ARTICLE

Specific Sequence Variations within the 4q35 Region Are Associated with Facioscapulohumeral Muscular Dystrophy

Richard J. L. F. Lemmers, Mariëlle Wohlgemuth, Kristiaan J. van der Gaag, Patrick J. van der Vliet, Corrie M. M. van Teijlingen, Peter de Knijff, George W. Padberg, Rune R. Frants, and Silvère M. van der Maarel

Autosomal dominant facioscapulohumeral muscular dystrophy (FSHD) is mainly characterized by progressive wasting and weakness of the facial, shoulder, and upper-arm muscles. FSHD is caused by contraction of the macrosatellite repeat D4Z4 on chromosome 4q35. The D4Z4 repeat is very polymorphic in length, and D4Z4 rearrangements occur almost exclusively via intrachromosomal gene conversions. Several disease mechanisms have been proposed, but none of these models can comprehensively explain FSHD, because repeat contraction alone is not sufficient to cause disease. Almost-identical D4Z4-repeat arrays have been identified on chromosome 10q26 and on two equally common chromosome 4 variants, 4qA and 4qB. Yet only repeat contractions of D4Z4 on chromosome 4qA cause FSHD; contractions on the other chromosomes are nonpathogenic. We hypothesized that allele-specific sequence differences among 4qA, 4qB, and 10q alleles underlie the 4qA specificity of FSHD. Sequence variations between these alleles have been described before, but the extent and significance of these variations proximal to, within, and distal to D4Z4 have not been studied in detail. We examined additional sequence variations in the FSHD locus, including a relatively stable simple sequence-length polymorphism proximal to D4Z4, a single-nucleotide polymorphism (SNP) within D4Z4, and the A/B variation distal to D4Z4. On the basis of these polymorphisms, we demonstrate that the subtelomeric domain of chromosome 4q can be subdivided into nine distinct haplotypes, of which three carry the distal 4qA variation. Interestingly, we show that repeat contractions in two of the nine haplotypes, one of which is a 4qA haplotype, are not associated with FSHD. We also show that each of these haplotypes has its unique sequence signature, and we propose that specific SNPs in the disease haplotype are essential for the development of FSHD.

Facioscapulohumeral muscular dystrophy (FSHD [MIM 158900]) is the third-most-common myopathy with an autosomal dominant pattern of inheritance. The disease is mainly characterized by progressive wasting and weakness of the facial, shoulder, and upper-arm muscles. FSHD displays wide clinical variability, ranging from asymptomatic individuals with minimal clinical signs apparent only on careful examination to patients who use wheelchairs.[1]

FSHD is caused by a contraction of the macrosatellite repeat D4Z4 in the subtelomeric region of chromosome 4q35.[2] The D4Z4 repeat array, which consists of single repeat units of 3.3 kb, is very polymorphic in length and is highly recombinogenic.[3] Control individuals carry D4Z4 repeats with 11–100 units, whereas the majority of patients have a D4Z4-repeat size of 1–10 units on one of their chromosomes 4. In 10%–30% of the cases, FSHD is caused by a de novo contraction of the D4Z4 repeat. Intrachromosomal gene conversions underlie these rearrangements.[4]

Detailed sequence analysis of the D4Z4 unit revealed the presence of a putative ORF in addition to repetitive sequences often found in heterochromatic domains of the human genome.[5,6] This intronless ORF, named "*DUX4*," encodes a 424-aa protein with two predicted homeobox sequences.[7] With consideration of the gradual spread in muscle weakness and the crucial role of homeobox proteins in body patterning during early development, *DUX4* is considered a strong candidate for FSHD. However, consistent evidence of expression emanating from D4Z4 is lacking.[7–9]

Therefore, other models were proposed in which D4Z4 plays an indirect role in FSHD pathogenesis.[10–13] All these models predict a change in chromatin structure on D4Z4 contraction. Several observations corroborate these models: a protein repressor complex has been identified that binds D4Z4 and suppresses the expression of genes in close vicinity of the repeat.[10] Moreover, D4Z4 is significantly hypomethylated in FSHD alleles.[13] Unfortunately, none of these models can comprehensively explain FSHD, because certain conditions, in addition to repeat contraction, need to be met for disease to occur.

First, the D4Z4 contraction needs to occur on chromosome 4q, since a very homologous and equally polymorphic D4Z4 repeat is located on chromosome 10q26, but FSHD-sized D4Z4 repeats on 10q have never been associated with disease.[14,15] Sequence analysis showed that the 10q26 repeat is ~98% identical to the 4q35 repeat, and the regions extending 40 kb proximal to D4Z4 and at least

10 kb distal to D4Z4 are highly homologous.[16] Specific sequence differences assist the molecular diagnosis of FSHD by discriminating between repeats from both chromosomes through the presence of chromosome-specific restriction-recognition sites within the repeat units: *Bln*I for 10q-derived repeat units and *Xap*I for 4q-derived repeat units.[15,17]

Second, the contraction needs to occur on a specific variant of 4qter. Two subtelomeric variations distal to D4Z4 were identified on chromosome 4, with alleles 4qA and 4qB.[16] The most prominent difference between 4qA and 4qB alleles is the presence of a 260-bp sequence (pLAM) followed by a 6.2-kb *β*-satellite repeat directly distal to D4Z4 on 4qA. Both variants are almost equally common in the control population. However, we have shown that FSHD alleles are exclusively associated with the 4qA variant.[18]

We hypothesized that allele-specific sequence differences are accumulating between 4qA and 4qB alleles because of a lack of allelic exchanges between 4qA and 4qB and that some of these differences underlie the 4qA specificity of FSHD. A first indication that 4qA and 4qB subtelomeres evolve relatively independently came from the observation that the distal A/B variation on 4qter was in linkage disequilibrium (LD) with a G/C SNP within the most proximal D4Z4 unit (hereafter designated as the "D4Z4 SNP").[4] We showed the presence of the C variant, which leads to an extra *Pvu*II restriction site in D4Z4, in 29% of the 4qB alleles, whereas this restriction site was almost absent in the 4qA alleles. This observation motivated us to analyze the allelic variation of the D4Z4 region on chromosomes 4 and 10 in more detail.

For this study, we used four polymorphisms in 4qter: the repeat-size variation of D4Z4, the D4Z4 SNP, the distal A/B variation, and a relatively stable simple sequence-length polymorphism (SSLP) located 3.5 kb proximal to D4Z4. We confirmed the LD between the region proximal and distal to D4Z4. On the basis of specific sequence variations within and close to D4Z4, the subtelomeric region of chromosome 4q can be subdivided into at least nine haplotypes. Most importantly, we show that D4Z4 contractions in one of the three 4qA haplotypes are not associated with FSHD.

Subjects and Methods
Control Individuals and Patients with FSHD

The polymorphic markers were determined in genomic DNA from 222 unrelated healthy individuals and 86 independent individuals with sporadic and familial FSHD. All control and patient DNA samples tested were randomly chosen from our collection of >1,000 individuals. Blood from all individuals was collected after informed consent was obtained.

Somatic Cell Hybrids and DNA Clones

The chromosome 4qA sources were the monochromosomal rodent somatic cell hybrids HHW1494 and SU10 (gift from S. Winokur, Irvine, CA) and phage clones λ42, λ68, and λ260201.[3]

All these sources represent FSHD alleles. As chromosome 4qB sources, we used the monochromosomal rodent somatic cell hybrids GM11687 (Coriell Institute for Medical Research, Camden, NJ), 4L 10 (gift from E. Stanbridge, Irvine, CA), and HHW416 (gift from M. Altherr, Los Alamos, NM). Chromosome 10qA sources were cosmid C85[3] and the monochromosomal rodent somatic cell hybrids 726 8a (U.K. Human Genome Mapping Project Resource Center) and GM11688 (Coriell Cell Repositories).

