Supplementary information

Materials and Methods

Patients' samples.

Blood samples were received at the Department of Medical Genetics, La Timone Children's Hospital in Marseille for molecular diagnosis and confirmation or exclusion of Facio-Scapulo Humeral Dystrophy. Individuals suspected of FSHD explored in this study were clinically assessed by neurologists with expertise in neuromuscular diseases who defined the presence or total absence of clinical signs and evaluated the involvement of the groups of muscle typically affected in the disease (facial, shoulder and pelvic girdle, upper and lower limbs and abdominal muscles). All patients described here displayed typical clinical signs of FSHD. A number of additional cases were also explored including volunteer at-risk relatives with various degrees of kinship. Informed consent was obtained from all patients for the genetic analyses.

DNA extraction and plug preparation.

For MC, PBMCs were prepared from fresh blood after red blood cell lysis. A 45 µL PBS suspension containing $5x10^5$ to $1x10^6$ PBMCs was mixed with an equal volume of 1.2 % Nusieve GTG agarose (Lonza, Basel, Switzerland) prepared in 1X PBS, previously equilibrated at 50°C. Plugs were left to solidify for 30 min at 4°C, then DNA was purified by an overnight incubation at 50°C in 250 µL of 0.5 M EDTA pH 8.0; 1 % Sarkosyl (Sigma-Aldrich, Saint Louis, MO, USA), 2 mg/mL proteinase K (Eurobio, Les Ulis, France), after which plugs were washed three times at room temperature in 10mM Tris, 1mM EDTA pH 8.0. Plugs were then either stored at 4°C in 0.5 M EDTA pH 8.0 or used immediately. Stored plugs were rinsed three times in 10 mM Tris, 1 mM EDTA pH 8.0 prior to use [1].

Hybridization.

Staining of DNA fibers was obtained by 1 h incubation in 40 mM Tris, 2 mM EDTA pH8 containing 3 µM Yoyo-1 (Life Technologies, Carlsbad, CA, USA) at room temperature. For

each patient, a single plug was transferred in 1 mL of 0.5 M MES pH 5.5, incubated at 68°C for 20 min to melt the agarose, and then incubated at 42°C overnight with 1.5 U beta agarasel (New England Biolabs, Ipswich, MA, USA). The solution was transferred to a combing vessel and DNA combing was performed with the Molecular Combing System on Combislide coverslips (Genomic Vision, Paris, France). After dehydration (4 h at 60°C), slides were either stored at -20°C for later use or immediately hybridized. Visual quality assessment of the combing (linearity, density, stretching and absence of breakage) is verified with an epifluorescence microscope equipped with an FITC filter and a 40x air objective. Prior to hybridization, coverslips are dehydrated by successive 3-min incubations in 70%, 90% and 100% ethanol baths and air-dried for 10 min at room temperature. Twenty µL of the probe mix were spread on the coverslip which was left to denature for 5 min at 90°C and hybridized overnight at 37°C on a SPoT-Light CISH Hybridizer. The coverslip was washed once for 5 min in 50% formamide, 1X SSC, then three times 3 min in SSC 2x. Detection was performed with three successive layers of fluorophore-conjugated antibodies or streptavidin, with a 20 min-incubation at 37°C in a humid chamber for every layer and three 3 min-washes in 2x SSC, 0.1% Tween at room temperature. Three additional 3 min-washes in PBS and dehydration by successive 3 min-washes in 70%, 90% and 100% ethanol were performed before mounting and visualization.

