# *DUX4*, a candidate gene of facioscapulohumeral muscular dystrophy, encodes a transcriptional activator of *PITX1*

**Manjusha Dixit*, Eugénie Ansseau†, Alexandra Tassin†, Sara Winokur‡, Rongye Shi*, Hong Qian*, Sébastien Sauvage†, Christel Mattéotti†, Anne M. van Acker§, Oberdan Leo§, Denise Figlewicz¶, Marietta Barro‖, Dalila Laoudj-Chenivesse‖, Alexandra Belayew†, Frédérique Coppée†,††, and Yi-Wen Chen*,**††**

*Center for Genetic Medicine Research, Children's National Medical Center, Washington, DC 20010; †Laboratory of Molecular Biology, University of Mons-Hainaut, 7000 Mons, Belgium; ‡Department of Biological Chemistry, University of California, Irvine, CA 92697; §Institut de Biologie et Médecine Moléculaires, Université Libre de Bruxelles, 6041 Gosselies, Belgium; ¶Department of Neurology, University of Michigan, Ann Arbor, MI 34295; ‖Institut National de la Santé et de la Recherche Médicale, ERI 25 Muscle et Pathologies, CHU A. de Villeneuve, University of Montpellier I, 34295 Montpellier, France; and **Department of Pediatrics, George Washington University, Washington, DC 48109

Communicated by Shirley M. Tilghman, Princeton University, Princeton, NJ, September 14, 2007 (received for review May 16, 2007)

**Facioscapulohumeral muscular dystrophy (FSHD) is an autosomal dominant disorder linked to contractions of the D4Z4 repeat array in the subtelomeric region of chromosome 4q. By comparing genome-wide gene expression data from muscle biopsies of patients with FSHD to those of 11 other neuromuscular disorders, paired-like homeodomain transcription factor 1 (*PITX1*) was found specifically up-regulated in patients with FSHD. In addition, we showed that the double homeobox 4 gene (*DUX4*) that maps within the D4Z4 repeat unit was up-regulated in patient myoblasts at both mRNA and protein level. We further showed that the DUX4 protein could activate transient expression of a luciferase reporter gene fused to the *Pitx1* promoter as well as the endogenous *Pitx1* gene in transfected C2C12 cells. In EMSAs, DUX4 specifically interacted with a 30-bp sequence 5′-CGGATGCTGTCTTCTAATTAGTTTGGACCC-3′ in the *Pitx1* promoter. Mutations of the TAAT core affected *Pitx1-LUC* activation in C2C12 cells and DUX4 binding *in vitro*. Our results suggest that up-regulation of both *DUX4* and *PITX1* in FSHD muscles may play critical roles in the molecular mechanisms of the disease.**

D4Z4 | expression profiling | homeodomain | atrophy

Facioscapulohumeral muscular dystrophy (FSHD) is the third most common inherited muscular dystrophy. Patients show progressive weakness and atrophy of the muscles in the face, upper arms, and shoulder girdle to lower limbs, and a right/left asymmetry of onset is common (1, 2). The disease is autosomal dominant and is associated with shortening of the D4Z4 repeat array from 11–100 to 1–10 copies in the 4q35 subtelomeric region (1, 3, 4). Several molecular mechanisms have been proposed to explain how the deletion might activate transcription of genes in the region (5–8, 13). Candidate genes that might contribute to the FSHD phenotype were reported in several studies. FSHD region gene 1 (*FRG1*), FSHD region gene 2 (*FRG2*), and adenine nucleotide translocator 1 (*ANT1*) were reported up-regulated in FSHD muscles (6), although other studies showed controversial results (7, 9, 10). Transgenic animals with *Frg1* overexpression show a muscular dystrophy phenotype (11). However, it is not clear whether the mouse is a FSHD model, and it is generally considered that other genes in the region contribute to the disease (1).

Although there is consensus that FSHD is a disorder of transcription regulation, the molecular pathways leading to muscular dystrophy and other unique clinical features of the disease are far from clear, including the asymmetric distribution of muscle weakness. Each D4Z4 repeat unit contains an ORF with a double homeobox putatively encoding the DUX4 protein (12, 40). Although initially considered as "junk" DNA, the DUX4 ORF was recently shown to be conserved in evolution for >100 million years (14). Because there are hundreds of homologous 3.3-kb elements with highly similar *DUX* genes dispersed in the human genome (15), it has been very difficult to specifically amplify *DUX4* mRNA against the background of other *DUX* mRNAs unlinked to FSHD (16). Therefore, it is not known whether *DUX4* is expressed and activated by the array contraction, and whether it is involved in FSHD.

