Facioscapulohumeral muscular dystrophy (FSHD) is the third most common inherited myopathy with an incidence of 1:20 000. Clinically, the disease is foremost characterised by a progressive weakness and atrophy of the facial, shoulder, and upper arm muscles, but extra-muscular manifestations are also part of the clinical spectrum. The disease displays an autosomal dominant mode of inheritance and its major locus (FSHD1) was mapped to the subtelomere of chromosome 4q35. This subtelomere is mainly composed of a chromosomally dispersed gene family, mapping only 37 kb proximal to the D4Z4 repeat array.

Over 95% of FSHD patients carry a deletion on chromosome 4q35. This subtelomere is mainly composed of a tandem repeat array consisting of 3.3 kb units (D4Z4). In the population, the D4Z4 repeat array is polymorphic and may contain between 11 and 150 units. Due to the deletion of an integral number of these repeated D4Z4 elements, the majority of patients with FSHD carry only 1 to 11 units. In general, an inverse relationship has been established between the residual repeat size and the severity and age at onset of the disease.

It has been proposed that FSHD may be explained by a position effect variegation model, or by a differential long distance cis loop model. In those models the contraction of the D4Z4 repeat leads to the transcriptional upregulation of 4qter genes by a spreading or looping mechanism originating from D4Z4. In another model, Winuk et al proposed that alterations in nuclear positioning of chromosomes may lead to general gene deregulation, resulting in a defect in myogenic differentiation in FSHD.

FRG2, an FSHD candidate gene, is transcriptionally upregulated in differentiating primary myoblast cultures of FSHD patients

T Rijkers*, G Deidda*, S van Koningsbruggen, M van Geel, R J L F Lemmers, J C T van Deutekom, D Figlewicz, J E Hewitt, G W Padberg, R R Frants, S M van der Maarel

**Background:** Autosomal dominant facioscapulohumeral muscular dystrophy (FSHD) is associated with partial deletion of the subtelomeric D4Z4 repeat array on chromosome 4qter. This chromosomal rearrangement may result in regional chromatin relaxation and transcriptional deregulation of genes nearby.

**Methods and results:** Here we describe the isolation and characterisation of FRG2, a member of a chromosomally dispersed gene family, mapping only 37 kb proximal to the D4Z4 repeat array. Homology and motif searches yielded no clues to the function of the predicted protein. FRG2 expression is undetectable in all tissues tested except for differentiating myoblasts of FSHD patients, which display low, yet distinct levels of FRG2 expression, partly from chromosome 4 but predominantly originating from its homologue on chromosome 10. However, in non-FSHD myopathy patients only distantly related FRG2 homologues are transcribed, while differentiating myoblasts from healthy controls fail to express any member of this gene family. Moreover, fibroblasts of FSHD patients and control individuals undergoing forced Ad5-MyoD mediated myogenesis show expression of FRG2 mainly originating from chromosome 10. Luciferase reporter assays show that the FRG2 promoter region can direct high levels of expression but is inhibited by increasing numbers of D4Z4 repeat units. Transient transfection experiments with FRG2 fusion-protein constructs reveal nuclear localisation and apparently FRG2 overexpression causes a wide range of morphological changes.

**Conclusion:** The localisation of FRG2 genes close to the D4Z4 repeats on chromosome 4 and 10, their transcriptional upregulation specifically in FSHD myoblast cultures, potential involvement in myogenesis, and promoter properties qualify FRG2 as an attractive candidate for FSHD pathogenesis.
chromosome that may vary in length between 1 and 150 units without pathological consequences renders a direct involvement of DUX4 questionable. In addition, in approximately 20% of the Dutch population, chromosome 4-type repeat arrays have been identified on chromosome 10, or vice versa. In contrast, short (partial) D4Z4 repeat arrays on chromosome 10 have never been reported in FSHD. 24–27

At 80 kb proximal to the D4Z4 repeat array, a member of the β-tubulin subfamily (TUBB4Q) has been identified. The high allelic sequence variability of TUBB4Q and the lack of expression of this gene suggest it may be a pseudogene. 28 In the past, another candidate gene for FSHD was identified at 120 kb proximal to D4Z4: FRG1 (FSHD Region Gene 1). This gene is ubiquitously transcribed and encodes a protein that seems to be involved in RNA processing. 29 It is highly conserved between vertebrates and non-vertebrates. 30 However, analysis of FRG1 expression levels in muscle biopsies of FSHD patients yields controversial results (also own observations).

Here, we characterise a novel gene, FSHD Region Gene 2 (FRG2), which maps only 37 kb proximal to the D4Z4 repeat array. In contrast to data presented by Gabellini et al., FRG2 expression was very low to undetectable both in control and FSHD muscle using quantitative real time RT-PCR. However, we demonstrate that FRG2 is upregulated in differentiating myoblast cultures of FSHD patients compared to healthy controls. Its expression in MyoD-driven myogenesis in fibroblasts indicates that FRG2 may play a role in the myogenic process. In addition, transient transfection experiments suggest that increasing numbers of D4Z4 units inhibit expression of FRG2. These results render FRG2 an attractive candidate gene contributing to FSHD pathogenesis.

