucccucucuccuucccucucccucucuccu

ucccucuccccucuuuccuucccucuccccucucuccuucccucuccccucuuuccuucccu
ccccccucucuccuucccucuccccucuuuccuucccuccccacucucucucucccucccuc
ccucccucccacucuccucuccccacccccguuucuuuccaugcucccccaugguuccuuuug
ucuuccucuuccacccccaggACUGGGGAAACCUGAUGAAGAAAAGGGUACUUUCCGCAGU
UCCAUCCGCCGU

Felis catus

AUGGCAUCUCUCCACCACAUCUUGGUACUCUGUGUGGGUUUCCUCGCCAUGGCCAACGCAGgugagucuggggguagguagcccgcuacccaccucagccccgagggaggggagggaagccuuucaagagaaccaaguuggagacuccaacuuaaacagagucagugggguaccaggaaguuggagguuucacgugggggagggagggaggaugcccagggcagacaaaguggagggugggauggagccccucuucccgcacaggacagaaucaaggacaacacgggaugaggcgggaagcuguagcaguuaagugaccaccugccccccccccaggacagugcagaugaggcagcuaagagccagagaggguggacagccuguucaggguccaauaguccccagacucucagcccaggacaggaaagcccccaaaaggucucuuuaugccccuuccugcccagggccccaggaagcauccaacccuugacuucuuucucuuuccagAGGCUCCACAGGAACACGAUCCAUUCACCUAUGgugagggagggaggggcauuuaugcuugggaguugggaggguggguggccuggguuugccgguccuguucccacuccuccugucucucucuuccuuuaucucugacucaguauuuauaucucugucagucuccuucucucucuuucucuugucucucagaaggcucucugacuccuucuuguuuuuaucuccccuugucuguguucucuguguuccggugugucuauuugacccucuccugucucucugauucucuuucuccuuuuaccucucaaugucuuacggcucauccucugcggccaucuccccuuucuuccccucuguuccccuccuccucuccaucauccuccccuacuuucuuguuuuccucccuccucuccccugcccugccuaccucuuccuuuuccuucccaccccucuccucucucucccaucccucuucucuuccuccccacaucuguuuccuccucucauugaucucucugcucuccucuccugaucuccuugccccuuccucccaauucugccccucuccuguccuuccuccuccucccucccuguuucucacucuccucccugccgcccacagACUACAAAACCCUGCGGAUCGGAGGCCUUAUCAUCGCCGGGAUCCUCUUCAUCCUGGGUAUCCUCAUCGUCCUGAGUGAGUACCCCCACCCUGCUACCUCCAGCCCCCGCAGGUGCCGUGUUUCGUCCCcgnnnnnnnnnnnnnnnnnnnnnNCAACCAGCAGCAGAGguaagaggcccccucggggcucucacucuccuccuuccgcucuagaagggcgcugggagugaggcagggaguccacucgacccacagccagagacccuucggcaaauacguauuaagcaccacuacgugccaagccgcaagccaggcccuggggaccaaacggcgaauaacugaacugcccucccugccuggcugagcucccagcccagugggggcggcggccguggguucccgcagccccacaaauauaaaaucacaaagcgugauaagugcugucaaagaaggagccacggcgaaaugagaucaccgcgcuuggcggcccaguuuagcucagaguccuguucccagcucuuugacucaucucuaaauaaaaauguaaagcacguaugaaucccuucuccgugugaaaacaauucuaaauguuauuuuacagaauagcuaauguuuagucccacgugcuaaaagguauuguuuuaucguguauuacguuauugacuccucacacccuugcaugagacugucguuaacagccucguuuuaucgaggaggaaaccgaggcgcagggaggugaagucacugcccgagaucacucuagaagaugaggguucggguuucuaaccuaggcugucuccaggccgucuguccucaagcccuucggccaucccucccccaccaacuuccagucucuuccccagcacuccgggaggcacuuggagcgcguccugccacacgcuuaaacugcuccugggugcgcccgauuuucucguuucuggcggccccgagccugcuccucgccggucucuggcccucaguacgggggucugugaaaugagggcguuacggggggguuccgggagcaucugccucucuuuuuucauccccacguccuucugccucuccuuucccagGACUGGGGAACCGGAUGAAGAGGAGGGAACUUUCCGCAGCUCCAUCCGCCgugagucugg<mark>gggagacuuuggggguuugggggugaggg</mark>cugguuccaagaacccccuuuccugguccucucugggcgugcggaggagggacuugacccgacagcgaagguccgggaguuccccugucgcucccccuccgcgggagcgggggugcccccuccugacacccgaacucuccguguuccccucagGUCUGUCCACCCGCAGGCGGUAG

Otolemur garnettii

AUGGCGUCCCUUGGCCACAUCUUGGUUCUCUGUGUGGGUCUUCUUGCCACAGCUGAUGCAGgugagucuagggggaggcagcccacuaugcaccucaacuccaaggaugacaggggguagggagaccucaugaaggaacccaacccacacuguuggggucaguuggggguaccaggaaauugagggggcuccacauggcagggauuuaaggggucuggccacaggcagggauuuaggggaggggagaugucaggcaaaguugggggugggaugaagucuucccaacuggacaauggaaucagacagccugcaguuagagagcuggggcaguuaggggcccccugcccaccccacccccgcccugaaagugccucugggcaggauggauccacagaggguggcagcauaccagacagcccccaaacucagcccaggaaagcagagcccggaagaauguuuuaugcccccucccugcccaggcccaggaagcauuugaccccugauuucucucucuuuccagAAAAUCCAAAGAAGGAACAUGACCCAUUCACCUAUGgugaggggcgucugggggaggguuugggucugacugcccuucccuauuucuucccucuuucucccacucucucucuuuuuucucucuccacccccucaccauuuauucucugucauucuccuucucucucuugcugucucucacugugucuagucuguuucucucccuucucucacucccugucucucugguugagcccucauggguucucucugucuuucccggggucucucaguaucucagggcccaucucccugggcuccugucuuuucuccccuccucaucucccuucuccuccccacacuccaccccugcucccugcccagcucccuucucuccucgcuuuccugggcauccuucuccucccugugguccuccuccccuuuagguuucucugaguaugua ugguuucuccucucauaucugcugcccucaucuucuuccuucccuccuccuccccugccccucugcucccauuguaccccucuccucuccuccuacccucacccugcuuccugccguccgcagACUACCGCUCCCUGCGCAUCGGAGGCCUCACCAUCGCUGGGAUCCUCUUCAUCCUGGGCAUCCUUAUCGUGCUCAgugagugcccccacccggcccccggcccccgccccggcuacgcccagccacgccccccgcgaggcgugcuggagcgagagcgccccgcagcgccagccgguggagucucagccuuucucuccucgcccacagGCAAAAGAUGCCGGUGCAAAUUCAACCAGCAGCAGAAguaagaggccuuucuccgccuugggugccuacuaagcucuagaa<u>gggcgcaggguggggugggugaggcgggg</u>aaguccaccccauccucggucagaucccugcagcgaaucuguauuaagcaccacuguguugccaaggcccacagucccaugguccccuugggaccaagcgaugaaaacaacucccacccucccuggccgagcucccaguccaguggggggcggcggccgugggguucugacagcuccacagauagaaaaucacaaagcgugauaacacaaagugccggaaagaaagggccgcggugaaaugagaucaccucacacuggcggcccaguuuagcuuggaguccuguuccuggcgcuuugauucagcuuuaaauaaaaauuuaaaagcacagcaguguaugg auaccuucgccaugcuaaagaaguacagauauuauaaguaacagcuaauauuuauugagug uguaccacgugucacacacuguuucaucaugucguuaggucucgugaggggaggugcuguuau uaaccccauuuuacaguugaggaaacugaggcacgggggagguaaagucgcuuugcccaagau cacucaacuggaagauggguaguucugggguuuccaccccgucaucgguggcagccugccag auuccuuugauccucccuccccccaacuucaggucccauccccaccucggcggagguacuc acguucaccugcugcgcauuuaaacugucccuggggugcuccucauuuccgacccugagccu gcuccucccacaccggucugggaaaugaggccuguauagggug000uuccuggaagggccgccu cuccccuuuuccaucccaauucccucugccucugccuucccagGACUGGGGAACCUGAUGA AGAGGAGGGAACUUUCCGCAGCUCCAUCCGCCgugaguuug<mark>gggagacugugggguauuuggg</mark><mark>gaagaagg</mark>cuaguuccaaggacuccuuuccuggcccgcccgggacacguugagggaggggcuugaucugggaguugccccgccgcgguccucccuccgugggagcggaggaugccccuucugacacccgaucucuuuguguugcccucagGUCUGUCCAGCCGCAGGCGGUAG