Molecular combing

The FSHD molecular combing procedure was reported in [1]. Briefly, for each sample, high molecular weight DNA in solution was combed on a glass coverslip using a motorized combing platform (Genomic Vision, France), with a uniform and constant stretching coefficient. A set of 7 fluorescent probes targeting the *D4Z4* array and its flanking regions including the qA/qB region on chromosomes 4 and 10, was hybridized on individual combed DNA fibers (V3 pink bar-code, Figure 1). The coverslip was entirely scanned by the ImageXpressMicro automated epifluorescence scanner at a 40X magnification (Molecular Devices, San Diego CA). Image analysis was performed automatically using the Combilog

Software (Genomic Vision, France) and reviewed manually by an experimented operator. For each allele, the ends of the segments are determined. Every signal is automatically attributed to a chromosome and/or haplotype, based on comparison of measurements to the theoretical lengths for the different motifs. All probe and gap sizes are automatically calculated. Only intact *D4Z4* signals were considered (i.e. signals comprising at least one probe at each end of the *D4Z4* signal, confirming that no fiber breakage had occurred within the *D4Z4* probe). Data were displayed as histograms for the determination of type of chromosome (4 or 10), haplotype and size. Further calculations, including average repeat array size for each allele, confidence intervals and number of signals for each independent allele were performed automatically. A minimum of 60 signals, 15 for each allele is analyzed per patient. A pathological allele was considered when a contracted *D4Z4* array with size inferior or equal to ten repeated units on a chromosome 4q is detected.

Characterization of 4qA-associated pLAM sequences

D4Z4-pLAM sequences were amplified using primers able to amplify either the short (4qAS) or long (4qAL) fragment distal to the last D4Z4 unit using described primers [2]. Amplification was done using the high fidelity Taq Platinum from Life technologies on 20ng of genomic DNA consisting in and initial denaturation step of 2 min at 98°C followed by 25 cycles of denaturation for 10 sec at 98°C, annealing for 30 sec at 64°C, elongation for 20 sec at 72°C followed by a final elongation of 2 min at 72°C. PCR fragments were analyzed on a 2% agarose gel and sequences were systematically verified by Sanger Sequencing. Sequences were aligned to the MF422078 (DUX4La) and MF422079 (DUX4Lb) sequences of reference [2].

Exome and Sanger sequencing

Exome capture with Agilent SureSelect All Exon kit V5 was applied to DNA samples from 10 probands in the families carrying the rearrangement, followed by paired-end sequencing by Illumina Technology (San Diego, CA, USA). Reads quality was assessed using fastqc 0.11.5

(FastQC; A Quality Control tool for High Throughput Sequence Data; http://www.bioinformatics.babraham.ac.uk/projects/fastqc/ by S. Andrews). After sequencing, data were mapped to the hg19 human reference assembly genome using the Maximum Exact Matches algorithm in Burrows–Wheeler Aligner (BWA) [3]. On average, 96% of target sequences were successfully covered with a depth ≥ 20X, with a mean depth > 65 X. Alignment quality was evaluated using qualimap 2.2.1 [4]. We used the Genome Analysis Software Kit (GATK) pipeline following GATK best-practice recommendation to performed variant calling [5]. PCR duplicates were marked using Sambamba tools (http://lomereiter.github.io/sambamba) [6]. The GATK base quality score recalibrator was applied to correct sequencing artifacts, and variants were called using haplotype caller. Variant annotation, mining and manual review were performed using VarAFT (http://varaft.eu) and ALAMUT software (Interactive Biosoftware, Rouen, France).

For *SMCHD1* direct sequencing primers were designed using the Primer3 software (http://frodo.wi.mit.edu/). PCR products were purified by Exonuclease I (New England Biolabs) and Rapid Alkaline Phosphatase (Roche) reactions, subject to Big Dye Terminator v3.1 sequencing reaction (Applied Biosystems) and analyzed by 3130xl Genetic Analyzer (Applied Biosystems) in accordance with manufacturer's instructions. Reactions were run for 35 cycles with the following amplification conditions: 95°C for 20 s, 59°C for 40 s and 72°C for 1 min and 40 s.

Bioinformatic predictions

Amino acid substitutions were predicted for their impact on transcript and protein function by UMD-Predictor [7]. Impact of variants on *SMCHD1* splicing was predicted by Human Splicing Finder, HSF, version 3.0 (cut-off for donor or acceptor sites is 55%). None of variants included for experimental analysis are reported from dbSNP149.

.

