Mammalian Genome 9, 131–135 (1998).

Mammalian Genome
Incorporating Mouse Genome

Genomic structure and complete sequence of the human FGFR4 gene

Markus Kostrzewa, Ulrich Müller

Institut für Humangenetik, Justus-Liebig-Universität Giessen, Schlangenzahl 14, D35392 Giessen, Germany

Received: 18 August 1997 / Accepted: 30 September 1997

Abstract. We report the genomic structure and entire sequence of the fibroblast growth factor receptor 4 (FGFR4) gene. The gene spans approximately 11.3 kb. It is composed of 18 exons ranging in size from 71 bp to 600 bp. Exon-intron boundaries follow the GT/AG rule. Exon 1 is untranslated and preceded by structural elements characteristic of a TATA-free promoter. Although there are promoter motifs in intron 4 as well, there is currently no evidence of alternative transcription of FGFR4. Comparison of exon-intron boundaries of FGFR4 with those of FGFR1 and 3 reveals a remarkable degree of homology. With the exception of four, exon boundaries are at identical positions in all three receptor genes. Short tandem repeat polymorphisms (STRPs) were identified in introns 2 and 16 of FGFR4. The STRPs together with the sequence information will facilitate the rapid analysis of FGFR4 in those human disorders in which this gene can be considered a candidate.

Introduction

Fibroblast growth factor receptors (FGFRs) are members of the receptor tyrosine kinase family. To date, four distinct FGFRs are known (Johnson and Williams 1993; Mason 1994). They share highly homologous structural elements, including three extracellular immunoglobulin-like domains (IgI, IgII, and IgIII), a transmembrane, and an intracellular tyrosine kinase domain. The Ig domains are preceded by a leader sequence, and IgI and IgII are separated by the acid box, a stretch of four to eight acidic amino acids. The tyrosine kinase domain is split into two catalytic domains by the insertion of 14 amino acids. Homology is greatest between FGFR1 and FGFR2 (72% amino acid identity), slightly less between FGFR1 and FGFR3, and least pronounced between FGFR1 and FGFR4 (55% identity). There are several isoforms of FGFR1, 2, and 3 that are generated by alternative splicing of the transcript. No isoforms are currently known of FGFR4.

cDNAs of the four receptors have been identified and the genes mapped. FGFR1 is located on Chromosome (Chr) 8 (8p11.2-p11.1; Wood et al. 1995), FGFR2 on Chr 10 (10q26; Mattei et al. 1991; Dionne et al. 1992), FGFR3 on Chr 4 (4p16.3; Thompson et al. 1991), and FGFR4 on Chr 5 (5q35.1-qter; Warrington et al. 1992). The expression patterns of FGFR1, 2 and 3 are distinct but overlapping (Johnson and Williams 1993). Thus, FGFR1 is predominantly expressed in the brain and in mesenchymal tissue, FGFR2 in brain and epithelial tissue, and FGFR3 in brain, spinal cord, and developing bone. In contrast, FGFR4 is primarily expressed in the developing endoderm and in skeletal muscle. Mutations in FGFR1, 2, 3 have been found in various human disorders involving growth and differentiation of bone such as achondroplasia and craniosynostotic syndromes (Park et al. 1995; Muenke and Schell 1995; Webster and Donoghue 1997; Müller et al. 1997). Mutations in FGFR4 have not yet been observed in any human disease.

Of the four receptor genes, the genomic structure of FGFR3 has been characterized best (Perez-Castro et al. 1997). The gene spans 16.5 kb and consists of 19 exons and 18 introns. The nucleotide sequence of 66% (11 kb) of the gene has been published. No comprehensive structural analysis has been reported on the FGFR1, 2, and 4 genes, and sequence data and structural information are especially sparse for the FGFR4 gene.

Here we report the genomic structure of human FGFR4, give its entire sequence, and describe two short tandem repeat polymorphisms (STRPs) within the gene. The findings obtained were applied to a comparative structural analysis of the four FGF receptor genes.

Materials and methods

PACS. PACs were obtained from the UK HGMP Resource Centre (Hinxton, UK).

DNA extraction. DNA was extracted from human peripheral blood according to standard procedures. PAC DNA was isolated with the QIAfilter Plasmid Maxi Kit (Qiagen, Hilden, Germany) according to manufacturer's instructions. Plasmids were isolated by the alkaline lysis method (Sambrook et al. 1989). PCR products for sequencing were excised from agarose gels and purified with the Prep-a-Gene system (Bio-Rad, Hercules, USA).

PCR. A fragment of human FGFR4 was amplified with previously described primers (Warrington et al. 1992). An initial denaturation at 94°C for 3 min was followed by 30 cycles at 94°C for 30 s, annealing at 56°C for 1 min, extension at 72°C for 1 min, and a final extension of 5 min at 72°C. Conditions for all additional PCRs were basically identical with the exception of the annealing temperature that was chosen specifically for each primer pair. For long-PCR, the Taq Extender (Stratagene, La Jolla, California) was used in addition to Taq polymerase, and extension at 72°C was increased to 4 min. Hot PCRs were run in the presence of 1 μCi [α-^{32}P]dCTP.