97

Tursiops truncatus

UGGCAUCUCUCAGCCACAUCUUGGUUCUUUGUGUGGGUCUCCUCGCCAUGGUCAACGCAGgu
gaguccggggaaggcagcccacuacccacuucagccccaaggguggcaaggguaggggaaac
ccugcaagagaaacaaguuggagacuccaaucugaacugagucaguugaggcaccaggaaau
uggggucucacgugggguucaguuacaggcugguaucuggggggaggggagaggaugccccag
gucaggcaaaggugguggagcgcgggauggaggccugugaaugacggaaucagagacaaccugua
augagacggagagcugaggcaguuaggugccugcuguuccacccaggacagugcagcggggc
agcuaagacccagagaggguaggcagccugcucaggguccagcaguccccagacucucggcc
caggaaugugaaacccccaaaaggucucguuaugcccuccuuguccaggaucccaggaagc
auccgacccugauuucuuucucucuuuccagAAGCGCCACAGGAACACGACCCAUUCACCUA
UGgugagggagggaggggcaccuuucuuugggaguuaggggggugcugggcugggUUcgccug
uccucuccucauuccucccuccccccaccucucucuguuucugucuccaucuaucucuaga
cucaguguuuaugucucugucauucucauucucucucuuucucuugaucuggcucucugacu
cuuucugucuugaucucccucugucucugucucucuguuggucucugucuauuucaucccuc
ucuguaucucugaaucuuucuccuuuuaccucuuaguaucuuauggcucauccucugcugcc
accugccuuucuccucucccugauccucaccaucauccuuccccacucccuugucuccuccc
uccuccuugccuugcccaccuccuccccucccuuuucuuuacauccucucuucucccugauc
ccccuucucccacaucucucccuccuccucucauugauuuccuuuccucuuccccauccuuc
cuccuugccccuuccuccccauucugccccucuaccauccuuccuccuccccucccugcauu
cacgcuccuccuugcugcccacagACUACCAAUCCCUGCGGAUCGGAGGCCUCAUCAUCGCC
GGGAUCCUUUUCAUCCUGGGCAUACUCAUCAUCUUGAgugaguaccccuacccagacuccag
ccgcgaaucugcgggugucugucccaaccccguccagucccgcuccgccuccaaccccgcc
ccuaucagagcccuuggccccgcccuuauuucugcccaccccggccucacgcccucggcccc
gccccuguuuaugccccuccccaucgcugcucuaggccccgcccccuucuacuccgcccucgg
ccccgcucccgcuucuaucaccgcccccgggcccgcccccguuuacggccccgcccucggccc
cgccccuguuucugccccuccccuuucccugccuaggcuccgcccugccccuucuacuccu
cccucgggccagccgcgucucuauugcaggcccagcacccccucgucccugccccacccuua
ucccccccgccgcuccagcccuggcgagcgcgagcucuggcgagcgcgagcucuggcucgcu
gugccgcguagagcguaaucucagcgcacuccuccccacgcccccacccacagGCAGAAGAU
GCCGGUGCAAAUUCAACCAGCAGCAGAGguaagaagccucucugggcucucacucaucuacu
cccgcucugaaggagcacuggggggcgaggccggaaguccacucaauccagagccagagagac
cccucggcaaauauguaucaagcaccuacuacguguccagccccaagccaggccuugggacc
aagccacgaauuaacugcccucccugccuggcugagcguccggcccagugggggcgguggcg
augggcuccaacagacccacaaauauaaaaucacgaagcgccguaagugcugugaagaaag
agccgcggcgaaaugagaucaccucacuuggcggcccaguuuagcucggaguccuguuucua
gcucucugauucauccuuaaauaaaaauguaaagcccaggcaaugugugaauaccuugucgg
uauuaagauacuuguagauguuauuauaaauaaagguaguauuuauugggugugucccacgu
gcaggaugcuguuucaugauguguuauuuaucccucacaccauaucuauaaggcaggugcuu
uuauuaaccucauuuuacagaugaggaaaccaaggcacgggaagguggucacacgccccaaa
ucgcucaacugcaaggugaaggcucuagguuucuaaccuaggccgucuccaggcaaucugcc
cucaagucgcuuuggcuauaccuccuuugcaaauucuaguuccuucuccagcauucggggag
guacucucuguuaccggccuggagcgccuccugccaaacacuuaaacugcuccuggggugcac
uggauuuucuggcggcccggagccugcuccucgcccaucucugggucucaguacacaagucu

gugaaaugggggcugcacuggggguuucccgggguguagcugccucucuccuuuuccaucccca
auuccuucugcaucucuuuucucagGACUGGGGAACCUGAUGAAGAGGAGGGAACUUUCCGC
AGCUCAAUCCGCCgugaguuug<mark>gggagacuucggguguuuggggaugcaag</mark>cugguuccaag
aacccccuuuccuggguccuccgugggcgcguggagcggggggcuugaucugaaacccgagguc
ggggaguugcccugcugugcucccaccuuccagaggagcuggggugcccccuccugcuauccg
aacucuauguguuuccccccuuagGUCUGUCCACCCGCAGGCGGUAG

