Dysregulation of 4q35- and muscle-specific genes in fetuses with a short D4Z4 array linked to facio-scapulo-humeral dystrophy

Natacha Broucqsault, Julia Morere, Marie-Cécile Gaillard, Julie Dumonceaux, Julia Torrents, Emmanuelle Salort Campana, André Maues de Paula, Marc Bartoli, Carla Fernandez, Anne Chesnais, et al.

Dysregulation of 4q35- and muscle-specific genes in fetuses with a short D4Z4 array linked to facio-scapulo-humeral dystrophy

Natacha Broucqsault[1,†], Julia Morere[1,†], Marie-Cécile Gaillard[1], Julie Dumonceaux[6], Julia Torrents[2], Emmanuelle Salort-Campana[1,3], André Maues De Paula[1,2], Marc Bartoli[1], Carla Fernandez[2], Anne Laure Chesnais[2], Maxime Ferreboeuf[6], Laure Sarda[2], Henry Dufour[3], Claude Desnuelle[7], Shahram Attarian[1,4], Nicolas Levy[1,5], Karine Nguyen[1,5], Frédérique Magdinier[1,†,*] and Stéphane Roche[1,†]

[1]Aix Marseille Université, INSERM UMR_S910, 27 boulevard J. Moulin, 13005 Marseille, France, [2]APHM, Laboratoire d'Anatomopathologie, [3]APHM, Service de Neurochirurgie, [4]APHM, Service de Neurologie, [5]APHM, Laboratoire de génétique médicale, Hôpital de la Timone, 264 Rue Saint-Pierre, 13385 Marseille, France, [6]UM76–UPMC, Inserm U974, CNRS UMR7215, 47, boulevard de l'Hôpital, La Pitié Salpétrière-Bâtiment Babinski, 75651 Paris Cedex 13, France and [7]Centre de Référence des Maladies Neuromusculaires, Nice University Hospital, 06000 Nice, France

Facio-scapulo-humeral dystrophy (FSHD) results from deletions in the subtelomeric macrosatellite *D4Z4* array on the 4q35 region. Upregulation of the *DUX4* retrogene from the last *D4Z4* repeated unit is thought to underlie FSHD pathophysiology. However, no one knows what triggers muscle defect and when alteration arises. To gain further insights into the molecular mechanisms of the disease, we evaluated at the molecular level, the perturbation linked to the FSHD genotype with no a priori on disease onset, severity or penetrance and prior to any infiltration by fibrotic or adipose tissue in biopsies from fetuses carrying a short pathogenic *D4Z4* array ($n = 6$) compared with fetuses with a non-pathogenic *D4Z4* array ($n = 21$). By measuring expression of several muscle-specific markers and 4q35 genes including the *DUX4* retrogene by an RT-PCR and western blotting, we observed a global dysregulation of genes involved in myogenesis including *MYOD1* in samples with <11 *D4Z4*. The *DUX4-fl* pathogenic transcript was detected in FSHD biopsies but also in controls. Importantly, in FSHD fetuses, we mainly detected the non-spliced *DUX4-fl* isoform. In addition, several other genes clustered at the 4q35 locus are upregulated in FSHD fetuses. Our study is the first to examine fetuses carrying an FSHD-linked genotype and reveals an extensive dysregulation of several muscle-specific and 4q35 genes at early development stage at a distance from any muscle defect. Overall, our work suggests that even if FSHD is an adult-onset muscular dystrophy, the disease might also involve early molecular defects arising during myogenesis or early differentiation.

INTRODUCTION

Facio-scapulo-humeral dystrophy (FSHD) is an autosomal-dominant disorder, ranked as the most prevalent muscular dystrophy with an incidence of 7 of 100 000 (http://www.orpha.net). Symptoms usually arise between the age of 20–40. Clinically, the disease manifests predominantly as facial weakness with progression to the upper body and then to the lower extremities with a characteristic asymmetric involvement of certain groups of skeletal muscles. There is a marked inter-

*To whom correspondence should be addressed at: Laboratory of Medical Genetics and Functional Genomics, UMR S_910, INSERM Aix-Marseille University 27, Bd Jean Moulin, 13005 Marseille, France. Tel: +33 491324908; Email: frederique.magdinier@univ-amu.fr
†The authors wish it to be known that in their opinion the first two and the last two authors should be regarded as joint authors.

and intra-familial heterogeneity in FSHD clinical expression, which probably depends on multiple genetic and environmental factors unknown yet. Importantly, FSHD is linked to deletion of an integral number of a 3.3 kb tandem macrosatellite repeat arranged as a head-to-tail array in the subtelomeric 4q35 region. Typically, non-affected individuals carry between 11-100 copies of this *D4Z4* element while patients with FSHD present between 1-10 units. Distal to *D4Z4*, a region defines two allelic forms, 4qA and 4qB. The 4qA sequence is characterized by the presence of an array of the 68 bp ß-satellite repetitive DNA and the pLAM sequence abutting the last *D4Z4* repeat and containing a polyadenylation site (1,2). Both qA and qB alleles are equally common in the population but FSHD is associated with the 4qA allele (3). In unaffected individuals, long *D4Z4* arrays harbor heterochromatin features, whereas FSHD-linked short *D4Z4* arrays contain epigenetic marks of non-transcribed euchromatin (4–6). *D4Z4*, which contains >70% of CpG sites is hypermethylated at the DNA level in normal cells but hypomethylated in both 4q-linked (FSHD1) and phenotypic (FSHD2) FSHD.

So far, the search for a FSHD candidate gene has been controversial. The expression of several 4q35 genes has been determined in biopsies and primary myoblasts derived from patients, but their involvement and dysregulation in FSHD remain unclear with some authors observing a transcriptional deregulation of several of them (7–10) while other do not (11–13). Among the candidate genes for FSHD, the *DUX4* retrogene localized within *D4Z4* is suspected to play a key role (1,2,14,15). Each *D4Z4* contains in its distal part an open reading frame encoding a putative protein containing two homeoboxes (1). In individuals carrying the 4qA allele downstream of the repeat, this *DUX4* sequence can be transcribed through the last *D4Z4* repeat and the proximal 4qA sequence leading to the production of a pre-messenger RNA with introns and a polyadenylation site (2,14,15). This pre-messenger RNA can be spliced into at least three different mRNAs of different sizes. In patient's samples, full-length *DUX4* transcripts (*DUX4-fl*) are detected at a very low level (in 1 of 1000 muscle cell nuclei) (15). In the current model, *DUX4-fl* might encode a toxic double homeobox transcription factor responsible for the activation of >500 other genes in skeletal muscle (16–18). Still, a major issue with the *DUX4* hypothesis is its extremely low abundance, the difficulty to detect the protein in patient's biopsies limiting its use as a biomarker for diagnosis or prognosis and the time window in which *DUX4* becomes pathogenic.