Supplementary tables

Supplementary table 1. Size distribution of D4Z4 arrays carried by chromosome 4 for the different subgroups of alleles: short (<35 kb) or long (>35 kb) D4Z4 arrays on A-type 4q arms; short (<35 kb) or long (>35 kb) D4Z4 arrays on B-type 4q arms. For each, the mean size and range are indicated in kb together with the number of D4Z4 units (Mean RU) and range.

	4qA (<35 kb)	**4qA (>35 kb)**	**4qB (<35 kb)**	**4qB (>35 kb)**
n	86	193	8	149
Mean size (kb)	18,025	108,9	23,925	83
Range (kb)	5.8-34.65	35-338	11.6-34.7	36-350
Mean RU	5	33	7	25
Range (RU)	2-10	11-102	3-10	11-106

Supplementary table 2. Size distribution of D4Z4 arrays carried by chromosome 10 for the different subgroups of alleles: short (<35 kb) or long (>35 kb) D4Z4 arrays on A-type 10q arms; short (<35 kb) or long (>35 kb) D4Z4 arrays on B-type 10q arms. For each, the mean size and range are indicated in kb together with the number of D4Z4 units (Mean RU) and range.

	10qA (<35 kb)	**10qA (>35 kb)**	**10qB (<35 kb)**	**10qB (>35 kb)**
n	114	299		25
Mean size (kb)	24.4	79.2		66
Range (kb)	5-35	36-318		38-166
Mean number (RU)	7	24		20
Range (RU)	1-10	11-96		11-50

Supplementary table 3. Analysis of the 4q35 distal region for the 4qA allele with short (<35 kb) or long (>35 kb) D4Z4 arrays. Mean size and size range of the pLAM and β satellite-containing region of the 4qA allele (β sat), telomeric distal probe, gap between the two probes and total size of the distal region are indicated in kb.

	Short (< 35 kb) 4qA alleles					**Long (> 35 kb) 4qA alleles**				
	β sat	Gap	β sat + gap	gap-Telo	Total	β sat	Gap	β sat + gap	β sat-gap-Telo	gap-Telo
n	38	38	38	38	38	76	76	76	76	
Mean size (kb)	7.5	6.6	14	16	22.5	6.42	8.8	15	24.3	17.2
Range (kb)	2.7-11.6	3-11	9-22	3-22	13-29.4	3.7-13	3.2-12	7.9-24	12.4-30	5.7-25

Supplementary table 4. Analysis of the 10q26 distal region for the qA allele with short (<35 kb) or long (>35 kb) D4Z4 arrays. Mean size and size range of the A type allele-specific probe allele (red) distal to D4Z4, telomeric distal probe, gap between the two probes and total size of the distal region are indicated in kb.

	Short (< 35 kb) 10qA alleles					**Long (> 35 kb) 10qA alleles**				
	β sat	Gap	β sat + gap	gap-Telo	Total	β sat	Gap	β sat + gap	gap-Telo	Total
n	38	38	38	38	38	120	120	120	120	120
Mean size (kb)	5.8	8.2	14.5	16.9	23.5	7.3	7.7	14.7	17.2	24.2
Range (kb)	3.7-12.3	3.6-13.3	9.4-23.5	6.1-24.6	11.8-29.8	2.9-12.5	1.9-12.5	9-22.3	3.6-24.2	11.1-31.8

Supplementary table 5. Analysis of the 4q35 distal region for the 4qB allele with short (<35 kb) or long (>35 kb) D4Z4 arrays. Mean size and size range of the B type allele-specific probe allele (blue) distal to D4Z4, telomeric distal probe, gap between the two probes and total size of the distal region are indicated in kb.

	Short (< 35 kb) 4qB alleles					**Long (> 35 kb) 4qB alleles**				
	Blue probe	Gap	Blue + gap	Gap-Telo	Total	Blue probe	Gap	Blue + gap	Gap-Telo	Blue + gap + Telo
n	-	-	-	-	-	57	51	57	51	57
Mean size (kb)	-	-	-	-	-	5.8	12.6	18.5	20.2	25.8
Range	-	-	-	-	-	3.9-7	4.7-	4.8-	4.7-	4.8-

(kb)							16.7	22	25	30.4

Supplementary table 6. Analysis of the 10q26 distal region for the qB allele with short (<35 kb) or long (>35 kb) D4Z4 arrays. Mean size and size range of the B type allele-specific probe allele (blue) distal to D4Z4, telomeric distal probe, gap between the two probes and total size of the distal region are indicated in kb.