Equus caballus

GGACGUUGGGGGCCUUCUUUCCGCAGGCGACAGCCUGAUUGGGgugagcgucccccgcccuc
cccuccaggccucccccugcgccuggcugggccgccuauuuugggagcaggaguggccagcc
uguggcuucccaggcaggccagacccaagaggaagggagugugguggggaaguuggggguuccc
ccaaaggugggagugugggugccgcuccuuccuucccugcggcuccgugugcaccccggcug
ggugugdguuuuggacacucacguguguaggccgguugcgugacuguguggggcaccaga
ggcccagaugacgaguaaugugugccugugucuccaccuacguccugcuccaaccagaagu
uggggagugggcguugucuccauucuaacccacugcgugugugugacugugug.uccgucc
cugccccuggcaucugucuucugucucucagccgcaccggccugggccuccugucacccgcc
cuguggugucugcucgggaaagaugucacccagagccgggcggaggagcuguuuuucccuuc
ucccauugcuuccaagagagcagcaaggguggccuucccacggggacagccuguggcacug
uggcagccgggccucaccccgcagggcugggcaugaccccgaguggggaaggaaggcugu
ugccauggugguccugugcgaggcaaauuccuccaggugaaguggagauauuuuuacccccg
ggggcaggcagagagcgggccggcgccgagggcaggagagcggggauaucgcagacacaaug
aggcugcgguguggcgugcaggugggcccuugggcgugüccuggccgugucaucccucacggg
gugucugucucuccaccgugggguguggcaccuugcuuucccacuagggcagcguggggaa
uucuggggccugcugcggggaggccccguggacggguaugücucacceccuguccuugcugu
caugugacacccaucauguggguuguuuuagugacacugüguucccaugaagcagaggguac
cccuguuucuugggguggccaagcuggaugacagaggcuuuuggaauuaaguccuucuacug
ucucugacuuaaccccuuggguücaggccccaguggüggagggacgagggüacaaauuuccu
gaccuuuggguguucccagGACAAUGGCAUCUCUUGGACACCUCUUGGUUCUCUGUGUGGGUC
UCCUCAGCAUGGCCAACGCAGgugagucuuggggagggagcccaccacccaccucacccccca
aggguggcaggggugagggaaaccuugcaagagaacugaguuggggacuucaaccuaaacug
aguuaguugaggcaccaggaaacuggggcucccacaugcaguccagucacaggcugguauuu
ggggaggggaucaggcaaaggugggcgacgggaugguggccagcuucccaugaaggaugga
aucuuagccuguaauaagaaggggagcuggggcagucaggugccccuugucccucucgggac
agugcagaggguccagguaagagccagagaggguggiggccgccugcucaggguccagucaccc
ccaaacucucagcccaggaagguaaagccccaagacggucacuuuaugcccucccuguccja
ggaucccaggaucccaggaagcauccugaucucucucucucuuuccagAAGCUCCACAGGAA
CAUGAUCCAUUCACCUAUGgugagggaggggcauccaucuuugggaguuggggaguggggüü
uggguuugccugúccugúccucauucacucucucuuucuuucuccccucucuccaugucucug
accccagcguuuauaucucugucacucuccuucucucucuuucccuugaucuggcucucug
acucugucuugaucuccccugucugucugcuucucucugucucuuucugugugucugccuc
cuuauuugaucucucccugucucucugauucuuucucuuucuccuuuaccucucaguaucu
cauugcccauccucugcugccuucugcccuuucucuccucccguuccccucauccucucca
ucacccccuuacucccuguuuccucucugcccugcccagcuucuccccucccuuucccuuc
caccucuccucucucuaaucucccuucucccuccucccauaucugcuccauccucucauug
acuuccuuccuccuuccccuccucccauucugcccucuccgauccuuccuccucucccucc
cugcccucacugüccuccuuguugcccguagACUACCAAUCACUGCGGAUCGGAGGCCUCAU
CAUCGCCGGGAUCCUCUUCAUCCUGGGCAUCCUCAUCGUCCUGAgugaguaccccgacccc
accgccuccagcccccucugcgggguuuccgccccagcuccgcccuguucccgccccgcuucca
accucgccgguauccacucggccucgccucuguucccgccccgcccccgucucccucggccc
cgccccuguucccguccugccucaucucccuagacccccgccccuauucccgccccgcccuau

cuccauaggccccgcccgcggccccgccucuaccccgcggccccgagucucgcaagcuccag
cuggagcacgagcgccgcgccgcgcccgccgggagccgacgcucagcccgguccuccccuuc
cccccacagGCAGAAAGUGCCGGUGCAAAUUCAACCAGCAGCAGAGguaggaggcccuccgg
cccgcacucaccuacucccgcuccagaggggcgcuggggggugagccagagccccgugcgcaa
aucugcauuaagcgccuacuaugugccuggcccccgaggcgggcgcugggacccagccgcgcc
uucaccgccgccggccuggcugagcucccagcccagcgcgggcggggggccguggguuccaac
aaccccacgaauguaaaaucacaaagcgugguaagcgcugcaaaggaaggagccgggcgaaa
ugagaucaccucacuugcuggcccaguuuagcucggaguccuauuccuagcucuuugauuca
ucuuuaaauaaaaauuuaaagcccaggcggaguaugaauaccuucuccaaguuaaagcaauu
agagauguuuuuauaaauaauagcuaauacugguuggggugugccuccgugccagauacugu
cugacugugugugguuaucuuagugacccucccgccacgucuaugagguaggugcugcucucag
ccucauuuuacagaugaggaaaccgaggcacgggaggugaagucacuagcccaggaucacuc
ggcuagaagacgaggauucgggguuucuaaccuaggccgucccugggcagggugcccccaag
ucccuuuggucaucccucccccgccagcucccaguccccuuccccccggcacucggcacucgg
gagaugcucccugucaccgcguggagcgccucccgccacgcccugaaccugcuccuggggggc
gccgccuccucgcccaccgcugggccucgggacacgggucuaugaaaugaggcugcccuggg
cauucggggagcggccgccucucuccuuuuccauccccaauuccuucugccucuccuuuccc
agGACUGGGGAACCUGAUGAAGAGGAGGGAACUCUCCGCAGCUCCAUCCGCCgugaguuug<mark>g</mark>
<mark>ggagacgucggguguuuggggggugaggg</mark>acguuuccgagaacccuuuccuggcccuccuug
gcugcguggagggaagggcuuggucugaaacccgaaggcggggaguucccccgcugcccacu
ccacagcagcggugugacucccugacacccgaacucucugugucccccccagGUCUGUCCACC
CGCAGGCGGUAGAAACAGCUGGCGCGAUGAACCCCAGCCAGguccugaacucugacugggcu
gcgggagcgagugaggggaggagggggcgguggcuggggggcugggggguggcugggaggg
cucuccugccucucacccguuucaccccauagGAUCCCCUGGCACAUGACGCCUCCCACCC
AGCCCCGAGCGCCCACCGCCACCAGGACUGUCAUCUUCCCGAGCCUGCCCCCACAGACUCG
UCUCUGCCGCCCAGACUUCCAAUAAAACGUGCUUUCUCUCCUG