Thus, the search for genes dysregulated in FSHD led so far to controversial conclusions and different reasons might explain these discrepancies. Furthermore, if the genetic defect is present constitutively in most of the FSHD cases, the clinical signs of the disease only appear late in life and nothing is known on the molecular features of the presymptomatic muscle, in particular during development. Therefore, we investigated whether biopsies from fetuses carrying a short *D4Z4* array display a developmental-specific phenotype compared with normal samples and quantified the expression level of different muscle-specific and 4q35 genes, including *DUX4*. Collectively, our data reveal changes in expression of several muscle-specific genes in fetuses with a short *D4Z4* array compared with normal specimen together with the dysregulation of several 4q35 genes, suggesting that muscle homeostasis is affected as early as the fetal stage in FSHD.

RESULTS AND DISCUSSION

Key questions with regard to FSHD is how and when the muscular phenotype arises, the identity of the gene(s) involved in the pathology and the link between *D4Z4* and the regulation of the FSHD-causing gene(s) in specific muscles. Therefore, to evaluate at the molecular level the perturbation linked to the FSHD genotype with no a priori on disease onset, severity or penetrance and prior to any infiltration by fibrotic or adipose tissue, we explored the expression level of several muscle-specific genes during the second trimester of gestation in six fetuses carrying a short pathogenic *D4Z4* array (FSHD carriers) compared with fetuses with a long non-pathogenic *D4Z4* array (controls, $n = 21$) at different gestational ages by measuring the expression of (i) several muscle-specific genes, (ii) *DUX4* and (iii) different 4q35 genes.

During the human intrauterine development, multinucleated primary myotubes appear around the 5th week of gestation, early muscle fibers around the 11th week while after the 20th week, most muscle fibers are packed with myofibrils and display peripheral nuclei similar to adult muscle. In fetuses, muscle mass increases predominantly by proliferative growth of myoblasts. Between the 20th and 24th weeks of gestation, innervation enhances muscle development and differentiation. Up to the 25th week, muscle displays a hyperplasic stage with an increase in cell number. Then, the cell size and muscle mass increase rapidly by hypertrophy while the postnatal growth of muscle is mostly characterized by remodeling of pre-existing fibers (19). Two types of muscle fibers are present during the fetal period, but the distinction cannot be made until the 18–20th week of gestation.

With regard to the developmental pattern of muscle-specific genes, our results in quadriceps indicate an increase in the expression of several genes encoding structural proteins such as sarcoglycan (7.7-fold), calpain 3 (4.4-fold), dysferlin (3.1-fold). Also upregulation of several genes encoding the myosin heavy chain was evidenced, either for genes corresponding to type I, slow oxidative red fibers (*MYH1*, 8.25-fold; *MYH7*, 3.75-fold) or type II, fast oxidative red fibers, *MYH2* (56-fold) in FSHD fetuses (Fig. 1A; Supplementary Material, Table S3; at least 2-fold change in expression; P-value <0.05). Furthermore, a significant upregulation of genes encoding proteins involved in muscle remodeling (*MURF1 (TRIM63)*, 5.3-fold) and calcium release was also observed at the RNA (*RYR1-1*, 10.7-fold; *CACNA1S*, 7.7-fold) and protein level (*RYR1*, 2.8-fold, Fig. 1B). Concerning transcription factors involved in muscle differentiation, both *MYOD1* and *MRF4* are upregulated in FSHD fetuses (6.8- and 21.4-fold, respectively, Fig. 1A; Supplementary Material, Table S3).

Using the same criteria, expression analysis was done on adult quadriceps biopsies from six non-carrier individuals and seven FSHD patients. Among the different genes mentioned above, only *MYOD* was found to be significantly upregulated (3.8-fold) in FSHD as described elsewhere (13) (Supplementary Material, Table S3,and Fig. S1). Interestingly, *MMP9* expression usually modulated in dystrophic muscle is not modified in FSHD fetuses (Fig. 1A), but seems decreased in adult FSHD patients (Supplementary Material, Fig. S1).

The development and plasticity of striated muscle are due to finely tuned networks acting at different levels from the prenatal

