The following full text is a publisher's version.

For additional information about this publication click this link.
http://hdl.handle.net/2066/21110

Please be advised that this information was generated on 2019-11-02 and may be subject to change.
Facioscapulohumeral muscular dystrophy (FSHD) is caused by deletions of 3.3-kb tandemly repeated units contained within a large polymorphic EcoRI fragment close to the telomere of chromosome 4q. Since the rearrangements were assumed to interfere with the structure or function of the putative FSHD gene, the gene search was focused on cosmids containing these repeat units and, in addition, cosmids spanning 75 kb of upstream sequences. cDNA selection hybridization was applied to four overlapping cosmid clones, yielding a total of 150 putative cDNA clones. These clones showed a random distribution across the cosmid contig, except for three regions which contained a much larger number of clones. Nine cDNA clones hybridized to a 2.2-kb EcoRI fragment, located 22 kb centromeric to the 3.3-kb repeated units. This 2.2-kb fragment showed evolutionary conservation, and analysis of the sequence by "GRAIL" predicted the presence of several exons. Transcripts homologous to this fragment could be identified but none of them originated from the 4q35 locus. Strikingly, most clones revealed 4-10 homologous loci, and no single copy clones could be isolated. These findings are in line with earlier observations by fluorescent in situ hybridization (FISH) showing hybridization of individual cosmid clones to multiple chromosomes. The presence of homologous regions on other chromosomes seriously complicates the cloning of the FSHD gene. © 1995 John Wiley & Sons, Inc.

Key words: facioscapulohumeral muscular dystrophy (FSHD) • chromosome 4qter • subtelomeric region • cDNA selection • positional cloning

SEARCH FOR THE FSHD GENE USING cDNA SELECTION IN A REGION SPANNING 100 kb ON CHROMOSOME 4q35

JUDITH C. T. van DEUTEKOM, MSc, MARTEN H. HOFKER, PhD, SILKE ROMBERG, BSc, MICHEL van GEEL, MSc, JOHANNA ROMMENS, PhD, TRACY J. WRIGHT, PhD, JANE E. HEWITT, PhD, GEORGE W. PADBERG, MD, PhD, CISCA WIJMENGA, PhD, and RUNE R. FRANTS, PhD

MGC-Department of Human Genetics (Ms. van Deutekom, Dr. Hofker, Ms. Romberg, Mr. van Geel, Dr. Wijmenga, and Dr. Frants) and Department of Neurology (Ms. van Deutekom, Ms. Romberg, Mr. van Geel, and Dr. Wijmenga), Leiden University, Leiden, The Netherlands; Department of Neurology, The Hospital for Sick Children, Toronto, Ontario, Canada (Dr. Rommens); Department of Structural and Cell Biology, University of Manchester, Manchester, United Kingdom (Drs. Wright and Hewitt); and Department of Neurology, University of Nijmegen, Nijmegen, The Netherlands (Dr. Padberg). Dr. Wijmenga's present affiliation is the National Center for Human Genome Research, NIH, Bethesda, MD.

Presented in part at the Eighth International Congress on Neuromuscular Diseases, Kyoto, Japan, 1994.

Acknowledgments: This study was supported by the Muscular Dystrophy Association (MDA), the Princes Beatrix Fonds, Association Française contre les Myopathies (AFM), Royal Netherlands Academy of Arts and Sciences (KNAW), and the European Neuromuscular Centre (ENMC).

Address reprint requests to Dr. R. R. Frants, PhD, MGC-Department of Human Genetics, Sylvius Laboratory, Leiden University, Wassenaarseweg 72, 2333 AL Leiden, The Netherlands.


Facioscapulohumeral muscular dystrophy (FSHD) is a neuromuscular disorder initially characterized by progressive weakness of the muscles of the face, shoulders, and upper arms.7,8 The gene was localized to chromosome 4q35,16,21 5 cM distal to the linkage group 4cen... D4S171-F11-D4S187-D4S163-D4S139.13,17,22 Subsequently, the FSHD gene region was further refined by the detection of de novo DNA rearrangements within a highly polymorphic EcoRI fragment using probe p13E-11 (D4F104S1).9,18,20,23 These rearrangements were shown to be deletions of an integral number of 3.3-kb tandemly repeated units.19,25 Several copies of this 3.3-kb repeated unit have been sequenced.5,26 Despite the presence of a large open reading frame containing a double homeobox motif,5 no evidence could be obtained that the complete FSHD gene resides within these repeats.
Therefore, the search for the FSHD gene was focused on a much larger area around the deletion region. Unfortunately, the region just telomeric of the repeats seems to be unclonable, probably due to the close proximity of the 4q telomere (about 50 kb) (Altherr, personal communication) and, likely, the presence of subtelomeric repeats. However, the region containing the 3.3-kb repeated units, as well as more upstream sequences, have been cloned into YACs and cosmids.24,27

