

Molecular combing reveals complex 4q35 rearrangements in Facioscapulohumeral dystrophy

Karine Nguyen, Francesca Puppo, Stéphane Roche, Marie-Cécile Gaillard, Charlene Chaix, Arnaud Lagarde, Marjorie Pierret, Catherine Vovan, Sylviane Olschwang, Emmanuelle Salort Campana, et al.

▶ **To cite this version:**

Karine Nguyen, Francesca Puppo, Stéphane Roche, Marie-Cécile Gaillard, Charlene Chaix, et al.. Molecular combing reveals complex 4q35 rearrangements in Facioscapulohumeral dystrophy. Human Mutation, 2017, 38 (10), pp.1432 - 1441. 10.1002/humu.23304 . hal-01614514

HAL Id: hal-01614514
https://hal.science/hal-01614514

Submitted on 14 Dec 2017

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

Molecular combing reveals complex 4q35 rearrangements in Facioscapulohumeral dystrophy

Karine Nguyen[1,2]* | Francesca Puppo[1] | Stéphane Roche[1] | Marie-Cécile Gaillard[1] | Charlène Chaix[2] | Arnaud Lagarde[1] | Marjorie Pierret[3] | Catherine Vovan[2] | Sylviane Olschwang[1,2,4] | Emmanuelle Salort-Campana[1,5] | Shahram Attarian[1,5] | Marc Bartoli[1] | Rafaëlle Bernard[1,2] | Frédérique Magdinier[1]* | Nicolas Levy[1,2,6]*

[1]Aix Marseille Université, INSERM GMGF UMR S_910, Marseille, 13385, France

[2]APHM, Département de Génétique Médicale, Hôpital d'enfants de la Timone, Marseille, 13385, France

[3]Genomic Vision, Bagneux, 92220, France

[4]Groupe Ramsay Générale de Santé, Hôpital Clairval, Marseille, France

[5]APHM, Centre de Référence des Maladies Neuromusculaires et de la SLA, Hôpital de la Timone, Marseille, 13385, France

[6]APHM, Centre de Ressources Biologiques, Hôpital de la Timone, Marseille, 13385, France

Correspondence
Karine Nguyen, Medical Genetics & Functional Genomics, UMR S_910, INSERM; Aix Marseille University, faculty of Medicine, 27, Bd Jean Moulin 13005 Marseille, France.
Email: karine.nguyen@ap-hm.fr; karine.nguyen@univ-amu.fr

*Frédérique Magdinier and Nicolas Levy are the co-last authors.

Contract grant sponsors: AFM Téléthon; FSH Society; Agence Nationale de la Recherche (ANR-13-BSV1-0001).

Communicated by Graham R. Taylor

Abstract

Facioscapulohumeral dystrophy (FSHD), one of the most common hereditary neuromuscular disorders, is associated with a complex combination of genetic variations at the subtelomeric 4q35 locus. As molecular diagnosis relying on Southern blot (SB) might be challenging in some cases, molecular combing (MC) was recently developed as an additional technique for FSHD diagnosis and exploration of the genomic organization of the 4q35 and 10q26 regions. In complement to the usual SB, we applied MC in a large cohort of 586 individuals with clinical FSHD. In 332 subjects, the two 4q alleles were normal in size, allowing exclusion of FSHD1 while we confirmed FSHD1 in 230 patients. In 14 patients from 10 families, we identified a recurrent complex heterozygous rearrangement at 4q35 consisting of a duplication of the *D4Z4* array and a 4qA haplotype, irresolvable by the SB technique. In five families, we further identified variations in the *SMCHD1* gene. Impact of the different mutations was tested using a minigene assay and we analyzed DNA methylation after sodium bisulfite modification and NGS sequencing. We discuss the involvement of this rearrangement in FSHD since all mutations in *SMCHD1* are not associated with *D4Z4* hypomethylation and do not always segregate with the disease.

KEYWORDS

D4Z4, FSHD, methylation, minigene complementation assay (pCAS), Molecular Combing, SMCHD1

1 | INTRODUCTION

With an estimated worldwide prevalence of 1:15,000, facioscapulohumeral dystrophy (FSHD) is one of the most common hereditary neuromuscular disorders. Nevertheless, clinical and genetic heterogeneity makes its diagnosis challenging. At the clinical level, the pathology is characterized by a progressive weakness of specific facial, shoulder, and upper arm muscles with a progression to the pelvic girdle and lower limbs (Padberg & van Engelen, 2009).

At the genetic level, FSHD is transmitted mostly as an autosomal dominant trait resulting from an unusual mechanism (Sarfarazi et al., 1992; Wijmenga et al., 1992a). The locus linked to the pathology is located in the subtelomeric region of the long 4q arm (van Deutekom et al., 1993; Wijmenga et al., 1992b). In 95% of the patients (FSHD1; MIM# 158900), the disease is associated with a reduction in the copy number of *D4Z4*, a 3.3 kb GC-rich macrosatellite tandem repeat. In control individuals, the number of repeated units ranges from 11 to 150 (41 to 350–500 kb), whereas in patients, one of the two 4q alleles is contracted and contains 1 to 10 units (van Deutekom et al., 1993).

Other genomic factors segregate with the disease. Distal to the *D4Z4* array, a polymorphic 10 kb sequence defines two alleles, 4qA and 4qB equally present in the population (Lemmers et al., 2002). The 4qter region also contains nine distinct haplotypes with some of them, termed permissive haplotypes, preferentially associated with the disease (Lemmers et al., 2007). A SNP in the pLAM sequence distal to the last *D4Z4* in the 4qA allele provides a polyadenylation signal for the *DUX4* transcript encoded by *D4Z4* (Gabriels et al., 1999; Lemmers et al., 2010a).