FSHD-Affected Families Rf10 and Rf204

Families Rf10 and Rf204 were ascertained via one of the Dutch Neuromuscular Centers. In Rf10, four affected members who all carry a 33-kb FSHD allele (4qA161; see the "Results" section) were identified. In addition, a healthy spouse (II-9) who carried a different 33-kb allele (4qA166) was identified. Also, his brother (II-11) carried this FSHD-sized 4qA166 allele and was not affected. Family Rf204 has been described before (as "family A")[19] and includes five patients with FSHD who carry a 17-kb FSHD allele (4qA161). In this family, a second FSHD-sized allele of 24 kb (4qA166) was found in an unaffected sister (II-11). Further analysis of this family revealed five unaffected children of the oldest, deceased brother (II-1), who carried the same 24-kb allele. All unaffected carriers of the FSHD-sized 4qA166 allele in both families have reached the age range at which the disease usually becomes manifest.

DNA Isolation

DNA was isolated from peripheral-blood lymphocytes (PBLs). PBLs were embedded in agarose plugs (InCert agarose [FMC]) at a concentration of 5×10^5 cells per plug and were treated with 600 $\mu g/\mu l$ pronase and 1% Sarkosyl for 40–48 h at 37°C. Next, plugs were washed in Tris-EDTA (TE)$^{-4}$ and were stored in 0.5 M EDTA at 4°C. Before they were used, plugs were successively equilibrated in TE^{-4} and the appropriate restriction-enzyme buffer.

Repeat Length and Distal Variation

For D4Z4 array sizing, DNA samples were double digested with *Eco*RI and *Hin*dIII or *Eco*RI and *Bln*I or with *Xap*I only. To determine the distal A/B variation at 4qter, DNA was digested with *Hin*dIII.[18] All digestions were performed according to the manufacturer's instructions. *Eco*RI, *Hin*dIII, and *Xap*I were purchased from MBI Fermentas, and *Bln*I was purchased from GE Healthcare. DNA was separated in a 22-h run on a 0.85% agarose gel (MP agarose [Roche]) by pulsed-field gel electrophoresis (PFGE) at 8.5 V/cm in four identical cycles, with a switch time increasing linearly from 1 s at the start to 20 s at the end of each cycle. The run was performed in 0.5 × Tris/borate/EDTA supplemented with 150 ng/ml ethidium bromide at 23°C.

Restriction Analysis of Proximal D4Z4 Unit

The D4Z4 SNP in the first (proximal) unit of the D4Z4 repeat (AF117653:g.6045G→C) was analyzed in genomic DNA samples by a double digestion with *Pvu*II and *Bln*I. *Pvu*II and *Bln*I were purchased from GE Healthcare. The Southern-blot analysis of the *Pvu*II polymorphism with probe p13E-11 (D4F104S1)[2] reveals chromosome 4–derived fragments of 2,849 bp (C variant, *Pvu*II+) or 4,559 bp (G variant, *Pvu*II−) in size, depending on the presence or absence of the *Pvu*II restriction site, whereas chromosome 10–derived fragments are 2,464 bp because of the presence of a *Bln*I restriction site.[4]
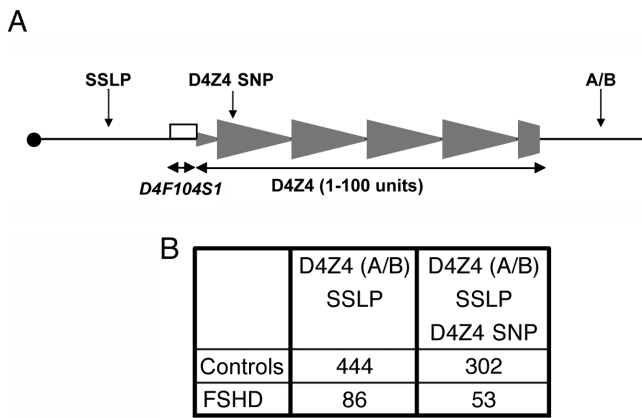
A

B

	D4Z4 (A/B) SSLP	D4Z4 (A/B) SSLP D4Z4 SNP
Controls	444	302
FSHD	86	53

Figure 1. *A,* Schematic representation of the D4Z4 repeat on chromosomes 4q35 and 10q26 with the localization of the four polymorphic markers used in this study. The SSLP marker is localized 3.5 kb proximal to D4Z4. The D4Z4 SNP, the D4Z4-repeat size variation, and the distal A/B variation have been described elsewhere.[4,18] The 475-bp *D4F104S1* sequence is localized immediately proximal to D4Z4. *B,* Overview of the total number of control and patient alleles that were tested for these polymorphisms.

Blotting and Hybridization

After digestion and gel electrophoresis, DNA was transferred to a Hybond XL membrane (GE Healthcare) by Southern blotting. Hybridization was performed in a buffer containing 0.125 M Na_2HPO_4 (pH 7.2), 10% PEG6000, 0.25 M NaCl, 1 mM EDTA, and 7% SDS for 16–24 h at 65°C.

Membranes for D4Z4 sizing were hybridized with probe p13E-11 and were washed in 2× saline sodium citrate (SSC) and 0.1% SDS. For A/B typing, membranes were sequentially hybridized with probes 4qA or 4qB and were washed in either 1×SSC and 0.1% SDS (4qA) or 0.3×SSC and 0.1% SDS (4qB). The linear gels for the *Pvu*II polymorphisms in the most proximal D4Z4 unit were hybridized with p13E-11 and were washed in 0.3×SSC and 0.1% SDS. All membranes were exposed for 16–48 h to phosphorimager screens and were analyzed with the Image Quant software program (Molecular Dynamics).

Genotyping of the SSLP

The SSLP proximal to D4Z4 is localized between positions 1532 and 1694 of AF117653 and was studied by PCR with the use of forward primer 5′-GGTGGAGTTCTGGTTTCAGC-3′ and reverse primer 5′-CCTGTGCTTCAGAGGCATTTG-3′. For fragment analysis, the forward primer was labeled with HEX. SSLP fragments were amplified using standard PCR methods, and size differences were determined with the use of an ABI Prism 3100 Genetic Analyzer. Primers were designed using Primer3 software.

Determination of SSLP and D4F104S1 *Sequences*

The majority of the SSLP and *D4F104S1* sequences were determined in the somatic cell hybrids and DNA clones described above. For the analysis of the SSLP sequence in other 4qA161 and 4qB163 alleles, so-called monosomic individuals were analyzed. Monosomic individuals carry a translocated chromosome 10–derived D4Z4 repeat on one of their chromosomes 4.[20] Se-

quence analysis of the monoallelic DNA sources showed that 10q-derived alleles include an 8-nt insertion in their SSLP sequence that is not present on 4qA161 and 4qB163 alleles. Consequently, DNA digestion in monosomic individuals who carry the 4qA161 or 4qB163 haplotype on their normal chromosome 4, with use of the 8-nt insertion-specific restriction enzyme *Fsp*BI, allows a discriminative PCR covering each of these SSLP sequences.

The D4Z4 SNP leading to a polymorphic *Pvu*II site within D4Z4 was used for the analysis of the SSLP and *D4F104S1* sequence of 4qB162, 4qB166, and 4qB168 alleles. This polymorphic *Pvu*II site is present in alleles encompassing the G variant (4qB162, 4qB164, 4qB166, and 4qB168 alleles) and is virtually absent in all other haplotypes. *Pvu*II digestion of the alleles carrying the G variant results in a 4,764-bp fragment containing this SSLP sequence, whereas alleles belonging to other haplotypes give rise to restriction fragments of 6,475 bp in size. Therefore, *Pvu*II-digested chromosomal DNA from individuals who carry a single allele with the G variant allows the isolation of a gel slice of ∼4,764 bp containing the SSLP and *D4F104S1* region. Purified DNA from this gel slice was used in a PCR specific to both regions.

Finally, the SSLP and *D4F104S1* sequences of 4qA166 alleles and the *D4F104S1* sequences of additional 4qA161 alleles were derived from gel slices containing *Eco*RI-digested chromosomal DNA fragments of appropriate size after PFGE analysis. The position of the D4Z4 repeat in the gel was determined using an appropriate molecular size standard (MidRange I PFG Marker [New England Biolabs]), after which the fragments were sliced out of the gel and the DNA was isolated.