Different techniques are currently available for the isolation of transcribed sequences using cosmid clones as starting material, such as screening cDNA libraries with complete cosmid inserts, cDNA selection, and exon trapping. In the cDNA selection strategy, the cDNA itself serves as a probe that will be hybridized to cosmid DNA. A great advantage of cDNA selection, as compared to for instance exon trapping, is the generation of genuine cDNA probes during the selection procedure. In the present study, cDNA selection was applied to a 100-kb cosmid contig, covered by four overlapping cosmids. This contig contains the tandem repeats and 75 kb of upstream sequences. Using two different pools of cDNA, a total of 150 putative cDNA clones was obtained. These clones as well as some genomic subclones were analyzed by DNA sequencing, and by hybridization to Southern and Northern blots for the identification of single copy and/or evolutionary conserved sequences and transcripts. The FSHD gene region was found to contain highly repetitive sequences as well as low copy sequences. Several putative cDNA clones hybridizing to these low copy sequences were found to originate from non-4q35 loci. This property of the FSHD region complicates cloning the FSHD gene.

MATERIALS AND METHODS

Cosmid Contig. A contig of four overlapping cosmids from the FSHD region was analyzed. This cosmid contig spans 100 kb, and includes cY3, cY34, and cY13, which were previously isolated from a library generated from YAC 25C2E (D4S1093),27 and cosmid c49, which was isolated from a chromosome 4 cosmid library.23 EcoRI and KpnI restriction maps of each cosmid were available (Fig. 1).

cDNA Selection. cDNA selection was performed as previously described.6,11 Hybridizations to the cosmids cY3, cY34, c49, and cY13 were performed using two different human cDNA pools: (1) tissue mix: total RNA of liver, lymphoblasts, fibroblasts, bone marrow, and intestine and poly A+ RNA or

![Figure 1](image_url)

FIGURE 1. Top: the map with loci from the FSHD linkage group on 4q35 is shown. Below: the EcoRI (E) and KpnI (K) restriction maps of the four overlapping cosmids cY3, cY34, c49, and cY13. The locations of several genomic clones and the putative cDNA (LCD) clones in the contig are shown above and below the restriction maps, respectively. The fragment sizes are given in kilobases. The 2.2-kb EcoRI fragment is indicated by a hatched box.
fetal liver; and (2) brain mix: total RNA of frontal cortex and poly A+ RNA of fetal brain. The eluted cDNAs were amplified using primers containing an EcoRI site, which facilitated cloning into the EcoRI site of the vector pBluescriptII/KS. Initially, all clones were hybridized to randomly primed cDNA (prepared from total human RNA) to detect and eliminate clones with highly repetitive sequences. Probes were prepared from plasmid inserts, which were obtained by digestion of plasmid DNA with EcoRI or by PCR.

**Analysis of Putative cDNA (LCD) Clones.** All LCD (Leiden cDNA) clones were initially mapped back to the parental cosmids cY3, cY34, c49, and cY13 by hybridization of the EcoRI inserts to Southern blots containing EcoRI and KpnI digests of each cosmid (1.5 μg/lane). Subsequently, positive clones were sequenced and analyzed for: (1) single copy sequences by hybridization to Southern blots with total human DNA digested with several restriction enzymes (7 μg/lane); (2) evolutionary conserved sequences by hybridization to “Zoo” blots containing EcoRI-digested DNA from several animals like Rhesus or Macaque monkey, cow, and mouse (10 μg/lane); and (3) 4q35 origin by hybridizations to chromosome 4-specific somatic cell hybrids HHW416 (kindly provided by Dr. Altherr) and GM11687 (NIGMS Human Genetic Mutant Cell Repository) (20 μg/lane) or to Y25C2E. Digestions were carried out with restriction enzymes of Pharmacia using recommended conditions. Inserts were labeled with α32P-dCTP, using the multiprime DNA labeling system (Amersham). Hybridizations were carried out for 16 h at 65°C as described.2 The filters were washed at 65°C and to a stringency of 0.3 x SSC/0.1% SDS for “Zoo” blots containing EcoRI-digested DNA from several animals like Rhesus or Macaque monkey, cow, and mouse (10 μg/lane); and 2 x SSC/0.1% SDS for “Zoo” blots, followed by autoradiography for 1–16 h at −70°C using Konica AX film with an intensifying screen.