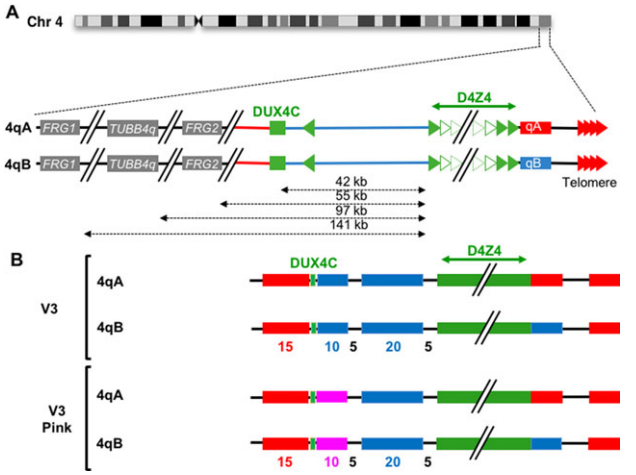


FIGURE 1 Schematic representation of the 4qA and 4qB subtelomeric regions and molecular combing bar coding. **A**: The most distal 4q region is represented (4q35 locus). The proximal *FRG1*, *TUBB4Q*, and *FRG2* genes are indicated. The *D4Z4* array is depicted by green triangles. Sequences starting with an inverted *D4Z4* repeat (green arrow) are specific to the 4q35 locus (red lines) while regions located between the *D4Z4* array and the inverted *D4Z4* repeat are also present on chromosome 10 (10q26 locus). The 4qA and 4qB haplotypes correspond to different genomic elements. The 4qA (red rectangle) is characterized by the presence of a sequence named pLAM immediately distal to the last *D4Z4* repeat and followed by an array of repeated β satellite elements associated with a 4qA haplotype upstream of the telomere (red arrows). The 4qB allele (depicted as a blue rectangle) differs from the 4qA by the absence of β satellite elements upstream of the telomere (red arrows). **B**: The V3 pink bar-code used to distinguish the different alleles (4qA/B, 10qA/B) is based on a combination of four different colors and different DNA probes encompassing the distal regions up to the telomeric sequence as previously described (Nguyen et al., 2011). This four-color barcode comprises one probe detected in blue and one in pink, which hybridize the proximal region common to chromosomes 4 and 10, one 6 kb probe detected in red, which hybridizes in the (TTAGGG)n telomeric ends, and a probe labeled in red that hybridizes the qA-specific β-satellite region, with a variable length (1–5 kb). The qB-specific probe, immediately adjacent to *D4Z4*, is detected in blue in this scheme. Proximal 4q-specific region is detected by a combination of red and pink probes while the proximal 10q-specific region is identified by hybridization with a blue probe. In a number of cases, the V3 bar-code was used, with two probes in blue for the proximal region common to chromosomes 4 and 10

Besides, in 2–3% of patients, a digenism mechanism linking mutation in the *SMCHD1* gene, the presence of pLAM-4qA has been proposed as causative of type 2 FSHD (FSHD2; MIM# 158901) (Lemmers et al., 2012b) while the cause of the pathology remains undetermined in the remaining 2–3% of FSHD patients. However, the current DUX4 model does not apply to all cases of FSHD patients and other pathogenic mechanisms might be involved (Broucqsault et al., 2013; Jones et al., 2012; Lemmers, O'Shea, Padberg, Lunt, & van der Maarel, 2012a; Ricci et al., 2013; Salort-Campana et al., 2015).

In most routine laboratories, FSHD diagnosis is based on the Southern blot (SB) technique using the hybridization with the p13E-11 probe that maps the proximal region adjacent to the first *D4Z4* repeat after digestion of genomic DNA with the *Eco*RI enzyme. However, in a number of cases, the resolution of this technique is not sufficient in providing a clear conclusion regarding the number of repeated units and haplotype and requires additional experiments to identify somatic mosaicism, 4q-10q translocations, p13E-11 deletion, and other non-canonical variants. To overcome this limitation, we recently developed a molecular combing (MC)-based approach providing an accurate FSHD diagnosis thanks to its high resolution (Nguyen et al., 2011). This methodology allows the direct visualization of hundreds of individual DNA molecules stretched on a glass slide at a 1 kb resolution after hybridization with a combination of DNA probes corresponding to the 4q35 and 10q26 loci. Thus, in a single step, MC allows direct and simultaneous visualization of the 4q35 and 10q26 loci and provides the size of the *D4Z4* arrays for the different alleles as well as their co-segregating haplotypes (Nguyen et al., 2011).

Here, we report 14 cases out of the 586 individuals referred by clinicians for FSHD diagnosis and explored by MC. In this subgroup of individuals, we describe an atypical and novel rearrangement of the distal end of the 4q35 locus consisting of a duplication of the *D4Z4* array located on a 4qA haplotype. In most cases, the rearrangement segregates with other features associated with FSHD1 or FSHD2 while for two individuals, duplication of the *D4Z4* array is the only molecular defect.

2 | MATERIALS AND METHODS

2.1 | Patients' samples

Individuals explored in this study were clinically assessed by neurologists with expertise in neuromuscular diseases who defined the presence or absence of clinical signs and evaluated the involvement of the groups of muscles typically affected in the disease (facial, shoulder and pelvic girdle, upper and lower limbs, and abdominal muscles) and level of impairment (Supp. Table S1). All 14 patients described here displayed typical clinical signs of FSHD. Blood samples were received at the Department of Medical Genetics, La Timone Children's Hospital Marseille, center of reference for molecular diagnosis of FSHD. Blood cells were embedded in agarose plugs for high-molecular-weight DNA extraction. After extraction, DNA was analyzed by both SB and MC (Nguyen et al., 2011). Informed consent was obtained from all patients for the genetic analyses.

Detailed materials and methods are provided as supplementary information.

3 | RESULTS

Between 2009 and 2014, a cohort of 586 individuals presenting with a pattern of muscular weakness consistent with FSHD was explored by MC. With this method, direct visualization of the DNA fiber from a region encompassing the telomere region to the chromosomes 4q- and 10q-specific subtelomeric sequences at a high resolution and at the scale of single molecules allows us to directly observe the physical structure of the region of interest and to determine the size of the *D4Z4* array (Figure 1). In 332 cases, the two 4q alleles were normal

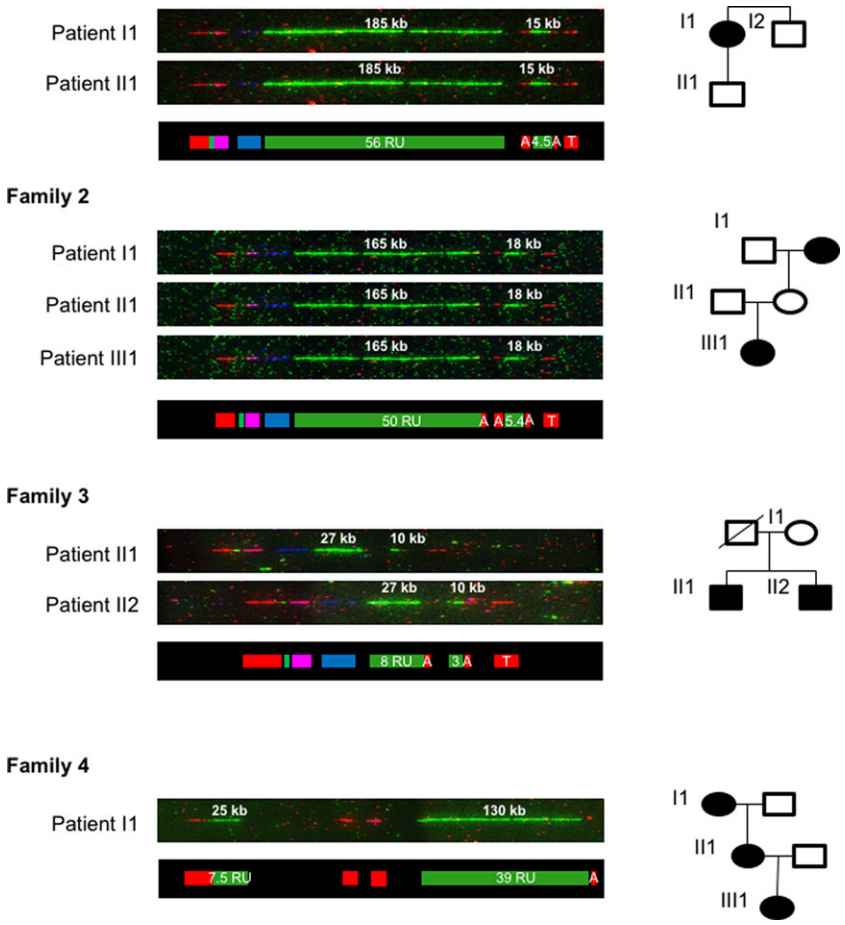


FIGURE 2 Molecular combing reveals complex and recurrent rearrangements of the 4q35 region in familial cases with FSHD. Results of MC are presented for eight individuals from four families, for which SB was inconclusive. Combed DNA was hybridized with the "V3 pink" combination of probes. Only the 4q allele carrying a complex rearrangement is presented. In each case, the DNA combing images are presented with the sizes of the *D4Z4* arrays given in kilobases (kb) together with a schematic representation of the rearranged allele with the number of repeated units (RU). Signals corresponding to the qA allele (A) or telomere (T) are indicated. In four families (numbered 1–4) for which the segregation of the rearrangement could be studied, the family tree is presented on the right of the figure