METHODS

Gene identification, characterisation, and genomic distribution

The sequence of PAC clone 226K22 (GenBank accession no. AF146191) was used for the in silico gene prediction by the Genotator browser. 31 A region of 3 kb just distal of D4S2463 containing four predicted exons preceded by a putative muscle specific promoter was termed FRG2 (FSHD Region Gene 2). The structure of FRG2 was initially characterised by RT-PCR on DNA of the monochromosomal rodent somatic cell hybrid GM11687 containing human chromosome 4 as its only human component (Coriell Cell Repositories, Camden, NJ, USA). GM11687 constitutively expresses the human component (gift from S Winokur, Irvine, CA), HWH416 (gift from E Stanbridge, Irvine, CA), SU10 (gift from S Winokur, Irvine, CA), and PAC clone RCPI-1-226K22 (all chromosome 4), cell lines 726-8a (UK HGMP Resource Center) and GM11688 (gift from E Stanbridge, Irvine, CA) and PAC clone RCPI-1-170K23, and cosmid 23D11 (all chromosome 10) was amplified using the primers 4f3 and 4r2, and the PCR product was sequenced and compared.

Myoblast culture

Myoblasts were isolated from two healthy males and one healthy female (myo 1, myo 5, and myo 7, respectively), three confirmed male FSHD patients (myo 2 carrying an allele of eight D4Z4 units, and myo 4 and myo 8 both carrying an allele of three D4Z4 units), a proximal myotonic myopathy (PRM) patient, and a desmin storage myopathy (DSM) patient. The site of biopsy was quadriceps, except for myo 4, which was derived from the biceps. Myoblasts were isolated essentially as described 32 and cultured in SKGM medium (Cambrex Bio Science, Walkersville, MD, USA) supplemented with 20% FBS according to the manufacturer’s instructions. Myoblasts of p. 6–10 were grown to 70% confluency and differentiation was initiated by changing to DMEM, high glucose, supplemented with 2% horse serum (HS; Gibco In Vitrogen, Merelbeke, Belgium). Differentiation was prolonged for 2–5 days.

Forced myogenesis

Dermal fibroblasts of an FSHD patient (91RD215) and a healthy individual (VH10) were cultured in DMEM high glucose, supplemented with 10% FBS. Fibroblasts were grown to 70% confluency for infection with Ad5 derived, replication defective adenoviral vector expressing the full length murine MyoD cDNA under the transcriptional control of the Rous sarcoma virus LTR. 33 34 The infection was carried out at a multiplicity of infection of 30 in DMEM/2% HS.

Figure 1 (A) Genomic structure of the FRG2 gene. FRG2 consists of four exons (boxes), and the ORF (black) starts in exon 1 and ends in exon 4. The LINE sequence in the ORF is in grey. The exon sizes (bp) are depicted above, and the primers used in this study are shown below the exons. (B) Exon sizes and splice sites of FRG2. Intronic sequences are in lower case, and exon sequences in capital letters.
Following an infection period of 2 h at 37°C, the medium was replaced by DMEM/4% Ultroser-G (Gibco Invitrogen). Myogenic differentiation was initiated after 2 days by changing the Ultroser-G content to 0.4%.

**RNA extraction**

Using the Instapure RNA extraction kit (Eurogentec, Seraing, Belgium) according to the manufacturer’s instructions, total RNA was extracted from four monochromosomal cell hybrids: GM11687, 4L10, and GM11688 (carry D4Z4 repeat arrays of 140, 85, and 60 kb, respectively), and 762-8a (mosaic for repeat arrays of 150 and 120 kb). Similarly, RNA was extracted from the Ad5-MyoD infected fibroblasts during proliferation, 2 days after infection (before differentiation) and at 2, 4, and 6 days after induction of differentiation. RNA extraction from myoblast cultures (p. 6–10) was done during proliferation, and after 1, 3, and 5 days of differentiation.

**RT-PCR**

In order to clone the FRG2 transcript, randomly primed first strand cDNA synthesis was performed using Superscript II (Invitrogen, Carlsbad, CA, USA). The transcription initiation site was determined by 5’ RACE. After first strand synthesis, cDNA was incubated with TdT (Amersham Pharmacia Biotech, Buckinghamshire, UK) in the presence of 2 mM dATP according to the manufacturer’s instructions. Both 5’ and 3’ RACE were performed using the primer dT-EXTV in combination with a gene specific primer (3r and 3f, respectively). Subsequently, nested PCR was performed using a gene specific primer (1r or 2r for 5’ RACE and 4f1 for 3’ RACE) in combination with the primer EXT2. Primers are depicted in fig 1 and all primer sequences are available upon request.