Pulsed-field gel electrophoresis. For determination of the human insert size, 300 ng of PAC DNA was cleaved with *Not*I. Fragments were separated on a 1% agarose gel by a CHEF system (Bio-Rad). Running buffer was 0.5 x TBE, and the running time was 14 h at 14°C with a ramp of 0.3 to 3 s at 6V/cm.

Filter hybridization. PCR products were labeled with [α-32P]dCTP and the Random Primers Labelling System (Gibco BRL, Gaithersburg, MD). Oligonucleotides were labeled with [γ-^{32}P]-ATP and T4 polynucleotide kinase (USB/Amersham, Buckinghamshire, UK). Hybridizations were performed at 65°C in aqueous solution according to standard procedures. Gridded human PAC filters were obtained from the UK HGMP Resource Centre (Hinxton).

Library construction. 500 ng of PAC 32C5 DNA was digested with *Eco*RI and cloned in pBluescript SK+ (Stratagene). Recombinant clones

Correspondence to: U. Müller

The nucleotide sequence data reported in this paper have been submitted to EMBL and have been assigned the accession number Y13901.

were gridded and transferred to nylon membranes (Hybond N+, Amersham) for subsequent hybridization with FGFR4 probes.

Sequencing. Direct sequencing of PCR products and sequencing of inserts cloned in plasmids were performed with the Thermosequenase fluorescent labeled primer sequencing kit (Amersham) or the SequiTherm EXCEL Long-Read Sequencing Kit-LC (Epicentre Technologies, Madison, WI) and 5′IRD-labeled primers. Sequencing products were separated and analyzed on an automated sequencing machine (LI-COR 4000L).

Analysis of STRPs. The PCR products of D5S2922 were separated on 6% denaturing polyacrylamide gels. An M13 sequencing ladder was used as size standard. For size determination of the CCT/AGG-repeat in intron 16 of FGFR4, one of the two primers was 5′IRD labeled, and the PCR product was separated and analyzed with the LI-COR sequencer.

Structural analysis. Sequence alignments were performed with the computer program Clustal V (Higgins et al. 1992). Promoter analysis of the FGFR4 gene was performed online. The CpG score was calculated applying the computer program GRAIL at the Oak Ridge National Laboratory server (http://avalon.epm.ornl.gov/GRAIL-bin/GRAILFORM_post). Transcription factor binding sites and possible transcription start sites were detected with the Gene Finder program package at the Baylor College of Medicine server (http://dot.imgen.bcm.tmc.edu:9331/gene-finder/gf.html; programs tssw and tssg).

Results

Genomic organization of FGFR4. Screening of the PAC library with an FGFR4-specific, ^{32}P-labeled amplification product (Warrington et al. 1992) resulted in the isolation of two PACs (#32C5 and #251C21). PAC #32C5 with an insert of 90 kb was used for further investigations. Both PACs had been assigned to 5q35 by fluorescence in situ hybridization (own unpublished results). An *Eco*RI plasmid library constructed from this PAC was screened with two oligonucleotides (5′-GGCTGGAGCTGGGAGTGAG-3′ from the 5′ noncoding region and 5′-AGGTCGAGCACTGTGTCAGG-3′ from the 3′ noncoding region) derived from the known cDNA sequences of FGFR4 (Partanen et al. 1991; Ron et al. 1993). Two nonoverlapping plasmids with inserts of 8.1 kb and 4.5 kb were isolated (Fig. 1). Comparison of the genomic sequence of the inserts with the known cDNAs revealed that the two plasmids contained the entire transcribed sequence of FGFR4. The remaining gap between the two plasmids was, therefore, thought to be intronic. It was bridged by Long PCR with primer 5′-CCTCGCAGGCAATTCCATC-3′ from exon 8 and the primer described above from the 3′ noncoding region. The amplification product of 5.1 kb was 1.0 kb larger than expected if both plasmid clones had been immediately adjacent to each other. The sequence of the intronic fragment bridging both plasmids was also determined. Comparison of the genomic sequence of FGFR4 with the cDNA revealed that FGFR4 is composed of 18 exons and 17 introns. The start codon ATG is in exon 2 and the stop codon in exon 18 (Fig. 1). The boundaries between exons and introns follow the GT/AG rule. Table 1 gives the exact sizes of exons and introns and the sequences of the intron-exon boundaries.

Table 1. Genomic structure of the FGFR4 gene.