	Short (< 35 kb) 10qB alleles					**Long (> 35 kb) 10qB alleles**				
	Blue probe	Gap	Blue + gap	Gap-Telo	Total	Blue probe	Gap	Blue + gap	Gap-Telo	Total
n	-	-	-	-	-	11	11	11	11	11
Mean size (kb)	-	-	-	-	-	5.7	12.8	19.2	20.7	27.1
Range (kb)	-	-	-	-	-	2.4-6.8	5.3-15.2	9.4-21.7	5.3-24.4	9.4-30.9

Supplementary table 7. Patients with FSHD with mosaicism of the 4q35 locus.

Patient ID	Gender	Clinical Status	% of short 4qA	qA/qB	10q	*SMCHD1* status
130402C	M	Affected		qA 2-3 RU qA 31 RU qB 18 RU	qA 16 RU qA 16 RU	ND
B2015-2148	M	Affected	22	4qA-4 RU 4qA-17 RU 4qA-19 RU	10qA-27 RU 10qA-32 RU	ND
130321A	F	Affected	ND	4qA-7 RU 4qA-19 RU 4qA-40 RU	10qA-17 RU 10qA-17 RU	ND
19273	F	Affected	42	4qA-2 RU 4qA-12 RU 4qA-44 RU	10qA-12 RU 10qA-32 RU	WT
19274	M	Affected	52	4qA-2 RU 4qA-20 RU 4qB-20 RU	10qA-5 RU 10q4-23 RU	WT
16111	F	Affected	31	4qA-6 RU 4qA-20 RU 4qB-23 RU	10qA-9 RU 10qA-26 RU	ND
19184	M	Affected	23	4qA-6 RU 4qA-23 RU 4qA-20 RU	10qA-6 RU 10qA-11 RU	WT
19183	M	Affected	14	4qA-2 RU 4qA-28 RU ?	10qA-18 RU 10qA-21 RU	WT
19182	M	Affected	34	4qA-3 RU 4qA-32 RU 4qA-19 RU	10qA-18 RU 10qA-18 RU	WT
19275	M	Affected	42	4qA-2 RU 4qA-36 RU 4qB-24 RU	10qA-14 RU 10qA-41 RU	WT
19272	F	Affected	45	4qA-4 RU 4qA-31 RU 4qA-34 RU	10qA-22 RU 10qA-16 RU	WT
19186	M	Affected	21	4qA-3 RU 4qA-32 RU 4qB-14 RU	10qA-19 RU 10qA-23 RU	WT
19276	M	Affected	24	4qA-6 RU	10qA-	WT

				4qA-15 RU 4qB-45 RU	14 RU 10qA- 27 RU	
19187	F	Affected	45	4qA-2 RU 4qA-22 RU 4qB-32 RU	10qA- 14 RU 10qA- Rearr	ND
120405B	F	Affected	10	4qA-2 RU 4qA-11 RU 4qB-22 RU	10qA- 38 RU 10qA- 14 RU	ND
19181	M	Affected	29	4qA-3 RU 4qA-48 RU 4qB-19 RU	10qA- 8 RU 10qA- 12 RU	WT
17706	F	Affected	25	4qA-2 RU 4qA-15 RU 4qB-41 RU	10qA- 6 RU 10qA- 14 RU	ND
19179	M	Affected	13	4qA-1 RU 4qA-39 RU 4qB-46 RU	10qA- 17 RU 10qA- 35 RU	WT
TalM	M	Non affected	6	qA 2 RU qA 15 RU qB 27 RU	qA 30 RU qA rearr.	ND
19180	M	Affected	29	4qA-3 RU 4qA-21 RU 4qB-14 RU	10qA- 6 RU 10qA- 35 RU	WT
19271	F	Affected	35	4qA-3 RU 4qA-20 RU 4qB-29 RU	10qA- 22 RU 10qA- 23 RU	WT
WARM	F	Affected	12	4qA- 2 RU 4qA- 17 RU 4qA 33 RU	10qA 9 RU 10qA 15 RU	ND