In this study, we tested the hypothesis that genes which show disease-specific changes in patients with FSHD are more likely to be involved in the early stages of disease progression. By comparing the genome-wide expression profiles of FSHD to those of 11 other neuromuscular disorders, we identified that paired-like homeodomain transcription factor 1 (*PITX1*) and four other genes were specifically up-regulated in muscle biopsies of FSHD patients. We further established the regulatory relationship between the *DUX4* gene within the D4Z4 repeat unit and the *PITX1* gene.

## Results

***PITX1* Is Specifically Up-Regulated in Muscles of Patients with FSHD.** In this study, we hypothesized that genes involved in the early steps of FSHD were likely to be disease-specific. By comparing expression profiles of FSHD to those of 11 other neuromuscular diseases, we can filter out expression changes reflecting secondary pathological processes, such as degeneration, regeneration, inflammation, and fibrosis formation. After expression profiling both clinically affected and less-affected muscles of patients with FSHD, we first identified genes differentially expressed in the FSHD muscles comparing with the normal controls ($P < 0.05$). We then compared the FSHD profiles vs. the other 11 neuromuscular disorders as a group and identified five genes specifically changed in FSHD [supporting information (SI) Table 1].

Among the five genes specifically changed in FSHD, *PITX1* showed the most dramatic up-regulation in both the unaffected (17-fold, $P = 2 \times 10^{-4}$) and affected (12-fold, $P = 7.7 \times 10^{-6}$)

GENETICS

Author contributions: M.D., E.A., and A.T. contributed equally to this work; M.D., E.A., A.T., S.W., R.S., H.Q., S.S., C.M., A.M.v.A., O.L., D.F., D.L.-C., A.B., F.C., and Y.-W.C. designed research; M.D., E.A., A.T., S.W., R.S., H.Q., S.S., C.M., A.M.v.A., M.B., D.L.-C., F.C., and Y.-W.C. performed research; A.T., A.M.v.A., O.L., and Y.-W.C. contributed new reagents/analytic tools; M.D., E.A., A.T., R.S., H.Q., O.L., D.L.-C., A.B., F.C., and Y.-W.C. analyzed data; and M.D., E.A., A.T., A.B., F.C., and Y.-W.C. wrote the paper.

The authors declare no conflict of interest.

Data deposition: The sequences reported in this paper have been deposited in the Gene Expression Omnibus (http://www.ncbi.nlm.nih.gov) (accession no. GSE9397) and Children's National Medical Center Public Expression Profiling Resources (http://pepr.cnmcresearch.org) databases.

Abbreviation: FSHD, facioscapulohumeral muscular dystrophy.

††To whom correspondence may be addressed. E-mail: frederique.coppee@umn.ac.be or ychen@cnmresearch.org.

This article contains supporting information online at www.pnas.org/cgi/content/full/0708659104/DC1.