Exon	Size [bp]	5′ splice donor	Intron size [bp]	3′ splice acceptor
1	⩾103	GAGGAGCCAGgtgag	2522	tttccctccctattttagGAAGG
2	144	GTGGAGCTTGgtatg	699	ctctccctctgcccacagAGCCC
3	264	ATTACAGGTGgtaag	90	cctctgtccctgatgtagACTCC
4	81	CCCCAGCAAGgtcag	111	ccttcccctgccctccagCACCC
5	167	AGGCATTCGGgtgag	580	cactctctctgcctgcagCTGCG
6	124	GATGTGCTGGgtgag	512	gtctcgcccggtccccagAGCGG
7	191	AGTCCTAAAGgtaaa	134	gcatgtcccccaccccagACTGC
8	139	GTGCTGCCAGgtgag	353	gtccatgtgcgagggcagAGGAG
9	194	GGCCCGACAGgtact	74	aagtctcccactttgcagTTCTC
10	146	CCCGGGACAGgtgcg	102	gactttctccatctccagGCTGG
11	122	ATGCTCAAAGgtgag	1553	cgtctgctgcccttacagACAAC
12	111	ACCCAGGAAGgtggg	92	cctccactccctctgcagGGCCC
13	191	GTCCCGGAAGgtata	333	agcccccgctccctgcagTGTAT
14	123	AACCAGCAACgtgag	107	gccctctctcccctccagGGCCG
15	71	AGAGTGACGTgtgag	246	ccctggccctgcctccagGTGGT
16	138	CCCCAGAGCTgtgag	673	ccgccccacctctcgcagGTACG
17	106	CTCTGAGGAGgtaca	129	cagctccgttccccacagTACCT
18	600			

Sequence analysis. The region surrounding exon 1 (position 997–1535) was found to be CG-rich with a CpG score of 0.758. Analysis of the region upstream of exon 5 with programs tssw and tssg indicated promoters with transcription start sites at nucleotide positions 1162 and 1097, respectively. No TATA box was found with either program, but numerous putative transcription factor binding sites were detected. These sites include several consensus binding site sequences of SP1. The findings are consistent with the presence of a typical TATA-less promoter (Pugh and Tjian 1991) 5′ to exon 1 of FGFR4. This corresponds well to the findings reported for FGFR3 by Perez-Castro et al. (1997).

The human FGFR4 cDNA reported by Partanen and associates (1991) starts within exon 2 of the gene. According to our findings this exon lacks the first 14 bp of the sequence published by Partanen et al. (1991). Homology starts only at nucleotide 13 of exon 2. In that respect, the genomic sequence is similar to that reported by Ron et al. (1993), who also did not find the first 14 bp of DNA described by Partanen and colleagues. The longer cDNA published by Ron and coworkers also includes exon 1 plus 13 additional bases that we did not detect. Hybridization of PACs 32C5 and 251C21 to the 13 bp oligonucleotide observed in the cDNA by Ron and associates (1993) did not result in a hybridization signal. This indicates that the most 5′ 13 bp described by Ron et al. are most likely an artefact and not part of FGFR4.

Analysis of the intronic sequences revealed ALU sequences in introns 1 and 11 of FGFR4. In addition, program tssw predicted a possible promoter within intron 4. This putative promoter has a TATA box at nucleotide position 5412–5417 and a possible transcription start site at position 5453.

Short tandem repeat polymorphisms. An increase in CA/GT repeats was observed in intron 2 and an interrupted trinucleotide (CCT/AGG) repeat was found in intron 16. Since short stretches of di-, and trinucleotides are frequently polymorphic, we tested the intron 2 locus for polymorphism in 49 unrelated controls. Primers

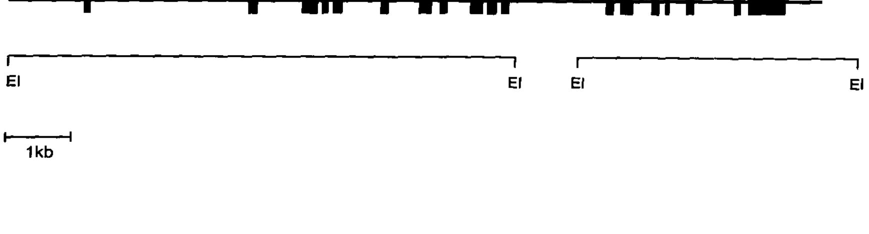

Fig. 1. Genomic structure of FGFR4. The plasmids used for sequencing are shown. The region between the two plasmids was bridged by long PCR and shown to contain much of intron 11 by DNA sequencing.

