European Journal of Human Genetics (2008) 16, 869–872
© 2008 Nature Publishing Group All rights reserved 1018-4813/08 $30.00
www.nature.com/ejhg

SHORT REPORT

Conservation of 5′-upstream region of the *FBN1* gene in primates

Krishna Kumar Singh[1,2], Praphulla Chandra Shukla[2] and Joerg Schmidtke*[,1]

[1]*Institute of Human Genetics, Hannover Medical School, Hannover, Germany;* [2]*Department of Cardiovascular Surgery, St Michael's Hospital, Toronto, Ontario, Canada*

Fibrillin-1 is a multifunctional extracellular protein encoded by the *FBN1* gene. *FBN1* is 237 kb in size and is located on chromosome 15q21. *FBN1* mutations are known to cause Marfan syndrome and other fibrillinopathies. *FBN1* is composed of 65 exons and 3 additional alternatively spliced exons at the 5′ end. The absence of the peptide sequence from the extreme N-terminus of the fibrillin-1 protein and the presence of in-frame and alternatively spliced exons at the 5′ end of the *FBN1* gene create some ambiguity about the translation start site and indicate a functional role of these alternatively spliced exons. We demonstrate here the conservation of 5′-upstream region of the *FBN1* gene among humans and non-human primates.
European Journal of Human Genetics (2008) **16,** 869–872; doi:10.1038/ejhg.2008.51; published online 19 March 2008

Keywords: *FBN1*; 5′ alternatively spliced exons; conservation; primates

Introduction

Fibrillin-1 is a 350-kDa extracellular protein, a principal constituent of 10 nm microfibrils,[1] and it has a variety of functions, including anchoring epithelial cells to the interstitial matrix and maintenance of elastic fibres.[2,3] Fibrillin-1 is encoded by the *fibrillin-1* (*FBN1*; MIM# 134797) gene, which is located on human chromosome 15q21, with an estimated size of 237 kb.[4,5] There are more than 600 mutations that have been identified so far (http://www.umd.be:2030/) in association with Marfan syndrome (MFS; MIM# 154700) and other so-called fibrillinopathies, including Weill–Marchesani syndrome (MIM# 608328), Marfanoid skeletal syndrome[6] and ectopia lentis (MIM# 129600).

At its 5′ end, the human *FBN1* gene carries three in-frame and alternatively spliced exons, called B, A and C, as well as a common exon M.[7] Exon M contains the putative initiating methionine, a Kozak consensus sequence (GGCATCATGCGG),[8] which is currently being referred as the beginning of exon 1 of the *FBN1* gene (http://www.umd.be:2030/). However, the presence of three alternatively spliced and in-frame exons and the absence of the peptide sequence from the extreme N-terminus of the protein generate some ambiguity regarding the site of translation initiation.[7] Furthermore, a functional role of these upstream exons/introns is indicated by the association of sequence variations therein with systemic sclerosis, scleroderma and MFS.[9–11]

We present here the data on the extent of sequence conservation of ~1.7 kb of the *FBN1* 5′-upstream region in humans and non-human primates, including rhesus monkey (*Macaca mulatta*), gorilla (*Gorilla gorilla*) and chimpanzee (*Pan troglodytes*). Our data demonstrate that there is a higher degree of conservation in the *FBN1* 5′-upstream exons/introns of humans and non-human primates.

Materials and methods

For analysing ~1.7 kb of the 5′-upstream region of the *FBN1* gene in humans and non-human primates

*Correspondence: Professor Dr J Schmidtke, Institute of Human Genetics, Hannover Medical School, Carl-Neuberg-Strasse 1, Hannover, Niedersachsen D-30625, Germany. Tel: +49 511 532 6537;
Fax: +49 511 532 5865; E-mail: schmidtke.joerg@mh-hannover.de
Received 5 November 2007; revised 28 December 2007; accepted 7 February 2008; published online 19 March 2008