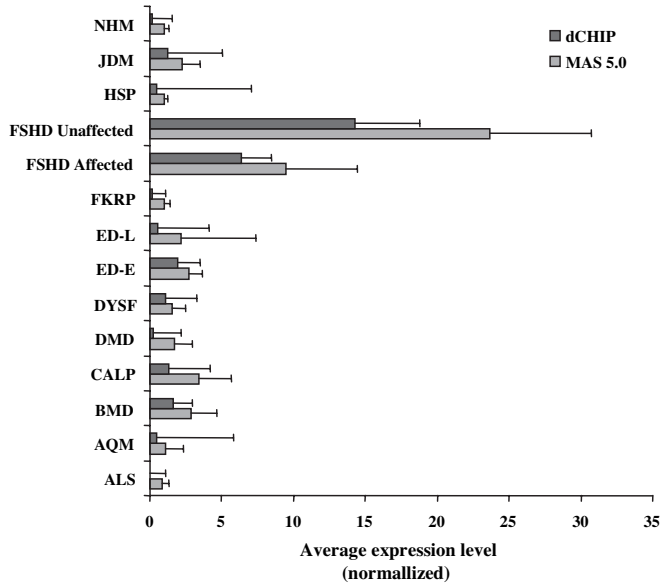


**Fig. 1.** *PITX1* is specifically up-regulated in FSHD compared with 11 other neuromuscular diseases. The expression level of the gene with arbitrary unit was determined by either dCHIP or Affymetrix MAS5.0. The sample size of each disease was normal healthy muscle (NHM), $n = 15$; juvenile dermatomyositis (JDM), $n = 25$; human spastic paraplegia (HSP), $n = 4$; FSHD, unaffected, $n = 5$, affected, $n = 9$; fukutin-related protein deficiency (FKRD), $n = 7$; Emery–Dreifuss muscular dystrophy, lamin A/C deficiency) (ED-L), $n = 4$; Emery–Dreifuss muscular dystrophy, emerin deficiency (ED-E), $n = 4$; dysferlinopathy (DYSF), $n = 10$; Duchenne muscular dystrophy (DMD), $n = 10$; Calpain-3 deficiency (CALP), $n = 10$; Becker muscular dystrophy (BMD), $n = 5$; acute quadriplegic myopathy (AQM), $n = 5$; and amyotrophic lateral sclerosis (ALS), $n = 9$.

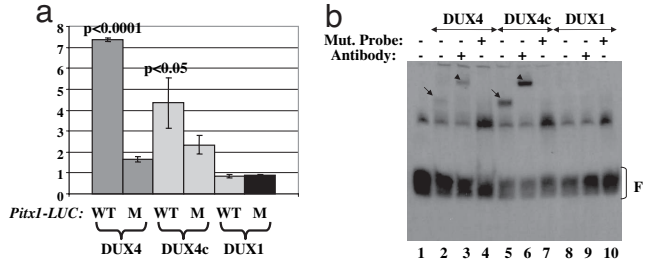


**Fig. 2.** DUX4 activates the *Pitx1* promoter in transient expression and specifically interacts with a cis element in the *Pitx1* promoter region. (*a*) C2C12 cells were cotransfected with the *pCIneo* expression vector encoding DUX4, DUX4c, or DUX1, as indicated, and a *Pitx1-LUC* reporter vector where the *Pitx1* promoter is either wild type (WT) or mutated (M; TAAT to TACC in the putative homeodomain-binding site). Luciferase activity was assayed on cells lysed 24 h after transfection. Data are provided as fold induction compared with luciferase activities obtained in cotransfection of the *Pitx1-LUC* vectors with the insertless *pCIneo vector*. (*b*) EMSA was performed with nuclear extract prepared from C2C12 cells at 16 h after transfection. Wild-type probe (lanes 1–3, 5, 6, 8, and 9) and mutated probes (lanes 4, 7, and 10) were end-labeled and incubated with nuclear extracts from cells transfected with *pCIneo-DUX4* (lane 2–4), -*DUX4c* (lanes 5–7), or -*DUX1* (lanes 8–10). Supershifts were determined by incubating probe and protein complexes with mAb 9A12 that recognizes DUX4 and DUX4c (lanes 3, 6, and 9). Small arrows in lanes 2 and 6 and arrowheads in lanes 3 and 6 indicate shifts and supershifts, respectively.

muscles of patients with FSHD (Fig. 1). We validated this result by real-time quantitative RT-PCR in FSHD muscle biopsies. This analysis showed *PITX1* was 11-fold up-regulated in FSHD unaffected muscles ($P = 2 \times 10^{-5}$) and 24-fold in affected muscles ($P = 2 \times 10^{-3}$) compared with muscles from healthy volunteers.

To determine whether the genes at the 4q35 region were differentially expressed in muscles of patients with FSHD, we examined the expression data of genes with probe sets designed on the U133A microarrays. The profile data showed that the mRNA expression levels of *DUX4*, *FRG1*, tubulin, $\beta$ polypeptide 4, member Q (*TUBB4Q*), and *ANT1* were not different among clinically affected FSHD, clinically unaffected FSHD, and control samples at the mRNA level. No expression of *DUX4* and *TUBB4Q* was detected by the microarrays, whereas the expression of *FRG1* and *ANT1* was detected (Affymetrix present calls). It is important to note that because of the existence of numerous homologues in the human genome, the design of probe sets specifically targeting the *DUX4* and *FRG1* genes on chromosome 4q35 is unlikely. To verify the specificity of the Affymetrix probe sets representing *DUX4* and *FRG1* genes, we BLAST searched the Affymetrix target sequences and individual probe sequences against the nonredundant (nr) database at National Center for Biotechnology Information (www.ncbi.nlm.nih.gov/BLAST) and confirmed that the probe sets were not specific to 4q35. The target sequence of *DUX4* aligned with homologues at chromosomes 2, 4, 10, 18, 22, and Y, whereas the target sequence of *FRG1* aligned with homologues on chromosomes 4, 9, 20, and 22. A BLAST search of individual probe sequences gave similar results.