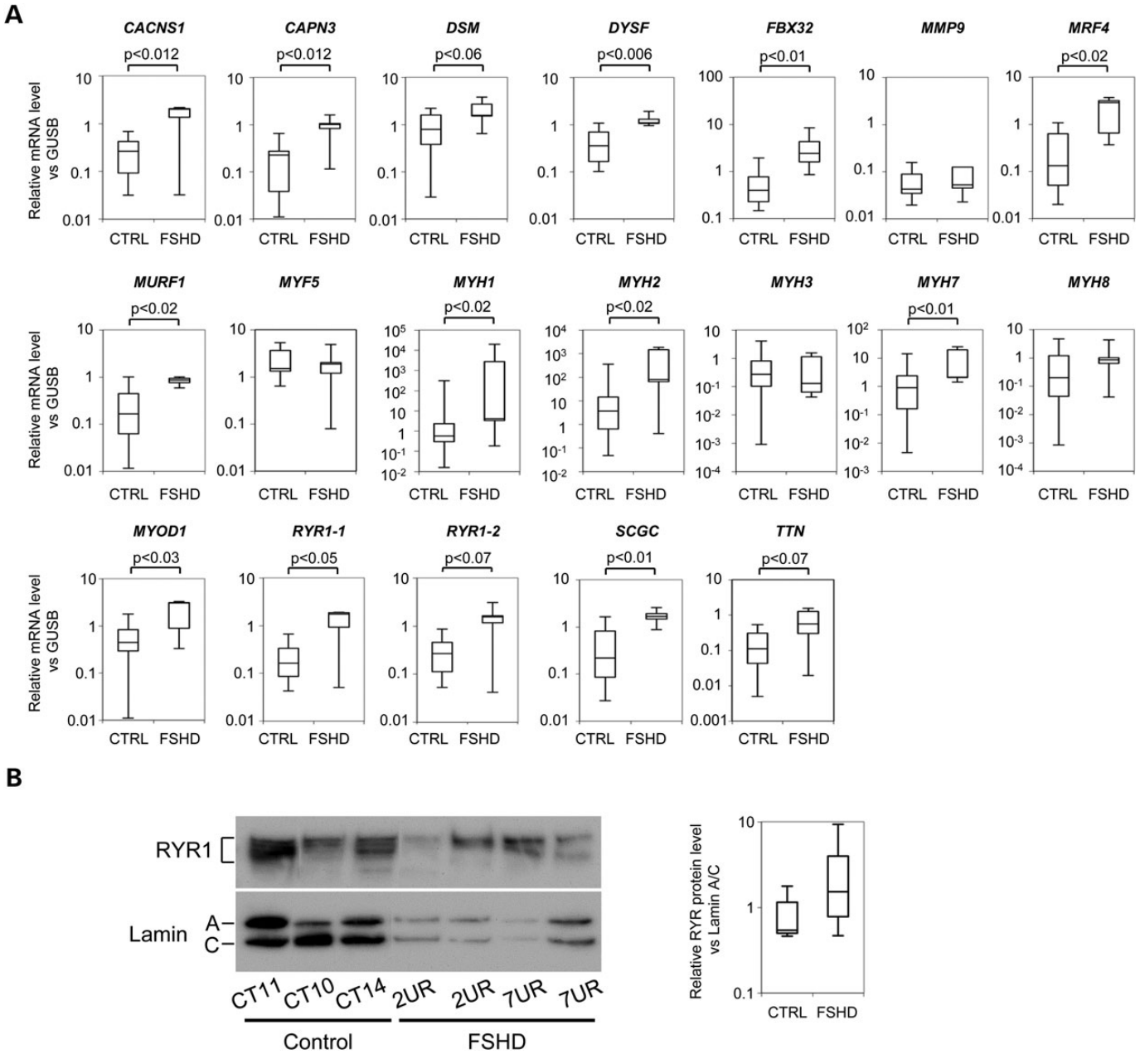

Figure 1. Transcriptional analysis of muscle-specific genes in fetal and adult muscle. (**A**) The box plots correspond to *GUSB* normalized data. *P*-values are indicated in each box. Experiments were done in triplicate. Horizontal lines are medians, the boxes correspond to the 50th percentile and the whiskers to the 75th percentile. Normalized transcription and *P*-values are given in Supplementary Material, Table S2. (**B**). Immunodetection of the RYR1 protein by western blotting (upper panel). Detection of Lamins A/C western was used as a reference (lower panel). The left panel corresponds to the quantification by densitometry of the RYR1/Lamins A/C signals. Boxes correspond to the 25th percentile and the whiskers to the 75th percentile. Horizontal lines correspond to the median.

to the post-natal period. Aberrant transcription of genes involved in the proliferation or differentiation of muscle cells is usually associated with intrinsic muscle defect. Histological observation of affected muscle from FSHD patients did not reveal any specific pathological change except, in some cases, changes in the fiber size, presence of angular fibers, infiltration by adipocytes, replacement of muscle fibers by fibrotic and inflammatory tissues or subtle sarcolemmal reorganization, which might change force generation by type II fibers (20,21). Also, from earlier work on primary cultures, necrotic features have been described for FSHD myoblasts, which are more sensitive to stress, while FSHD myocytes fuse at a faster rate compared with controls but retain the appearance of an undifferentiated state. We report here, for the first time, global molecular changes in the fetal myogenic differentiation program, prior to any sign of dystrophic phenotype, suggesting that subtle molecular changes precede the dystrophic phenotype. Our work suggests that the clinical FSHD-specific phenotype might find its origin at very early developmental stages prior to any clinical sign of the disease. As suggested by others, even if FSHD is an

adult-onset muscular dystrophy, the disease might also depend on defects during myogenesis or early differentiation. Type-I slow-twitch oxidative fibers are preferentially found at early developmental stages, while type-II (fast twitch glycolytic) increases after the 26th week (22,23). Then, developmental myosin isoforms are progressively replaced by the adult isozymes. The marked increase in *MYH2* in FSHD fetuses compared with controls suggests either modification in fiber conversion as observed in adults affected with FSHD (20,24) or a change in the timing of replacement since the change from type I to type II is usually seen after the 26th week of gestation. Furthermore, in FSHD, the presence of immature fibers expressing fetal myosin has been observed (25,26), suggesting also that maintenance or reactivation in the post-natal or adult muscle of factors acting during the intrauterine period might contribute to the cascade of alterations leading to the pathology.

The identity of the gene(s) involved in the pathology remains a matter of debate, but ectopic expression of the *DUX4* retrogene from the last *D4Z4* repeat and the distal qA region has been proposed as the main cause of the FSHD phenotype (1,2,15,16). To address whether *DUX4* was present at early developmental stages, we used oligodT primed complementary DNA (cDNA) and PCR amplification with the primers previously described (15) to determine the presence of the full-length *DUX4* transcript (*DUX4-fl*), blindly to the genetic background, in different fetal muscles from 5 FSHD fetuses and 21 controls (Fig. 2A, Table 1), but also in other tissues normally not affected in FSHD (diaphragm, spinal cord, skin, brain, and kidney). We detected the *DUX4-fl* mRNA in four of five FSHD samples either in biceps (one of four tested) or quadriceps (four of four) (Table 1), but also in non-affected somatic tissues such as skin, brain or kidney (Supplementary Material, Table S4), confirming that *DUX4* transcription can be observed in tissues (muscular and non-muscular) from individuals carrying a shortened *D4Z4* array as observed in muscular biopsies from adult FSHD patients (15). Moreover, as observed in healthy adult muscle (27), we also report the presence of the *DUX4-fl* in fetal biopsies from genetically unaffected fetuses although at a lower frequency (in 5 out of 17 biceps and 5 out of 21 quadriceps tested, Table 1), but also in non-muscular samples, not affected in the disease (Supplementary Material, Table S4). In adult samples, we confirmed the presence of the *DUX4-fl* in quadriceps of both controls (three of six) and FSHD patients (three of seven).

In FSHD patients carrying the 4qA allele downstream of the repeat, the transcription of *DUX4* through *D4Z4* and the proximal 4qA sequence produces a pre-messenger RNA with introns (1,2,15) spliced into different mRNAs varying in size and containing a polyadenylation site. A short *DUX4* transcript (DUX4-s) can be detected in myoblasts from patients and 50% of controls. The long *DUX4* transcript (*DUX4-fl*) is detected at a very low level (in 1 of 1000 muscle nuclei) in 50% of FSHD patient's biopsies, 40% expressing the *DUX4-s* and the others 10% presenting no expression of either form (15), but also in primary myoblasts and muscle biopsies from non-carrier unaffected adults (27).

In order to confirm specificity and sequence, we sequenced all PCR products amplified with the *DUX4-fl* primers and observed striking differences in splicing between FSHD carriers and controls. In most of the DUX4-positive samples from FSHD carriers (four out of five), we amplified the non-spliced *DUX4-fl* isoform, while eight out of nine controls express the spliced *DUX4-fl*. Indeed, two alternative forms have been described for the *DUX4-fl* pathogenic transcripts differing by alternative splicing of a 135 bp-long intron in the 3′ UTR (Fig. 2B). The termination codon of the DUX4 protein is located before the first alternative intron, and splicing is not expected to modify the frame or sequence of the pathogenic DUX4 protein.

Our data on *DUX4* are consistent with the recent report showing the prevalence of a long *DUX4-fl* transcript in approximately half of the FSHD biopsies tested (15), but also in unaffected individuals (27). However, we extend here these findings by demonstrating that (i) *DUX4* is expressed as early as fetal development, together with *DUX4* targets (Ferreboeuf et al., submitted), in muscular and also non-muscular somatic tissues; (ii) both spliced and non-spliced *DUX4-fl* isoforms are found in adult muscle (iii) in fetal FSHD samples, RNA processing of the first intron is rarely observed, suggesting a FSHD-specific control of splicing. However, after their initial description (2,14,15), the respective regulation, the presence or role of the two variants has not been further investigated or discussed, and recent reports focusing on the biological function of the DUX4 protein describe a *DUX4-fl* encoding vector without any precision on whether it corresponds to the spliced or non-spliced form. After cloning of the PCR products and sequencing, we did not observe any co-expression of the spliced or non-spliced *DUX4-fl* suggesting that, as observed for *DUX4-s* and *DUX4-fl*, the two isoforms might be mutually exclusive. Nevertheless, the underlying mechanisms regulating alternative splicing and their respective role remain to be investigated.