Ailuropoda melanoleuca

AUGGCGCCUCUCUACCGCAUCUUGGUUCUCUGUGUGGGUUUUCUCACCACGGCCAACGCAGg
ugagucuggggguacuuagcccccacccaccccagccccaagggaggcaaggggugagggaag
cccuucaagagaaccaaguuggagacuccaaccuaaaugagucaguuggggguaccaggaaau
uggggguuucacauggggcccacucacaagcuaguauuuggggguagaggagcggaugcaagc
ucaggcgaagguggacgguggauggaggcccacuucccucagaugacagaaucacugacaa
cacggguaagacaggauuuggggggcaguuaggugcccccugccccacccaagauggugcag
augaggcagcuaagacccagacaggguggggagccugcucagggcccaacagucgccagacu
cucagcccaggagaguaaagcccccagaaguucucuuuauguccuucucuauccagaguccc
aggaagcauccaaccccugacuucuuucucuuuccagAAGCUCCACAGGAACACGACCCAUU
CACCUAUGgugaggcaggaagaggugucuauccgcgggaggugggagggcgccuggccuggg
uuugcccauccuguccucauuacuccucuucucuuucuccaucucucuuucugacucagugu
cucucucucucucucucagaucucugacuccuuguuuugaucuccccuugucugucucuc
ugugcgccugucuacuugaucccucucugucucucugagucuccuucuccuuuaaccucuca
guaucuuauggcucauccucugcucugccucccuuucucccucucuguuccccucauccuc
uccaccauccuucacuucuuccuuguuuccucucuccucccugcccugcccaccuccuccu
uuccuucccacccucuucucucccugaucccucuccuccuuccucccuuaucuucuucuucc
ccugacugaucuccuuaccccuccuccucccugguccuccuugcccacuccuccccauuc
ugccccucuccuauccucccccucccucccugcccucacucuccuccccugccguccacag
ACUACCAAUCCCUGCGGAUCGGAGGCCUCAUCAUCGCCGGGAUCCUCUUCAUCCUGGGUAUU
CUCAUCGUCCUGAgugaguacccccaccccgcuaccuccagcccccgcaaccucgcgagcaa
gagccggcgcacgagcgccacucggugcccgccuagagcugaaucucagccucgucucccuc
cuacccccacccacagGCAGAAGGUGCCGGUGCAAAUUCAACCAGCAGCAGAGguaagaggc
cccuccgggcucucacucaccuacuuucgcucuagaggggcgcagggagugaggcggggagu
ccacccgccccgcagccagagacccuucagcaacuauguauuaagcaccuacuacguuccaa
accccaagccaggcccuggggaccaaauggugaauagauaaacugccucccugccuggcug
agcucccagccccgugggggcggcggccgugggguuccaacagccccacaaauauaaaaucac
aaagcgugauaagugcugugagagaaagagccacggcgaaaugagaucaccgcacuuggcgg
cccaguuuagcucagcguccuguuccuagcucuuugauucaucuuuaaauaaaaauuuaaag
caauguaugaauaccuucuccauguuaaaacaauuauagacaugcuuuuaaauaauagcuaa
uauuuagucccaaaugccagauacuguuuuauugguguauuaucuuauuuacuccucacaccc
cuuguaugaggucggugcuguuguuaaccucauuuuacugaugaggaaaccgaggcacaggg
aggugaagucacuugcccaagaucacucaacuaggagauaaggguucggguuucuagccua
ggcugucuccaggccggcugcccucaaguccuucagcacccucccccaccaacuucccggc
ucuuccccaacccuccgggaggcacuggagcaccuccugccacaagcuuaaaccgcuccug
ggugcgccugauuaucuagcggaccccgcgccugcuccucgcccaucucuggcccucaguac
aggggggugugugaaaugagggcugcacuggggcauucagggggcggcugccucucuccuuuu
ucaucccuaauucuuucugccagGACUGGGGAACCUGAUGAAGAGGAGGGAACUUUCCGCAG
CUCCAUCCGCCgugaguuug<mark>gggagacuucggguguuuggggugagg</mark>cagguuccaagaa
ccccuuuucuggcccggacaggggcgcggaggagggcuugaucugaaaccagaggucugg
gaguuccccugcuguuccccgcuccgcaggaacuggggugcccuccugacacccgaacucu
cuguguuccccucagGUCUGUCCACCCGCAGGCGGUAG

Pongo abelii

AGGCUUCCCAGGCAGGCCAACCCAAGAGGGAGGGAGUGUGGUUGAGGCAGUGGGUUCUGCAG
GGUGGGAUGUGGGCGACUCCUCCCUGCCCUGCUGGUGCGUGUGCACCCUGGCAGGGUGUGGA
GUUUGGACACACACGUGUGUAGGGCUGGUUGCGUCACUGCGUGGGGGGCACCGGAGGCCCAGA
GGAGGAGUACUGGAUGCCUGACGGUGUUUACACCCCACGUCCUGCUCCAACCAGAAGUUUGG
GGAGAGguuguuguccauguccauuccggccccacugugugugugugugugugugugugguaucc
cugccccagcauguguuuucuaucucucaggcccacugggcugggccucaugucacuugccu
gacauccgauugugaaagaugucacccagaggcgggcagaggggcugucuuuuucuuuucuu
guugcugcccagggaggagauggggguggacuuucccacaggggcagccuguggcgauguggc
agcugggccucaccccggcagggcugugcgugaccccugagugggggaaggcaggcuguug
ccaugguggccugagcgagcagaauuccuccaggugaagugggagauauuuauaccccggg
ucaggccgggagcgggcgggcggagagggcagggagcugggauuucgcggggcauagugagg
ccgggcauguaggcaggugggacuugggcgugcccugcugucuccugcuccguguuugugug
aggcagcgccuccucugcccugccaggguaggucugggaaucgggggccugcugcgggaggu
ggaggccuaagggaggcccccagggacugugugucucaccccugucccugcuacguuguguu
guugugugaccccaucguggagguuguuuuggugacacugugucccacgaagcuggggaua
cccguuucucuagcuuggagccaccaagcuagaggacgaacgcuucugugauucggucccca
gacugucucugacuuaaucccuuggguucaagcccugugugggagagcaagggcacacacug
ccuaauccgugguguccccccccagGACAAUGGCAUCUCUUGGCCACAUCUUGGUUUUCUGUG
UGGGUCUCCUCACCAUGGCCAAGGCAGgugagugcaggggaggcugcccgcuacccaccuca
gccccagggauggcggcggggaccgaagaaccaaguuggagaccccaaccuagacugagucg
gcuggguuaccaagaaguuugggggucgccacauggggguccagucacaggcugguauuuggg
ggaggggagaggaagccccagaucaggcaaagauggggguggggauggggggcugaauuccgguu
gggauaacugggucacagacagccugccgugagucagggagcugggggcaguuaggugccgcc
ugccccuucgggacagugcaaaggaggcagcugggacccagagaggguggggcagccugccu
agacacccucagacucucagcccagcacggcagagcccccaguggucuccuuaugccccucu
cugccaggaccccaggaagcauucaaccccugauuucucucucuuuccagAAAGUCCAAAGG
AACACGACCCGUUCACUUAUGgugagcgggggggucuaauuuugagucuuggggggagagccug
gcuuugcuggucguuugauuccccucgcccucccccagagucccaguauugauaucuguca
uucuccuucccucuauuuuguccuuccucucugauuccaccugucugcaucuucuccugucu
gugucuaucugugucgcugucugugugauaccucucugguugucuuucucuugccugggucu
gucucagccucucguggcccauccucugcuucuucccaucuucucuccccgcugucuccuc
ucccugccccguccucccuuuccauacacccucuccucucccuggucccccacuuucc
uccuuccauaucugcuccccuuaauuaucuuauuuccccccuucugccugcugguccuuuc
ucccuguucccuccuccccaauuuacccucuccuauucucccuccucuccuccccugcccuc
accuucccugcucugcugcucacagACUACCAGUCCCUGCAGAUAGGAGGCCUCGUCAUCGC
CGGGAUCCUCUUCAUCCUGGGCAUCCUCAUCGUGCUGAgugagugccccuagcccccgcccu
cuaccccgccucucccuggccccgccucucccuagccccgccccucccgccccaacccccucc
caggccuugccccgccuacccugccuuggcucccuggcccccggucucgccucuagccccgc
cccgucccccaagccccgccccccgcgcgggcgagcuggagcgacagcgccgcuuggugccc
gccaggagggagccucagccucuccuaccucuccacgcccacagGCAGAAGAUGCCGGUGCA
AGUUCAACCAGCAGCAGAGguaagacgcccauccacgcccuccuucgcccgcuccugcucua
gagggggccgcgggugaggcggggaguaccccugacccgcagcccgauccccaucagcgacu
auguauuaagcaccuacuaugugccauggcccaagccuggcccugggaucaagcgaggaaaa