For both *Pvu*II and *Eco*RI sliced agarose fragments, the DNA was extracted with a gel extraction kit (Machery Nagel Gel Extraction NS Extract II) that allows the extraction of DNA fragments <50 kb. The SSLP sequence was determined with the primers disclosed in the "Genotyping of the SSLP" section. The *D4F104S1* sequence was analyzed using forward primer 5′-CCCAGTTACTGTTCTG-GGTGA-3′ and reverse primer 5′-GAAAGCCCCCTGTGGGAG-3′. Primers were designed using Primer3 software.

Statistical Analyses

Differences among major haplogroups with respect to the D4Z4 SNP were tested by means of a nonparametric Wilcoxon-Mann-Whitney test with Monte-Carlo simulation (10,000 randomizations). To obtain a global estimate of D4Z4-repeat number differences among the haplotypes, we used a Kruskall-Wallis test with Monte-Carlo simulation (10,000 randomizations). Both tests were performed using StatXact 4 from Cytel Software. Descriptive statistics (mean, median, and counts) were obtained using the descriptive data–analysis tool included in Microsoft Excel.

Results
Allelic Subdivision of 4q Subtelomere

We demonstrated elsewhere that mitotic D4Z4 rearrangements occur at a much higher frequency intrachromosomally.[4] Furthermore, the apparent suppression of recombination between 4qA and 4qB type alleles was confirmed by the different distribution of a SNP within the proximal D4Z4-repeat unit on 4qA and 4qB chromosome ends.[4] In this study, the sequence variation within the FSHD locus was further analyzed using four different polymorphisms (fig. 1*A*). Proximal to D4Z4, we identified a novel, relatively stable SSLP
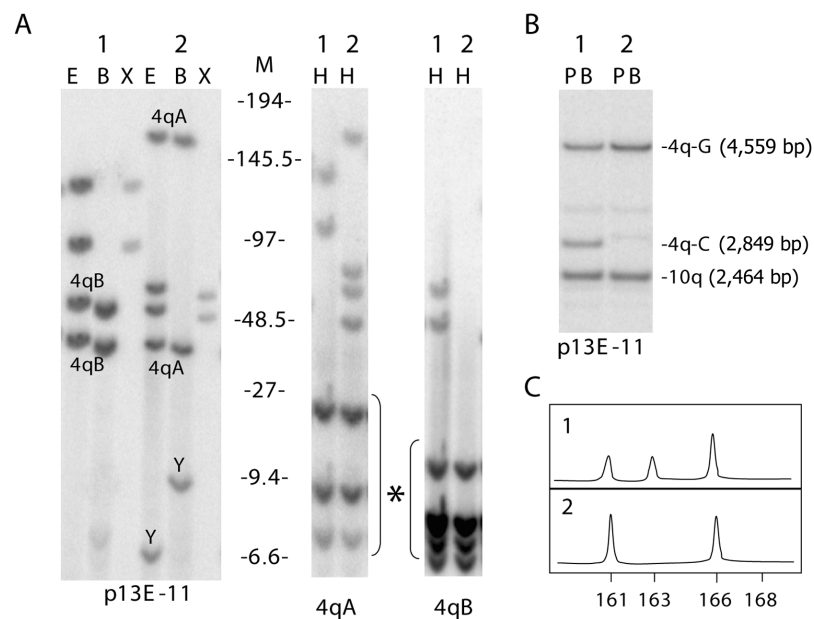
Figure 2. Examples of D4Z4 genotyping, including the three polymorphic markers in control individual (1) and FSHD-affected patient (2). *A,* D4Z4 sizing and A/B typing after PFGE and Southern blotting. For D4Z4 sizing, DNA was digested with restriction enzymes *Eco*RI and *Hin*dIII (*lane E*), *Eco*RI and *Bln*I (*lane B*), or *Xap*I (*lane X*), and the blotted DNA was hybridized with probe p13E-11 located immediately proximal to D4Z4 (*left panel*). The four fragments present in lane E represent two 4q- and two 10q-derived repeat arrays with different repeat sizes. The 4q or 10q origin of the D4Z4 repeats can be determined by digestions with *Bln*I or *Xap*I, which fragmentize 10q- or 4q-derived repeat arrays, respectively. Consequently, individual 1 carries 4q D4Z4 repeats of 40 kb and 55 kb and 10q repeats of 95 kb and 130 kb, whereas individual 2 carries 4q alleles of 35 kb and 165 kb and 10q repeats of 50 kb and 65 kb. For A/B typing, DNA was digested with *Hin*dIII (*lane H*) and was hybridized serially with probes 4qA and 4qB, located immediately distal to D4Z4 (*right panel*). Both 4q alleles (40 kb and 55 kb) in individual 1 carry the B variant, and both 4q alleles (35 kb and 165 kb) in individual 2 carry the A variant. As was shown elsewhere, 10q alleles almost always carry the A variant.[18] The molecular weight marker (M) in kilobases is shown in the middle. Cross-hybridizing Y-chromosome–derived fragments with probe p13E-11 in individual 2 (male) are indicated with "Y." Non-4q and -10q cross-hybridizing fragments with probes 4qA and 4qB are marked with an asterisk (*). *B,* Analysis of the G/C SNP within D4Z4 (for description, see the work of Lemmers et al.[4]). Double digestion of genomic DNA with *Pvu*II and *Bln*I (PB) reveals 10q fragments of 2,464 bp upon hybridization with p13E-11. 4q alleles reveal fragments of either 4,559 bp or 2,849 bp, depending on the presence (C variant) or absence (G variant) of the *Pvu*II restriction site in D4Z4, respectively. One 4qB allele in individual 1 contains the C variant, whereas all other 4q alleles in both individuals carry the G variant. *C,* Example of the fragment run of the SSLP sequence. Altogether, this analysis demonstrates that individual 1 carries a 4qB161 and a 4qB163 allele next to the two 10qA166 alleles and that individual 2 carries two 4qA161 alleles and two 10qA166 alleles.

that was shown to be moderately polymorphic on 4q and 10q alleles. In the D4Z4 repeat, the D4Z4 SNP within the most proximal unit was analyzed. Furthermore, the length of the repeat array was determined in addition to the distal A/B variation. All these analyses were done using Southern blotting (for D4Z4 length, A/B variation, and D4Z4 SNP) or PCR (for SSLP variation). In total, 222 independent control individuals and 86 independent patients with FSHD were analyzed for the D4Z4-repeat size, the A/B polymorphism, and the SSLP variation. Of these individuals, 151 controls and 53 patients with FSHD were also analyzed for the D4Z4 SNP (fig. 1*B*). An example of the genotyping of 4qter and 10qter alleles in two unrelated individuals is depicted in figure 2. Figure 2*A* shows the analysis of the D4Z4 size variation and determination of the A/B polymorphism by PFGE. Figure 2*B* shows Southern-blot analysis of the D4Z4 SNP, and figure 2*C* shows the results of the SSLP analysis by PCR in the same individuals.

The size of the SSLP sequence was next established in six independent 10q sequences from GenBank (accession numbers AL845259, AY028079, BX649463, BX294170, BX005259, and AL954635) and three monoallelic DNA sources (M&M), and they were found to be 165 bp without sequence variation (fig. 3). Further analysis showed that the peak at 166 bp in the SSLP fragment run represents the 165-bp 10q fragment and that the running pattern is disturbed by a specific 8-nt sequence within this SSLP (see below).