Amplification of FRG2 transcripts for sequence analysis was done in 35 cycles of 40 s at 94°C, 40 s at 55°C, and 90 s at 72°C, using primers 1p and 4r2. Nested amplifications of 25 cycles were performed using 1 μl of the first amplification products and primers 3f and 4r1.

For radioactive and real time RT-PCR, RNA was treated with RNase free DNase I (Promega Benelux, Leiden, The Netherlands) prior to cDNA synthesis. RNA isolated from the monochromosomal cell line GM11687, which expresses FRG2, was used to test all primersets and as a positive control in RT-PCR reactions. To obtain radioactively labelled FRG2 PCR products, initially 40 cycles (reaction volume 25 μl, primers 3f and 4r1) were performed: 40 s at 94°C, 40 s at 55°C, and 40 s at 72°C. Then 12.5 μl of this PCR product was amplified for one cycle in the presence of 2.5 μCi [32P]dCTP, adding fresh primers, buffer, and Taq polymerase to a final volume of 25 μl. GAPDH was amplified similarly in 25 cycles using GAPDHf and GAPDHz primers. The products were separated by 5% PAGE and detected by autoradiography.

Real time RT-PCR was performed on an ABI-Prism 7700 Sequence Detector, running 40 cycles of 15 s at 95°C, 60 s at 60°C (initial denaturation 10 min at 95°C), using SYBR Green PCR Master Mix (Applied Biosystems, Foster City, CA, USA). Primersets spanning an intron (sequences available upon request) were designed using Primer Express 1.0 and results were analysed and quantified using SDS 1.9.1 software (Applied Biosystems).

**Sequence analysis**

PCR products were either cloned in the pCR2.1-TOPO cloning vector (Invitrogen) and sequenced using the M13 forward and reverse vector primers, or directly purified using the PCR purification kit (Qiagen, Hilden, Germany) and sequenced using gene specific primers. Sequence reactions were performed with the BigDye Terminator cycle sequencing kit (Applied Biosystems) and analysed on an ABI377 DNA sequencer.

**Expression vector constructs and transfection**

The full length FRG2 cDNA was cloned in frame in pSUPERCATCH, pSG8, and pEGFP-C1 (Clontech, Palo Alto, CA, USA). These vectors fuse the FRG2 protein to different aminoterminal tags: FLAG, vesicular stomatitis virus glycoprotein (VSV-G), and enhanced green fluorescent protein (EGFP), respectively.

Then 20 h prior to transfection, 2×10⁶ COS-1, U2OS human osteosarcoma cells, TE671 human rhabdomyosarcoma cells (gift from Dr A Belajew, Mons, Belgium), or human myoblasts were plated on coverslips in 24-well tissue culture plates. Cells were transiently transfected with Fugene (Roche, Basel, Switzerland) according to the manufacturer’s instructions. Briefly, 1 μl Fugene was mixed with 20 μl DMEM and left for 5 min at room temperature. Subsequently, 0.2 μg Qiagen purified plasmid DNA was added and incubated for 15 min at room temperature. The cells were refreshed with 0.3 ml of medium and the Fugene-DNA mix was added dropwise to the medium. COS-1 and TE671 cells were fixed 24–48 h after transfection, human myoblasts were allowed to differentiate for 3 days prior to fixation.

**Promoter reporter constructs**

To generate luciferase reporter constructs, a 3.4 kb BglII fragment immediately preceding the ORF of FRG2 was cloned into plgL3-Basic (Promega), both in the forward and the inverted orientation. A smaller construct containing 800 bp including the CAAT and TATA boxes was generated by internal deletion of the BglII-BglII fragment. Next, a 3.3 kb KpnI D4Z4 repeat fragment was isolated from lambda42 phage DNA (access. no. AF117653) and inserted into the KpnI restriction site of pGL3 basic and of the pGL3 vector containing the 3.4 or 0.8 kb FRG2 promoter fragment, respectively. Constructs in which the D4Z4 unit was inserted in a forward orientation were selected. Further D4Z4 repeat units were then introduced into the constructs carrying FRG2 promoter fragments by inserting (partially digested) SfiI fragments (again isolated from lambda42 phage DNA) into the unique and directional SfiI site of the D4Z4 KpnI unit. This resulted in pGL3 vectors in which no, one, two, or three D4Z4 units were directionally cloned 5’ to the FRG2 promoter driving luciferase expression. A construct carrying one D4Z4 unit but no FRG2 promoter was used as a control for the activity of D4Z4 alone. Transfection in TE671 human rhabdomyosarcoma and in U2OS cells was performed as described under “Expression vector constructs and transfection”, but in six-well plates (Greiner Bio-One, Longwood, FL, USA) and with 0.5 μg of the reporter construct; efficiency was controlled by co-transfection of 0.5 μg pCMVLacZ. For each construct three to six wells per cell type were transfected and analysed independently. Luciferase and β-galactosidase activity were determined using the appropriate assay systems (Promega). Activity of a construct was expressed as luciferase units per unit of β-galactosidase and compared to the empty plgL3-Basic vector.