Supplementary table 8. Patients affected with FSHD carrying a 4q35 deletion encompassing the p13E11 (D4F104S1) region. We tested the presence of the short or long pLAM sequence in patients 121108A, 14668, 19185 (underlined). All three are carriers of the short pLAM sequence associated to chromosome 4q. Data are presented in supplementary figure 5

Patient ID	Gender	Clinical Status	qA/qB	pLAM	10q	*SMCHD1* status
121108A	F	Affected	4qA-3 RU ; delp13E11 4qB-39 RU	Short 260bp 4qA fragment	10qA-6 RU 10qA-11 RU	WT
14668	M	Affected	4qA-12 RU ; delp13E11 4qA-21 RU	Short 260bp 4qA fragment	10qA-8 RU 10qA-14 RU	WT
19185	M	Affected	4qA-12 RU ; delp13E11 4qA-51 RU	Short 260bp 4qA fragment	10qA-21 RU 10qA-30 RU	WT
100519A-III$_1$ (Index case)	F	Affected	4qA-4 RU ; delp13E11 4qA-37 RU	N/A	10qA-21 RU 10qA-28 RU	WT
100519A-II$_1$ (Father)	M	Affected	4qA-4 RU ; delp13E11 4qA-47 RU	N/A	10qA-32 RU 10qA-32 RU	ND
100519A-III$_5$ (Sister)	F	Non affected	4qA-42 RU 4qA-42 RU	N/A	10qA-29 RU 10qA-29 RU	ND
100519A-III$_6$ (Brother)	M	Affected	4qA-4 RU ; delp13E11 4qA-37 RU	N/A	10qA-23 RU 10qA-28 RU	ND

Supplementary table 9. Rearrangement of the proximal 4q region

Patient ID	Sex	Clinical Status	qA/qB	10q	*SMCHD1* status
110718J	M	Affected	4qA-25 RU 4qB-36 RU + Presence of a D4Z4-qB signal in the proximal region	10qA-10 RU 10qA-21 RU	WT

Supplementary table 10. Analysis of the D4Z4 proximal region for the 4qA, 4qB, 10qA and 10qB alleles with short (<35 kb) or long (>35 kb) D4Z4 arrays. Mean size and size range of the centromeric probes (blue and pink) proximal to D4Z4, the gap between D4Z4 probe and the blue centromeric probe and the gap between the blue and the pink probes (kb).

	4qA (<35kb)			
	Gap 1	Centromeric probe (blue)	Gap2	Centromeric probe (pink)
n	38	38	38	38
Mean size (kb)	4	33	9.4	18
Range (kb)	2.8-30	11-37	5-25.3	3.5-21.9
	4qA(>35kb)			
	Gap1	Centromeric probe (blue)	Gap2	Centromeric probe (pink)
n	78	78	75	75
Mean size (kb)	4.6	33.3	7.6	18.3
Range (kb)	2.6-7.1	17.4-37.7	4.4-25.4	11.8-21.5
	4qB			
	Gap1	Centromeric probe (blue)	Gap2	Centromeric probe (pink)
n	61	61	59	59
Mean size (kb)	4.6	33.3	9	18.5
Range (kb)	3-7.4	24.8-36	3.7-26.6	10.1-22.6