in size, allowing the exclusion of FSHD1. In 230 patients, we were able to confirm a positive diagnosis of FSHD1 with individuals showing one 4qA-contracted *D4Z4* allele below or equal to 10 repeated units. In 17 patients, we detected a mosaic 4qA-contracted allele, and in six patients, a deletion of *D4Z4* encompassing the p13E-11 region. Three patients carried a homozygous 4qA-contracted allele. Interestingly, MC also revealed complex rearrangements of the *D4Z4* array in 14 patients from 10 unrelated families (Figure 2). In family 4, two genomic variants segregate: a canonical contraction of *D4Z4* on a 4qA haplotype, and a complex rearrangement. The index case (4-I1) affected with FSHD transmitted the short *D4Z4* array (6 units) to her affected daughter (4-II1) and her affected grand-daughter (4-III1). In addition, the index case (4-I1) carries a complex rearrangement consisting of two *D4Z4*-qB arrays (25 kb at the proximal end and 130 kb at the telomeric end) separated by a complex organization of sequences hybridized by the bar code probes (Figure 2). In this case, the rearrangement revealed by MC corresponds to the SB results in patient 4-I1 (Supp. Figure S1). However, this rearrangement does not segregate with the disease (Figure 2) suggesting that it is not causative of FSHD.

In the other families, the *D4Z4* array rearrangement consists of a repetitive pattern of two arrays of different size containing a different number of repeated units but carried by the same chromosome, each one being flanked on the distal side by a qA sequence (Figure 2). In all families but one, the proximal array is large (above 50 kb, i.e. >11 units) whereas the distal array is contracted (under 25 kb). In one family (family 3), the two arrays are contracted (27 and 10 kb). In one family (family 4), out of 10, the proximal array is contracted (25 kb, 7 units), whereas the distal array is above the threshold of 11 units (130 kb, 39 units). In sporadic cases 5 and 8 (Figure 3), we detected the same rearrangement corresponding to a large *D4Z4* array (180 kb, 54 units) followed by a distal short *D4Z4* array (20 kb, 6 units) suggesting that these two individuals might descend from a common ancestor. However, this would require additional investigations.

In order to test the frequency of these rearrangements in the population, 50 control individuals with no muscular disease were analyzed by MC. Combed DNA molecules hybridized with the 4q-specific probes were analyzed using the Combilog software that gives the distribution of signals corresponding to the *D4Z4* array and hence, average size of the array. Histograms were plotted as the average number of

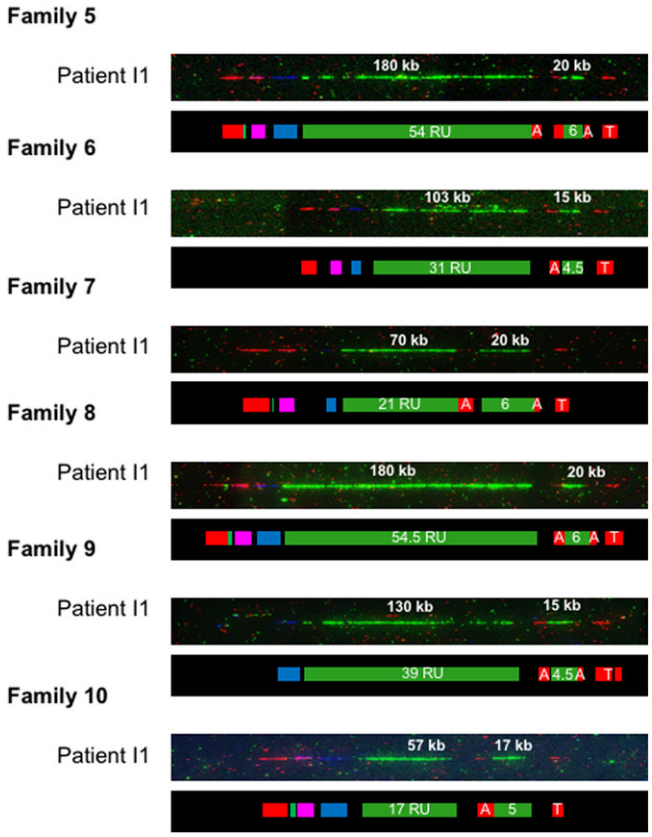


FIGURE 3 Molecular combing reveals complex and recurrent rearrangements of the 4q35 region in sporadic cases with FSHD. Results of MC are presented for six sporadic cases with clinical signs of FSHD and for which SB was inconclusive. Combed DNA was hybridized with the "V3 pink" combination of probes. Only the 4q allele carrying a complex rearrangement is presented. In each case, the DNA combing images are presented with the sizes of the *D4Z4* arrays given in kilobases (kb) together with a schematic representation of the rearranged allele with the number of repeated units (RU). Signals corresponding to the qA allele (A) or telomere (T) are indicated

signal for the size of the *D4Z4* array (Supp. Figure S2). The recurrent 4q rearrangement was found in one out of the 100 chromosomes 4 analyzed. Several 4qB-contracted alleles were seen as well as several 10qA-contracted alleles but none pathological 4qA-contracted allele (data not shown).

3.1 | *SMCHD1* status

Mutations in the*SMCHD1* gene (MIM# 614982) have been recently described in a subset of individuals affected with FSHD but carrying more than 11 *D4Z4* repeated units (FSHD2) and selected on the low level of *D4Z4* methylation (Lemmers et al., 2012b). Moreover, SMCHD1 loss of function might act as a modifier and increase disease severity in FSHD1 patients with "borderline" *D4Z4* arrays (9–11 repeated units) (Sacconi et al., 2013).

Hence, to determine whether clinical signs of FSHD are associated with *SMCHD1* mutation in the 14 patients with complex 4q35 rearrangements or exclude the involvement of this factor, we performed exome sequencing in 10 probands, one in each family, and confirmed *SMCHD1* variants by Sanger sequencing. We found heterozygous *SMCHD1* variants in five out of the 10 patients analyzed (Figure 4A). Four variants out of five are predicted to affect splicing, and all the variants are predicted to be damaging, three of them being amino acid substitutions, one being a stop codon, and one being a splice mutation (Table 1 and Figure 4B). In order to verify the impact of the variants on splicing, we performed a Complementation Assay using a splicing reporter minigene pCAS assay to evaluate the effect of variants affecting splicing sites (Gaildrat et al., 2010). Results are recapitulated in Table 1 and Figure 4.