(five rhesus monkeys, two chimpanzees and two gorillas), genomic DNA was extracted from venous blood, collected from humans (healthy blood donors), rhesus monkeys (Deutsches Primatenzentrum, Gottingen, Germany), chimpanzees (own DNA facility) and gorillas (own DNA facility), using standard protocols. The primers used in this study were taken from Singh *et al.*[11] Standard PCR conditions were as follows: initial denaturation at 95°C for 10 min followed by 33 cycles of 96°C for 1 min, 58–62°C for 1 min and 72°C for 1 min with final elongation at 72°C for 10 min in a 50-μl reaction mixture containing 1 × buffer (Qiagen, Germany), 1 × Q solution (Qiagen), 20 pM each primer and 2.5 U *Taq polymerase* (Qiagen). PCR products were purified with ExoSAP-IT (USB, USA), and both strands were sequenced with BigDye Terminator chemistry version 1.1 by standard protocol (ABI, USA). Sequencing reactions were carried out at 96°C for 10 s, 50°C for 5 s and 60°C for 4 min (25 cycles) (Biometra, Germany). The reaction mixtures were further purified using DyeEx™ 2.0 Spin kit (Qiagen) and analysed on the 3100-*Avant* Genetic Analyzer, according to the supplier's instructions, using the sequence analysis software (ABI).

A search for transcription factor binding site was performed for human 1.7 kb of the 5′-upstream region containing alternatively spliced exons/introns using TFSEARCH (http://www.cbrc.jp/research/db/TFSEARCH.html).

Results

Homology between humans and non-human primates

A high degree of homology was seen in the 5′-upstream region of the *FBN1* gene of humans and non-human primates (Figure 1 and Supplementary Figure 1). Striking features seen in the *FBN1* 5′-upstream exons/introns of humans and non-human primates were as follows: (1) presence of GC at splice donor site for exon B in rhesus monkeys, with all other known splice sites being conserved in humans and non-human primates (Figure 1 and Supplementary Figure 1); (2) insertion of a nucleotide C in exon A of non-human primates in comparison to humans (Supplementary Figure 1); (3) variations in the number of CGCCG repeats in intron A of humans ((CGCCG)2), chimpanzees ((CGCCG)2), gorillas ((CGCCG)5) and rhesus monkeys ((CGCCG)3) (Supplementary Figure 1); (4) an indel (indel A) polymorphism in exon C of rhesus monkeys and chimpanzees (Supplementary Figure 1); (5) conservation of Kozak consensus sequence in humans and non-human primates (Supplementary Figure 1); (6) two SNPs in the intron A +42C>G and intron C +58T>A of rhesus monkey; and (7) presence of intron C +48C in chimpanzees and gorillas and intron C +48T in the rhesus monkey sequence (Supplementary Figure 1).

Putative transcription factor binding sites identified within 5′-upstream region of human *FBN1* gene

The search for transcription factor binding site in the 5′-upstream region of the human *FBN1* gene revealed the presence of several putative transcription factor binding sites for MZF1, GATA-1/2/3, deltaE, Sp1, SRY, AP-4, c-Rel and Nkx-2.5 transcription factors (Table 1). The majority of the putative transcription factor binding sites identified were in intron regions of the alternatively spliced exons (Table 1).

Discussion

Direct sequencing revealed a high degree of conservation in the 5′-upstream exons/introns of the *FBN1* gene in humans and non-human primates, which supports the functional relevance of this gene region. However, the presence of GG at the splice donor site of exon C of porcine (Figure 1), an indel (indel A) polymorphism in exon C of rhesus monkeys and chimpanzees, the insertion of a nucleotide C in exon A of non-human primates and the presence of GC at the splice donor site for exon B in rhesus monkeys (Figure 1) create some ambiguity about