**Sequence Analysis.** Large parts of genomic sequence were obtained by subcloning the 2.2-kb, 8.5-kb, and 6.5-kb EcoRI fragments just centromeric to the repeated units (Fig. 1). All genomic subcloning was performed in the pBluescriptII/KS vector.

Sequence reactions were carried out according to the dideoxy method,15 using a T7 Sequencing Kit (Pharmacia) and the universal and reverse M13-20 primers. Genomic and cDNA sequences were compared and analyzed by the DNAsis Computer Pro-
113, and 302 appeared to map to multiple restriction fragments, suggesting that they contained either several exons or repetitive sequences. Most clones were also hybridized to total human DNA. Despite the initial screening against repeats, some LCD clones still contained highly repetitive sequences (LCD 19, 57, 59, 67, 127, and 323). Genuine single copy probes could not be obtained. Hence, 26 low copy clones (about 4–10 copies) showing relatively simple hybridization patterns were selected for subsequent analysis.

LCD clones were hybridized to chromosome 4–specific somatic cell hybrids (HHW416 and GM11687) or to the YAC 25C2E. The hybridization pattern on chromosome 4–specific DNA was compared to the complex hybridization pattern generated by the LCD clones on Southern blots containing total human DNA. Each of the LCD clones 2, 10, 21, 51, 64, and 83 generated appropriate unique bands, indicative of a chromosome 4 origin. All other clones, in particular the clones located just centromeric of the 9b6a locus on cY3, showed a prominent non-4q35–specific hybridization pattern. These clones are homologous to the FSHD candidate region, but are probably derived from homologous loci outside the cosmid contig.

As previous mapping studies had indicated the target region to show considerable homology to several other chromosomal regions, sequence data of the LCD as well as the corresponding genomic clones were generated to identify the most homologous LCD clones. In particular, large parts of the 2.2-kb, 8.5-kb, and 6.5-kb EcoRI fragments (Fig. 1), which are located just proximal to the repeated units on cosmid c49, were sequenced. Of ten LCD clones mapping to these EcoRI fragments, the percentage of homology ranged between 80% and 97%. LCD clones 21, 60, 64, 72, 80, and 83 show the highest homologies of 97%, 91%, 96%, 93%, 93%, and 91%, respectively, over the entire insert. Hence, these clones could originate from 4q35. In addition, all LCD and genomic sequences were compared to sequences in the EMBL or GenBank data bases. Several clones show homology to repetitive sequences, such as LINE-1 and β-satellite sequences (Table 1). With the exception of LCD 60 (see below), no other significant homologies were found.

The search for potential coding sequences was carried out by “GRAIL” (Table 1), and by testing the clones for evolutionary conserved sequences (Table 1). About 25% of the clones reveal conserved sequences at nine distinct fragments. Interestingly, clones 21, 60, and 83 show both evolutionary conservation, and the presence of an excellent or good exon as predicted by “GRAIL.”

Analysis of the 2.2-kb EcoRI Fragment. The gene search focused on the 2.2-kb EcoRI fragment, because it contains the LCD clones 21, 60, and 83, which are highly homologous to the genomic sequence, evolutionary conserved, and coinciding with predicted exons (Fig. 2). Moreover, using the entire 2.2-kb EcoRI fragment as a probe, hybridization was detected in mouse, cow, dog, and chicken.

LCD 21 is a more specific probe than the entire 2.2-kb EcoRI fragment and was used to screen human cDNA libraries of skeletal muscle, placenta, and fetal brain. A placenta cDNA, clone CD25, with a length of 1800 bp was isolated and partly sequenced, showing 80% homology to the genomic sequence. CD25 detected highly conserved sequences in the monkey, cow and mouse (Fig. 3). However, no transcript could be detected in poly A+ RNA of several monkey tissues. Both the relatively low sequence homology to the 2.2-kb EcoRI fragment and the complex hybridization pattern on total human DNA hampered further studies with CD25.