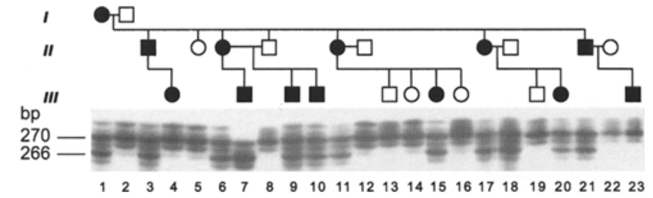

Fig. 2. Segregation of D5S2922 in a large three generation family. Affected members have craniosynostosis, Boston type (Warman et al. 1993). Several recombinations are found between D5S2922 and the disease locus.

used for amplification were 5′-GTGGTGAGCTCCTTGCCTG-3′ and 5′-TGGGGCATGTCACACACAC-3′. Three alleles of 266, 270, and 272bp were observed and a heterozygosity of 0.41; and a PIC value of 0.34 were determined. Mendelian inheritance was shown in a large three-generation family with craniosynostosis, Boston type (Müller et al. 1993). Several recombinations were found between the disease locus, *MSX2* (Jabs et al. 1993) and the polymorphic marker (D5S2922, Fig. 2). Primers 5′-CCTCGACCCCACTTTCCAG-3′ and 5′-GTGGGGAGTTTGATGAGGGA-3′ were applied to the amplification of the repeat in intron 16. Polymorphic products were obtained, but they were markedly smaller than expected on the basis of the DNA sequence. For example, the amplification product of the intron 16 locus of PAC 32C5 was 383 bp but should have been 566 bp based on the sequence. The 383 bp fragment was shown to contain intron 16-specific sequences. Currently we cannot explain this discrepancy. Although Mendelian inheritance of this polymorphic marker was documented (not shown) we had no D5S number assigned to this locus. It appears, however, that this polymorphism is identical to the *Eco*RI polymorphism described by Armstrong et al. (1991).

Comparison with other FGFR genes. FGFR4 is composed of 18 exons, one less than the homologous gene FGFR3 (Perez-Castro et al. 1997) and most likely FGFR 1 and 2 as well (Johnson and Williams 1993), provided the latter are also preceded by an untranslated exon. The extra exons of FGFRs 1, 2, and 3 are utilized for the generation of alternative splice products of IgIII (Johnson et al. 1991; Avivi et al. 1992; Miki et al. 1992). Comparison of FGFR4 with FGFR3 showed the highest degree of homology for amino acids encoded by exons 6, 7 of both receptors and by exons 8/9, 11/12, 12/13, 14/15, 15/16, and 16/17 of FGFR4 and FGFR3 (Fig. 3). Exons 6, 7, and 8/9 code for parts of IgIII and the domain connecting IgII and IgIII, and exons 11–16 (12–17 of FGFR3) encode the tyrosine kinase domain. The lowest degree of homology is found for the transmembrane domain (exon 9 and 10, respectively) and the extracellular structures of the receptor (exons 2–4). Exon by exon comparison was also possible between most of FGFR1 and FGFR4, and highest percentages of homology were found at the same positions as observed between FGFR4 and FGFR3. No comprehensive comparison was possible with FGFR2 since exact information on exon-intron boundaries is available for exons 8 and 9 of FGFR2 only.

Sequences of FGFR1, FGFR3, and FGFR4 are aligned in Fig. 4 to demonstrate the location of the coding exons with respect to the amino acid sequence of three receptors. The intron-exon boundaries are at identical positions for most exons and are shifted by several amino acids at four exon boundaries only.

Discussion

We have sequenced the entire FGFR4 gene and determined its structural organization. Homology is pronounced between FGFR4 and FGFRs 3 and 1 and the known portions of FGFR2 even at the genomic level. Exon-intron boundaries follows the GT/AG rule in FGFR3 and FGFR4 and probably in FGFR1 as well. In the latter sequence, however, one intron (#10) has been described (Johnson et al. 1991) that lacked the GT consensus sequence. For the remaining exon/intron boundaries of FGFR1, the GT/AG rule applies as well.

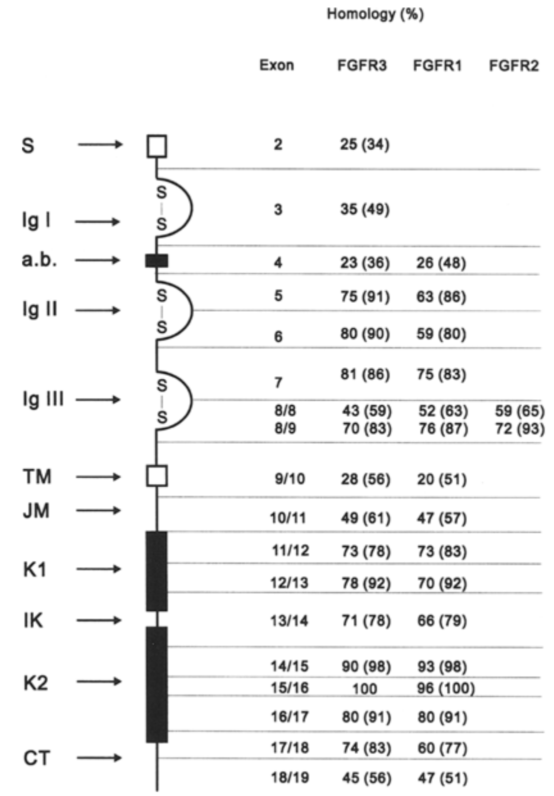

Fig. 3. Exon by exon comparison between FGFR4 and known portions of FGFR1, 2, and 3 at the amino acid level. Percentage of identity is given. The percentage of homology including conserved amino acid exchanges is shown in brackets. S = signal peptide; a.b. = acid box; IgI–IgIII = immunoglobulin-like loops I–III; TM = transmembrane domain; JM = juxtamembrane domain; K1 and K2 = catalytic domains of tyrosine kinase; IK = interkinase domain; CT = C-terminal tail.