After analysis of all alleles with the above-described markers, all alleles from the control individuals were categorized and counted. An overview of the different 4q and 10q haplotypes is depicted in figure 4*B*. On the basis of the proximal SSLP, 4qA alleles (*n* = 200) are able to be subdivided into three haplotypes (4qA161, 4qA163, and 4qA166). Of these haplotypes, 4qA161 is the most prevalent (86%). All but one 4qA allele analyzed for the D4Z4

4q sources

	Haplotype	SSLP sequence				D4Z4 SNP	D4Z4	4qA/4qB
		CA_n-MA-CA_n	G/T	CT_n	8 nt	(G/C)		
λ42	4qA161	CA_{10}-AA-CA_{10}	G	CT_6	-	G	15	A
λ68	4qA161	CA_{10}-AA-CA_{10}	G	CT_6	-	G	15	A
λ260201	4qA161	CA_{10}-AA-CA_{10}	G	CT_6	-	G	17	A
SU10	4qA161	CA_{10}-AA-CA_{10}	G	CT_6	-	G	10	A
HHW1494	4qA161	CA_{10}-AA-CA_{10}	G	CT_6	-	G	20	A
4L-10	4qB163	CA_{10}-AA-CA_{11}	G	CT_6	-	G	85	B
HHW416	4qB163	CA_{10}-AA-CA_{11}	G	CT_6	-	G	96	B
GM11687	4qB168	CA_{19}	T	CT_7	+	C	140	B

10q sources

	Haplotype	SSLP sequence				D4Z4 SNP	D4Z4	4qA/4qB
		CA_n-MA-CA_n	G/T	CT_n	8 nt	(G/C)		
GM11688	10qA166	CA_9-AA-CA_8	T	CT_7	+	G	60	A
762-8A	10qA166	CA_9-AA-CA_8	T	CT_7	+	G	120	A
C85	10qA166	CA_9-AA-CA_8	T	CT_7	+	G	27	A
AL845259	10qA166	CA_9-AA-CA_8	T	CT_7	+	G	-	-
AL954635	10qA166	CA_9-AA-CA_8	T	CT_7	+	G	-	-
AY028079	10qA166	CA_9-AA-CA_8	T	CT_7	+	G	-	-
BX649463	10qA166	CA_9-AA-CA_8	T	CT_7	+	G	-	-
BX294170	10qA166	CA_9-AA-CA_8	T	CT_7	+	G	-	-
BX005259	10qA166	CA_9-AA-CA_8	T	CT_7	+	G	-	-

Figure 3. Haplotype analyses of individual 4q and 10q DNA sources, of which the last six 10q sequences are obtained from GenBank. The identity of the DNA source is shown in the first column. The haplotype and the SSLP sequence are listed in columns 2 and 3, whereas the D4Z4 SNP is shown in column 4. The last two columns display the D4Z4-repeat length (in kb) and the distal variation.

SNP (127 of 128) carry the G variant of the D4Z4 SNP. 4qB alleles ($n = 244$) are more polymorphic for the proximal SSLP and are able to be subdivided into six haplotypes: 4qB161 (2%), 4qB162 (5%), 4qB163 (68%), 4qB164 (1%), 4qB166 (3%), and 4qB168 (21%). The 4qB163 haplotype represents the most common haplotype, and the majority of the alleles within this haplotype (107 of 108) have the G variant of the D4Z4 SNP. In the small number of 4qB161 alleles ($n = 4$), we found only the G variant of the D4Z4 SNP. By contrast, almost all alleles belonging to other 4qB haplotypes (61 of 62) carry the C variant of the D4Z4 SNP. There is a significant difference in distribution of the D4Z4 SNP between 4qA and 4qB haplotypes ($P < .0001$) and between 4qB163 and 4qB162, 4qB166, and 4qB168 haplotypes ($P < .0001$). 10q alleles seemed more homogeneous than 4q alleles and were subdivided into the haplotypes 10qA166 (96%) and 10qA164 (4%). All GenBank 10qA166 sequences mentioned above and monoallelic 10qA166 sources show the G variant in D4Z4 (fig. 3).

The almost invariable presence of the 166-bp SSLP on chromosome 10 was concluded from segregation analysis in a subset of families with FSHD ($n = 52$ segregations) and from the observation that, on SSLP analysis, individuals homozygous for 4qA ($n = 41$) invariably presented with a peak of double intensity at 166 bp (10qA166), in addition to peaks of single or double intensity at 161 bp

(4qA161) or 163 bp (4qA163) or increased intensity of the peak at 166 bp (4qA166).

FSHD Alleles and Individual 4q Sources

Since contractions of the D4Z4 repeat on 4qA alleles are associated with FSHD, carriers of a pathogenic D4Z4 repeat were also studied for these four polymorphisms. In total, 86 independent FSHD alleles were analyzed, and all were shown to belong to the 4qA161 haplotype, on the basis of the SSLP and the distal A/B variation. Of these alleles, 53 were analyzed for the D4Z4 SNP, and all carried the G variant.

In addition, eight monosomic 4q sources were completely genotyped (fig. 3). Five 4qA sources that represent pathogenic alleles from patients with FSHD display the characteristics of the 4qA161 haplotype. Furthermore, three monochromosomal rodent somatic cell hybrids that carry 4qB alleles were analyzed and were shown to belong to the 4qB163 and 4qB168 haplotypes.

Nonpathogenic 4q Haplotypes

It was surprising that all 86 FSHD alleles analyzed belong to the 4qA161 haplotype (fig. 4C). No FSHD alleles that belong to the 4qA166 haplotype were found, even though they represent 11% of the total 4qA alleles in control in-
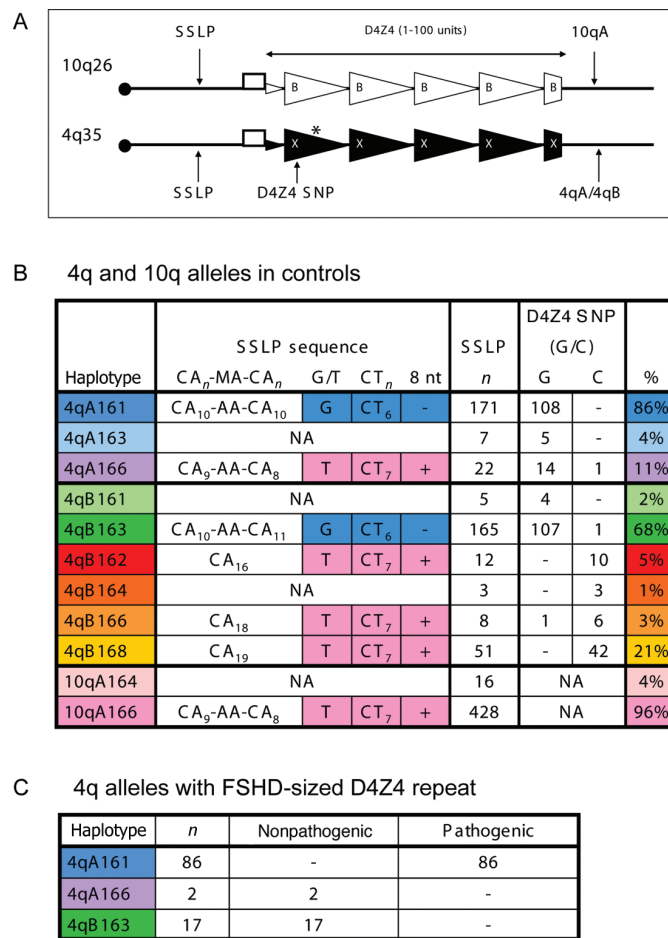
A

B 4q and 10q alleles in controls

Haplotype	SSLP sequence CA$_n$-MA-CA$_n$	G/T	CT$_n$	8 nt	SSLP n	D4Z4 SNP (G/C) G	C	%
4qA161	CA$_{10}$-AA-CA$_{10}$	G	CT$_6$	-	171	108	-	86%
4qA163	NA				7	5	-	4%
4qA166	CA$_9$-AA-CA$_8$	T	CT$_7$	+	22	14	1	11%
4qB161	NA				5	4	-	2%
4qB163	CA$_{10}$-AA-CA$_{11}$	G	CT$_6$	-	165	107	1	68%
4qB162	CA$_{16}$	T	CT$_7$	+	12	-	10	5%
4qB164	NA				3	-	3	1%
4qB166	CA$_{18}$	T	CT$_7$	+	8	1	6	3%
4qB168	CA$_{19}$	T	CT$_7$	+	51	-	42	21%
10qA164	NA				16	NA		4%
10qA166	CA$_9$-AA-CA$_8$	T	CT$_7$	+	428	NA		96%