3.2 | DNA methylation

D4Z4 is extremely GC-rich (>70%) (Lyle, Wright, Clark, & Hewitt, 1995). In FSHD1 patients, the *D4Z4* array is slightly hypomethylated (de Greef et al., 2010; de Greef et al., 2009; Gaillard et al., 2014; Sacconi et al., 2012; van Overveld et al., 2005; van Overveld et al., 2003). In individuals with a shortened *D4Z4* array, hypomethylation in affected individuals is lower than in asymptomatic carriers suggesting that this epigenetic modification might act as a modifier in the clinical expression of the disease (Gaillard et al., 2014; Lemmers et al., 2015). This hypomethylation is more pronounced in FSHD2 patients (Gaillard et al., 2014; Hartweck et al., 2013) and has been associated in a number of cases to *SMCHD1* mutations (Lemmers et al., 2012b).

In order to further characterize the molecular alterations associated with the disease in patients with complex 4q35 rearrangements, we analyzed DNA methylation by sodium bisulfite sequencing across the *D4Z4* repeat and especially at two subregions (DR1 and 5′) for which hypomethylation has been described (Gaillard et al., 2014; Hartweck et al., 2013) in blood DNA. In the different cases, the percentage of CpG methylation was analyzed by deep-sequencing of thousands of DNA molecules per sample after sodium bisulfite modification and PCR amplification using primers amplifying methylated and unmethylated DNA with the same efficiency (Table 2 and Supp. Figure S3). This highly quantitative method provides the percentage of methylation per CpG, the global methylation level per sequence and an accurate determination of the number of hypomethylated allele for each sequence analyzed (Supp. Figure S3) (Roche et al., submitted). In earlier reports, *D4Z4* was considered as hypomethylated when the percentage of methylated molecules was below 30% at DR1 (Hartweck et al., 2013) and below 55% in 5′ in FSHD2 patients compared with controls (Gaillard et al., 2016; Gaillard et al., 2014). The average methylation level is indicated for each patient in Table 2 and Figure 4C for index cases. Based on this analysis, we observed *D4Z4* hypomethylation in family 2 (23% at DR1 in the proband, 2-III1) that correlates with a mutation in exon 28 leading to a partial exon skipping and introduction of a stop codon; in family 9 (13% methylation at DR1) that correlates with an amino acid substitution, encoded by exon 11, predicted as damaging and in family 6 (28% at DR1) associated with an amino acid substitution encoded by exon 39 and causing partial exon skipping (Figures 4B and 4C and Table 2). The two affected individuals in family 3 carry a variant in exon 45 predicted as damaging at the protein level and associated with hypomethylation (27% at DR1) in 3-II1 but with a more moderate effect on the methylation level

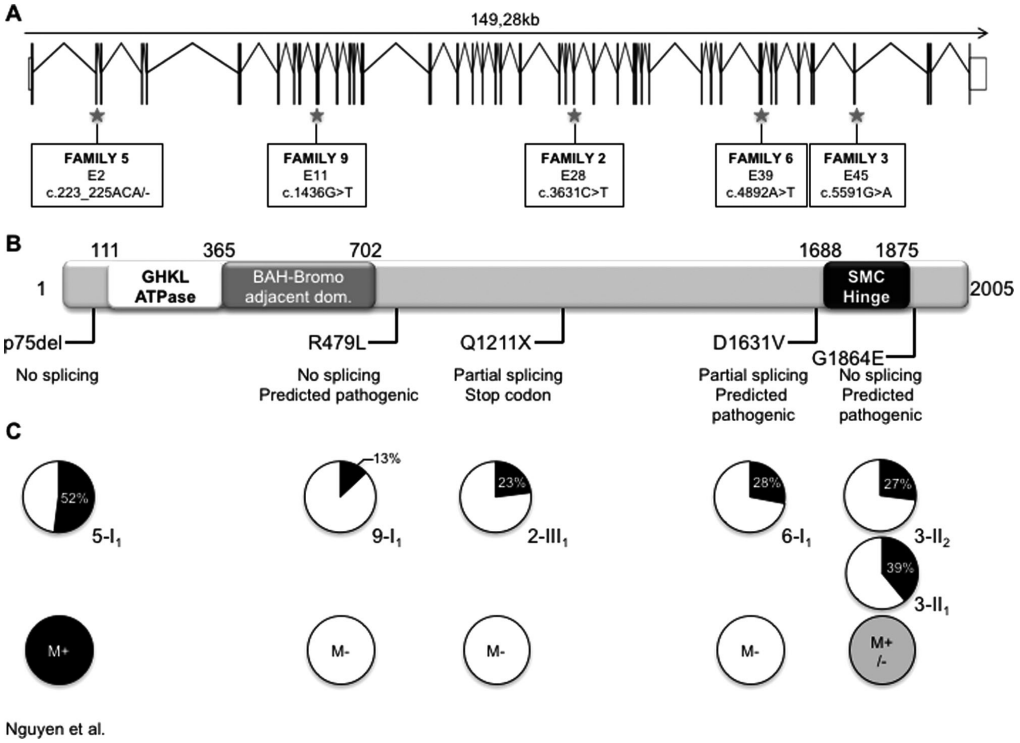


FIGURE 4 Distribution of *SMCHD1* variants in patients with 4q35 rearrangements and DNA methylation levels. **A**: Schematic representation of the *SMCHD1* gene. Exons are represented as vertical lines. Splicing sites are depicted. Position of the different variants identified in patients with 4q35 rearrangements is indicated for the different exons using the Genbank NM_015295.2 entry as reference. **B**: Schematic representation of the SMCHD1 protein with position of the main domains. The variants and corresponding consequences are indicated. **C**: Pie charts showing the DNA methylation level (black) for the DR1 sequence for the different probands. Individuals showing residual methylation below 30% are considered as hypomethylated (M-)

in the other brother (39% at DR1; 3-II2) consistent with the inter-individual variability of this epigenetic modification. In family 5 carrying a variant in exon 2 of *SMCHD1*, predicted as a splicing defect, we observed neither hypomethylation nor splicing defect suggesting that this *SMCHD1* variant is not associated with *D4Z4* methylation and is likely not pathogenic.

4 | DISCUSSION

Using MC in the diagnosis strategy of FSHD in a cohort of 586 individuals, we describe for the first time the recurrence of a heterozygous novel complex rearrangement of the 4q35 locus involving the *D4Z4* array in patients clinically affected with FSHD (Figure 5). Of note, among the samples received for FSHD diagnostic request, a large amount of cases, do not display clinical features typical of FSHD. In these cases, clinicians aim at ruling out FSHD rather than confirm it explaining the high number of negative diagnoses in this cohort. The different patients who carry the rearrangement share similar genomic features with a tandem "duplication" of the *D4Z4*-qA region consisting of two arrays of *D4Z4* separated by a qA sequence and perhaps some telomeric sequence. We describe for the first time an in cis duplication of the *D4Z4* array that occurred in 14 patients from 10 independent families. In a number of cases, the rearrangement co-exists with other genomic variants associated with FSHD (a short *D4Z4* array, *SMCHD1* variant or *D4Z4* hypomethylation). In two cases, the rearrangement was the only molecular alteration found at the FSHD locus. In the different patients, we did not find variants in other genes potentially associated with the disease (*DNMT3B*, MIM# 602900; *FRG1*, MIM# 601278; *FRG2*, MIM# 609032; *FAT1*, MIM# 600976).