Based on the hypothesis of a position effect mechanism, gradual changes in the expression of genes located at the 4q35 locus have been proposed for FSHD. In this model, relaxation and long-range chromatin-dependent mechanisms might be associated with increased and variable expression of one to several genes at distance of the *D4Z4* array. So far, upregulation of the 4q35 genes has been a matter of debate with some authors showing an increase in expression of several of them (8–10) and others disputing these findings (11–13) in the absence of a gradient correlating distance from *D4Z4* and gene expression (11,12,28). The reasons for such discrepancies are not known, but might depend on the type of sample (biopsies versus primary myoblasts or myotubes), type of muscle but also inter-individual variability linked to genetic or epigenetic factors. To limit the impact of such variability and elucidate the regulation of the 4q35 genes in individuals carrying a short *D4Z4* array, we selected 36 genes located in the 4q35 region and measured their expression at the RNA level in muscular biopsies from 6 FSHD and 21 control fetuses. For at least 14 of the 4q35 genes, a significant increase in expression was observed in quadriceps (Fig. 2C; Supplementary Material, Table S4; at least 2-fold change in expression; P-value <0.05). We were not able to detect *FRG2* transcripts in fetal biopsies, but observed *FRG1* (10.9-fold; $P < 0.04$) and *SLC25A4* (*ANT1*, 6-fold, $P < 0.005$) upregulation. In addition, the *PDLIM3* gene (2.17-fold) involved in skeletal muscle function is also upregulated in FSHD fetuses. In adults, upregulation of *ANT1* has already been reported (8,29) and proposed as an early event in the pathological process (9), while no noticeable changes were reported for *PDLIM3* (*ALP*) (12,30) involved in Z-disc organization and regulation of the myogenic differentiation transcriptional network(31). Distal to

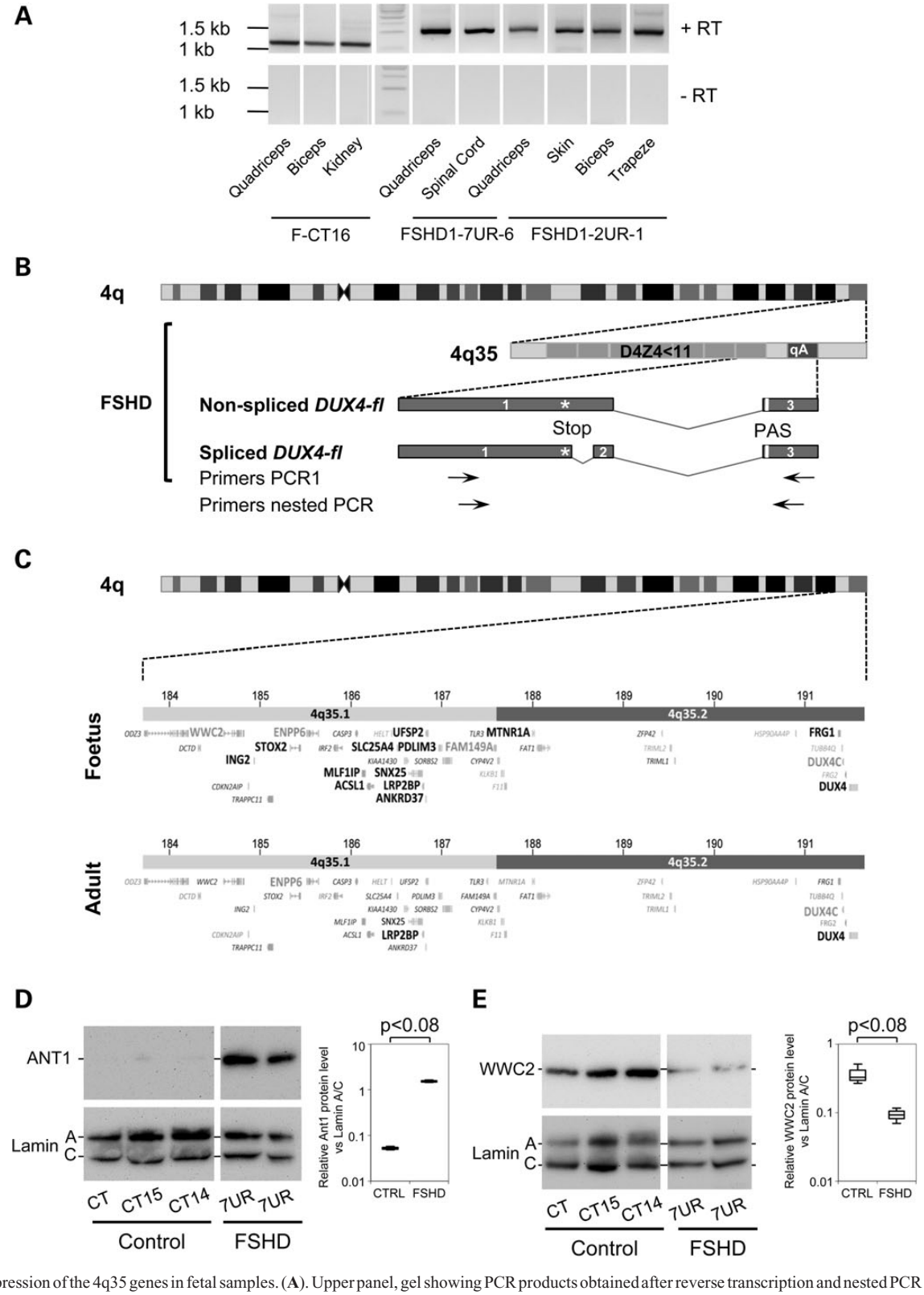

Figure 2. Expression of the 4q35 genes in fetal samples. (**A**). Upper panel, gel showing PCR products obtained after reverse transcription and nested PCR amplification on RNA extracted from muscle biopsies (biceps and quadriceps) and other somatic tissues of fetuses carrying a short *D4Z4* array (FSHD1-7UR-6; FSHD1-2UR-1) or non-carriers (F-CT16). Lower panel corresponds to amplification from samples not treated with reverse transcriptase (RT) (**B**). Schematic representation of the last *D4Z4* repeat encoding the two *DUX4-fl* isoforms. The *DUX4* exons indicated have grey boxes and numbered (1–3). The position of the polyadenylation site (PAS), stop codon (star) and primers used for RT–PCR are indicated. (**C**). Diagram showing changes in gene expression across the 4q35 region in quadriceps from fetuses ($n = 6$; upper panel) and adults ($n = 7$; lower panel) carrying a shortened *D4Z4* array compared with controls (fetuses, $n = 21$; adults, $n = 6$). Black and upper cases: upregulated genes; Grey upper case: downregulated genes; black: no significant change; grey, not expressed. Experiments were done in triplicate. Normalized