**Immunohistochemistry**

Then 24–48 h after transfection, cells cultured on coverslips were washed with PBS and fixed in 2% paraformaldehyde/0.1% Triton X-100 for 20 min. After washing twice in PBS/0.1% Triton X-100, GFP transfected cells were dehydrated and embedded in Vectashield (Vector Laboratories, Burlingame, CA, USA). Cells transfected with FLAG or VSV fusion constructs were washed twice in PBS/0.1% Triton X-100, once in PBS/1% BSA (fraction V), blocked in 0.1 M...
FRG2 is a candidate gene for FSHD

---

**Figure 2A** cDNA sequence of the chromosome 4 copy of FRG2. The predicted FRG2 protein sequence is shown below the DNA sequence. Arrowheads (►) indicate the first nucleotide following the splice junctions. A splice variant obtained by a second CAG triplet at the splice acceptor site of exon 4 is also indicated. The polymorphisms resulting in amino acid changes between the chromosome 4 and chromosome 10 copies in the ORF are depicted above the chromosome 4 sequence. The potential NLS and PTS1 signals are in bold while the polyadenylation signals are underlined. The LINE sequence is shaded.
NH₄Cl for 10 min, and washed again in PBS/1% BSA. After incubation with the first antibody in PBS/1% BSA for 1 h, coverslips were washed with PBS/1% BSA and incubated with the second antibody for 30 min. Subsequently, they were washed once with PBS/1% BSA and once in PBS, dehydrated and embedded in Vectashield. The entire procedure was done at room temperature. Antibodies were used as follows: monoclonal M2 anti-FLAG (Sigma-Aldrich, St Louis, MO, USA) 1:5000, monoclonal anti-VSV-tag P5D4 (gift from Dr J Franssen, Nijmegen, The Netherlands) 1:3000, monoclonal anti-desmin (Monosan, The Netherlands) 1:100, and rabbit-anti-mouse Alexa 594 (Molecular Probes, Eugene, OR, USA) 1:1000.

**Fluorescence in situ hybridisation (FISH) on metaphase chromosomes**

Metaphase chromosome spreads of cultured control lymphocytes were hybridised with a 5.2 kb SmaI-XhoI fragment from cosmid 23D11 on which the FRG2 gene resides. The probe was labelled by nick translation with biotin-16-dUTP (Roche). Hybridisation, washing, and staining were performed as described previously.39

**RESULTS**

**Identification of FRG2**

Computer aided exon prediction programs revealed the presence of four putative exons within a 3 kb genomic interval approximately 37 kb proximal to the D4Z4 repeat array. This predicted exon cluster was named FRG2, for FSH Region Gene 2. In each predicted exon, primers were designed for RT-PCR analysis. The intron-exon structure of FRG2 was defined using the constitutively expressed FRG2 (see under “Expression analysis of FRG2”) in the hamster-human hybrid cell line GM11687, which retained only human chromosome 4. The cDNA products were sequenced and compared to the genomic sequence of the chromosome 4 specific PAC clone RCPI-1-226K22 (acc. no. AF146191).

FRG2 is composed of four exons and encodes an mRNA of 2084 bp (figs 1 and 2). The ORF starts in exon 1, ends in exon 4, and encodes a putative protein of 278 amino acids. By 5' and 3' RACE, the transcription initiation and two polyadenylation sites were identified, respectively. Both polyadenylation signals, 48 bp apart, are used. No alternative splicing was observed except for the use of both acceptor sites created by a CAG doublet in exon 4. This splice variation does not change the ORF, but creates an additional alanine codon. Using PSORT (http://psort.nibb.ac.jp), two potential nuclear localisation signals (NLS) at amino acid positions 96 (RKKR) and 157 (KRHR) were predicted as well as a peroxisomal targeting signal (PTS1) at the carboxyterminal end of the protein.40 41 The complete cDNA sequence of FRG2 is shown in fig 2A and B. The putative FRG2 protein does not show significant homology to known proteins in the databases. A partial LINE sequence was identified within the ORF of FRG2. Significant homology outside this LINE element was found with only a few ESTs and IMAGE clones isolated from various tumour tissues and blood lymphocytes.