In three out of the 10 families analyzed (families 1, 4, and 10), the rearrangement on one 4q chromosome co-exists with a contracted *D4Z4* array on the other 4qA chromosome. Moreover, the rearrangement does not segregate with the disease in families 1 and 4, suggesting that it is not the cause of the disease in these two examples. In family 4, the rearrangement consists of the duplication of the *D4Z4* array, flanked distally by a qB sequence, which is expected to be non-permissive. Thus, in these cases, the pathological 4qA contracted allele probably accounts for the clinical manifestation and the rearrangement is unlikely causative of the disease. In three other families (families 2, 6, and 9), the probands carry the rearrangement along with an *SMCHD1* variant that is associated with *D4Z4* hypomethylation. Family 2 carries a "borderline" 4qA *D4Z4* array besides the rearrangement, as described in FSHD1-FSHD2 patients carrying both a borderline *D4Z4* array (i.e., 9–11 units) and a mutation in *SMCHD1* (Sacconi et al., 2013) but in family 2, the different genomic features only segregate with the disease in the grand-mother (2-I_1) and the proband (2-III_1) while the mother (2-II_1) is unaffected, illustrating the high degree of

TABLE 1 Analysis of *SMCHD1* variants in individuals carrying 4q35 rearrangements by exome sequencing

Family ID	Patient ID	*SMCHD1* status	Prediction	pCAS2 assay results
1	1-I_1	Not mutated		
	1-I_2			
	1-II_1			
2	2-I_1	Exon 28 c.3631C > T p.Q1211X	Creation of an exonic ESS site. Potential splicing alteration. Nucleotide substitution creates a STOP codon.	Partial exon skipping (ex28 119 bp), with ORF loss.
	2-II_1	Exon 28 c.3631C > T p.Q1211X		
	2-III_1	Exon 28 c.3631C > T p.Q1211X		
3	3-I_1	Not mutated		
	3-II_1	Exon 45 c.5591G > A p.G1864E	Activation of an exonic cryptic acceptor site in the presence of one or more cryptic branch point(s). New exonic ESS site. Exonic ESE site broken. Potential alteration of splicing. AA substitution predicted as damaging.	No splicing difference.
	3-II_2	Exon 45 c.5591G > A p.G1864E		
4	4-I_1	Not mutated		
	4-II_1			
	4-III_1			
5	5-I_1	Exon 2 c.223_225delACA p.T75del	Exonic ESE site broken. Potential alteration of splicing	No splicing difference.
6	6-I_1	Exon 39 c.4892A > T p.D1631V	New donor site and alteration of an exonic ESE site. Potential splicing defect. AA substitution predicted as damaging.	Partial exon skipping (ex39 120 bp) without ORF loss.
7	7-I_1	Not mutated		
8	8-I_1	Not mutated		
9	9-I_1	Exon 11 c.1436G > T p.R479L	Probably no impact on splicing. AA substitution predicted as pathogenic.	No splicing difference.
10	10-I_1	Not mutated		

The presence of *SMCHD1* sequence variant was analyzed by exome sequencing. *SMCHD1* variants numbering is based on the GenBank NM_015295.2 cDNA sequence. We used +1 as the A of the ATG translation initiation codon in this reference sequence, with the initiation codon as codon 1. The corresponding variant and prediction are indicated for each individual. The impact of all mutations was tested using a functional splicing assay (pCAS).

non-penetrance and intra-familial variability in the disease despite the presence of the FSHD-associated molecular features, including *SMCHD1* mutation and hypomethylation.

In family 3, the affected individuals carry the rearrangement on one 4q chromosome, and a "borderline" 4qA *D4Z4* array (10 units) on the other 4q chromosome. This borderline *D4Z4* array is transmitted from the unaffected mother to the two affected sons who carry an *SMCHD1* gene variant and variable level of methylation. In family 5, the patient carries the rearrangement and an *SMCHD1* variant in exon 2 for which we did not observe splicing defect and that is not associated with *D4Z4* hypomethylation, suggesting that this variant might be non-pathogenic. Finally, in families 7 and 8, the rearrangement is the only alteration on chromosomes 4q, and we did not detect any *SMCHD1* variation by exome sequencing suggesting that in these two patients, the rearrangement might be pathogenic.

The frequency of the rearrangement in one out of 50 control individuals does not argue against a pathogenic effect in FSHD as the *D4Z4* array contraction carried by 95% of FSHD patients is also seen with a frequency of 3% in the general population (Scionti et al., 2012). This emphasizes the recombinogenic potential of this subtelomeric region and suggests that the rearrangement may participate in the pathogenesis of FSHD, perhaps with other defects yet unknown.

In most routine laboratories, FSHD diagnosis is based on the SB technique using the *Eco*RI enzyme and hybridization with the p13E-11 probe (D4F104S1) on genomic DNA, to determine *D4Z4* copy number located on chromosomes 4. However, in many cases, the interpretation remains difficult, and often requires additional manipulations (digestion with other restriction enzymes, use of additional probes) (Lemmers et al., 2012a; Ricci et al., 2013). Other limitations of the SB technique include non-canonical deletions extending into the p13E-11 probe-covered region in 1% of FSHD patients (Lemmers et al., 2003) or somatic mosaicism originating from mitotic *D4Z4* rearrangements which accounts for about 40% of de novo cases (van der Maarel et al., 2000) and might be underestimated both in patients and in the

TABLE 2 Summary of clinical and molecular data in individuals from 10 families investigated by MC and sodium bisulfite sequencing

Family ID	Patient ID	Sex	Clinical status	qA/qB	10q	Methylation (%) DR1/5′/Mid/3′
1	1-I_1	F	Affected	6A/rearr	32A/28B	62/69/66/60
	1-I_2	M	Unaffected	19A/18B	19A/31A	73 (5′)
	1-II_1	M	Unaffected	29B/rearr	16A/25B	82 (5′)
2	2-I_1	F	Affected	22A/rearr	28A/35A	24/52/63/52
	2-II_1	F	Unaffected	19A/rearr	5A/42A	26.5/52/62/55
	2-III_1	F	Affected	15A/rearr	8A/30A	23/48/60/55
3	3-I_1	F	Unaffected	10A/26B	34A/9A	67/70/71/57
	3-II_1	M	Affected	10A/rearr	34A/35A	27/46/58/57
	3-II_2	M	Affected	10A/rearr	10A/57A	39/48/62.5/56
4	4-I_1	F	Affected	6A/rearr	18A/28A	53.5/63/63/58
	4-II_1	F	Affected	6A/19A	10A/28A	56 (5′)
	4-III_1	F	Affected	6A/33A	29A/28A	75 (5′)
5	5-I_1	M	Affected	> 400 kb/rearr	6A/22A	52/64/66/58
6	6-I_1	M	Affected	21A/rearr	2A/39A	28/55/60/58
7	7-I_1	M	Affected	51A/rearr	85A/ > 300 kb-A	74/80/73/60
8	8-I_1	F	Affected	38A/rearr	4A/29A	67/72/68.5/58
9	9-I_1	M	Affected	14A/rearr	10A/29A	13/44/59/54
10	10-I_1	M	Affected	7A/rearr	20A/27A	53/66/64.5/56

Patients (corresponding to 10 different families) were examined by a neurologist who evaluated the different group of muscles specifically affected in FSHD: face, shoulder and pelvic girdles, upper limbs, lower legs, and abdominal muscle and level of impairment. At the molecular level, patient's genotype was evaluated by molecular combing and the number of *D4Z4* units together with the haplotype (qA or qB) was determined for each chromosome 4 or 10. We also analyzed *D4Z4* methylation level by sodium bisulfite followed by deep sequencing using a PGM Ion torrent. Values correspond to the sequence methylation score, calculated as the average ratio of methylated CpG on all aligned CpG for a given sequence for four regions within *D4Z4*. All sequences were analyzed in both orientations for a complete coverage. On average, the number of sequences analyzed is 1,440 for DR1, 4,200 for 5P, 1,200 for Mid, and 7,000 for 3P

general population. Somatic mosaicism was excluded in the 14 patients reported here.