LCD 60 was found to share a 96% homologous region of 86 bp with a skeletal muscle cDNA clone (HSBB8H051, accession no. Z28907), which was isolated by Génethon in a random sequencing project (“The Genexpress cDNA Program”). Interestingly, this clone itself showed high homology (94%) to another, larger cDNA clone isolated by Génethon HSB11B021, accession no. Z19285 (Fig. 2). Using primers based on the sequence of clone HSB11B021, a 235-bp product of the predicted size was obtained by PCR from a skeletal muscle cDNA library. This product, B021, detected a 9-kb transcript in poly A+ RNA of monkey shoulder muscle (Fig. 4a). However, B021 did not hybridize to Y25C2E and HHW416 DNA (Fig. 4b), suggesting that the transcript did not originate from 4q35. Subsequent analysis of a single chromosome cell hybrid mapping panel (kindly provided by Dr. Spurr) allowed the localization of B021 to chromosome 8.

DISCUSSION

The cDNA selection method is a simple, rapid, and effective tool for the generation of a regional transcription map. When pools of cDNAs are utilized, transcripts can be detected originating from different tissues within a single experiment. The main advantage of this method is the immediate
Table 1. List of putative cDNA clones in the FSHD gene region.

<table>
<thead>
<tr>
<th>LCD clone</th>
<th>Insert length (bp)</th>
<th>Genomic homology*</th>
<th>Hybridization pattern†</th>
<th>Evolutionary conservation</th>
<th>4q35 origin‡</th>
<th>Predicted exons§</th>
<th>Significant homologies‖</th>
</tr>
</thead>
<tbody>
<tr>
<td>2</td>
<td>475</td>
<td>84%</td>
<td>low</td>
<td>—</td>
<td>+</td>
<td>—</td>
<td>—</td>
</tr>
<tr>
<td>10</td>
<td>400</td>
<td>88%</td>
<td>low</td>
<td>NT</td>
<td>+</td>
<td>—</td>
<td>G</td>
</tr>
<tr>
<td>19</td>
<td>750</td>
<td>high</td>
<td>NT</td>
<td>NT</td>
<td>—</td>
<td>—</td>
<td>LINE 1</td>
</tr>
<tr>
<td>21</td>
<td>625</td>
<td>97%</td>
<td>low</td>
<td>+/−</td>
<td>+</td>
<td>—</td>
<td>—</td>
</tr>
<tr>
<td>23</td>
<td>780</td>
<td>83%</td>
<td>low</td>
<td>NT</td>
<td>—</td>
<td>—</td>
<td>—</td>
</tr>
<tr>
<td>51</td>
<td>800</td>
<td>88%</td>
<td>low</td>
<td>+</td>
<td>+</td>
<td>—</td>
<td>G</td>
</tr>
<tr>
<td>53</td>
<td>500</td>
<td>low</td>
<td>low</td>
<td>−</td>
<td>—</td>
<td>—</td>
<td>—</td>
</tr>
<tr>
<td>57</td>
<td>700</td>
<td>50%</td>
<td>high</td>
<td>NT</td>
<td>NT</td>
<td>—</td>
<td>βsat</td>
</tr>
<tr>
<td>59</td>
<td>425</td>
<td>65%</td>
<td>high</td>
<td>NT</td>
<td>NT</td>
<td>—</td>
<td>βsat, HSB11B021</td>
</tr>
<tr>
<td>60</td>
<td>950</td>
<td>91%</td>
<td>low</td>
<td>+</td>
<td>+/−</td>
<td>G</td>
<td>HSB88H051</td>
</tr>
<tr>
<td>64</td>
<td>655</td>
<td>96%</td>
<td>low</td>
<td>—</td>
<td>+</td>
<td>—</td>
<td>—</td>
</tr>
<tr>
<td>67</td>
<td>850</td>
<td>high</td>