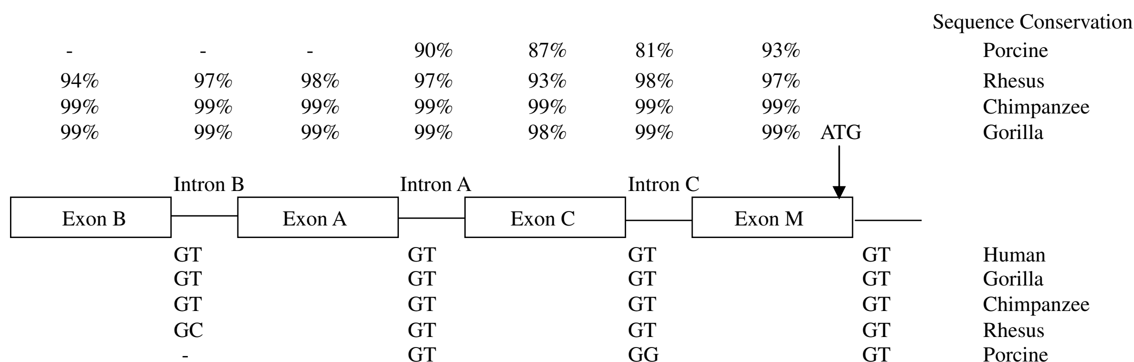

Figure 1 Schematic representation of human *FBN1* upstream exons/introns. Exons B, A and C are alternatively spliced exons.[5] The presumptive initiation codon is indicated by ATG in exon M. The sequence homology among humans,[7] non-human primates (gorillas, chimpanzees and rhesus monkeys) and porcine[7] is given above each exon/intron and splice donor sites are shown below.

Table 1 Putative transcription factor binding sites identified by using TFSEARCH for vertebrate transcription factors (threshold score 90) within alternatively spliced exons and associated introns of human *FBN1* gene

Location	*Sequence*	*Strand*	*Transcription factor*	*Entry*	*Score*
Exon B	AGCGGGGA	+	MZF1	M00083	94.8
Intron B	CGCGATAGGT	–	GATA-1	M00075	97.6
Intron B	CGCGATAGGT	–	GATA-2	M00076	92.9
Intron B	GGTGGGGA	–	MZF1	M00083	98.3
Intron B	GGAGGGGA	–	MZF1	M00083	93.0
Intron B	CCCCACCTCAA	+	deltaE	M00073	90.0
Intron B	GGTGATGAGG	+	GATA-1	M00075	91.8
Intron B	GGGGTGGGGT	+	Sp1	M00008	90.8
Intron B	AAAGAAA	–	SRY[a]	M00148	90.0
Exon A	GGTGGGGA	+	MZF1	M00083	98.3
Intron A	AGCGCCAGCTGTGGACGT	–	AP-4	M00005	91.7
Intron A	GGCGGGGA	+	MZF1	M00083	93.0
Exon C	CGCGATGCGC	–	GATA-1	M00075	91.8
Exon C	GGGGATATGG	+	GATA-2	M00076	92.9
Exon C	GCTGGGGA	+	MZF1	M00083	90.4
Exon C	GGGGATATGG	+	GATA-1	M00075	90.2
Intron C	TGGGCTTTCC	+	c-Rel	M00053	95.9
Intron C	AATGGGGA	+	MZF1	M00083	93.9
Intron C	GGCGGGGA	–	MZF1	M00083	93.0
Intron C	GAGGCGGGGA	–	Sp1	M00008	91.8
Intron C	CGGGATGCTG	–	GATA-1	M00075	91.4
Exon M	TGAAGTG	+	Nkx-2.5	M00240	90.7
Exon M	CGGGATAGCG	+	GATA-2	M00076	99.2
Exon M	CGGGATAGCG	+	GATA-1	M00075	95.5
Exon M	GGGATAGCG	+	GATA-3	M00077	90.0
Exon M	CATGATGCCG	–	GATA-1	M00075	90.6

[a]Three similar putative binding sites were identified in intron B for transcription factor SRY.