Table 1. Amplification of the DUX4-fl isoform in fetal and adult muscles

	Gestational age (weeks)	Biceps	Quad.
Fetuses			
FSHD1-2UR-1	16	$(+)^2$	$(+)^2$
FSHD1-4UR-2	16	–	$(+)^2$
FSHD1-2UR-4	25	N.A.	$(+)^1$
FSHD1-7UR-5	26	–	N.A.
FSHD1-7UR-6	26	–	$(+)^2$
F-CT1	12	–	–
F-CT2	13	–	–
F-CT3	14	–	$(+)^1$
F-CT4	14	$(+)^1$	–
F-CT5	14	–	–
F-CT6	14	–	–
F-CT7	15	–	$(+)^1$
F-CT8	18	N.A.	$(+)^1$
F-CT9	18	$(+)^1$	–
F-CT10	20	–	–
F-CT11	21	N.A.	–
F-CT12	23	–	–
F-CT13	24	N.A.	–
F-CT14	25	–	$(+)^1$
F-CT15	25	$(+)^1$	–
F-CT16	26	$(+)^2$	$(+)^2$
F-CT17	34	N.A.	–
F-CT18	37	$(+)^1$	–
F-CT19	37	–	–
F-CT20	37	–	–
F-CT21	37	–	–
	Age at enrollment (yrs)	Biceps	Quad.
Adults			
A-CT1	51	–	N.A.
A-CT2	77	–	N.A.
A-CT3	69	N.A.	$(+)^1$
A-CT4	45	N.A.	$(+)^1$
A-CT5	49	N.A.	$(+)^1$
A-CT6	52	N.A.	–
FSHD1-1	54	N.A.	$(+)^2$
FSHD1-2	66	N.A.	–
FSHD1-3	61	N.A.	–
FSHD1-4	61	N.A.	–
FSHD1-5	57	N.A.	–
FSHD1-6	57	N.A.	$(+)^1$
FSHD1-7	45	N.A.	$(+)^2$

The table recapitulates the results of at least three independent experiments. Samples positive in at least two out of the three RT–PCR assays have been sequenced. (−) indicates that the *DUX4-fl* transcript was not detected. (+) corresponds to samples where *DUX4-fl* was detected (either biceps or quadriceps). Each amplicon was sequenced and presence of either the spliced $(+)^1$ or non-spliced $(+)^2$ *DUX4-fl* isoform is indicated.

PDLIM3, the *SORBS2* gene regulated by the *PDLIM3-SORBS2* distal enhancer (32) during muscle development is not differentially expressed in FSHD (Fig. 2C, Supplementary Material, Fig. S4). In the vicinity of *PDLIM3A* and *SCL25A4*, four other genes are upregulated (*LRP2BP*, *ACSL1*, *ANKRD37* and *UFSP2*). *LRP2BP* (6.11-fold) regulates LRP2 function and HDL endocytosis and *ACSL1* (4.3-fold) encodes the acyl-coenzyme A synthetase long chain 1 implicated in the adaptation of several tissues including muscle to aerobic metabolism. The function of *ANKRD37* is not known, but might be activated by *HIF1* in response to hypoxia while *UFSP2* encodes an ubiquitin-fold modifier protease. Their respective function in muscle remains obscure. Two additional genes are also upregulated in FSHD: *MNTR1A* (5.43-fold) encoding a melatonin receptor and *ING2* (2.12-fold) encoding a chromatin-remodeling factor implicated in muscle differentiation. Beside *DUX4C (*7.3-fold, $P < 0.05$*),* two genes *ENPP6* (ectonucleotide pyrophosphatase/phosphodiesterase 6), *WWC2* (WW and C2 domain containing protein) appear to be slightly down-regulated in FSHD compared with control muscles (2.3-; -1.9- respectively, $P > 0.05$; Figure 2C; Supplementary Material, Table S5, Supplementary Material, Fig. S2C). SCL25A4 and WWC2 were immunodetected by western blotting and densitometry analysis confirmed their respective up- (29.6-fold, $P < 0.1$; Fig. 2D) or down-regulation (3.5-fold; $P < 0.1$; Fig. 2E) in fetal quadriceps FSHD samples compared with age-matched controls. As a comparison, we examined the expression status of the 36 genes in seven FSHD adult biopsies and six controls (Supplementary Material, Fig. S2 and Table S5). In agreement with the literature but in contrast to fetuses, we did not observe any significant dysregulation for these genes in adult muscles.

In fetuses, shortening of the *D4Z4* array is accompanied with dysregulation in a set of genes involved in myogenesis suggesting an altered myogenesis as early as the fetal stage. The genetic defect linked to FSHD might directly or indirectly disrupt the myogenic program as discussed before (33). Wasting and weakening in adult FSHD muscle might thus result from the long-term accumulation of defects such as structural changes, fiber conversion or response to stress arising as early as the fetal life. In FSHD, the rationale for the selective and asymmetrical involvement of specific groups of muscle remains unclear, since all the affected muscle do not originate from the same somite or same region in the embryo. However, one can speculate that common regulatory pathways regulating migration or differentiation of muscle cells are affected during muscle differentiation, and that the abnormality leading to the pathology begins during development. Strikingly, many of the 4q35 genes upregulated in muscles of FSHD carriers are also implicated at different levels in muscle function and homeostasis. Thus, dysregulation of the 4q35 locus together with changes in muscle-specific genes might alter the capacity of the muscle fiber to respond to stress during adult life leading, in turn, to the progressive weakening observed at the clinical level. We did not find any evidence of a common regulatory pathway for the different genes upregulated in FSHD fetuses (Supplementary Material, Figure S3) but, interestingly, many of the genes strongly upregulated in fetuses (*ACSL1*, *LRP2BP*, *PDLIM3*, *SLC25A4*, *UFSP2* and *MNTR1A*) are clustered in a region located ∼5–5.5 Mb upstream of the most proximal *D4Z4* element, flanked by two genomic regions located to the nuclear periphery (lamin attachment domains). This observation, in light of our previous work (4,34,35) and data from others (36,37), suggests that

transcription values are indicated in Supplementary Material, Table S6. We also observed upregulation of *PDLIM3, UFSP2, STOX2, ING2, ANT1* in biceps from FSHD fetuses. The corresponding box plots are shown in Supplementary Material, Figure S3. Upregulation is not correlated either with orientation of the gene or with the presence of a CpG island at the promoter region. (**D**) Western blot of ANT1/SLC25A4 and (**E**) WWC2 protein. Lamins A/C were used as a reference. On the left of each gel, quantification by densitometry of the protein-specific band/Lamins A/C signals. Horizontal lines are medians, the boxes correspond to the 25th percentile and the whiskers to the 75th percentile.

conformational and topological changes mediated by *D4Z4* array shortening might modify the regulation of the whole 4q35 region. Furthermore, with regard to expression of the 4q35 genes and muscle-specific genes, our results suggest that FSHD might not simply result from overexpression of a single gene or retrogene, but more likely from a cascade of dysregulations at different stages of the muscle development and maturation associated with the DUX4-dependent cascade or on other *D4Z4*-dependent modification.