C 4q alleles with FSHD-sized D4Z4 repeat

Haplotype	n	Nonpathogenic	Pathogenic
4qA161	86	-	86
4qA166	2	2	-
4qB163	17	17	-

Figure 4. Overview of the different haplotypes that were defined after complete genotyping of 4qA, 4qB, and 10q alleles. *A,* All D4Z4-repeat units on 10q alleles encompass the *Bln*I restriction site (B), which is absent in D4Z4 in all 4q alleles. Likewise, all D4Z4-repeat units on 4q alleles encompass the *Xap*I restriction site (X) that is absent in D4Z4-repeat units on 10q alleles, except for D4Z4 repeats in the 4qA166 haplotype, which lack the *Xap*I restriction site in their most proximal D4Z4 unit (indicated with an asterisk [*]). *B,* Genotyping of 4q and 10q alleles in control individuals. The first column lists the different haplotypes that have been identified, and the second shows the sequence of the SSLP in each haplotype (M = A or C). Those haplotypes for which we could not obtain SSLP sequence information are indicated by "NA" (not analyzed). The third and fifth columns represent, respectively, the number of alleles found in each haplotype (SSLP n) and the prevalence (%) of each haplotype among 4qA, 4qB, and 10q alleles. The distribution of the G and C variants in the D4Z4 SNP is shown in the fourth column (D4Z4 SNP (G/C)). In total, 200 4qA alleles were genotyped, and, on the basis of SSLP length, they were subdivided into three haplotypes: 4qA161, 4qA163, and 4qA166. 4qB alleles ($n = 244$) were subdivided into six haplotypes, designated as 4qB161, 4qB162, 4qB163, 4qB164, 4qB166, and 4qB168, and 10q alleles into haplotypes 10qA166 and 10qA164. There is a significant difference in D4Z4 SNP distribution between 4qA and 4qB haplotypes ($P < .0001$) and between 4qB163 and 4qB162, 4qB166, and 4qB168 haplotypes ($P < .0001$). *C,* Genotyping of alleles with an FSHD-sized D4Z4 repeat. Without exception, all FSHD alleles belong to the 4qA161 haplotype. All alleles with an FSHD-sized repeat that belong to the 4qA166 or 4qB163 haplotypes are nonpathogenic.

dividuals. However, we identified two patients from independent families with FSHD in whom we found evidence of the nonpathogenicity of D4Z4 contractions on 4qA166 alleles. Both of them carried 4qA166 alleles having D4Z4-repeat sizes in the pathogenic range, in addition to an FSHD-sized 4qA161 D4Z4 repeat. This latter allele we considered to be pathogenic, on the basis of its cosegregation with disease.

Family Rf10 is a 3-generation FSHD-affected family with patients who are moderately affected (fig. 5). The pre-vious molecular diagnosis of this family was very complex because the 33-kb (8 D4Z4 units) 4qA allele was detected not only in the affected family members but also in a healthy spouse (individual II-9). Our current study reveals that this healthy spouse carries a 33-kb 4qA166 allele, whereas a 33-kb 4qA161 allele was detected in all patients with FSHD in this family, indicating its causal relation to disease. Additional genotyping of the family of individual II-9 showed that his brother (II-11) also carries the 33-kb 4qA166 allele and is not affected with FSHD.
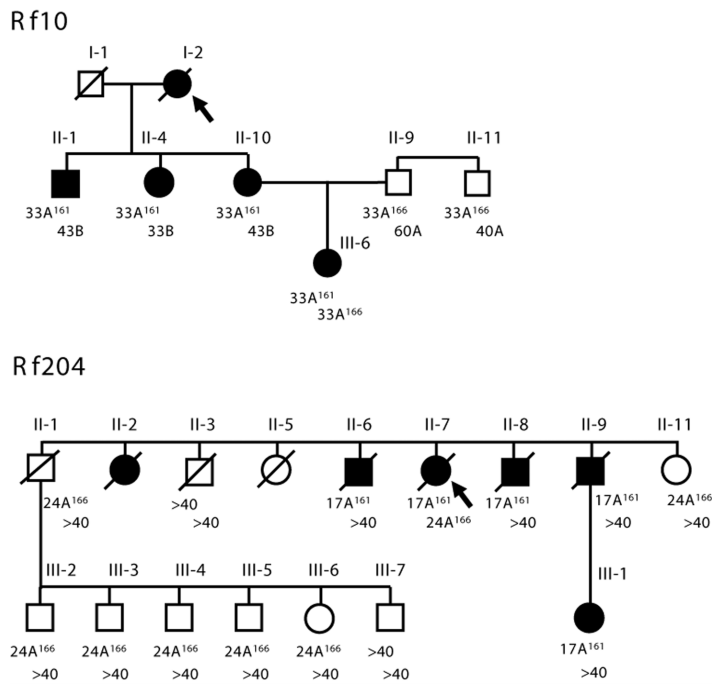
Figure 5. Pedigrees of FSHD-affected families Rf10 and Rf204. The patients with FSHD in family Rf10 carry a 33-kb 4qA161 (33A[161]) allele and are moderately affected. Individuals II-9 and II-11 carry a 33-kb 4qA166 (33A[166]) allele and do not display any clinical signs of muscular dystrophy. In family Rf204, all patients with FSHD carry a 17-kb 4qA161 (17A[161]) allele and are moderately to severely affected. Six individuals carry an FSHD-sized allele of 24 kb (24A[166]), which belongs to the 4qA166 haplotype, and they are not affected (II-11, III-2, III-3, III-4, III-5, and III-6). Index cases are indicated with an arrow.

In family Rf204, a 17-kb (3 D4Z4 units) 4qA161 allele is causing FSHD. The patients of this family are moderately to severely affected. A second FSHD-sized allele (24 kb, 5 D4Z4 units) in this family has been reported elsewhere[19] but was now shown to belong to the 4qA166 haplotype. Although D4Z4 repeats of 24 kb on 4qA alleles would be expected from prior information to cause FSHD, six individuals (II-11, III-2, III-3, III-4, III-5, and III-6) who carry this allele have no clinical signs of muscular dystrophy. Only one individual (II-7) who carries this allele developed FSHD, but she also carries the 17-kb pathogenic 4qA161 allele.

In addition to these nonpathogenic D4Z4 contractions on 4qA166 alleles, we further substantiated our previous observation that contractions on 4qB alleles do not cause FSHD.[21] At present, we identified 17 FSHD-sized D4Z4 repeats on 4qB alleles in independent individuals. In line with previous findings on 4 of these alleles, none of these 17 alleles were shown to be pathogenic. We determined the haplotype of 13 of these nonpathogenic alleles and showed that they all belong to the 4qB163 haplotype (fig. 4C).

Haplotype-Specific Sequences

To find further evidence of the accumulation of haplotype-specific sequence variations at 4qter, we determined the sequence of the SSLP and a 475-bp fragment in *D4F104S1* in the different haplotypes. To this end, the SSLP and *D4S104F1* fragment were PCR amplified and were sequenced in at least three independent sources of the most common 4q and 10q haplotypes. As shown in figure 6, each haplotype analyzed is defined by a unique combination of 15 SNPs within the *D4F104S1* fragment.

The SSLP was shown to consist of a number of short, rather stable microsatellite repeats (CA and CT). In addition, the SSLP contains an A/C SNP, a G/T SNP, and a polymorphic insertion of 8 nt (fig. 4B). On the basis of these results, the 10qA166 haplotype can be distinguished from the most abundant 4q haplotypes (4qA161 and 4qB163) by the presence of SNP variant T, a CT stretch of 7 instead of 6 dinucleotides, and the presence of the 8-nt insertion.