As observed previously, copies of sequences from this subtelomeric chromosomal region are dispersed throughout the genome.12 31 32 42 Homologues of FRG2 were assigned to multiple chromosomal loci by PCR amplification using FRG2 primers 2f and 3r (fig 1). Closely related sequences were identified on chromosomes 1, 4, 8, 10, 18, and 20 (data not shown). FISH analysis on metaphase spreads of human fibroblasts using a highly homologous 5 kb genomic FRG2 probe derived from chromosome 10 showed strong hybridisation to the 4qter and 10qter regions, as well as weaker signals on chromosome 19 or 20q12. Sometimes weak signals were also detected on chromosome 1 (centromeric) and the p arms of the acrocentric chromosomes (data not shown). BLAST searches using the genomic sequence of FRG2 revealed additional homologues sequences on chromosomes 3, 7, 16, and 22. It must be noted here that the assignment of BAC and PAC sequences to specific chromosomes are often

---

**Figure 2B** cDNA sequence of the chromosome 4 copy of FRG2 (continued).

```cda
912 GAG ATG CAG CTC TGG GAA TGA GAA CAA GGA TCT GCT CCT TCT CAG ATT 959
960 TCA GAG CAC CAG TGG CAA TTT TAG ACA CAC TGG GTT AAT AAA 1007
1008 TAG CAG CAC ATG TCA TGA TAG CAC TCA TAG GAG ACA AAG CTT TGG CAG GCT ATA CTC 1055
1056 AGA ATG GCG AGA GCC CTT AAT TTT GCA GAC CGC TAA GAC TAT AGA CAA 1103
1104 ATT TTA TAT TTT ATG TTA GAC ATT TGA TGG CTT TGG CAT GTC TGA TGA 1151
1152 CAG TCA TGC ATT CCT TCT TTA TTA TCA GAA AAA CAT TAG AAT GTA ATC GTG 1199
1200 AAT TGG CAT ATT TTA GAT TGG AGA AAA GAA GTA AAT ATA AAA TTA TCT GCT 1247
1248 CTC TTT TGG TTT TTT TTT TTT TTT GAG ACA GTC TGG CTA TGT CAC CCA 1295
1296 GCC TGC AGT GCA GTG GCA CTA CCT TAG CTC ACT GCA ACC TCT GCT CCT 1343
1344 TGG GTT CAA ACA ATT CTC ATC CTC CCT CCC AAG CAG CTC GTA CTA 1391
1392 CAG GCA GTA ACT GCT ATG GCC TCA TAT TTT TTT TTT TCT TGG ATT GCT 1439
1440 AGT AGT AGA GAC AGA GAT TGC TCA TGG GCA GGC TCT TGC CCA GCT CAG ACT CAG 1487
1488 GTT ATC TGC CCT CCC CCC CCC CAC GTC CTC TGG ACA TCA CAG GCA TGA 1535
1536 GCC GCC TTA CCA AGA AAT TGC TCT TCT TCT TTT AAT GCA GAA AAG GCA 1583
1584 GCC TCT CAC TCT TCG AGC CTC CCA CCC AGT ACT ATC CAC AAA 1631
1632 CCA TTA ATA GCA CTC CCT GTG GGA AAA TGT CTA TAT ATT TTT AGT TGT 1679
1680 ATG STA TTA TAG TAA AAT TAC TAT GCA AGC TGT TTA TCT TTA ATA TCT 1727
1728 CTA CAT AAA TAT ATG AAG TAT TAT CTA TGA AAT GTT ACT GTC 1775
1776 TAT TTA TGT ACC TGC CTC ATG TTT CAT TTC ATT TTA AAG ATC CTA AAT 1823
1824 TTA TAT TTT ATT TAT TAT ACA GCT GTA CTA CAA AAT ATC AGT CAA 1871
1872 AGG ATA TGG AGA TTT CAT GAC CTA CAA TAC AGT GTA TGT TGT TAT 1919
1920 ATT TGA CGT ATA TCT TAC TGC TAT TTT GTA CTA CGA CTA TCA ACT ATT 1967
1968 TCA TTA TCT AAG TTT ATT TCT TCT TGT GCT TTA TAA TTT TCA GCT 2015
2016 ATT GTA ATG AAA TAA ACA ATG TGG TTT GCA ATT TTA AAT TCC CAT 2063
2064 ATG GAA TTT GTA TTT AAT AAA 2084
```
ambiguous due to pericentromeric and subtelomeric plasticity. Sequence comparisons demonstrated that the FRG2 copies on chromosomes 4 and 10 are highly homologous to each other, whereas homologues assigned to chromosomes 3 and 22 encode identical proteins but display a total of 20 amino acid substitutions compared to the chromosome 4 variant (fig 3).

Since only the copies on chromosomes 4 and 10 are closely associated with intact D4Z4 repeat arrays, we further focused on these two copies. Sequence comparison of the chromosome 4 and 10 loci revealed five nucleotide mismatches in the ORF. All of these, except for the silent nucleotide difference at position 584, result in amino acid substitutions (chromosome 4 vs 10): a Ser8Pro at position 75, Ala131Pro at position 444, His148Arg at position 496, and His155Arg at position 517, respectively (fig 3).