aaccucccgcccuuccuggccgagcucccagccuaguggaggcgguggccguggguuccaac
agccccacagauagaaaaaucacaaagcgugauaacacaaaaugcaggaaagaagaaacggc
ggugaaaugagaucaucucacacgcggcccaguuuagcuuagaguccuguuccuagcucuuu
gauuccucuuugaauaaaauguuaaagcauggacaauguaugaauauguuagaacaauuaua
gauauuaucauaaguaguagcuaauauuuauugggguguguaccacgugucagauacgguuuc
acuuccucugggagggaggugcuguuauuaaccccauuugacagaugaggaaacugaggcac
agggagguuaaagucacuuuguucaagaucacucaaguggaauauggggaauucgggguuuc
caacccaggccaucucauggcagucugccaaguccccacgacuaucccuccccuaccaacuu
cacaucccugcccccaaauccgaggaguacucaccguuaaccagcuuagaagcccccuguc
agcacuuaagcugcuccugggugcuccucauuucuagcggaccccgagcccgcucuucgucc
auaucugggccuaguuacaccaaucugggaaaggaggcuuguacugggggguuccuagaagg
gcagccucuccccuuuccaucccgaaaucccucugccucugucuucccagGACUGGGGAAC
CCGAUGAAGAGGAGGGAACUUUCCGCAGCUCCAUCCGCCgugagucug<mark>gggagacugcgggu</mark>
<mark>auuuugggggagaggg</mark>cugguuccaaggaacccuuuuccuggcccucccuggcugcguagagg
gaggggcuggaucugaaagcugaggguggggaguugccccgccgcgggccccaccugcccag
gagcuggggaugccucucaugaaugaccccgaucuccguguuccccccagGUCUGUCCACC
CGCAGGCGGUAGAAACACCUGGCGCGAUGCAAUCCGGCCAGgugcugcagcucugacacggc
ggugggagggaaggaggaggaaggaaaggcgggagagggagggggccaagugcca<u>ggguug</u>
<u>aagggcggcgaggggugggg</u>cuggacguccccccucgccucucacccuuuuuccccucacag
GACUCCCCUGGCACCUGACGUCUCCCACGCUCCACCUGCGCGCCCACCGCCCCCUCCGCCGC
CCCUUCCCCAGCCCUGCCCCCGCAGACUCCCCCUGCAGCCCAGAAUUCCAAUAAAACGUGCG
UUCCUCUCGACUGCACUUUGUCGGUCUCGGUCCCUCAGCGCGAAAGCCCAGCGCCCCUGGAC
CCCAGCAGGGGGCGCCCCACCCUAGAGGAUUGUGCGGGGACGACGGUGGUGGGCGGGGGCGG
GGGNNNNNNNNNNNNNNNN

Oryctolagus cuniculus

cgcggguggGcggaggcacacccuugcugagcacgcugucccacaggacAUGGCCUAUCUC
CAUCACACCUUGCUUGUCUGCAUGGGUCUCCUCGCCAUGGCCAAUGCAGgugagucgcggg
cggcacGgggaggagaaaccccaugagaaccacaucggagaccccaaccucgagcuggggu
auaagcuggggugcuaaacugaggacccccacaugaggcugagcugcaggcgaggggaggcc
aagguaggGGgaggGaugcuggcagaggcaccugccacgagucagagugcugggacccagag
aagguggGcggcccacccaggaagguggagccccauaggGucuucagaggGucgcccucucu
gagacgcgaccccgaugccucucuuuccagAAGCGCCGCAGGAGCAGGAUCCAUUCACCUAC
GGgugagcaggGucccucuguuggGggGagugGaggGgagggGacuggacuugcccgucuagucu
caugccgcccgcaccccuguccucgcucucuccugcuggGgggucucucccucucagcaucuc
gggGcucaugGucagcugcuuccaaggccgcucccguucccguacaccuucuccugcgugcu
cccccccccuccccccgcccagggaucuuucuccccuucuccugccaccucuuuccucccccgc
cacccacagACUACCAGUCCCUGCGGAUCGGAGGCCUCAUCAUCGCCGGGAUCCUCUUCAUC
CUGGGCAUCCUCAUCAUCCUCAgugagagcccgcccggucgcgcccccugcccagccugugu
cccagccaccccaccccgccucgcccugucccaccccaccuccgcguccugacuccagc
ccugcccguguucuacccgcccuacgaccacccaccccgugccucgccugccucggcccccu
cuuccggacaccaguucugccccgccuccugcccgccuaguuaaguccgccccugcccu
cgcucugcccugcuannnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnn
nnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnn
nnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnn
nnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnn
nnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnn
nnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnn
nnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnn
nnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnccuuA
GAGAGGGGCGUGGGgugaggccggcaguccgcagccccgcggccagaauccuucggcgaacg
ugcgugacgcaccugcugugugccaagcccccggcccgggguggcgcgcgguggcagcggcg
cccgccugggaggucaccccggcgaggacacgcagcggcaggggGccacgcggggucgcucg
gcggccugguuuagccccgaguccuguucagAGCGCUUUGACACAGCGCGAAGgaaagcgac
accugcgcgcgcucacaggGuguucuuacggcgggagccagagcugccggggcgcgcgccacg
ugcgagaccccgcugcauugcaugcggucuuacugaugccucacaccgcccgcacuaggcag
gccgccacgcggggagauggaggcaccugcccgagaucaugcc<u>gggagaguggGugggGguc
cuggg</u>ucuacgacucgagucccugcggagggacucgccgcuccugucacccgcucuccugcc
cggccgGggaggcugcuccucgcgcgcccugggccucccgauaccggcuugggaaacgggGu
uccgcggggcuuccuggaagcaguugccucuccacccccaacucccugcgccucucccuuc
cagGACCGGGGAGCCCGACGAAGAGGAGGGGACUUUCCGCAGCUCCAUCCGCCgugaguccg
g<mark>ggacgggGcgcggaagccuagggGcugaggg</mark>cuggacccggaaucccuuccccggcgcuugcc
gggcgcccgggggGgagggGcuugaucugaaaugggGuugcgggGgaguagcccgcagccgccg
cgggcccccuccugacgccagaacucuccgugugcccucagGUCUGUCCACUCGCAGGCGGUA
A