On the basis of the SSLP sequence, most haplotypes can now be distinguished from each other. Similar to the 10qA166 SSLP sequence, other alleles that carry the 8-nt insertion (4qA166, 4qB162, 4qB166, and 4qB168) were also shown to be 1 nt smaller on sequencing than in the SSLP fragment run.

In retrospect, we carefully reanalyzed D4Z4 sizing by PFGE of 4qA166 alleles with restriction enzymes *Eco*RI/*Hin*dIII, *Eco*RI/*Bln*I, and *Xap*I. In total, we analyzed 15 different 4qA166 alleles and observed that, as expected, all D4Z4 units lack the *Bln*I restriction site within their repeats, as is characteristic of 4q repeats (data not shown). Remarkably, all these alleles also lack the *Xap*I restriction

```
4qA161    AGCCAGACTTCATCAAATTTCTGCACCAATGAAAAAAAAATTTACAAGAGAAAAACAAAA
4qB163    AGCCAGACTTCATCAAATTTCTGCACCAATGAAAAAAAAATTTACAAGAGAAAAACAAAA
4qA166    AGCCAGACTTCATCAAATTTCTGCAGCAATGAAAAAAAAATTTACAAGAGAAAAACAAAA
10qA166   AGCCAGACTTCATCAAATTTCTGCAGCAATGAAAAAAAAATTTACAAGAGAAAAACAAAA
4qB168    AGCCAGACTTCATCAAATTTCTGCAGCAATGAAAAAAAAATTTACAAGAGAAAAACAAAA
          *********************** ************************************

4qA161    AACCCTATTAAACGTCACGGACAAGGCCAGAGTTTGAATATACTGTGGTCATCTCTGCTC
4qB163    AACCCTATTAAACGTCACGGACAAGGCCAGAGTTTGAATATACTGTGGTCATCTCTGCTC
4qA166    AACCCTATTAAACGTCACGGACAAGGCCAGAGTTTGAATATACTGTGGTCATCTCCGCTC
10qA166   AACCCTATTAAACGTCACGGACAAGGCCAGAGTTTGAATATACTGTGGTCATCTCCGCTC
4qB168    AACCCTATTAAACGTCACGGACAAGGCCAGAGTTTGAATATACTGTGGTCATCTCCGCTC
          ****************************************************** ****

4qA161    CAGTGCAAACTGTTTCCAGAAAGCCTACTTCTATTTTCCTTGCTGTAACAGAGGAACATT
4qB163    CAGTGCAAACTGTTTTCAGAAAGCCTACTTCTATTTTCCTTGCTGTAACAGAGGAACATT
4qA166    CAGTGCAAACTGTTTTCAGAAAGCCTGCTTTTATTTTCCTTGCTGTAACAGAGGAACATT
10qA166   CAGTGCAAACTGTTTTCAGAAAGCCTGCTTTTATTTTCCTTGCTGTAACAGAGGAACATT
4qB168    CAGTGCAAACTGTTTTCAGAAAGCCTGCTTTTATTTTCCTTGCTGTAACAGAGGAACATT
          *************** ********** *** *****************************

4qA161    TCCTGTCTTATGTTTATTCTACTCTGCAATCCCCTAAGGCTTTTTCTCTCCCTCCCAGAA
4qB163    TCCTGTCTTATGCTTATTCTACTCTGCAATCCCCTAAGGCTTTTTCTCTCCCTCCCAGAA
4qA166    TCCTGTCTTATGCTTATTCTACTCTGCAGTCCCCTAAGGCTTTTTCTCTCCCTCCCAGAA
10qA166   TCCTGTCTTATGCTTATTCTACTCTGCAGTCCCCTAAGGCTTTTTCTCTCCCTCCCAGAA
4qB168    TCCTGTCTTATGCTTATTCTACTCTGCAGTCCCCTAAGGCTTTTTCTCTCCCTCCCAGAA
          ************ *************** *******************************

4qA161    TCTTAAAGTGCATTCGAACTCACAGGCAAAATCCTCCCAGAATCTTGTGAGAACATAAAT
4qB163    TCTTAAAGTGCATTCGAACTCACAGGCAAAATCCTCCCAGAATCTTGTGAGAACATAAAT
4qA166    TCTTAAAGTGCATTCGAACGCACAGGCAAAATCCTCCCAGAAACTTGTGAAAACATAAAT
10qA166   TCTTAAAGTGCATTCGAACGCACAGGCAAAATCCTCCCAGAAACTTGTGAAAACATAAAT
4qB168    TCTTAAAGTGCATTCGAACGCACAGGCAAAATCCTCCCAGAAACTTGTGAAAACATAAAT
          ******************* ********************** ******* *********

4qA161    GATCTGACTAGTTTGGCATTGCTTTTGGGGATCTGGGAAAATCTGTGCACACTTCTGGAG
4qB163    GATCTGACTAGTTTGGCATTGCTTTTGGGGATCTGGGAAAATCTGTGCACACTTCTGGAG
4qA166    GATCTGACTAGTTTGGCATTGCTTTTGGGGATCTGGGAAAATCTGTGCACACTTCTGGAG
10qA166   GATCTGACTAGTTTGGCATTGCTTTTGGGGATCTGGGAAAATCTGTGCACACTTCTGGAG
4qB168    GATCTGACTAGTTTGGCATTGCTTTTGGGGATCTGGGAAAATCTGTGCACACTTCTGGAG
          ************************************************************

4qA161    ACCCTTGTCATGCCATTTTTTATAAATCTATTGTGCCTCAAGTCAGAAGTGTGTGAGGGG
4qB163    ACCCTTGTCATGCCATTATTTATAAATCTATTGTGCCTCAAGTCAGAAGTGTGTGAGGGG
4qA166    ACCCTTGTCAGGCCATTTTTTATAAATCTATTGTGCCTCAAGTCAGAAGTGTGTGAGGGG
10qA166   ACCCTTGTCAGGCCATTTTTTATAAATCTATTGTGCCTCAAGTCAGCAGTGTGTGAGGGG
4qB168    ACCCTTGTCAGGCCATTTTTTATAAATCTATTGTGCCTCAAGTCAGAAGTGTCTGAGGGG
          ********** ****** ***************************** ***** *******

4qA161    AGATGGGGAGACATTGGGATGCGCGCGCCTGGGGCTCTCCCACAGGGGGCTTTCG
4qB163    AGATGGGGAGACATTGGGATGCGCGCGCCTGGGGCTCTCCCACAGGGGGCTTTCG
4qA166    AGATGGGGAGACATTGGGATGCGCGCGCCTGGGGCTCTCCCACAGGGGGCTTTCG
10qA166   AGATGGGGAGACATTGGGATGCGCGCGCCTGGGGCTCTCCCACAGGGGGCTTTCG
4qB168    AGATGGGGAGACATTGGGATGTGCGCGCCTGGGGCTCTCCCACAGGGGGCTTTCG
          ********************* *********************************
```