Despite some limitations (integrity of the combed DNA, overlapping signals, very large size alleles), direct visualization of the DNA fiber from the telomeric region to the specific chromosomes 4 and 10 sequences at the high resolution of 1 kb and at the scale of single molecules allowed us to simultaneously explore the genomic organization of the 4q35 and 10q26 loci in a single experiment, and measure the size of the *D4Z4* arrays on the four 4q and 10q alleles (Nguyen et al., 2011). In addition, by MC, we directly visualize the qA or qB-specific sequence with a combination of specific probes which is not feasible by SB in a single test and requires additional manipulations (Lemmers et al., 2002). In most cases, chromosome 10 has been described as type A. The majority of patients described here carry 10qA alleles. However, 10qB alleles were detected in two individuals in family 1 (one is affected, 1-I1; the other one is not, 1-II1). According to previous reports, 10qB haplotypes might account for approximately 5–14% of all 10q chromosomes but might be overlooked since 10q alleles are usually not studied (Lemmers et al., 2010b; Nguyen et al., 2011) during the routine diagnostic procedure. The bar coding of the 4q and 10q ends thus permits the visualization of the physical structure of these two subtelomeric regions. SBs might thus be erroneously interpreted with 10qB alleles considered as 4q alleles. Even if SB limitations can be overcome using different restriction enzymes and probes, MC provides a new tool to detect mosaicism (Nguyen et al., 2011), p13E-11 deletions and complex rearrangements in a single step experiment representing a reliable controlled and standardized genetic testing method.

Besides the shortening of the *D4Z4* array, hypomethylation of the repeated sequence has been described in FSHD1 and FSHD2 patients while methylation level is similar to controls in asymptomatic carriers (Gaillard et al., 2014; Lemmers et al., 2015). Although not analyzed for diagnostic purpose, this epigenetic modification might act as a modifier in the clinical manifestation of the disease. In addition, higher severity was associated with hypomethylation in FSHD patients with borderline repeat number (i.e., 9–11 *D4Z4* units) carrying a heterozygous *SMCHD1* mutation (Sacconi et al., 2013). We detected *SMCHD1* mutations in five cases. Four out of five of these variants are predicted to affect splicing. We tested the impact of these sequence variations on splicing using a minigene assay and correlated the impact of the variant with DNA methylation. We show that these variants are not systematically associated with *D4Z4* hypomethylation or with the presence of clinical signs. Overall, this underlies the complexity of the disease and the challenge in the interpretation of the various molecular features such as mutation in *SMCHD1* and *D4Z4* methylation and the importance of functional testing of the *SMCHD1* variants for a more complete understanding of the genomic defects associated with the disease.

Overall, for FSHD, it is thus critical to understand and evaluate the advantages and limitations of the different molecular tests available and their yield for diagnosis but also for understanding the complexity of the 4q and 10q subtelomeric regions and the implication of the

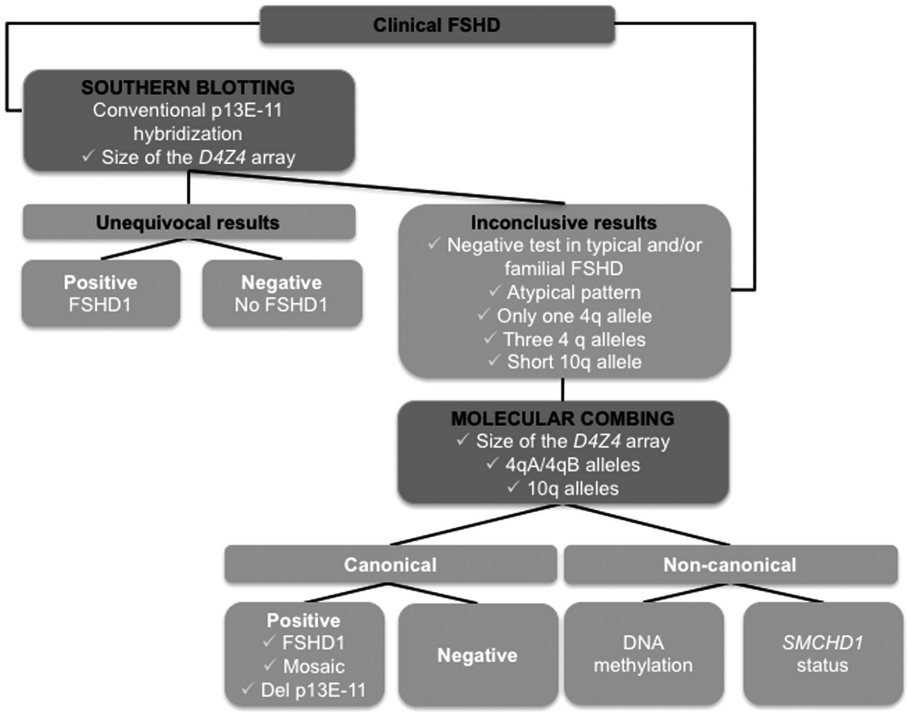


FIGURE 5 Molecular combing strategy for FSHD diagnostic. When molecular diagnosis is requested in patients presenting with FSHD phenotype, conventional SB technique is in most cases applied first to provide an unequivocal positive (one short 4q allele with other alleles on 4q and 10q visible) or negative (four 4q and 10q alleles visible, none 4q short allele, if possible with available data from index case analysis) diagnosis. In case of inconclusive results, including discordance between clinical and molecular data, and atypical pattern, MC provides a clear conclusion in case of canonical FSHD1, mosaics, p13E-11 deletion, or true negative diagnosis. When complex rearrangements are identified, further explorations can be performed including methylation and *SMCHD1* gene analysis

different haplotypes in the pathogenesis of this complex genetic disorder. This work also opens new perspectives in the investigation of the molecular mechanism associated with 4q rearrangements and on the putative role of SMCHD1. The current model associates chromatin decompaction of the *D4Z4* array with subsequent reactivation of the DUX4 retrogene encoded by the most distal *D4Z4* repeat suggesting a key role for the chromatin structure of the *D4Z4* array. Based on our observations, one can speculate that the organization of the rearranged locus maintains an open chromatin structure for the distal shortened array but might also modify the topology of the entire region (Arnoult et al., 2010; Bodega et al., 2009; Ottaviani et al., 2009; Pirozhkova et al., 2008; Robin et al., 2015). However, this remains to be characterized at the molecular level.