translation of these exons in these species. It is, however, noteworthy that all other confirmed human splice donor sites were conserved in non-human primates (Figure 1). Variations in the number of CGCCG repeats in intron A of humans and non-human primates show the polymorphic nature of these repeats (Supplementary Figure 1). Two SNPs seen in rhesus monkey were intron A +42C>G and intron C +58T>A; intron A +42C>G appears to be a result of evolutionary divergence, as G is present in humans and other non-human primates at the respective position, whereas the other SNP intron C +58T>A appears to be specific for rhesus monkey, and both alleles, were seen in the same frequency (allele frequency 0.50) (Supplementary Figure 1). Known human intronic variation, intron C +48 C>T, also appears to be a result of the evolutionary divergence, as intron C +48C was found in the wild-type sequence of the chimpanzees and gorillas and intron C +48T was found in the wild-type sequence of rhesus monkeys (Supplementary Figure 1). Our data on *P. troglodytes* were in full consensus with the available sequence from NCBI (Ptr15_WGA16806_2), except a variation where nucleotide A at position 3775016 was replaced by nucleotide G in our sequence.

The search for transcription factor-binding sites revealed the presence of several putative vertebrate transcription factor binding sites within the 5′-upstream region of the human *FBN1* gene. These findings further indicate that 5′-upstream alternatively spliced exons/introns may control the *FBN1* expression in a tissue-specific manner. This, however, remains speculative, as the data need to be confirmed functionally (Table 1).

In summary, the presence of 5′ alternatively spliced exons in humans, a higher degree of homology and conservation of the splicing sites across species borders indicate that this gene region may play a functional role. We were unable to define the 5′ end of the *FBN1* gene; therefore, the presence of more upstream and in-frame exons cannot be ruled out. We are planning to study the expression of different 5′ alternatively spliced exons in different tissues, investigate their role in transcription and translation, identify the 5′-upstream end and confirm the translation initiation site of the *FBN1* gene in humans.

References

1 Sakai LY, Keene DR, Engvall E: Fibrillin, a new 350-kD glycoprotein, is a component of extracellular microfibrils. *J Cell Biol* 1986; **103**: 2499–2509.
2 Pereira L, Andrikopoulos K, Tian J *et al*: Targeting of the gene encoding fibrillin-1 recapitulates the vascular aspect of Marfan syndrome. *Nat Genet* 1997; **17**: 218–222.
3 Mecham RP, Heuser JE: The elastic fiber; in Hay ED (ed): *Cell Biology of Extracellular Matrix*. New York: Plenum, 1991, pp 79–109.

4 Magenis RE, Maslen CL, Smith L, Allen L, Sakai LY: Localization of the fibrillin (FBN) gene to chromosome 15, band q21.1. *Genomics* 1991; **2**: 346–351.
5 Biery NJ, Eldadah ZA, Moore CS, Stetten G, Spencer F, Dietz HC: Revised genomic organization of FBN1 and significance for regulated gene expression. *Genomics* 1999; **56**: 70–77.
6 Milewicz DM, Grossfield J, Cao SN, Kielty C, Covitz W, Jewett T: A mutation in FBN1 disrupts profibrillin processing and results in isolated skeletal features of the Marfan syndrome. *J Clin Invest* 1995; **95**: 2373–2378.
7 Corson GM, Chalberg SC, Dietz HC, Charbonneau NL, Sakai LY: Fibrillin binds calcium and is coded by cDNAs that reveal a multidomain structure and alternatively spliced exons at the 5′ end. *Genomics* 1993; **17**: 476–484.
8 Kozak M: An analysis of vertebrate mRNA sequences: intimations of translational control. *J Cell Biol* 1991; **115**: 887–903. Review.
9 Tan FK, Wang N, Kuwana M *et al*: Association of fibrillin 1 single-nucleotide polymorphism haplotypes with systemic sclerosis in Choctaw and Japanese populations. *Arthritis Rheum* 2001; **44**: 893–901.
10 Kodera T, Tan FK, Sasaki T, Arnett FC, Bona CA: Association of 5′-untranslated region of the fibrillin-1 gene with Japanese scleroderma. *Gene* 2002; **297**: 61–67.
11 Singh KK, Shukla PC, Rommel K, Schmidtke J, Arslan-Kirchner M: Sequence variations in the 5′ upstream regions of the FBN1 gene associated with Marfan syndrome. *Eur J Hum Genet* 2006; **14**: 876–879.

Supplementary Information accompanies the paper on European Journal of Human Genetics website (http://www.nature.com/ejhg)