Figure 6. Alignment of a 475-bp consensus sequence within *D4F104S1* of the most common haplotypes (4qA161, 4qB163, 4qA166, 4qB168, and 10qA166). The consensus is based on the sequence of at least three independent alleles for each specific haplotype. The highlighted nucleotides indicate the presence of haplotype-specific SNPs. Similar to the SSLP, the 4qB163 sequence is very homologous to the pathogenic 4qA161 sequence, whereas the 4qA166 sequence is more homologous to the 10qA166 and 4qB168 sequences.
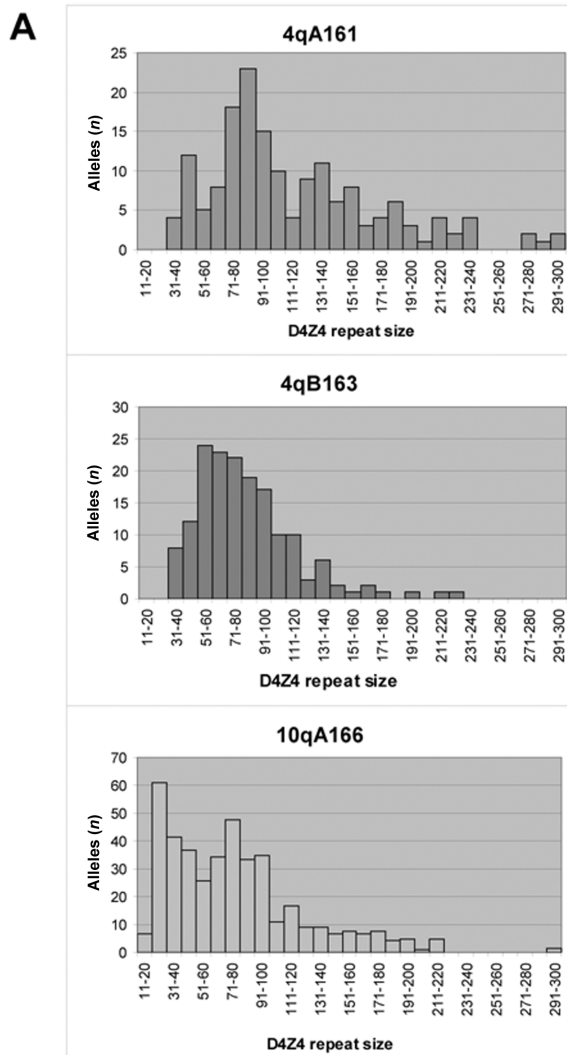
Figure 7. *A,* Overview of D4Z4-repeat length distribution in the most common haplotypes, 4qA161, 4qB163, and 10qA166. Statistical analysis revealed that the D4Z4-repeat size distributions differ significantly between these haplotypes (*P* < .0001). *B,* Mean and median of the D4Z4-repeat length and number of alleles in the major haplotypes.

site in their first unit. This insensitivity for *Bln*I and *Xap*I in the first repeat unit of the D4Z4 repeat could not be observed in alleles of the other 4q haplotypes. Since the *Xap*I restriction site is localized 2.7 kb distal to *D4F104S1*, this finding suggests the presence of additional haplotype-specific sequence variations in D4Z4 and reinforces our

D4Z4-Repeat Size Distribution

Finally, the D4Z4-repeat size distribution was studied in the different haplotypes. It was shown elsewhere that both 4q and 10q alleles display a multimodal repeat size distribution.[20] Because of a limited number of alleles in the low-abundance haplotypes, the size distribution was analyzed only for the most common types on chromosomes 4 and 10. This analysis showed that these haplotypes show a significantly different distribution (*P* < .0001) of D4Z4-repeat sizes (fig. 7*A*). In addition, the mean and median of the different D4Z4 alleles were calculated for these haplotypes (fig. 7*B*). The observed sequence variations in the SSLP and *D4F104S1* sequences and the differences in the mean and median D4Z4-repeat size and size distribution corroborate the concept that the different haplotypes have evolved independently of each other.

Discussion

Contractions of the macrosatellite repeat D4Z4 cause FSHD, but the molecular mechanism underlying this myopathy is largely unknown. Transcription emanating from D4Z4 has never been consistently established, which led to different disease models in which the contraction of D4Z4 causes a change in chromatin conformation and loss of spatiotemporal control over gene expression in *cis* or in *trans*.[7–13] However, none of these models can comprehensively explain the epigenetic mechanisms underlying FSHD.

To study the role of D4Z4 in the pathogenesis of FSHD, allelic variants of the D4Z4 region have been analyzed. Previously,[16] D4Z4 repeats were detected on three different chromosome ends—on two variants of chromosome 4 (4qA and 4qB) and on chromosome 10q. In addition, a polymorphic sequence variation within D4Z4 on 4qB alleles suggested the existence of additional allelic D4Z4 variants.[18] The aim of this study was to investigate the allelic variation in D4Z4 alleles in relation to the 4qA specificity of FSHD. Therefore, we examined sequence variations on 4qA, 4qB, and 10q chromosomes, by analyzing four different polymorphic markers in patients and controls of European descent. The markers are localized proximal to (for SSLP), within (for D4Z4 SNP and D4Z4-repeat number variation), and distal to (for A/B) the D4Z4 repeat.

As shown in figure 4*B*, we were able to unequivocally subdivide all 4qter and 10qter chromosomes in nine different 4q and two distinct 10q haplotypes. The presence of rare alleles within a haplotype that carry a D4Z4 SNP variant different from the majority of alleles indicates that exchanges between the different haplotypes do occur but at a very low frequency and supports the observation that D4Z4 contractions generally occur intrachromosomally.[4]

Subsequently, we stratified for alleles with D4Z4-repeat

sizes <38 kb, which usually fall into the pathogenic range (fig. 4*C*).[18,21] In total, 86 FSHD alleles were analyzed, and all could be assigned to the 4qA161 haplotype. Importantly, in two independent families (Rf10 and Rf204), we were able to show the presence of 4qA166-type FSHD-sized D4Z4 repeats in multiple unaffected relatives, indicating that these alleles are nonpathogenic. Finally, we identified 17 FSHD-sized D4Z4 repeats on 4qB alleles, which were all nonpathogenic. Further genotyping of these short 4qB alleles showed that they all belong to the 4qB163 haplotype. These two findings show that D4Z4 contractions alone are insufficient to cause disease.

Sequence comparison of the SSLP and *D4F104S1* sequences in the different haplotypes showed the presence of only a few consistent sequence variations between the nonpathogenic 4qB163 and the pathogenic 4qA161 haplotypes. In alleles from the nonpathogenic 4qA166 haplotype, we observed the absence of the *Xap*I restriction site in the proximal D4Z4-repeat unit, a feature characteristic of chromosome 10–derived repeat units and not observed in other 4q haplotypes. Furthermore, 4qA166 and 10qA166 haplotypes share high sequence similarities in their SSLP and *D4F104S1* sequences, which differ considerably from the sequence of the 4qA161 haplotype (figs. 4 and 6).

Previously, the nonpathogenicity of FSHD-sized D4Z4 repeats on chromosome 10 was explained by its chromosomal context: only the transcriptional activity of genes on 4q, not that of genes on 10q, would be sensitive to D4Z4-repeat contractions. Moreover, the A/B variation distal to D4Z4 was thought to play a role in the nonpathogenicity of short 4qB alleles. This study shows that neither of these explanations comprehensively addresses the unique association of FSHD with D4Z4 contractions on specific haplotypes, since 4qA166 alleles are now shown to be nonpathogenic as well.

Apparently, because of suppression of interchromosomal rearrangements, each haplotype evolves relatively independently, which allows the accumulation of haplotype-specific sequence variations that may underlie the 4qA161 specificity of FSHD. These sequence variations may be important for the chromatin structure or transcriptional characteristics of the 4q subtelomere and changes thereof in FSHD and must be located close to (distal or proximal) or within D4Z4.

Interestingly, the region immediately proximal to D4Z4 has been implicated to function as a nuclear matrix attachment region (MAR) separating the D4Z4 repeat and upstream genes in two DNA loops.[22] This MAR was shown to be weakened on one chromosome 4 in FSHD myoblasts, suggesting the coexistence of the contracted D4Z4 repeat and upstream genes in a single DNA loop on disease chromosomes. It was postulated that these differences in loop organization may account for the transcriptional deregulation of the upstream genes *FRG1* and *FRG2* in FSHD.[10] Since we found haplotype-specific sequence variations in the region coinciding with this MAR, it will be interesting to study the functionality of the MAR in the different haplotypes and to determine whether the MAR weakening in contracted alleles is specific to the 4qA161 haplotype.

Alternatively, D4Z4 homologues recently have been identified in the genomes of rodents and Afrotheria.[23] The hypothetical *DUX4* gene was shown to be conserved during evolution, which might support a coding function for D4Z4. Therefore, it is eminent to further investigate the presence of haplotype-specific sequence variations in *DUX4* in relation to FSHD.