	10qA (<35kb)			
	Gap1	Centromeric probe (blue)	Gap2	Centromeric probe (pink)
n	37	37	28	28
Mean size (kb)	3.9	34.1	8.9	16.4
Range (kb)	2.7-10.4	24.8-62.9	4.6-13.2	5.5-23.8
	10qA (> -35kb)			
n	Gap1	Centromeric probe (blue)	Gap2	Centromeric probe (pink)
Mean size (kb)	123	123	102	102
Range (kb)	4.5	34.1	9.1	16.5
n	2.3-11.7	22.7-65.1	3.8-36.8	5.3-36.8
	10qB			
n	Gap1	Centromeric probe (blue)	Gap2	Centromeric probe (pink)
Mean size (kb)	11	11	11	11
Range (kb)	4.2	34.6	9.3	19.7
n	2.5-5.3	33.2-45	6.2-11.2	6.5-24.4

Supplementary table 11. Analysis of the D4Z4 proximal region for the 4qA, 4qB, 10qA and 10qB alleles with short (<35 kb) or long (>35 kb) D4Z4 arrays. Mean size and size range of the centromeric region probes (blue and pink probes) proximal to D4Z4 are indicated in kb.

	4qA (<35 kb)	**4qA (>35 kb)**	**4qB**	**10qA (<35kb)**	**10qA (>35 kb)**	**10qB**
n	38	78	61	37	123	11
Mean size (kb)	65.1	63.4	65.2	59	62.6	67.8
Range (kb)	52.9-77.9	22.3-80.1	36.6-78.5	34.5-73.2	32.3-115.6	61.6-77

Legends to supplementary figures.

Supplementary figure 1. D4Z4 array size at 4q and 10q subtelomeres carrying long D4Z4 arrays (>35 kb).

A. We determined the distribution in size of long D4Z4 array (>35 kb) carried by 4qA or 10qA chromosomes by analyzing signals obtained by molecular combing. Scattergrams display the size distribution in kilobases (kb); 4qA chromosomes, n= 193; mean size=108.9kb (range: 11-102 RU); 10qA chromosomes, n= 299; mean size=79.2 (range: 36-318 kb). 4qA alleles are on average longer than 10q alleles ($p<0.0001$). **B.** We determined the distribution of D4Z4 array size carried by 4qB or 10qB chromosomes by analyzing signals obtained by molecular combing. Scattergrams display the size distribution in kilobases (kb); 4qB chromosomes, n= 149; mean size= 83kb (range: 36-350); 10qB chromosomes, n= 25; mean size= 66kb (range: 38-166 kb). 4qB alleles are on average longer than 10q alleles ($p<0.003$). For the different panels, detailed analyses are provided in Supplementary tables 1 and 2. In the different panel, the size is indicated in kilobases (kb). The mean size is indicated by a red line. Differences in size distribution were determined using a non-parametric Kruskal Wallis test with pair-wise comparison and Bonferroni correction for false positive. ***, $p<0.0001$; *, $p=0.003$.

Supplementary figure 2. Sequence length variation at the distal 4q and 10q subtelomeres.

Scattergrams of the qA distal region containing βsatellite element for chromosome 4 and 10 with short (< 35 kb) or long (>35 kb) D4Z4 arrays. The size is indicated in kilobases (kb). Mean size is shown by a red line. Detailed analyses are provided in supplementary table 3 and 5. The estimated size of the red probe is 6kb. Differences in size distribution were determined using a non-parametric Kruskal Wallis test with pair-wise comparison and Bonferroni correction for false positive. Scattergram highlight the large variability in size of the A-type haplotype containing arrays of variable sizes of tandemly repeated β satellite

elements. For short 4qA alleles (D4Z4 array <35kb), the size of the distal region ranges between 2.7 and 11.6 kb (mean size: 7.5kb; n= 38). For long D4Z4 arrays (>35 kb, 4qA), the size of the distal region ranges between 3.7 and 13kb (mean size: 6.42 kb, n= 76). The distal 4qA is significantly longer in patients affected with FSHD (p value= 0.03). On the opposite, size of the distal region is smaller for short 10qA alleles (mean size, 5.8 kb, range= 3.7-12.3 kb, n= 38) compared to long 10qA alleles (mean size, 7.3kb, range= 2.9-12.5 kb, n=120) (p value= 0.038).

Supplementary figure 3. Analyses of 4q35 and 10q26 proximal regions.