Expression analysis of FRG2
Expression analysis by RT-PCR of nine monochromosomal cell hybrids containing either human chromosome 4 (n = 5) or human chromosome 10 (n = 4) as their only human component, revealed constitutive FRG2 expression only in GM11687 (chromosome 4, data not shown). Peripheral blood lymphocytes and fibroblasts from FSHD patients and controls, and control brain and fetal muscle were used as a template to test for FRG2 expression by RT-PCR using primers in each exon (fig 1). Muscle biopsies of seven independent FSHD patients and six healthy controls were analysed by
and a higher percentage of cells with a necrotic appearance. Swollen cytoplasm, large vacuoles adjacent to the nucleus, healthy myoblasts. FSHD myoblasts are characterised by a FRG2 RT-PCR using primers that recognise all homologues show patients (PROMM and DSM), and healthy controls. Nested cultures of FSHD patients, non-FSHD muscular dystrophy serum starvation, expression of myoblasts. However, after induction of differentiation by serum deprivation (day 2, 3, and 5). FRG2 and desmin transcript levels were normalised against 18S rRNA transcript levels. FRG2 primers were specific for chromosomes 4 and 10. Expression of FRG2 expression levels were found between control (mean Ct = 38.4 (SD 1.5)) and FSHD (mean Ct = 38.3 (SD 1.5)) Moreover, if any expression was detected, no differences in FRG2 expression levels were found between control (mean Ct = 38.6 (SD 1.6)) and FSHD (mean Ct = 38.3 (SD 1.5)) muscle, contradicting the specific and repeat-length dependent upregulation in FSHD muscle recently reported by Gabellini et al.19 Next, myoblast cultures isolated from skeletal muscle biopsies were analysed. As described before,26 myoblast cultures of FSHD patients are phenotypically different from healthy myoblasts. FSHD myoblasts are characterised by a swollen cytoplasm, large vacuoles adjacent to the nucleus, and a higher percentage of cells with a necrotic appearance.27 Furthermore, FSHD myoblasts appear to have a reduced replicative capacity and a morphology resembling senescence.4 Expression of FRG2 was analysed in myoblast cultures of FSHD patients, non-FSHD muscular dystrophy patients (PROMM and DSM), and healthy controls. Nested RT-PCR using primers that recognise all homologues show barely or no detectable FRG2 expression in proliferating myoblasts. However, after induction of differentiation by serum starvation, expression of FRG2 was demonstrated in myoblasts of FSHD patients and to a lesser extent in the non-FSHD myopathy patients, but not in healthy controls (results not shown). These data were confirmed and quantified by real time RT-PCR using FRG2 primers specific for chromosome 4 and 10. Expression of desmin demonstrated the presence of myogenic cells in the differentiating primary myoblast cultures (fig 4). Additionally, immunohistochemical analysis revealed that virtually all (80–100%) cells in the proliferating myoblast cultures of patients and controls stained positive for desmin46 (data not shown).

Chromosomal origin of transcription
To determine in more detail the chromosomal origin of the FRG2 transcripts in differentiating myoblasts, the RT-PCR products of all patient cell lines were directly analysed on an ABI377 DNA sequencer. Interestingly, the FRG2 transcripts in differentiating myoblasts of FSHD patients were mainly derived from chromosome 10 and to a lesser extent from chromosome 4. FRG2 transcripts in non-FSHD myopathies were however not derived from chromosome 4 or 10, but compatible with FRG2 sequences derived from chromosome 3 or 22, which are identical. In order to show that the sequence variations that we observed between the different chromosomes are not merely randomly occurring polymorphisms, exon 4 of FRG2 was amplified by PCR from independent sources. Direct sequence analysis of the PCR products revealed that in five independent sources of chromosome 4, the exon 4 sequences are 100% identical. The same holds true for four independent sources of chromosome 10. Moreover, among all the homologous sequences identified by BLAST search, none was 100% homologous to the chromosome 4 or 10 sequence, indicating that the chromosome 4-type and 10-type FRG2 sequences are indeed chromosome specific.

Promoter study
In order to characterise the putative promoter region preceding the FRG2 gene, we designed and generated luciferase reporter constructs. In TE671 human rhabdomyosarcoma cells, a transiently transfected reporter construct carrying the 3.4 kb region immediately preceding the FRG2 gene on chromosome 4 in the forward orientation displayed 125 times higher activity than the same region cloned in the inverted orientation. Assays using only 800 bp predicted promoter region containing the CAAT and TATA boxes showed 254-fold induction of luciferase activity as compared to the 3.4 kb inverted construct (results not shown).
FRG2 is a candidate gene for FSHD

To test their influence on the activity of the FRG2 promoter, up to three D4Z4 repeat units were directionally cloned 5' of the 0.8 and 3.4 kb FRG2 promoter sequences in the luciferase reporter constructs and transiently transfected into TE671 and U2OS human osteosarcoma cells. Interestingly, the highest luciferase activity was observed when transfecting constructs containing one D4Z4 unit 5' of an FRG2 promoter sequence. Both for the 0.8 and the 3.4 kb promoters tested, increasing the numbers of D4Z4 units resulted in a stepwise reduction of luciferase activity, indicating that the presence of D4Z4 repeat units can inhibit the activity of the FRG2 promoter (fig 5).