Even if in a large number of cases, FSHD is associated with a shortening of the *D4Z4* array on a 4qA allele, the high number of exceptions described in the literature (Ricci et al., 2013; Scionti et al., 2012) cannot be ignored for a full picture of this complex pathology, the subsequent interpretation of the genetic data but also the molecular modeling of the pathology. In addition, our results highlight the variability of molecular data among patients and within families and the complexity one might face in testing and counseling FSHD families. Furthermore, our data support the added value of MC compared with SB for the primary diagnosis of the disease but also for the elucidation of complex cases such as rearrangements or unresolved genotypes and the systematic detection of 4qA/qB and 10q alleles. Overall, our results, together with a recent study (Vasale et al., 2015), show that MC has a lower false-negative rate compared with SB and thus can be reliably used for molecular diagnosis of FSHD in general but also for the resolution of complex cases in clinical practice, the most undisputable feature in the disease being the clinical phenotype unifying FSHD1, FSHD2, and more complex cases (Figure 5).

ACKNOWLEDGMENTS

We are indebted to all families for participating in this study.

DISCLOSURE STATEMENT

The authors declare no conflict of interest.

AUTHOR CONTRIBUTIONS

K.N. designed the study, supervised, conducted, analyzed the combing experiments, and wrote the manuscript. F.P. designed, conducted, and analyzed the mutation analysis. S.R. and M.G.C. designed, conducted, and analyzed the DNA methylation analysis. A.L. conducted the DNA methylation sequencing analysis and contributed to the design of the assay. S.O. contributed to the DNA methylation analysis and development of the deep sequencing methodology. C.V. and C.C. conducted and analyzed the Southern blotting and MC, respectively. M.B. designed the Exome experiments, provided valuable discussion, intellectual input, and edited the manuscript. E.S.C. and S.A.

provided patients' samples and clinical data. R.B. analyzed the data and edited the manuscript. F.M. analyzed the data, wrote and edited the manuscript. N.L. designed the study, wrote and edited the manuscript.

A patent application (No. EP08165310.7) dedicated to MC for the diagnosis of FSHD1 and exploration of *D4Z4* has been registered by Genomic Vision, University of the Mediterranean, and Public Assistance of the Hospitals of Marseille. N.L. is coinventor of the patent.

REFERENCES

Arnoult, N., Schluth-Bolard, C., Letessier, A., Drascovic, I., Bouarich-Bourimi, R., Campisi, J., … Londono-Vallejo, A. (2010). Replication timing of human telomeres is chromosome arm-specific, influenced by subtelomeric structures and connected to nuclear localization. *PLoS Genetics*, *6*(4), e1000920, 1–15.

Bodega, B., Ramirez, G. D., Grasser, F., Cheli, S., Brunelli, S., Mora, M., … Ginelli, E. (2009). Remodeling of the chromatin structure of the facioscapulohumeral muscular dystrophy (FSHD) locus and upregulation of FSHD-related gene 1 (FRG1) expression during human myogenic differentiation. *BMC Biology*, *7*, 41, 1–15.

Broucqsault, N., Morere, J., Gaillard, M. C., Dumonceaux, J., Torrents, J., Salort-Campana, E., … Roche, S. (2013). Dysregulation of 4q35- and muscle-specific genes in fetuses with a short D4Z4 array linked to facioscapulo-humeral dystrophy. *Human Molecular Genetics*, *22*(20), 4206–4214.

de Greef, J. C., Lemmers, R. J., Camano, P., Day, J. W., Sacconi, S., Dunand, M., … Tawil, R. (2010). Clinical features of facioscapulohumeral muscular dystrophy 2. *Neurology*, *75*(17), 1548–1554.

de Greef, J. C., Lemmers, R. J., van Engelen, B. G., Sacconi, S., Venance, S. L., Frants, R. R., … van der Maarel, S. M. (2009). Common epigenetic changes of D4Z4 in contraction-dependent and contraction-independent FSHD. *Human Mutation*, *30*(10), 1449–1459.

Gabriels, J., Beckers, M. C., Ding, H., De Vriese, A., Plaisance, S., van der Maarel, S. M., … Belayew, A. (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*(1), 25–32.

Gaildrat, P., Killian, A., Martins, A., Tournier, I., Frebourg, T., & Tosi, M. (2010). Use of splicing reporter minigene assay to evaluate the effect on splicing of unclassified genetic variants. *Methods in Molecular Biology*, *653*, 249–257.

Gaillard, M. C., Puppo, F., Roche, S., Dion, C., Campana, E. S., Mariot, V., … Bartoli, M. (2016). Segregation between SMCHD1 mutation, D4Z4 hypomethylation and Facio-Scapulo-Humeral Dystrophy: A case report. *BMC Medical Genetics*, *17*(1), 66, 1–9.

Gaillard, M. C., Roche, S., Dion, C., Tasmadjian, A., Bouget, G., Salort-Campana, E., … Magdinier, F. (2014). Differential DNA methylation of the D4Z4 repeat in patients with FSHD and asymptomatic carriers. *Neurology*, *83*(8), 733–742.

Hartweck, L. M., Anderson, L. J., Lemmers, R. J., Dandapat, A., Toso, E. A., Dalton, J. C., … Kyba, M. (2013). A focal domain of extreme demethylation within D4Z4 in FSHD2. *Neurology*, *80*(4), 392–399.

Jones, T. I., Chen, J. C., Rahimov, F., Homma, S., Arashiro, P., Beermann, M. L., … Jones, P. L. (2012). Facioscapulohumeral muscular dystrophy family studies of DUX4 expression: Evidence for disease modifiers and a quantitative model of pathogenesis. *Human Molecular Genetics*, *21*(20), 4419–4430.

Lemmers, R. J., de Kievit, P., Sandkuijl, L., Padberg, G. W., van Ommen, G. J., Frants, R. R., & van der Maarel, S. M. (2002). Facioscapulohumeral muscular dystrophy is uniquely associated with one of the two variants of the 4q subtelomere. *Nature Genetics*, *32*(2), 235–236.

Lemmers, R. J., Goeman, J. J., van der Vliet, P. J., van Nieuwenhuizen, M. P., Balog, J., Vos-Versteeg, M., … van der Maarel, S. M. (2015). Inter-individual differences in CpG methylation at D4Z4 correlate with clinical variability in FSHD1 and FSHD2. *Human Molecular Genetics*, *24*(3), 659–669.

Lemmers, R. J., O'Shea, S., Padberg, G. W., Lunt, P. W., & van der Maarel, S. M. (2012a). Best practice guidelines on genetic diagnostics of Facioscapulohumeral muscular dystrophy: Workshop 9th June 2010, LUMC, Leiden, The Netherlands. *Neuromuscular Disorders*, *22*(5), 463–470.

Lemmers, R. J., Osborn, M., Haaf, T., Rogers, M., Frants, R. R., Padberg, G. W., … Upadhyaya, M. (2003). D4F104S1 deletion in facioscapulohumeral muscular dystrophy: Phenotype, size, and detection. *Neurology*, *61*(2), 178–183.

Lemmers, R. J., Tawil, R., Petek, L. M., Balog, J., Block, G. J., Santen, G. W., … van der Maarel, S. M. (2012b). Digenic inheritance of an SMCHD1 mutation and an FSHD-permissive D4Z4 allele causes facioscapulohumeral muscular dystrophy type 2. *Nature Genetics*, *44*(12), 1370–1374.