At this step, it remains difficult to connect into a single model the mechanism of the disease and the respective contribution of the different *DUX4-fl* isoforms and other genes in the pathogenesis of this complex muscular dystrophy. Transcriptional mechanisms governing *D4Z4* transcription are likely to be important for understanding the pathology since a large number of small, polyadenylated or long non-coding RNA emanating from the repeat have been described (14,38,39). In general, our data suggest that reduction in the number of *D4Z4* might lead to the pathology by enhancing *in cis* expression of different genes including *DUX4*. However, the absence of the pathogenic *DUX4-fl* in a number of FSHD cases (15), its presence non-muscular somatic tissues in FSHD carriers but also in unaffected individuals (27) suggest that other alterations or pathways might contribute to the muscular phenotype as a direct or indirect consequence of this *cis* effect. FSHD is a highly variable disease with a high variability in penetrance within families even in individuals carrying the same number of repeats or monozygotic twins (25,40,41), suggesting that environment and lifestyle modify expression of the gene(s) involved in the disease. Overall, our work does not exclude a role for a stable *DUX4* transcript linked to a short or relaxed *D4Z4* array and the presence of a distal functional polyadenylation site, but suggests that *DUX4* detection is not the only marker of FSHD. Disease onset and progression might be associated with either dosage of the pathogenic transcript, stability or maintenance in the muscle fiber or burst of activation at critical stages. Of note, we were able to detect *DUX4* as early as the 14th week of gestation in normal fetuses and 16th week in FSHD samples, corresponding to the formation of fibers and myofibrils (11th–20th week). Recent data indicate that DUX4 is regulated during embryogenesis (42), and that muscle is sensitive to *DUX4* dosage at very early stages (43). Hence, the window of time in which *DUX4* is activated might be critical for FSHD. However, dysregulation of different 4q35 and muscle-specific genes in fetal muscles also suggests that a cascade of events probably precedes FSHD symptoms, and that accumulation of defects at different stages of muscle development might contribute to the pathomechanisms.

MATERIALS AND METHODS

Biological samples

Fetal biopsies from FSHD ($n = 6$) and non-carrier fetuses ($n = 21$) were obtained after therapeutic abortion. Research was approved by the biomedicine agency (PFS13-006). The parents have provided written informed consent for the use of biopsies for medical research in accordance with the Declaration of Helsinki. Controls are neither carrier of any known genetic mutation nor affected by a muscular pathology. Adults have provided informed consent and muscles of clinically affected and control individuals were obtained using a standardized muscle biopsy protocol. Controls were selected in the same age range and sex representation.

Quantitative RT–PCR

Total RNA was extracted from biopsies using the classical Chomczynski and Sacchi method (44). Reverse transcription of 1 μg of total RNA was performed using the Superscript II kit and oligo dT following manufacturer's instructions at 42°C for 50 min followed by inactivation at 70°C for 15 min (Life Technologies). Primers were designed using Primer Blast and Primer 3 (Supplementary Material, Table S1). Real-time PCR amplification was performed on a LightCycler 480 (Roche) using the SYBR green master mix. All PCRs were performed using a standardized protocol, and data were analyzed with the Lightcycler 480 software version 1.5.0.39 (Roche). Primer efficiency was determined by absolute quantification using a standard curve. For each sample, fold change was obtained by comparative quantification and normalization to expression of the *GUSB* standard gene. Similar results were obtained after normalization to β-actin or *ß2M*. For statistical analysis, data are expressed as means ± SEM. Statistical significance was assessed by the non-parametric Wilcoxon–Mann–Whitney statistical test using the Gnumeric spreadsheet version 1.10.16.

RT–PCR for *DUX4*

One microgram of total RNA was used for first-strand cDNA synthesis using SuperScript II reverse transcriptase and oligo dT as described above. Primary PCRs were performed with Taq DNA polymerase (Euromedex) using 7% of the first-strand reaction as template in a total reaction volume of 30 μl. Nested PCRs were performed on 1 μl of the primary reaction. For *DUX4-fl* detection, the following primers were used: PCR1-forward: 5′-CCC CGA GCC AAA GCG AGG CCC TGC GAG CCT-3′; PCR1-reverse: 5′-GTA ACT CTA ATC CAG GTT TGC CTA GA-3′; PCR2-forward: 5′-CGG CCC TGG CCC GGG AGA ACG CGG CCC GC-3′; PCR2-reverse 5′-TCT AAT CCA GGT TTG CCT AGAV CAG C-3′ as described in (15). PCR products were examined on 3% Molecular Biology Grade Agarose gels stained with ethidium bromide. In order to discard any risk of contamination by genomic DNA, each RNA sample was treated with *DNAse* I and PCR amplification was performed on RNA samples incubated in the absence of reverse transcriptase (RT-) or without addition of cDNA. In all cases, amplification products were only observed when RNA samples were reverse-transcribed. All PCR products were sequenced.

Western blot

Whole protein extracts were obtained from biopsies disrupted in 200 μl extraction buffer (Tris-HCl pH8 100 mM, 10% SDS, 10 mM EDTA, 10% glycerol, protease inhibitor) using a T18 Ultra Turax. Proteins were separated by electrophoresis and transferred onto a PVDF membrane following the protocol recommended by the supplier for the Life Technologies NuPAGE system (including MOPS running buffers, Bis/Tris 4-12% gels and NuPAGE transfer buffer). PVDF membranes

were blocked for 1 h in 5% (w/v) non-fat dry milk in PBS-Tween (0.1% Tween-20 in PBS) and incubated for 90 min with the following primary antibodies: Lamin A/C (1 of 10 000, Clone 4C11, Sigma-Aldrich), WWC2 (1 of 1000, Sigma-Aldrich), SLC25A4 (1 of 1000, Sigma-Aldrich), RYR1 (1 of 1000, ref: ab2868, Abcam). After four washes in PBS-T, an anti-mouse IgG secondary antibody coupled to horseradish peroxidase (ThermoFisher) was incubated for 90 min (1/20 000). After washes, the signal was revealed by enhanced chemiluminescence (ECL, SuperSignal West Pico, Pierce) using a biospectrum imaging system (UVP).

ACKNOWLEDGEMENTS

We are indebted to all families for participating in this study. We wish to thank Mr Armand Tasmadjian for technical help. We also thank Dr Sabrina Sacconi for providing adult biopsies and critical reading.

Conflict of Interest statement. N.B. is the recipient of a fellowship from AFM (Association Française contre les Myopathies). S.R. is the recipient of a fellowship from ANR (FSHDecrypt, ANR-09-GENO-038), M.-C.G. is the recipient of a fellowship from the FSH Society and M.F. is the recipient of a fellowship from FSHD Global (Australia).

FUNDING

This study was funded by FSHD Global (Australia), the FSH Society (USA) and an ANR grant (FSHDecrypt, ANR-09-GENO-038) (to F.M.).