105

Gorilla gorilla gorilla

AUGGCGUCUCUUGGCCACAUCUUGGUUUUCUGUGUGGGUCUCCUCACCAUGGCCAAGGCAGg
ugagugcaggggaggcugcccgcuacccaccucagccccagg<u>ggguggcggugggg</u>accgaag
aaccaaguuggagaccccaaccuagacuaagucggcugggguaccaagaaguuuggggggucu
ccacgugggguccagucacaggcugguauuuggggggaggggagaggaagccccagaucaggc
aaagaugggguggggaggggcugaaucccgaugugauaacugggucacagacagccugccg
ugagucagggagcuggggcaguuaggugccaccugacccaucugggacagugcagagggggc
agcugggacccagagagugugggcagccugccuagacacccucagacucuaagcccagcaag
gcagagcccccaguggucuccuuaugcccucccugccaggaccccaggaagcauccaaccc
cucauuucucucuuuccagAAAGUCCAAAGGAACACGACCCGUUCACUUACGgugagcgggg
ggggucuaauuuugaguccuggggggagagccuggcuuugcuggguccuuugauucccccucgc
ccucccccagaguccccaguauugauaucucugucauucuccuucccucuauuuuguccuucg
ucucugauuccaccugucugcaucuuuuccugucuguggucuaucugugucacugucuaugug
auaccucucugguucucuuucucuugccugcgucugucucagcaucucauggcccaucccucu
gcuucuucccaucuucucuccccccguccuccuccucccugucccccuccuucccuuuccuau
acacccccuuuccucucccugguaccccacuuuccuccucccauaucugcuccccccuuaauua
ucuuauuuccccccuucugccugcuggguccuuucucccuuccuccuucccaauuuaccccuc
uccuauucucccuccucuccucccugcccucaccuucccugcucugcugcucacagACUACC
AGUCCCUGCAGAUCGGAGGCCUCGUCAUCGCCGGGAUCCUCUUCAUCCUGGGCAUCCUCAUC
GUGCUGAgugagugccccuagcccccgcccucuaccccgccucucccuggccccgccucucc
cnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnn
nnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnn
nnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnn
nnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnn
nnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnn
nnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnn
nnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnn
nnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnn
nnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnggguggccguggguuccaacagcccccacagAUAG
AAAAAUCACAAAGCGUGAUAACACAAAGUGCAGgaaagaagaaaccgcggugaaaugagauc
aucucacacgcggcccaguuuagcuuagagucuuguuccuagcucuuugauuccucuucgaa
uaaaauguuaaagcauggacaauguaugaauauguuagaacaauuauagauauuaucauaag
uaguagcuaauauuuacuggguguguaccacgugucagauacgguuucacuuccucugggag
ggaggugcuguuauuaaccccauuugacagaugaggaaacuaaggcacagggagguaaaguc
acuuuguucaagaucacucaaguggaagaugggggguucggguuuccaacccaggccaucu
cauggcagucugccaaguccccaugacuauccucccuaccaacuucacauccugcccccc
aaauccgcggagguacucacuguuaaccagcuuagaagcccccugccagcacauaagcugcu
ccuggggugcuccucauuucuagcggaccccgagccugcucuucguccauaucgggccuagu
uacaccagucugggaaaggaggcuuguacgggggguuccuagaagggcagccucucccccu
uuccaucccgaaaucccucugccucugucuucccagGACUGGGGAACCCGAUGAAGAGGAGG
GAACUUUCCGCAGCUCCAUCCGCCgugagucug<mark>gggagacugcggguauuuuggggagacgg</mark>
cugguuccaaggacccccuuuuccuggcccucccuggcugcguagagggaagggcuggaucug

aaagcggagggcggggaguugccccgccgcgggcccccaccugcccaggagcuggggaugccu
cuccagaaugaccccccgaucuccguguuccccccagGUCUGUCCACCCGCAGGCGGUAG