Lemmers, R. J., van der Vliet, P. J., Klooster, R., Sacconi, S., Camano, P., Dauwerse, J. G., … van der Maarel, S. M. (2010a). A unifying genetic model for facioscapulohumeral muscular dystrophy. *Science*, *329*(5999), 1650–1653.

Lemmers, R. J., van der Vliet, P. J., van der Gaag, K. J., Zuniga, S., Frants, R. R., de Knijff, P., & van der Maarel, S. M. (2010b). Worldwide population analysis of the 4q and 10q subtelomeres identifies only four discrete interchromosomal sequence transfers in human evolution. *The American Journal of Human Genetics*, *86*(3), 364–377.

Lemmers, R. J., Wohlgemuth, M., van der Gaag, K. J., van der Vliet, P. J., van Teijlingen, C. M., de Knijff, P., … van der Maarel, S. M. (2007). Specific sequence variations within the 4q35 region are associated with facioscapulohumeral muscular dystrophy. *The American Journal of Human Genetics*, *81*(5), 884–894.

Lyle, R., Wright, T. J., Clark, L. N., & Hewitt, J. E. (1995). The FSHD-associated repeat, D4Z4, is a member of a dispersed family of homeobox-containing repeats, subsets of which are clustered on the short arms of the acrocentric chromosomes. *Genomics*, *28*(3), 389–397.

Nguyen, K., Walrafen, P., Bernard, R., Attarian, S., Chaix, C., Vovan, C., … Lévy, N. (2011). Molecular combing reveals allelic combinations in facioscapulohumeral dystrophy. *Annals of Neurology*, *70*(4), 627–633.

Ottaviani, A., Rival-Gervier, S., Boussouar, A., Foerster, A. M., Rondier, D., Sacconi, S., … Magdinier, F. (2009). The D4Z4 macrosatellite repeat acts as a CTCF and A-type lamins-dependent insulator in facio-scapulo-humeral dystrophy. *PLoS Genetics*, *5*(2), e1000394, 1–11.

Padberg, G. W., & van Engelen, B. G. (2009). Facioscapulohumeral muscular dystrophy. *Current Opinion in Neurology*, *22*(5), 539–542.

Pirozhkova, I., Petrov, A., Dmitriev, P., Laoudj, D., Lipinski, M., & Vassetzky, Y. (2008). A functional role for 4qA/B in the structural rearrangement of the 4q35 region and in the regulation of FRG1 and ANT1 in facioscapulohumeral dystrophy. *PloS One*, *3*(10), e3389, 1–9.

Ricci, G., Scionti, I., Sera, F., Govi, M., D'Amico, R., Frambolli, I., … Tupler, R. (2013). Large scale genotype-phenotype analyses indicate that novel prognostic tools are required for families with facioscapulohumeral muscular dystrophy. *Brain*, *136*(Pt 11), 3408–3417.

Robin, J. D., Ludlow, A. T., Batten, K., Gaillard, M. C., Stadler, G., Magdinier, F., … Shay, J. W. (2015). SORBS2 transcription is activated by telomere position effect-over long distance upon telomere shortening in muscle cells from patients with facioscapulohumeral dystrophy. *Genome Research*, *25*(12), 1781–1790.

Sacconi, S., Camano, P., de Greef, J. C., Lemmers, R. J., Salviati, L., Boileau, P., … Desnuelle, C. (2012). Patients with a phenotype consistent with

facioscapulohumeral muscular dystrophy display genetic and epigenetic heterogeneity. *Journal of Medical Genetics*, *49*(1), 41–46.

Sacconi, S., Lemmers, R. J., Balog, J., van der Vliet, P. J., Lahaut, P., van Nieuwenhuizen, M. P., … van der Maarel, S. M. (2013). The FSHD2 Gene SMCHD1 is a modifier of disease severity in families affected by FSHD1. *The American Journal of Human Genetics*, *93*(4), 744–751.

Salort-Campana, E., Nguyen, K., Bernard, R., Jouve, E., Sole, G., Nadaj-Pakleza, A., … Attarian, S. (2015). Low penetrance in facioscapulohumeral muscular dystrophy type 1 with large pathological D4Z4 alleles: A cross-sectional multicenter study. *Orphanet Journal of Rare Diseases*, *10*, 2, 1–8.

Sarfarazi, M., Wijmenga, C., Upadhyaya, M., Weiffenbach, B., Hyser, C., Mathews, K., … Padberg, G. W. (1992). Regional mapping of facioscapulohumeral muscular dystrophy gene on 4q35: Combined analysis of an international consortium. *The American Journal of Human Genetics*, *51*(2), 396–403.

Scionti, I., Greco, F., Ricci, G., Govi, M., Arashiro, P., Vercelli, L., … Tupler, R. (2012). Large-scale population analysis challenges the current criteria for the molecular diagnosis of fascioscapulohumeral muscular dystrophy. *The American Journal of Human Genetics*, *90*(4), 628–635.

van der Maarel, S. M., Deidda, G., Lemmers, R. J., van Overveld, P. G., van der Wielen, M., Hewitt, J. E., … Frants, R. R. (2000). De novo facioscapulohumeral muscular dystrophy: Frequent somatic mosaicism, sex-dependent phenotype, and the role of mitotic transchromosomal repeat interaction between chromosomes 4 and 10. *The American Journal of Human Genetics*, *66*(1), 26–35.

van Deutekom, J. C., Wijmenga, C., van Tienhoven, E. A., Gruter, A. M., Hewitt, J. E., Padberg, G. W., … Frants, R. R. (1993). FSHD associated DNA rearrangements are due to deletions of integral copies of a 3.2 kb tandemly repeated unit. *Human Molecular Genetics*, *2*(12), 2037–2042.

van Overveld, P. G., Enthoven, L., Ricci, E., Rossi, M., Felicetti, L., Jeanpierre, M., … van der Maarel, S. M. (2005). Variable hypomethylation of D4Z4 in facioscapulohumeral muscular dystrophy. *Annals of Neurology*, *58*(4), 569–576.

van Overveld, P. G., Lemmers, R. J., Sandkuijl, L. A., Enthoven, L., Winokur, S. T., Bakels, F., … van der Maarel, S. M. (2003). Hypomethylation of D4Z4 in 4q-linked and non-4q-linked facioscapulohumeral muscular dystrophy. *Nature Genetics*, *35*(4), 315–317.

Vasale, J., Boyar, F., Jocson, M., Sulcova, V., Chan, P., Liaquat, K., … Higgins, J. (2015). Molecular combing compared to Southern blot for measuring D4Z4 contractions in FSHD. *Neuromuscular Disorders*, *25*(12), 945–951.

Wijmenga, C., Hewitt, J. E., Sandkuijl, L. A., Clark, L. N., Wright, T. J., Dauwerse, H. G., … Frants, R. R. (1992a). Chromosome 4q DNA rearrangements associated with facioscapulohumeral muscular dystrophy. *Nature Genetics*, *2*(1), 26–30.

Wijmenga, C., Sandkuijl, L. A., Moerer, P., van der Boorn, N., Bodrug, S. E., Ray, P. N., … Frants, R. R. (1992b). Genetic linkage map of facioscapulohumeral muscular dystrophy and five polymorphic loci on chromosome 4q35-qter. *The American Journal of Human Genetics*, *51*(2), 411–415.

SUPPORTING INFORMATION

Additional Supporting Information may be found online in the supporting information tab for this article.