There is one striking difference between FGFRs 1, 2, and 3 and FGFR4. While the former genes appear to be composed of 19 exons, FGFR4 has 18 exons only. The extra exon is located between exons 8 and 9 (relative to the FGFR4 sequence) and is utilized as part of an alternatively spliced transcript that codes for isoform IgIIIb of FGFR1, 2, and 3 (Johnson and Williams 1993; Avivi et al. 1993). In contrast to FGFR1, 2, and 3, there is no evidence of alternative splicing of the FGFR4 transcript. For example, only one transcript of 3 kb has been detected on Northern blots containing RNA from various tissues (Partanen et al. 1991; Ron et al. 1993). There is no evidence of a FGFR1, 2, 3 exon 9 equivalent within intron 8 of the FGFR4 gene. Furthermore, there is no evidence of a soluble IIIa-only form of FGFR4, as has been described for FGFR1 (Johnson and Williams 1993). In contrast to FGFR1, there is no stop codon in intron 7 and no polyadenylation site in FGFR4 that could be used for alternative processing of the transcript and the generation of a IgIIIa form (Vainikka et al. 1992). We also failed to detect such sites in intron 7 of the gene but found a few sequence discrepancies, including five base changes and absence of a GC and a C in the sequence of Vainikka and associates (1992).