Sus scrofa

GAGGGGUGGGGUGGGGCUGGGAGGGGACACCGCUGAGGGCGGCGCGUCCAGCUCAGGGCCAG
GGGGUCCAGCCGGCCGUUUGGGGGCCUUCUUCCAGCAGGGGACAGCCUGACUGGGgugagcg
uccccccuccccuccaggccucacccguggccuggccgggagccuauuuugggagcagaag
uggcgcccaggcaggccagacccaagaggaagggugugugguuugggucguuggggcuccccc
agggguggggagugugggugcagcuccuccuuucgccuguggcugcgugugcacccuggcugg
gugugcgguuuggacacucacgugugagcucgggcugcguccuuguuugggggccccggga
ggcccagaugaggaguacugugugccuguggcuccacccgagucuugcuccaaccagaagu
uugggggagugggguguugucuccacuccagccccacugugcgcgcgugcauguaugcgugagu
gcgccccggccccaggcaucugccuugggucucucagcccugcuggccugggccuccugguca
cucgccuggguguuguccacuguugaaagaugucacccagagacaggcaaaggggauuuuuu
ccccguucuccucacuuccaggagaggagaagaggugggccuuucccacggggggcagccugug
gcguguggcagccaggccucacccccggcaaugacccccgaggcaggaggggggggauaggaaa
gcuguugccauggugggccugugcgaggcaaauuccuccagggugaagugggagauauuuaua
cccaggggucaggcagagagugggccaguggccgagggcaggagagcugggaucuuucagaca
cagcgaggcaguuuugugcaagcaggugggguccuuggggugugucuccuggcugucuggggcucug
gcaggcuggccgucucucugccaggggggggugaggcaccuccaccaccggaguaggugugg
ggaaucuggggccugcugugggggaagcccgggggccuucgggaggucccuguggauuuugugu
gucccccccugccugguccuugguccuugcugccauaugaccccccccccccauggguggggaugu
uugggugagacugugccugggugaagcaggggggugccuuguuucccagggcgaccaagcuag
auggaugcuuuuggaauuaagucucuggacuguaucuuaaccccuuggguucaggccccagu
ggguaggggaacaaggguacacacuuccugacccuuggggugcucccccagGACAAUGGCAUCUCU
CAGCCACAUCUUGGUUCUCUGGGUCGGGAUCCUCACCGUGGUCAACGCAGgugagucugcgg
gaggcagcccaccacucaccucagccccaaggggguggcaagcguagggggggaaagccugcgaga
gaacuaaguuggagacucggucuacucuaagucccguugaggagccaggaaauuuggggggucuc
acauggggguccagucacaggcuggugugugggggggaggggagaagaugcccccaggucaggcaa
aaaauggagggugggggauggaggcgggcuuccgugaacgauggagucagucgauggacaaccug
uaaugagaugggagcugaggcaguuacauggcccugcccugcccugcccaggacagugcaga
gggguggcugggacccagagcc<u>ggggguggggguggggguggggg</u>ggcgcggccugcucaga
guccaacaguccccagacucucagcccagaaaugcgaacccccaaaaggaccucguuacgcc
ccucccuguccaggaucccagggagcauccaccucugaugucuuucucucuuuccagAAGCU
CCACAGGAACACGACCCAUUCACCUAUGgugagggaggggagggggccucuuucuuggggagcu
ggcuggugggggugggggggcugggcugaguuagccuguccucuccuccccccaucucucucug
uuucugucuccaccugucucuagacucucaguguuuauccccucguucauucucauucucucu
cuuucucaugaucuggcucuaucuugaucucuuuggcuguuucucucugucucucagugagu
cccugucucucugucucucugauucucuuucucccuuucccccucucaguaucccaugggcccag
ccuccacuuccuccuggcuuuucuccucuuccugcccccccucauccuuuccaucaccucccc
aucccuuguuuuccuccccuccuccuugcccugcccaccuccucuuucuccuuuuucuuuac
auccucuuuucuucccaucuccugucucucgucucucccuccuccucuuauugaucuccuuu
cuccucuucccggauccuuccuccucuggccccuuccucccuauucugcuccucuccuggcuuuu
ccuccucccucccugcauucaccgucccuccugcuguccacagACUACCAAUCCCUUCGGA
UCGGAGGCCUCAUCAUCGCCGGAAUCCUCUUCAUCCUGGGCAUACUCAUCGUCUUGAguggg
uaccccacacugccuccagccaugcuucugcgcgugucugcuccagccccgcccugucccc

gccccgccuccaaccccgccucuaucccagcccuuggccccgcccuuguuucugcccugcac
cuguucugcccugggccccgucucugucucgccuucggacccgcccucgccccuuccuaucc
ccgcccucggccccgccccuguuccugcccccacccuuaucccgccccccagccucuucugc
ccuggcucgcgcgcgagcucuggcuggcugugcggcacgagcagaauaucagcgcgcucuu
ccucacacccacccacag**GCAGAAGAUGCCGGUGCAAGUUCAACCAGCAGCAGAG**guaaga
ggccucuguggaccucacuccccuacuccggcucuaaaggggagcugggggugaggccgga
aguccacucaauccacagccagagacaccggcaaauauguacuaagcaccugcuaugugccc
agccccgagguagccugggaccaagccacgaaauaaguaaacuccuugccuggcugagcucc
cagcccagugggggcgguggcaguggguuccaagagacccacaaauauaaaaucacaaagcg
ugauaagugcugugaaagag CCUCGGCAAAAUGAGAUCACCUCACUUGGCGGCCCAGUUUAG
CUCGGAGUCCUGU

Ovis aries

AUGGCAUCUCUCAGCCACAUCUUGGUUCUUUGUGUGGGUCUCCUUGCCAUGGUCAACGCAGg
ugagucuggggggaggcagcccacuaccaccucagccacaggggugggcaagggguaggggaaaa
uucugcaagagaaccaagcuggagccuccagucugaacugagucucagugggggucucacgug
gggucuagucacaggcggggagcgggggaaggggagaggaugcccagggcuggggcaaaggg
ggaggguggggauggaggccggcuucccgugaacgauggaaucauagacaaccuggaaugaga
uggagagcugaggcagucaggugccuccugcccuacuccggacaguacagagaggcagccaa
gacccagagagggugcgcagccuacucagaguccaacaguccccagacuccugccccgagaa
uguaaaacucccacaggggucucgugaugcugcucucuguccagggucccaggaagcauccaa
cccugauuucuucuucucuuuccagAAGCACCACAGGAACACGACCCAUUCACCUACGguga
gggaggggcaucuuucucuggggugcuaggugguuagccugcccucuucucauuccuccccuc
ucccuaccucucucuguuucugucuccaucauucucuagacucucggguguuuauaucuuugu
cauucucauucuuucucucuugauuuggcucucugacucuuucugucucgaucucccuuugu
cucugucucucuguggugugagucucugucucuccaugugguccucucugucucucugauu
cuuucucuucuuaccucuuaguaucccacggcccauccucuguugccaccugccuuucuccu
cucccugaccccuuucccucuccaucauucuuccuuacucgauugcuuccucccuccucccu
ugcccuguccaccuccucccuuuucucccugauccccuucucccacaucgcuccuccucc
ucucacugauguccuuucucuccucuuccccauccucccgcuuugccccuuccucccauuc
ugccccuuuccuauccuuccuccuccccucccugcauucacccucccucccugcauucacccg
ccucccugcugcccacagACUACCAAUCCCUGCGGAUCGGAGGCCUUAUAAUCGCCGGAAUU
CUCUUCAUCCUGGGCAUACUCAUCGUCUUGAgugagucccccuauccugccuucagucgcuu
ccggugucugccccagcccugcccaaccucccgcuccgccuucaaccccgcccccuucugugc
ccuuggccccgccuuuguuucuguccacccaacuuuccgcccucagccucgcaccuguucuu
ccccugcccucagccacgucccaguuucugcccuuccccccauucccgccuccggccccgccc
cguccuuauccugccugcccucgggcccgcccccgccucgaunnnnnnnnnnnnnnnnnnnnn
nnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnn
nnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnn
nnnnnnnnnnnnnnnnnnnnnnnnggggaaucucagcgcacuccucuccaugcccccacccca
cagGCAGAAGAUGCCGGUGCAAAUUCAACCAGCAGCAGAGguaagaagccucucuggacccu
caucuucucucgcucuaaaggagcguuggggggugaggccagaaguccacuugauccacagac
ccgucggcaaauauguauuaaacaccuacuacgugcccagccccaagccaggccuuggcacc
aagccacgaguuaacugcccucucguaugucugagcuuccagccaaguggggggcgguggcc
auggguuccaacagacccacaaauauaaaaccacaaagcaugauaagugcugugaaagaguc
ggggugaaaugagaucaccucacuuggcggcccaguuuagcucagagcccguuccuagcuc
uuugauuuaucguuaaauaaaauuuuaaagcccaagcaacauaugaauaucuucuucccauu
aagauacuuguagauguuauuauaaauaaagguaguauuuuauugugucccucguucaagau
gcuauuucauuaugguguuauuuaccucucacaccaucucuauaaaguagguguuauaa
ccucauuuuacagaugaggaaaccgagacaugggaaggguuucacacgccccagaucauuca
acuggaagaugagggcucuaguuucuaaccuaggcugucuccaggcaaucugcccucaagu
ugcuuuggcuguauaccuccucuuccaaauucuaguuccuuccucagcacuguggggagguac
ucgcuguuaucggcuuggagcgccuccugccaaacacuuaaacugcuccgggguugcacugg
auuuucuggcggcuccaaccugcuccuugcccaucgugggggucucaguggauuggucuguga
aaugggggccggcauuggggucuuccgaguguagcugucucucuccuuuuccacccccccaauu