REFERENCES

1. Gabriels, J., Beckers, M.C., Ding, H., De Vriese, A., Plaisance, S., van der Maarel, S.M., Padberg, G.W., Frants, R.R., Hewitt, J.E., 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.
2. Lemmers, R.J., van der Vliet, P.J., Klooster, R., Sacconi, S., Camano, P., Dauwerse, J.G., Snider, L., Straasheijm, K.R., van Ommen, G.J., Padberg, G.W. *et al.* (2010) A unifying genetic model for facioscapulohumeral muscular dystrophy. *Science*, **329**, 1650–1653.
3. Lemmers, R.J., Wohlgemuth, M., Frants, R.R., Padberg, G.W., Morava, E. and van der Maarel, S.M. (2004) Contractions of D4Z4 on 4qB subtelomeres do not cause facioscapulohumeral muscular dystrophy. *Am. J. Hum. Genet.*, **75**, 1124–1130.
4. Ottaviani, A., Rival-Gervier, S., Boussouar, A., Foerster, A.M., Rondier, D., Sacconi, S., Desnuelle, C., Gilson, E. and Magdinier, F. (2009) The D4Z4 macrosatellite repeat acts as a CTCF and A-type lamins-dependent insulator in facio-scapulo-humeral dystrophy. *PLoS Genet.*, **5**, e1000394.
5. Tsumagari, K., Qi, L., Jackson, K., Shao, C., Lacey, M., Sowden, J., Tawil, R., Vedanarayanan, V. and Ehrlich, M. (2008) Epigenetics of a tandem DNA repeat: chromatin DNaseI sensitivity and opposite methylation changes in cancers. *Nucleic Acids Res.*, **36**, 2196–2207.
6. Zeng, W., de Greef, J.C., Chen, Y.Y., Chien, R., Kong, X., Gregson, H.C., Winokur, S.T., Pyle, A., Robertson, K.D., Schmiesing, J.A. *et al.* (2009) Specific loss of histone H3 lysine 9 trimethylation and HP1gamma/cohesin binding at D4Z4 repeats is associated with facioscapulohumeral dystrophy (FSHD). *PLoS Genet.*, **5**, e1000559.
7. Gabellini, D., D'Antona, G., Moggio, M., Prelle, A., Zecca, C., Adami, R., Angeletti, B., Ciscato, P., Pellegrino, M.A., Bottinelli, R. *et al.* (2006) Facioscapulohumeral muscular dystrophy in mice overexpressing FRG1. *Nature*, **439**, 973–977.
8. Gabellini, D., Green, M.R. and Tupler, R. (2002) Inappropriate gene activation in FSHD: a repressor complex binds a chromosomal repeat deleted in dystrophic muscle. *Cell*, **110**, 339–348.
9. Laoudj-Chenivesse, D., Carnac, G., Bisbal, C., Hugon, G., Bouillot, S., Desnuelle, C., Vassetzky, Y. and Fernandez, A. (2005) Increased levels of adenine nucleotide translocator 1 protein and response to oxidative stress are early events in facioscapulohumeral muscular dystrophy muscle. *J. Mol. Med.*, **83**, 216–224.
10. Rijkers, T., Deidda, G., van Koningsbruggen, S., van Geel, M., Lemmers, R.J., van Deutekom, J.C., Figlewicz, D., Hewitt, J.E., Padberg, G.W., Frants, R.R. *et al.* (2004) FRG2, an FSHD candidate gene, is transcriptionally upregulated in differentiating primary myoblast cultures of FSHD patients. *J. Med. Genet.*, **41**, 826–836.
11. Jiang, G., Yang, F., van Overveld, P.G., Vedanarayanan, V., van der Maarel, S. and 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. Winokur, S.T., Chen, Y.W., Masny, P.S., Martin, J.H., Ehmsen, J.T., Tapscott, S.J., van der Maarel, S.M., Hayashi, Y. and Flanigan, K.M. (2003) Expression profiling of FSHD muscle supports a defect in specific stages of myogenic differentiation. *Hum. Mol. Genet.*, **12**, 2895–2907.
13. Cheli, S., Francois, S., Bodega, B., Ferrari, F., Tenedini, E., Roncaglia, E., Ferrari, S., Ginelli, E. and Meneveri, R. (2011) Expression profiling of FSHD-1 and FSHD-2 cells during myogenic differentiation evidences common and distinctive gene dysregulation patterns. *PLoS One*, **6**, e20966.
14. Snider, L., Asawachaicharn, A., Tyler, A.E., Geng, L.N., Petek, L.M., Maves, L., Miller, D.G., Lemmers, R.J., Winokur, S.T., Tawil, R. *et al.* (2009) RNA transcripts, miRNA-sized fragments and proteins produced from D4Z4 units: new candidates for the pathophysiology of facioscapulohumeral dystrophy. *Hum. Mol. Genet.*, **18**, 2414–2430.
15. Snider, L., Geng, L.N., Lemmers, R.J., Kyba, M., Ware, C.B., Nelson, A.M., Tawil, R., Filippova, G.N., van der Maarel, S.M., Tapscott, S.J. *et al.* (2010) Facioscapulohumeral dystrophy: incomplete suppression of a retrotransposed gene. *PLoS Genet.*, **6**, e1001181.
16. Vanderplanck, C., Ansseau, E., Charron, S., Stricwant, N., Tassin, A., Laoudj-Chenivesse, D., Wilton, S.D., Coppee, F. and Belayew, A. (2011) The FSHD atrophic myotube phenotype is caused by DUX4 expression. *PLoS One*, **6**, e26820.
17. Wallace, L.M., Garwick, S.E., Mei, W., Belayew, A., Coppee, F., Ladner, K.J., Guttridge, D., Yang, J. and Harper, S.Q. (2011) DUX4, a candidate gene for facioscapulohumeral muscular dystrophy, causes p53-dependent myopathy in vivo. *Ann Neurol*, **69**, 540–552.
18. Kowaljow, V., Marcowycz, A., Ansseau, E., Conde, C.B., Sauvage, S., Matteotti, C., Arias, C., Corona, E.D., Nunez, N.G., Leo, O. *et al.* (2007) The DUX4 gene at the FSHD1A locus encodes a pro-apoptotic protein. *Neuromuscul. Disord.*, **17**, 611–623.
19. Ahn, S., Riccio, A. and Ginty, D.D. (2000) Spatial considerations for stimulus-dependent transcription in neurons. *Annu. Rev. Physiol.*, **62**, 803–823.
20. Lassche, S., Stienen, G.J., Irving, T.C., van der Maarel, S.M., Voermans, N.C., Padberg, G.W., Granzier, H., van Engelen, B.G. and Ottenheijm, C.A. (2013) Sarcomeric dysfunction contributes to muscle weakness in facioscapulohumeral muscular dystrophy. *Neurology*, **80**, 733–737.
21. Reed, P., Porter, N.C., Strong, J., Pumplin, D.W., Corse, A.M., Luther, P.W., Flanigan, K.M. and Bloch, R.J. (2006) Sarcolemmal reorganization in facioscapulohumeral muscular dystrophy. *Ann. Neurol.*, **59**, 289–297.
22. Barbet, J.P., Thornell, L.E. and Butler-Browne, G.S. (1991) Immunocytochemical characterisation of two generations of fibers during the development of the human quadriceps muscle. *Mech. Dev.*, **35**, 3–11.
23. Butler-Browne, G.S., Barbet, J.P. and Thornell, L.E. (1990) Myosin heavy and light chain expression during human skeletal muscle development and precocious muscle maturation induced by thyroid hormone. *Anat. Embryol. (Berl)*, **181**, 513–522.
24. Celegato, B., Capitanio, D., Pescatori, M., Romualdi, C., Pacchioni, B., Cagnin, S., Vigano, A., Colantoni, L., Begum, S., Ricci, E. *et al.* (2006) Parallel protein and transcript profiles of FSHD patient muscles correlate to the D4Z4 arrangement and reveal a common impairment of slow to fast fibre