For each panel, the region analyzed is indicated by an arrow. Scattergrams display the size of the different regions in kilobases (kb). Mean size is shown by a red line. Detailed analyses are provided in supplementary table 10. **A.** The gap upstream of the first D4Z4 repeat was analyzed for 348 signals. This gap is not significantly different for the different conditions with a mean size ranging between 2.3 to approximately 10kb, except for one sample with a size of 30kb. Of note, this region is more variable than expected (5kb). **B.** Size distribution of the p13E11 blue probe. The size is estimated to be 20 kb. The calculated mean size ranges between 33 to 34.6 kb for 348 signals analyzed. **C.** Distribution of the size between the blue and magenta probes ranging between 3.7 and 36.8 kb. **D.** Distribution of the size of the magenta probe, estimated to be 10 kb. The mean size ranges between 3.5 to 36.8 kb for the different types of alleles with a broad distribution between the different situations.

Supplementary figure 4. Pedigree of the different families carrying an additional 10q allele. Details are provided in table 12.

Supplementary figure 5. Analysis of the pLAM sequence in patients with atypical genomic features. PCR genotyping for discrimination between short (260bp) or long (330bp) pLAM sequences in different patients carrying an atypical 4q or 10q genomic structure.

1. 14668, p13E11 deletion (sup table 8). **2.** Family 3-II2, not affected. **3.** Family 3-II1, affected. **4.** Family 3-I2, not affected. **5.** Family 3-I2, not affected. **6.** B2015-085 (Table 1). **7.** Family 2-II3, affected (Table 1). **8.** Family 1-II1, affected. **9.** 19185, p13E11 deletion (sup table 8). **10.** Family 5-II1, affected. **11**. Family 5-I1, not affected. **12.** Family 5-I2, not affected. **13.** Family 4-I2, not affected. **14.** Family 4-I3, not affected. **15.** Family 4-II2, affected. **16.** B2013-1393 (Table 1). **17.** B2016-1344 (Table 1). **18.** B2016-2376 (Table 1). **19.** B2016-2473 (Table 1). 20. 121108A, p13E11 deletion (sup table 8). **21.** FSHD1 (8RU). **22.** FSHD1 (8RU).

References

1 Nguyen, K. *et al.* Molecular combing reveals allelic combinations in facioscapulohumeral dystrophy. *Ann Neurol* **70**, 627-633, doi:10.1002/ana.22513 (2011).

2 Lemmers, R. J. *et al.* Deep characterization of a common D4Z4 variant identifies biallelic DUX4 expression as a modifier for disease penetrance in FSHD2. *Eur J Hum Genet* **26**, 94-106, doi:10.1038/s41431-017-0015-0 (2018).

3 Li, H. & Durbin, R. Fast and accurate short read alignment with Burrows-Wheeler transform. *Bioinformatics* **25**, 1754-1760, doi:10.1093/bioinformatics/btp324 (2009).

4 Okonechnikov, K., Conesa, A. & Garcia-Alcalde, F. Qualimap 2: advanced multi-sample quality control for high-throughput sequencing data. *Bioinformatics* **32**, 292-294, doi:10.1093/bioinformatics/btv566 (2016).

5 Van der Auwera, G. A. *et al.* From FastQ data to high confidence variant calls: the Genome Analysis Toolkit best practices pipeline. *Curr Protoc Bioinformatics* **43**, 11 10 11-33, doi:10.1002/0471250953.bi1110s43 (2013).

6 Tarasov, A., Vilella, A. J., Cuppen, E., Nijman, I. J. & Prins, P. Sambamba: fast processing of NGS alignment formats. *Bioinformatics* **31**, 2032-2034, doi:10.1093/bioinformatics/btv098 (2015).

7 Salgado, D., Bellgard, M. I., Desvignes, J. P. & Beroud, C. How to Identify Pathogenic Mutations among All Those Variations: Variant Annotation and Filtration in the Genome Sequencing Era. *Hum Mutat* **37**, 1272-1282, doi:10.1002/humu.23110 (2016).