Altogether, our study demonstrates that the subtelomeric region of chromosome 4q35 can be divided into nine different haplotypes. Our data strongly support the hypothesis that interchromosomal repeat exchanges between the different 4q haplotypes are rare, which allows each of the haplotypes to accumulate specific sequence variations. Only D4Z4 contractions in specific haplotypes, most notably 4qA161, seem to be associated with FSHD. Therefore, future research should be focused on identifying consistent sequence variations for the different haplotypes, since these variations may be essential to FSHD pathogenesis.

Web Resources

Accession numbers and URLs for data presented herein are as follows:

GenBank, http://www.ncbi.nlm.nih.gov/Genbank/ (for SSLP sequences [accession numbers AF117653, AL845259, AY028079, BX649463, BX294170, BX005259, and AL954635])

Online Mendelian Inheritance in Man (OMIM), http://www.ncbi.nlm.nih.gov/Omim/ (for FSHD1A)

Primer3, http://frodo.wi.mit.edu/cgi-bin/primer3/primer3_www.cgi

References

1. Padberg GW (1982) Facioscapulohumeral disease. Leiden University, Leiden
2. Wijmenga C, Hewitt JE, Sandkuijl LA, Clark LN, Wright TJ, Dauwerse HG, Gruter AM, Hofker MH, Moerer P, Williamson R, et al (1992) Chromosome 4q DNA rearrangements associated with facioscapulohumeral muscular dystrophy. Nat Genet 2:26–30
3. van Deutekom JC, Wijmenga C, van Tienhoven EA, Gruter AM, Hewitt JE, Padberg GW, van Ommen GJ, Hofker MH, Frants RR (1993) FSHD associated DNA rearrangements are due to deletions of integral copies of a 3.2 kb tandemly repeated unit. Hum Mol Genet 2:2037–2042
4. Lemmers RJ, van Overveld PG, Sandkuijl LA, Vrieling H, Padberg GW, Frants RR, van der Maarel SM (2004) Mechanism and timing of mitotic rearrangements in the subtelomeric D4Z4 repeat involved in facioscapulohumeral muscular dystrophy. Am J Hum Genet 75:44–53

5. Hewitt JE, Lyle R, Clark LN, Valleley EM, Wright TJ, Wijmenga C, van Deutekom JC, Francis F, Sharpe PT, Hofker M, et al (1994) Analysis of the tandem repeat locus D4Z4 associated with facioscapulohumeral muscular dystrophy. Hum Mol Genet 3:1287–1295

6. Winokur ST, Bengtsson U, Feddersen J, Mathews KD, Weiffenbach B, Bailey H, Markovich RP, Murray JC, Wasmuth JJ, Altherr MR, et al (1994) The DNA rearrangement associated with facioscapulohumeral muscular dystrophy involves a heterochromatin-associated repetitive element: implications for a role of chromatin structure in the pathogenesis of the disease. Chromosome Res 2:225–234

7. Gabriels J, Beckers MC, Ding H, De Vriese A, Plaisance S, van der Maarel SM, Padberg GW, Frants RR, Hewitt JE, Collen D, et al (1999) Nucleotide sequence of the partially deleted D4Z4 locus in a patient with FSHD identifies a putative gene within each 3.3 kb element. Gene 236:25–32

8. Winokur ST, Chen YW, Masny PS, Martin JH, Ehmsen JT, Tapscott SJ, van der Maarel SM, Hayashi Y, Flanigan KM (2003) Expression profiling of FSHD muscle supports a defect in specific stages of myogenic differentiation. Hum Mol Genet 12:2895–2907

9. Alexiadis V, Ballestas ME, Sanchez C, Winokur S, Vedanarayanan V, Warren M, Ehrlich M (2007) RNAPol-ChIP analysis of transcription from FSHD-linked tandem repeats and satellite DNA. Biochim Biophys Acta 1769:29–40

10. Gabellini D, Green M, Tupler R (2002) Inappropriate gene activation in FSHD: a repressor complex binds a chromosomal repeat deleted in dystrophic muscle. Cell 110:339–248

11. Jiang G, Yang F, van Overveld PG, Vedanarayanan V, van der MS, Ehrlich M (2003) Testing the position-effect variegation hypothesis for facioscapulohumeral muscular dystrophy by analysis of histone modification and gene expression in subtelomeric 4q. Hum Mol Genet 12:2909–2921

12. Masny PS, Bengtsson U, Chung SA, Martin JH, van Engelen B, van der Maarel SM, Winokur ST (2004) Localization of 4q35.2 to the nuclear periphery: is FSHD a nuclear envelope disease? Hum Mol Genet 13:1857–1871

13. van Overveld PG, Lemmers RJ, Sandkuijl LA, Enthoven L, Winokur ST, Bakels F, Padberg GW, van Ommen GJ, Frants RR, van der Maarel SM (2003) Hypomethylation of D4Z4 in 4q-linked and non-4q-linked facioscapulohumeral muscular dystrophy. Nat Genet 35:315–317

14. Bakker E, Wijmenga C, Vossen RH, Padberg GW, Hewitt J, van der Wielen M, Rasmussen K, Frants RR (1995) The FSHD-linked locus D4F104S1 (p13E-11) on 4q35 has a homologue on 10qter. Muscle Nerve 2:S39–S44

15. Deidda G, Cacurri S, Piazzo N, Felicetti L (1996) Direct detection of 4q35 rearrangements implicated in facioscapulohumeral muscular dystrophy (FSHD). J Med Genet 33:361–365

16. van Geel M, Dickson MC, Beck AF, Bolland DJ, Frants RR, van der Maarel SM, de Jong PJ, Hewitt JE (2002) Genomic analysis of human chromosome 10q and 4q telomeres suggests a common origin. Genomics 79:210–217

17. Lemmers RJL, de Kievit P, van Geel M, van der Wielen MJ, Bakker E, Padberg GW, Frants RR, van der Maarel SM (2001) Complete allele information in the diagnosis of facioscapulohumeral muscular dystrophy by triple DNA analysis. Ann Neurol 50:816–819

18. Lemmers RJ, de Kievit P, Sandkuijl L, Padberg GW, van Ommen GJ, Frants RR, van der Maarel SM (2002) Facioscapulohumeral muscular dystrophy is uniquely associated with one of the two variants of the 4q subtelomere. Nat Genet 32:235–236

19. Wohlgemuth M, Lemmers RJ, van der Kooi EL, van der Wielen MJ, van Overveld PG, Dauwerse H, Bakker E, Frants RR, Padberg GW, van der Maarel SM (2003) Possible phenotypic dosage effect in patients compound heterozygous for FSHD-sized 4q35 alleles. Neurology 61:909–913

20. van Overveld PG, Lemmers RJ, Deidda G, Sandkuijl L, Padberg GW, Frants RR, van der Maarel SM (2000) Interchromosomal repeat array interactions between chromosomes 4 and 10: a model for subtelomeric plasticity. Hum Mol Genet 9:2879–2884

21. Lemmers RJ, Wohlgemuth M, Frants RR, Padberg GW, Morava E, van der Maarel SM (2004) Contractions of D4Z4 on 4qB subtelomeres do not cause facioscapulohumeral muscular dystrophy. Am J Hum Genet 75:1124–1130

22. Petrov A, Pirozhkova I, Carnac G, Laoudj D, Lipinski M, Vassetzky YS (2006) Chromatin loop domain organization within the 4q35 locus in facioscapulohumeral dystrophy patients versus normal human myoblasts. Proc Natl Acad Sci USA 103:6982–6987

23. Clapp J, Mitchell LM, Bolland DJ, Fantes J, Cornocan AE, Scotting PJ, Armour JA, Hewitt JE (2007) Evolutionary conservation of a coding function for D4Z4, the tandem DNA repeat mutated in facioscapulohumeral muscular dystrophy. Am J Hum Genet 81:264–279