<td>high</td>
<td>NT</td>
<td>—</td>
<td>—</td>
<td>—</td>
</tr>
<tr>
<td>68</td>
<td>925</td>
<td>low</td>
<td>low</td>
<td>−</td>
<td>—</td>
<td>—</td>
<td>—</td>
</tr>
<tr>
<td>72</td>
<td>500</td>
<td>93%</td>
<td>low</td>
<td>−</td>
<td>+/−</td>
<td>—</td>
<td>—</td>
</tr>
<tr>
<td>80</td>
<td>350</td>
<td>93%</td>
<td>low</td>
<td>+</td>
<td>+/−</td>
<td>—</td>
<td>—</td>
</tr>
<tr>
<td>81</td>
<td>850</td>
<td>87%</td>
<td>low</td>
<td>−</td>
<td>NT</td>
<td>—</td>
<td>G</td>
</tr>
<tr>
<td>83</td>
<td>600</td>
<td>91%</td>
<td>low</td>
<td>−</td>
<td>+</td>
<td>E</td>
<td>—</td>
</tr>
<tr>
<td>104</td>
<td>600</td>
<td>low</td>
<td>low</td>
<td>+</td>
<td>+/−</td>
<td>—</td>
<td>—</td>
</tr>
<tr>
<td>113</td>
<td>1000</td>
<td>low</td>
<td>low</td>
<td>—</td>
<td>—</td>
<td>—</td>
<td>—</td>
</tr>
<tr>
<td>117</td>
<td>550</td>
<td>low</td>
<td>low</td>
<td>NT</td>
<td>NT</td>
<td>—</td>
<td>—</td>
</tr>
<tr>
<td>118</td>
<td>600</td>
<td>84%</td>
<td>low</td>
<td>NT</td>
<td>NT</td>
<td>—</td>
<td>E, G</td>
</tr>
<tr>
<td>120</td>
<td>585</td>
<td>low</td>
<td>low</td>
<td>NT</td>
<td>NT</td>
<td>—</td>
<td>—</td>
</tr>
<tr>
<td>127</td>
<td>650</td>
<td>high</td>
<td>high</td>
<td>NT</td>
<td>NT</td>
<td>—</td>
<td>—</td>
</tr>
<tr>
<td>128</td>
<td>800</td>
<td>low</td>
<td>low</td>
<td>+</td>
<td>+/−</td>
<td>—</td>
<td>—</td>
</tr>
<tr>
<td>129</td>
<td>450</td>
<td>87%</td>
<td>low</td>
<td>+</td>
<td>NT</td>
<td>—</td>
<td>G</td>
</tr>
<tr>
<td>302</td>
<td>600</td>
<td>74%</td>
<td>low</td>
<td>—</td>
<td>—</td>
<td>—</td>
<td>—</td>
</tr>
<tr>
<td>320</td>
<td>450</td>
<td>85%</td>
<td>low</td>
<td>+</td>
<td>—</td>
<td>—</td>
<td>G</td>
</tr>
<tr>
<td>322</td>
<td>500</td>
<td>53%</td>
<td>high</td>
<td>NT</td>
<td>NT</td>
<td>—</td>
<td>βsat</td>
</tr>
<tr>
<td>328</td>
<td>800</td>
<td>90%</td>
<td>low</td>
<td>—</td>
<td>—</td>
<td>—</td>
<td>(CA)n</td>
</tr>
<tr>
<td>331</td>
<td>400</td>
<td>55%</td>
<td>low</td>
<td>NT</td>
<td>NT</td>
<td>—</td>
<td>—</td>
</tr>
<tr>
<td>332</td>
<td>550</td>
<td>51%</td>
<td>low</td>
<td>NT</td>
<td>NT</td>
<td>—</td>
<td>—</td>
</tr>
<tr>
<td>340</td>
<td>700</td>
<td>84%</td>
<td>low</td>
<td>+</td>
<td>—</td>
<td>—</td>
<td>G</td>
</tr>
</tbody>
</table>

*Percentage of homology of the LCD clone to the genomic sequence.
†Low copy (4–10 loci) or high copy hybridization pattern on digested total human DNA.
‡Hybridizations to chromosome 4–specific DNA (HHW416, GM11687, Y25C2E).
§Sequence analysis by GRAIL.
§§More than 80% homology over more than 100 bp to sequences in EMBL or Genbank data bases.
‖Not tested.

Availability of cDNA clones without screening cDNA libraries. However, the method is dependent on selecting the appropriate tissue expressing the gene of interest at a suitable level.