ccugauacaucuccuuucccagGACUGGGGAACCUGAUGAAGAGGAGGGAACUUUCCGCAGU
UCAAUCCGCCgugagucug<mark>gggagacuucggguguuugggaaugcgag</mark>uugauuccaagaag
cccuuugcgggcccucccugggcgaggggggucgcggggagggggcuugaucugcccugccgu
gccccgaccuuucagaggagcuggggugcccucuugacuuccggauucucugucuuccucc
ucagGUCUGUCCACCCGCCGGCGGUAG

APPENDIX II

***FXYD1* variant 009 pre mRNA sequence ENSEMBL Transcript ID: ENST00000589121**

```
Key:    UTR region
        Intronic sequence
        Exonic translated sequence
```

UUUUCUGUGUGGGUCUCCUCACCAUGGCCAAGGCAGgugagugcaggggaggcugcccgcua
cccaccucagccccaggggguggcgguggggaccgaagaaccaaguuggagaccccaaccuag
acuaagucggcuggggguaccaagaaguuuggggggucuccacgugggguccagucacaggcug
guauuugggggaggggagaggaagccccagaucaggcaaagauggggugggaugggggcugaa
uccccgaugggauaacugggucacagacagccugccgugagucagggagcugggggcaguuag
gugccaccugccccaucugggacagugcagaggggggcagcugggacccagagagugugggca
gccugcccagacaccucagacucuaagcccagcaaggcagagccuccaguggucuccucau
gccccucccugccaggaccccaggaagcauucaaccccugauuucucucucuuuccagAAAG
UCCAAAGGAACACGACCCGUUCACUUACGgugagcggggggucuaauuuugaguccuggggg
agagccuggcuuugcugguccuuugauucccccucgcccuccccccagagucccaguauugau
aucucugucauucuccuucccucuauuuuguccuuccucucugauuccaccugucugcaucu
uuuccugucugugucuaucuguaucacugucuaugugauaccucucuggguucucuuucucuu
gccugcgucugucucagcaucucguggcccauccucugcuucuucccgucuucucuccccccc
uguccuccuccucccuguccccucccucccuuuccuauacaccccuuuccucucccuggguac
cccacuuuccuccucccauaucugcuccccuuaauuaucuuacuucccccccuucugccugc
ugguccuuucucccuguucccuccuucccaauuuaccccucuccuauucucccuccugucuu
cccugcccucaccuucccugcucugcugcucacagACUACCAGUCCCUGCAGAUCGGAGGCC
UCGUCAUCGCCGGGAUCCUCUUCAUCCUGGGCAUCCUCAUCGUGCUGAgugagugcccccuag
cucccgcccucuaccccgccucucccuggccccaccucucucuggccccgccucucccuggc
cccgccucucccuagcccccccucucccuggccccgcuucucccuggucccgccccuccccugg
ccccgccccgccccaaccccucccaggccuugccccgccuacccugccuugguuccccggcc
cccggucucgccucuagccccgcccccgucccccaagCCCCGCCCCUCGCGAGGGCGAGCUGG
AGCUACAGCGCCGCUUGGCGCCCGCCGGGAGGGAGCCUCAGCUUCUCCUACCUCUCCACGCC
CACAGGCAGAAGAUGCCGGUGCAAGUUCAACCAGCAGCAGAGguaagacgcccccucccccgcc
cuccuucgcccgcuccugcucuggagggcgccgcggggugaggcggggaguaccccugacccg
cagcccgauccccgucagcgacuauguauuaagcaccuacuaugugccauggcccaagccug
gcccugggaccaagcgaggaaaaaccucccgcccuuccuggccgagcucccagccuagugg
aggcgguggccguggguuccaacagccccacagauagaaaaaucacaaagcgugauaacaca
aagugcaggaaagaagaaacggcggugaaaugagaucaucucacacgcggcccaguuuagcu
uagagucuuguuccuagcucuuugauuccucuucgaauaaaauguuaaagcauggacaaugu
augaauauguuagaacaauuauagauauuaucauaaguaguagcuaauauuuacuggguguug
uaccacguguuagauacgguuucacuuccucugggaggaggugcguguuauuaaccccauuu
gacagaugaggaaacuaaggcacagggagguaaagucacuuuguucaagaucacucaagugg
aagaugggggguucuggguuuccaacccaggccaucucauggcagucugccaaguccccaug
acuaucccucccccaccaacuucacaucccugcccccaaauccgcggagguacucacuguua
accagcuuagaagcccccugccagcacauaagcugcuccuggggugcuccucauuucuggcgg

accccgagccugcucuucguccauaucugggccuaguuacaccaaucugggaaaggaggcuu
guacuggggggguuccuagaagggcagccucucccccuuuccaucccgaaaucccucugccuc
ugucuucccagGACUGGGGAACCCGAUGAAGAGGAGGGAACUUUCCGCAGCUCCAUCCGCCg
ugagucuggggagacugcggguauucuggggagagggcugguuccaaggaccgcuuuucccg
gcccucccuggcugcguagagggaagggcuggaucugaaagcggagggcggggaguugcccc
gccgcgggccccaccugcccaggagcugggggaugccucuccagaaugacccccgaucuccgu
guucccccccagGUCUGUCCACCCGCAGGCGGUAGAAACACCUGGAGCGAUGGAAUCCGGCCA
Ggugcugcagcucugacacggcggugggagggaaggagggaggaaggaaaggcgggagaggg
aggggggccaagugccagaguugaagggcggcgaggggugggggcuggacgucccccccucgccu
cucacccuuuucacccucacagGACUCCCCUGGCACCUGACAUCUCCCACGCUCCACCUGCG
CGCCCACCGCCCCCUCCGCCGCCCCUUCCCCAGCCCUGCCCCCGCAGACUCCCCCUGCCGCC
AAGACUUCCAAUAAAACGUGCGUUCCUCUCGACA