FRG2 expression in forced myogenesis

To study the potential involvement of FRG2 in myogenesis, fibroblast cultures of FSHD patients and healthy controls were analysed for FRG2 expression before and after infection with an adenovirus carrying the murine MyoD gene. To confirm the initiation of forced myogenesis, expression of muscle specific desmin was analysed by RT-PCR. Before infection and at four time points after infection and differentiation, RNA was isolated and analysed for the presence of FRG2. Fibroblasts did not express FRG2 during proliferation or upon infection with MyoD. However, at all time points after induction of differentiation by serum starvation, strong FRG2 expression was observed in FSHD patient and healthy control fibroblasts (fig 6). This expression of FRG2 originated predominantly from chromosome 10 and to a lesser extent from chromosome 4 (data not shown).

Subcellular localisation and morphology upon transfection of FRG2

Since the putative FRG2 protein contains a potential carboxyterminal peroxisomal PTS1 signal and two putative nuclear localisation (NLS) signals, its subcellular localisation was studied. The FRG2 protein was fused to aminoterminal VSV, FLAG, and GFP tags, respectively, to minimise the artificial effects of the tag on the cellular localisation. We transiently transfected several cell lines including COS-1, TE671, and human myoblasts. All transfections showed a nuclear, but no peroxisomal localisation of FRG2. The staining varied from a homogeneous to a granular nuclear pattern (fig 7), but no specificity for a particular subnuclear compartment was established. Interestingly, despite the low number of transfected myoblasts, overexpression of EGFP-FRG2 apparently causes morphological changes in FSHD and control myonuclei ranging from necrotic, pyknotic to fused and clumped nuclei (fig 7).
predict a promoter including CAAT and TATA boxes preceding the gene. This putative promoter was shown to be very active in a luciferase reporter assay. The presence of increasing numbers of D4Z4 units resulted in gradual reduction of FRG2 promoter activity, supporting the concept that D4Z4 is repressing transcriptional activity in cis, possibly through the action of the D4Z4 repressor complex. However, in concordance with previously published data, also transactivating elements should be present in D4Z4 as a single D4Z4 unit already drives luciferase activity (fig 5). Our transient transfection assays with FRG2 expression constructs suggested a nuclear localisation for the FRG2 protein, as predicted by the NLS signals. Non-consensus peroxisomal targeting signals (PTS-1), such as the AKL motif in FRG2, only function in combination with specific additional amino acid residues. Their absence in FRG2 may explain its nuclear rather than peroxisomal localisation.

As shown by our PCR, FISH, and BLAST searches, homologous genomic FRG2 segments are dispersed over the genome. The D4Z4 linked FRG2 sequences from chromosomes 4 and 10 are almost identical to each other, whereas other homologues are incomplete or only remotely related to the FRG2 sequence originating from chromosome 4. These homologues are often found in pericentric or subtelomeric regions that are frequently involved in ectopic chromosomal rearrangements. The presence in these regions of genes, such as members of the human olfactory gene family, has prompted others to suggest that subtelomeres serve to generate gene diversity.

Despite the widespread presence of FRG2-like sequences, the FRG2 genes on chromosome 4 and 10 have several properties that make them attractive genes for involvement in FSHD pathogenesis. First, the FRG2 genes on both chromosome 4 and 10 are closely linked to the D4Z4 repeat. Deletion of repeat units from chromosome 4 is the primary event in FSHD. Second, the FRG2 promoter can, in principle, drive expression to very high levels and the gene is constitutively expressed in the monochromosomal cell line GM11687, but undetectable in a variety of human tissues. This demonstrates that the copies on chromosome 4 and 10 have the potential to be expressed under tight regulatory control in vivo. In concordance with this, our promoter studies demonstrate that the FRG2 promoter is sensitive to the presence of D4Z4 repeat units. Third, the expression profile: our studies in myoblast cultures show that copies of FRG2 compatible with chromosome 4 and 10 origins are expressed only in differentiating, but not proliferating myoblasts of FSHD patients, while healthy control myoblasts failed to express any copy of FRG2. In contrast, differentiating myoblasts of non-FSHD myopathies express a distantly related copy of FRG2 (see below). Fourth, its putative involvement in myogenesis: FRG2 expression is absent in all tissues tested except differentiating myoblasts and an evident transcriptional upregulation is detected after adenoviral MyoD expression in fibroblasts. Forced expression of MyoD and other members of the myogenic family of basic helix-loop-helix transcription factors activates a myogenic program in fibroblasts. It has also been demonstrated that MyoD can remodel chromatin in the regulatory region of skeletal muscle genes and initiate endogenous gene transcription. This may explain how FRG2 expression can be detected upon strong adenoviral MyoD overexpression in healthy fibroblasts, but not in normal, untreated myoblasts.