In addition to intron 7, introns 8 and 9 of FGFR4 have also been sequenced before (Vainikka et al. 1992). Our sequence deviates from the intron 8 sequence reported by four insertions of one or two nucleotides and by insertions of one and eight nucleotides in the sequence of Vainikka and coworkers (1992). In addition,

```
                                      Ex2><Ex3
FGFR4    ---MRLLLALLGILLSVPGPPVLSLEASEEVELEPCLAPSLE--QQEQELTVALGQPVRLCCGRAERG----GHWYKEGSRLAPAGRVRGWRGRL
                                         Ex2><Ex3
FGFR3    MGAPACALALCVAVAIVAGASSESLGTEQRVVGRAAEVPGPEPGQQEQ-LVFGSGDAVELSCPPPGGGPMGPTVWVKDGTGLVPSERVLVGPQRL

FGFR1    MWSWKCLLFWAVLVTATLCTARPSPTLPEQAQPWGAPVEVES------FLVHPGDLLQLRCRLRDDVQS--INWLRDGVQLAESNRTRITGEEV
                *     .         *    .      .                   *  . * *           * ..*  *  . *       .
                                      Ex3><Ex4                     Ex4><Ex5
FGFR4    EIASFLPEDAGRYLCLAR-GSMIVLQNLTLITGDSLTSSN-DDEDPKSHRDLSNRHSYPQQ----APYWTHPQRMEKKLHAVPAGNTVKFRCPAA
                                      Ex3><Ex4                    Ex4><Ex5
FGFR3    QVLNASHEDSGAYSCRQR-LTQRVLCHFSVRVTDAPSSG--DDEDGE---DEAEDTGVDTG----APYWTRPERMDKKLLAVPAANTVRFRCPAA
                                      Ex3><Ex4                      Ex4><Ex5
FGFR1    EVQDSVPADSGLYACVTSSPSGSDTTYFSVNVSDALPSSEDDDDDDDSSSEEKETDNTKPNRMPVAPYWTSPEKMEKKLHAVPAAKTVKFKCPSS
         ..      *.* * *     .       ..   *.  *   **.*     .              ***** *..*.*** ****  **.*.**..
                               Ex5><Ex6                                Ex6><Ex7
FGFR4    GNPTPTIRWLKDGQAFHGENRIGGIRLRHQHWSLVMESVVPSDRGTYTCLVENAVGSIRYNYLLDVLERSPHRPILQAGLPANTTAVVGSDVELL
                               Ex5><Ex6                                Ex6><Ex7
FGFR3    GNPTPSISWLKNGREFRGEHRIGGIKLRHQQWSLVMESVVPSDRGNYTCVVENKFGSIRQTYTLDVLERSPHRPILQAGLPANQTAVLGSDVEFH
                               Ex5><Ex6                                Ex6><Ex7
FGFR1    GTPNPTLRWLKNGKEFKPDHRIGGYKVRYATWSIIMDSVVPSDKGNYTCIVENEYGSINHTYQLDVVERSPHRPILQAGLPANKTVALGSNVEFM
         *.*.*.. *** *. *. ..**** ..*   **..*.******.*.***.***  ***  .* ***.**************** *  .** **
                                         Ex7><Ex8                                        Ex8><Ex9
FGFR4    CKVYSDAQPHIQWLKHIVINGSSFGADGFPYVQVLKTADINSS--EVEVLYLRNVSAEDAGEYTCLAGNSIGLSYQSAWLTVLPEEDPTWTAAAP
                                         Ex7><Ex9                                        Ex9><Ex10
FGFR3    CKVYSDAQPHIQWLKHVEVNGSKVGPDGTPYVTVLKTAGANTTDKELEVLSLHNVTFEDAGEYTCLAGNSIGFSHHSAWLVVLPAEEELVEADEA
                                         Ex7><Ex9                                        Ex9><Ex10
FGFR1    CKVYSDPQPHIQWLKHIEVNGSKIGPDNLPYVQILKTAGVNTTDKEMEVLHLRNVSFEDAGEYTCLAGNSIGLSHHSAWLTVLEALE-ERPAVMT
         ****** *********. .***  * *. *** .****  *..  *.*** *.**. *************** * .**** **   .    *
                                                              Ex9><Ex10
FGFR4    EARYTDIILYASGSLALAVLLLLAGLYRGQALHGRH--PRPPATVQKLSR-FPLARQFSLESGSSGKSSS--SLVRGVRLSSSGPALLAGLVSLD
                                                             Ex10><Ex11
FGFR3    GSVYAGILSYGVGFFLFILVVAAVTLCRLRSPPKKG---LGSPTVHKISR-FPLKRQVSLESNASMSSNT--PLVRIARLSSGEGPTLANVSELE
                                                                Ex10><Ex11
FGFR1    SPLYLEIIIYCTGAFLISCMVGSVIVYKMKSGTKKSDF-HSQMAVHKLAKSIPLRRQVTVSADSSASMNSGVLLVRPSRLSSSGTPMLAGVSEYE
            *  *. *  *       ..   . . ..   .        .*.*...  ** ** .. . .*    .   ***  ****     **..   .
                 Ex10><Ex11                              Ex11><Ex12                           Ex12><Ex13
FGFR4    LPLDPLWEFPRDRLVLGKPLGEGCFGQVVRAEAFGMDPARPDQASTVAVKMLKDNASDKDLADLVSEMEVMKLIGRHKNIINLLGVCTQEGPLYV
                 Ex11><Ex12                              Ex12><Ex13                           Ex13><Ex14
FGFR3    LPADPKWELSRARLTLGKPLGEGCFGQVVMAEAIGIDKDRAAKPVTVAVKMLKDDATDKDLSDLVSEMEMMKMIGKHKNIINLLGACTQGGPLYV
                 Ex11><Ex12                              Ex12><Ex13                           Ex13><Ex14
FGFR1    LPEDPRWELPRDRLVLGKPLGEGCFGQVVLAEAIGLDKDKPNRVTKVAVKMLKSDATEKDLSDLISEMEMMKMIGKHKNIINLLGACTQDGPLYV
         ** ** **  * ** ************** *** *.*  .  .   *******  *..***.**.****.**.**.********* *** *****
                                                               Ex13><Ex14
FGFR4    IVECAAKGNLREFLRARRPPGPDLSPDGPRSSEGPLSFPVLVSCAYQVARGMQYLESRKCIHRDLAARNVLVTEDNVMKIADFGLARGVHHIDYY
                                                               Ex14><Ex15
FGFR3    LVEYAAKGNLREFLRARRPPGLDYSFDTCKPPEEQLTFKDLVSCAYQVARGMEYLASQKCIHRDLAARNVLVTEDNVMKIADFGLARDVHNLDYY
                                                               Ex14><Ex15
FGFR1    IVEYASKGNLREYLQARRPPGLEYCYNPSHNPEEQLSSKDLVSCAYQVARGMEYLASKKCIHRDLAARNVLVTEDNVMKIADFGLARDIHHIDYY
         .** *.******.*.****** . .    .  *  *.   ************.** *.***************************** .*..***
         Ex14><Ex15              Ex15><Ex16                                    Ex16><Ex17
FGFR4    KKTSNGRLPVKWMAPEALFDRVYTHQSDVWSFGILLWEIFTLGGSPYPGIPVEELFSLLREGHRMDRPPHCPPELYGLMRECWHAAPSQRPTFKQ
         Ex15><Ex16              Ex16><Ex17                                    Ex17><Ex18
FGFR3    KKTTNGRLPVKWMAPEALFDRVYTHQSDVWSFGVLLWEIFTLGGSPYPGIPVEELFKLLKEGHRMDKPANCTHDLYMIMRECWHAAPSQRPTFKQ
         Ex15><Ex16              Ex16><Ex17                                    Ex17><Ex18
FGFR1    KKTTNGRLPVKWMAPEALFDRIYTHQSDVWSFGVLLWEIFTLGGSPYPGVPVEELFKLLKEGHRMDKPSNCTNELYMMMRDCWHAVPSQRPTFKQ
         ***.*****************.***********.***************.****** **.******.* .*  .** .**.**** *********
                    Ex17><Ex18
FGFR4    LVEALDKVLLAVS-EEYLDLRLTFGPYSPSGGDAS-STCSSSD-SVFSHDPLPLGSSSFPFGSGVQT------
                  Ex18><Ex19
FGFR3    LVEDLDRVLTVTSTDEYLDLSAPFEQYSPGGQDTP-SSSSSGDDSVFAHDLLPPAP---PSSGGSRT------
                  Ex18><Ex19
FGFR1    LVEDLDRIVALTSNQEYLDLSMPLDQYSPSFPDTRSSTCSSGEDSVFSHEPLPEEP-CLPRHPAQLANGGLKRR
         ***    ** ...    *   *****        ***   *.   *..**  . ***.*.  **         *
```