differentiation and a general deregulation of MyoD-dependent genes. *Proteomics*, **6**, 5303–5321.
25. Tupler, R., Barbierato, L., Memmi, M., Sewry, C.A., De Grandis, D., Maraschio, P., Tiepolo, L. and Ferlini, A. (1998) Identical de novo mutation at the D4F104S1 locus in monozygotic male twins affected by facioscapulohumeral muscular dystrophy (FSHD) with different clinical expression. *J. Med. Genet.*, **35**, 778–783.
26. Rahimov, F., King, O.D., Leung, D.G., Bibat, G.M., Emerson, C.P. Jr., Kunkel, L.M. and Wagner, K.R. (2012) Transcriptional profiling in facioscapulohumeral muscular dystrophy to identify candidate biomarkers. *Proc Natl Acad Sci USA*, **109**, 16234–16239.
27. Jones, T.I., Chen, J.C., Rahimov, F., Homma, S., Arashiro, P., Beermann, M.L., King, O.D., Miller, J.B., Kunkel, L.M., Emerson, C.P. Jr. *et al.* (2012) Facioscapulohumeral muscular dystrophy family studies of DUX4 expression: evidence for disease modifiers and a quantitative model of pathogenesis. *Hum. Mol. Genet.*, **21**, 4419–4430.
28. Masny, P.S., Chan, O.Y., de Greef, J.C., Bengtsson, U., Ehrlich, M., Tawil, R., Lock, L.F., Hewitt, J.E., Stocksdale, J., Martin, J.H. *et al.* (2010) Analysis of allele-specific RNA transcription in FSHD by RNA-DNA FISH in single myonuclei. *Eur. J. Hum. Genet.*, **18**, 448–456.
29. Tsumagari, K., Chang, S.C., Lacey, M., Baribault, C., Chittur, S.V., Sowden, J., Tawil, R., Crawford, G.E. and Ehrlich, M. (2011) Gene expression during normal and FSHD myogenesis. *BMC Med Genomics*, **4**, 67.
30. Bouju, S., Pietu, G., Le Cunff, M., Cros, N., Malzac, P., Pellissier, J.F., Pons, F., Leger, J.J., Auffray, C. and Dechesne, C.A. (1999) Exclusion of muscle specific actinin-associated LIM protein (ALP) gene from 4q35 facioscapulohumeral muscular dystrophy (FSHD) candidate genes. *Neuromuscul. Disord.*, **9**, 3–10.
31. Pomies, P., Pashmforoush, M., Vegezzi, C., Chien, K.R., Auffray, C. and Beckerle, M.C. (2007) The cytoskeleton-associated PDZ-LIM protein, ALP, acts on serum response factor activity to regulate muscle differentiation. *Mol. Biol. Cell*, **18**, 1723–1733.
32. McCord, R.P., Zhou, V.W., Yuh, T. and Bulyk, M.L. (2011) Distant cis-regulatory elements in human skeletal muscle differentiation. *Genomics*, **98**, 401–411.
33. Winokur, S.T., Barrett, K., Martin, J.H., Forrester, J.R., Simon, M., Tawil, R., Chung, S.A., Masny, P.S. and Figlewicz, D.A. (2003) Facioscapulohumeral muscular dystrophy (FSHD) myoblasts demonstrate increased susceptibility to oxidative stress. *Neuromuscul. Disord.*, **13**, 322–333.
34. Ottaviani, A., Schluth-Bolard, C., Gilson, E. and Magdinier, F. (2010) D4Z4 as a prototype of CTCF and lamins-dependent insulator in human cells. *Nucleus*, **1**, 30–36.
35. Ottaviani, A., Schluth-Bolard, C., Rival-Gervier, S., Boussouar, A., Rondier, D., Foerster, A.M., Morere, J., Bauwens, S., Gazzo, S., Callet-Bauchu, E. *et al.* (2009) Identification of a perinuclear positioning element in human subtelomeres that requires A-type lamins and CTCF. *EMBO J.*, **28**, 2428–2436.
36. Bodega, B., Ramirez, G.D., Grasser, F., Cheli, S., Brunelli, S., Mora, M., Meneveri, R., Marozzi, A., Mueller, S., Battaglioli, E. *et al.* (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 Biol.*, **7**, 41.
37. Pirozhkova, I., Petrov, A., Dmitriev, P., Laoudj, D., Lipinski, M. and 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**, e3389.
38. Cabianca, D.S., Casa, V., Bodega, B., Xynos, A., Ginelli, E., Tanaka, Y. and Gabellini, D. (2012) A long ncRNA links copy number variation to a polycomb/trithorax epigenetic switch in FSHD muscular dystrophy. *Cell*, **149**, 819–831.
39. Block, G.J., Petek, L.M., Narayanan, D., Amell, A.M., Moore, J.M., Rabaia, N.A., Tyler, A., van der Maarel, S.M., Tawil, R., Filippova, G.N. *et al.* (2012) Asymmetric bidirectional transcription from the FSHD-causing D4Z4 array modulates DUX4 production. *PLoS One*, **7**, e35532.
40. Griggs, R.C., Tawil, R., McDermott, M., Forrester, J., Figlewicz, D. and Weiffenbach, B. (1995) Monozygotic twins with facioscapulohumeral dystrophy (FSHD): implications for genotype/phenotype correlation. FSH-DY Group. *Muscle Nerve Suppl.*, **2**, S50–S55.
41. Tawil, R., Storvick, D., Feasby, T.E., Weiffenbach, B. and Griggs, R.C. (1993) Extreme variability of expression in monozygotic twins with FSH muscular dystrophy. *Neurology*, **43**, 345–348.
42. Krom, Y.D., Thijssen, P.E., Young, J.M., den Hamer, B., Balog, J., Yao, Z., Maves, L., Snider, L., Knopp, P., Zammit, P.S. *et al.* (2013) Intrinsic epigenetic regulation of the D4Z4 macrosatellite repeat in a transgenic mouse model for FSHD. *PLoS Genet.*, **9**, e1003415.
43. Mitsuhashi, H., Mitsuhashi, S., Lynn-Jones, T., Kawahara, G. and Kunkel, L.M. (2013) Expression of DUX4 in zebrafish development recapitulates facioscapulohumeral muscular dystrophy. *Hum. Mol. Genet.*, **22**, 568–577.
44. Chomczynski, P. and Sacchi, N. (2006) The single-step method of RNA isolation by acid guanidinium thiocyanate-phenol-chloroform extraction: twenty-something years on. *Nat Protoc.*, **1**, 581–585.