Sequence comparisons of various FRG2 copies revealed that the FRG2 protein from chromosome 10 differs in only four amino acids from the chromosome 4 sequence. Two of these changes introduce prolines that may change the conformation of the protein. In contrast, the homologues identified on chromosomes 3 and 22 encode putative proteins with 20 amino acid substitutions compared to the chromosome 4 variant. A number of these amino acid replacements involve polarity changes, possibly affecting protein function. FRG2 proteins from chromosome 4 (or 10) may therefore function (or even dysfunction) differently from the 3/22-like FRG2 copy that is expressed in non-FSHD myopathies. A recently published gene expression profiling study suggested that FSHD exhibits a unique defect in myogenic differentiation. These and our observations that the nuclear morphology of healthy and FSHD myoblasts transfected with FRG2 tends to deteriorate, leads us to speculate that aberrant expression of the chromosome 4 or 10 copy of FRG2 may very well be detrimental to myogenic differentiation and can thus play a causative role in the loss of muscle function seen in FSHD patients. The low levels of FRG2 expression from chromosome 3 or 22 in non-FSHD myopathies may then merely represent general muscular dystrophy related processes in these myoblasts. However, we cannot exclude that FRG2 expression is not a primary effect of D4Z4 deletion, but a secondary reaction of the FSHD myoblasts, albeit different from non-FSHD muscular dystrophies.

Figure 8  Transfection hypothesis. Partial deletion of the D4Z4 repeat array on the chromosome 4a subtelomeric region changes the local chromatin structure and its interaction with the nuclear envelope, enabling expression of FRG2 to a low level. Upon pairing of chromosomes 4 and 10, transcription of FRG2 from chromosome 10 is transactivated.

Radioactive RT-PCR failed to detect any expression of FRG2 in RNA isolated from seven independent FSHD and six healthy control muscle biopsies, while analysis of β-actin, desmin, GAPDH, and 18S rRNA transcript levels confirmed the integrity of our RNA samples. Real time RT-PCR confirmed the (almost complete) absence of FRG2 transcripts in all of our biopsies, thereby failing to reproduce the repeat-length dependent upregulation of FRG2 expression in FSHD muscle as reported by Gabellini et al. Assuming a role of FRG2 in the differentiation of myoblasts, one possible explanation for this discrepancy may be that the level of regeneration in a biopsy is critical for the detection of FRG2 expression. It must, however, be noted here that others have likewise been unsuccessful in reproducing the upregulation of FRG1 and ANTI expression in FSHD muscle also described by Gabellini et al. A challenging finding is that in FSHD patient myoblasts FRG2 is expressed both from chromosome 4 and 10, while FSHD is only associated with a partial deletion of the D4Z4 repeat array on chromosome 4. It has been demonstrated that chromosome pairing during interphase can lead to transcriptional activation of homologous genes in trans by transvection (reviewed in Pirrotta14). Interestingly, somatic pairing between the subtelomeric regions of chromosomes 4q and 10q is increased in FSHD patients compared to controls, possibly facilitating gene regulation by trans sensing effects. Moreover, D4Z4 repeats transfected into C2C12 myoblasts
were shown to have a significant trans effect on myoblast differentiation. In a recent model, Winokur et al and S Winokur, personal communication) suggest that aberrant gene expression in FSHD results from disturbed interaction of 4qtel with the nuclear envelope following contraction of the D4Z4 repeat array. Elaborating on this, we propose that partial deletion of the D4Z4 repeat array leads to a change in local chromatin structure and aberrant interaction of 4qtel with the nuclear envelope, enabling FRG2 expression in cis on chromosome 4. In those FSHD cells where chromosomes 4 and 10 are somatically paired, activation of FRG2 expression may also occur in trans on chromosome 10, possibly via transvection (fig 8). The higher responsiveness of the chromosome 10 copy for FRG2 expression upon pairing and transvection may be explained by the close proximity of euchromatic sequences at 60 kb distance, whereas on chromosome 4 non-subtelomeric sequences are at least 500 kb further upstream.

A partial deletion of D4Z4 on chromosome 10 however does not result in FSHD. A possible explanation lies in the observation that, unlike chromosome 4qtel, 10qtel is not normally localised to the nuclear periphery (S Winokur, personal communication). It is currently unknown whether contracted D4Z4 repeats on chromosome 10 can also activate FRG2 in cis, and in turn on chromosome 4. FRG2 expression studies in patients with a deletion of the repeat array on chromosome 4 extending proximal to FRG2 and in healthy individuals with a FSHD sized repeat array on chromosome 10 would support such a transvection mechanism for FRG2. Unfortunately, myoblasts of these individuals are currently not available to test this hypothesis.

ACKNOWLEDGEMENTS

The authors thank E Cuppen for providing us with the pSG8-WSV vector and H Dauverse for performing FISH analysis.

ELECTRONIC-DATABASE INFORMATION

Information on PSORT is available from http://psort.nibb.ac.jp

REFERENCES