Fig. 4. Alignment of amino acid sequences of FGFR4, FGFR3, and FGFR1. Location of known exon boundaries is given. * marks identical amino acid residues in all three sequences; . indicates conserved amino acid exchanges.

three bases differed between the two sequences. The sequence of intron 9 was identical with the exception of one base change. Comparison of the genomic sequence established here to known cDNA sequences (Partanen et al. 1991; Ron et al. 1993) revealed two amino acid changes. At amino acid position 10 we detected an isoleucine instead of a valine, and at position 136 we found a leucine instead of a proline. There is also a discrepancy between the cDNA sequences reported. While Partanen et al. found a valine at position 297, Ron et al. (1993) detected aspartic acid as did we.

A FGFR4 gene has also been described in other vertebrates including rat (Horlick et al. 1992), mouse (Stark et al. 1991), chicken (Marcelle et al. 1994), pleurodeles (Shi et al. 1992), and danio (Thisse et al. 1995). These studies have demonstrated that FGFR4 has been less conserved during evolution than the other three receptors. Thus, in danio an extra Ig loop was reported, and in the rat no signal peptide, IgI loop nor acid box, was found. Interestingly, a methionine is also found in exon 5 of the human sequence that is homologous to the start codon (ATG) of the rat. Although we did find a possible promoter in intron 4, it remains unclear whether this promoter is used for alternative processing of the FGFR4 transcript. In any case, such transcript would occur in low quantities only, since it could not be detected on Northern blots (Partanen et al. 1991; Ron et al. 1993).

While mutations in FGFR1, 2, and 3 were shown to cause various human diseases of bone differentiation, no disease has yet been associated with mutations in FGFR4. Given that FGFR4 is expressed in skeletal muscle and may play an important role in the maintenance of the undifferentiated state of myoblasts (Shaoul et al. 1995), FGFR4 mutations might be detected in diseases of muscle. Both the sequence and STRP information provided here will facilitate rapid mutation screening in disorders in which FGFR4 is considered a candidate gene.

Acknowledgment. This work was supported by the European Union (Contract GENE-CT93-0050 (DG 12 SSMA).

References

Armstrong E, Hästbacka J, Partanen J, Huebner K, Alitalo K (1991) RFLPs in the fibroblast growth factor receptor-4 locus (FGFR4) in 5q33-qter. Nucleic Acids Res 19, 5096

Avivi A, Yayon A, Givol D (1993) A novel form of FGF receptor-3 using an alternative exon in the immunoglobulin domain III. FEBS Lett 330, 249–252

Dionne CA, Modi WS, Crumley G, O'Brien SJ, Schlessinger J, Jaye M (1992) BEK, a receptor for multiple members of the fibroblast growth factor (FGF) family, maps to human chromosome 10q25.3–q26. Cytogenet Cell Genet 60, 34–36

Higgins DG, Bleasby AJ, Fuchs R (1992) Clustal V: improved software for multiple sequence alignment. CABIOS 8, 189–191

Horlick RA, Stack SL, Cooke GM (1992) Cloning, expression and tissue distribution of the gene encoding rat fibroblast growth factor receptor subtype 4. Gene 120, 291–295

Jabs EW, Müller U, Li X, Ma L, Luo W, Haworth I, Klisak I, Sparkes R, Warman ML, Mulliken JB, Snead M, Maxson R (1993) A mutation in the homeodomain of the human MSX2 gene in a family affected with autosomal dominant craniosynostosis. Cell 75, 443–450

Johnson DE, Williams LT (1993) Structural and functional diversity in the FGF receptor multigene family. Adv Cancer Res 60, 1–41

Johnson DE, Lu J, Chen H, Werner S, Williams LT (1991) The human fibroblast growth factor receptor genes: a common structural arrangement underlies the mechanism for generating receptor forms that differ in their third immunoglobulin domain. Mol Cell Biol 11, 4627–4634

Marcelle C, Eichmann A, Halevy O, Bréant C, Le Douarin NM (1994) Distinct developmental expression of a new avian fibroblast growth factor receptor. Development 120, 683–694