Using cDNA selection, a cosmid contig spanning 100 kb of the FSHD gene region was screened with cDNA from the frontal cortex, fetal brain, liver, lymphoblast, fibroblast, bone marrow, intestine, or fetal liver. Some 150 LCD clones were isolated, which were distributed across all four overlapping cosmids. All clones were evaluated by hybridization to total human DNA, YAC DNA, cosmide DNA, and somatic cell hybrid DNA containing chromosome 4 only. Further, evolutionary conservation of most clones was investigated, and sequences were tested for homology using sequence data bases (Table 1). Strikingly, none of the putative cDNA clones was single copy. A total of 26 clones showed a low copy hybridization pattern, and 9 of those were evolutionary conserved. The most promising LCD clones mapped to the 2.2-kb EcoRl fragment, 22 kb upstream of the 3.3-kb repeated unit. However, independent evidence for transcribed sequences originating from this fragment was not obtained.

The difficulty of isolating cDNA clones derived from 4qter seems to parallel the observation that the FSHD gene region shows homology to several other chromosomal regions. For instance, in FISH experiments, cosmids from this region show cross-hybridization to several different chromosomes like 1q12, 10q26, 13p12, 14p12, 15p12, 21p12, and 22p12. Likewise, probe p13E-11 recognizes several loci. Recently, one of these loci was...
mapped to 10qter (Bakker et al., this issue). These observations could imply that any gene in this region is a member of a gene family, which is dispersed throughout the genome. For this reason, genomic sequences were generated to efficiently check the authenticity of the LCD clones. Although 16 of 22 LCD clones showed a homology of more than 80% to the genomic sequence, we have not been able to isolate clones from cDNA libraries which map to the FSHD region using these LCD clones as probes. However, since it is not known where, when, and how the putative FSHD gene is expressed, the region under investigation cannot be unambiguously excluded to harbor the FSHD gene. More sensitive mRNA detection or expression independent techniques, such as RT-PCR and exon trapping respectively, may have to be applied to detect rare or transiently expressed transcripts.

An alternative explanation for our results is that the cDNA selection was biased yielding preferably low copy repetitive sequences. To investigate whether the absence of single copy probes was due to an inherent property of the cosmid contig, CY34 was subcloned as random SauIII A fragments. Thirty clones were tested, showing similar properties as the LCD clones. This indicates that the cosmid itself is indeed devoid of real single copy sequences.

Other investigators focused on the 3.3-kb repeated unit. The DNA rearrangements, which are due to deletions of an integral number of this repeated unit, suggested that the FSHD gene could be disrupted through a structural truncation. Despite considerable efforts, no transcriptional activity has been found in these repeats. However, the structure and vicinity to the telomere of this region opens the possibility that the dele-

---

**FIGURE 2.** Location of LCD clones within the 2.2-kb EcoRI fragment. Sequence analysis by GRAIL revealed three exons, indicated by open boxes (E: excellent exon; G: good exon). CD25, isolated from a placenta cDNA library, is indicated by the dashed line. The Généthon cDNA clone HSBB8H051 (open box) shows a 96% sequence homology to the distal part of LCD 60; another cDNA clone of Généthon, HSBB11B021, shows a 94% homology to the proximal part of HSBB8H051. The arrows indicate the primers used to generate clone B021 (235 bp).

**FIGURE 3.** Hybridization pattern of the placenta cDNA clone CD25 on a "Zoo" blot containing EcoRI-digested DNA of human, monkey, cow, and mouse.
FIGURE 4. (a) Hybridization pattern of clone B021 to a Northern blot containing total RNA as well as poly A+ DNA (indicated by "+ "). A 9-kb transcript was detected in the lane with poly A+ RNA isolated from the shoulder muscles of a Rhesus monkey. (b) Hybridization of B021 to digested human, YAC (c) or somatic cell hybrid HHW416 DNA. B021 detects a single fragment in human DNA but does not hybridize to chromosome 4-specific DNA.

sitions have an effect on the transcription, rather than on the structure of the FSHD gene. It is possible that such a position effect may act over a longer distance than was anticipated in this study. The availability of a much larger cosmid/YAC contig, than was used so far, allows extending the gene search to a region of several megabases in the centromeric direction.

REFERENCES