Mason IJ (1994) The ins and outs of fibroblast growth factors. Cell 78, 547–552

Mattei M-G, Moreau A, Gesnel M-C, Houssaint E, Breathnach R (1991) Assignment by in situ hybridization of a fibroblast growth factor receptor gene to human chromosome band 10q26. Hum Genet 87, 84–86

Miki T, Bottaro DP, Fleming TP, Smith CL, Burgess WH, Chan AM-L, Aaronson SA (1992) Determination of ligand-binding specificity by alternative splicing: Two distinct growth factor receptors encoded by a single gene. Proc Natl Acad Sci USA 89, 246–250

Müller U, Warman ML, Mulliken JB, Weber JL (1993) Assignment of a gene locus involved in craniosynostosis to chromosome 5qter. Hum Mol Genet 2, 119–122

Müller U, Steinberger D, Kunze S (1997) Molecular genetics of craniosynostotic syndromes. Graefe's Arch Clin Exp Ophthalmol 235, 545–550

Muenke M, Schell U (1995) Fibroblast-growth-factor receptor mutations in human skeletal disorders. Trends Genet 11, 308–313

Park W-J, Bellus GA, Jabs EW (1995) Mutations in fibroblast growth factor receptors: phenotypic consequences during eukaryotic development. Am J Hum Genet 57, 748–754

Partanen J, Mäkelä TP, Eerola E, Korhonen J, Hirvonen H, Claesson-Welsh L, Alitalo K (1991) FGFR-4, a novel acidic fibroblast growth factor receptor with a distinct expression pattern. EMBO J 10, 1347–1354

Perez-Castro AV, Wilson J, Altherr MR (1997) Genomic organization of the human fibroblast growth factor receptor 3 (*FGFR3*) gene and comparative sequence analysis with the mouse *Fgfr3* gene. Genomics 41, 10–16

Pugh BF, Tjian R (1991) Transcription from a TATA-less promoter requires a multisubunit TFIID complex. Genes Dev 5, 1935–1945

Ron D, Reich R, Chedid M, Lengel C, Cohen OE, Chan AM-L, Neufeld G, Miki T, Tronick SR (1993) Fibroblast growth factor receptor 4 is a high affinity receptor for both acidic and basic fibroblast growth factor but not for keratinocyte growth factor. J Biol Chem 268, 5388–5394

Sambrook J, Fritsch E, Maniatis T (1989) *Molecular Cloning: A Laboratory Manual,* (2nd ed.) (Cold Spring Harbor, N.Y.: Cold Spring Harbor Laboratory Press)

Shaoul E, Reich-Slotky R, Berman B, Ron D (1995) Fibroblast growth factor receptors display both common and distinct signaling pathways. Oncogene 10, 1553–1561

Shi D-L, Feige J-J, Riou J-F, DeSimone DW, Boucaut J-C (1992) Differential expression and regulation of two distinct fibroblast growth factor receptors during early development of the urodele amphibian Pleurodeles waltl. Development 116, 261–273

Stark KL, McMahon JA, McMahon AP (1991) FGFR-4, a new member of the fibroblast growth factor receptor family, expressed in the definitive endoderm and skeletal muscle lineages of the mouse. Development 113, 641–651

Thisse B, Thisse C, Weston JA (1995) Novel FGF receptor (Z-FGFR4) is dynamically expressed in mesoderm and neurectoderm during early zebrafish embryogenesis. Dev Dyn 203, 377–391

Thompson LM, Plummer S, Schalling M, Altherr MR, Gusella JF, Housman DE, Wasmuth JJ (1991) A gene encoding a fibroblast growth factor receptor isolated from the Huntington disease gene region of human chromosome 4. Genomics 11, 1133–1142

Vainikka S, Partanen J, Bellosta P, Coulier F, Basilico C, Jaye M, Alitalo K (1992) Fibroblast growth factor receptor-4 shows novel features in genomic structure, ligand binding and signal transduction. EMBO J 11, 4273–4280

Warman ML, Mulliken JB, Hayward P, Müller U (1993) Newly recognized autosomal dominant disorder with craniosynostosis. Am J Med Genet 46, 444–449

Warrington JA, Bailey SK, Armstrong E, Aprelikova O, Alitalo K, Dolganov GM, Wilcox AS, Sikela JM, Wolfe SF, Lovett M, Wasmuth JJ (1992) A radiation hybrid map of 18 growth factor, growth factor receptor, hormone receptor, or neurotransmitter receptor genes on the distal region of the long arm of chromsome 5. Genomics 13, 803–808

Webster MK, Donoghue DJ (1997) FGFR activation in skeletal disorders: too much of a good thing. Trends Genet 13, 178–182

Wood S, Schertzer M, Yaremko ML (1995) Sequence identity locates CEBPD and FGFR1 to mapped human loci within proximal 8p. Cytogenet Cell Genet 70, 188–191