**DUX4 Is a Transcription Activator of *PITX1*.** PITX1 is a member of the paired family of homeodomain transcription factors (17). That the *PITX1* gene is located on chromosome 5q31, not 4q35, suggested that the change of *PITX1* and other genes with FSHD-specific changes may occur downstream of one or few genes adjacent to the D4Z4 arrays, namely *FRG1*, *FRG2*, and *DUX4*. Homeobox-containing genes are often part of regulatory networks involved in embryonic development and express transcription factors that can regulate other homeobox genes in the network (18). Because both PITX1 and DUX4 are homeodomain transcription factors, we hypothesized that DUX4 was an upstream regulator of *PITX1*. To test this hypothesis, we first searched the *PITX1* promoter region and looked for putative DUX4-binding site. Three guidelines were used: (*i*) the site might have a core motif (TAAT) often found in homeodomain-binding site, (*ii*) the sequence would be in an evolutionary conserved region, and (*iii*) the sequence might share similarity with the homologous DUX1-binding site (19). We identified a putative binding site ≈500 bp upstream from the transcription start site (SI Fig. 6). The "Vertebrate Multiz Alignment and Conservation" function of the University of California, Santa Cruz, genome browser (http://genome.ucsc.edu) was used to determine that the region was highly conserved among six species (mouse, dog, rabbit, rat, human, and opossum). In addition to the TAAT core motif, the site shares an adjacent 2 bp of exact match with the DUX1-binding sequence (SI Fig. 6).

To test whether the putative DUX4-binding site was functional, we first amplified a 369-bp mouse *Pitx1* promoter fragment containing the site and cloned it into the *pGL3-basic* luciferase reporter vector. Cotransfection of C2C12 cells with the *pGL3-Pitx1* promoter construct (*Pitx1-LUC*) and a *pCIneo-DUX4* expression vector resulted in 7.4-fold increase in luciferase activity at 24 h (Fig. 2*a*) compared with samples cotransfected with the insertless *pCIneo* vector ($n = 4$, $P = 1.6 \times 10^{-21}$). Cotransfection of *Pitx1-LUC* and *pCIneo-DUX4c* expressing a 4q35 homologue of DUX4 led to a 4.3-fold ($P = 0.002$) up-regulation of the luciferase activity, but no significant change was observed with *pCIneo-DUX1* that expressed a non4q35 homologue.

To determine the specificity of the interaction between DUX4 and the putative binding site, we mutated the TAAT sequence to TACC in the *Pitx1-LUC*. Cotransfection of this mutated vector with *pCIneo-DUX4* reduced luciferase activity 4-fold ($P = 1.7 \times 10^{-17}$). Cotransfection with *pCIneo-DUX4c* also led to an ≈2-fold decrease

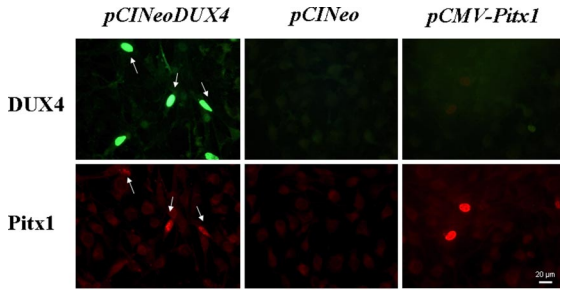


**Fig. 3.** DUX4 induces expression of endogenous *Pitx1* gene in C2C12 cells. C2C12 cells were transfected with the *pCIneo-DUX4* expression vector (*Left*), the insertless *pCIneo* as a negative control (*Center*), or a Pitx1 expression vector as a positive control (*Right*). Double immunofluorescence was performed 24 h after transfection with mAb 9A12 and a secondary antibody coupled to FITC for DUX4 (green, *Upper*) and with a rabbit serum raised against a Pitx1-specific peptide and secondary antibodies coupled to Texas red (*Lower*). Arrows point to three nuclei coexpressing DUX4 and Pitx1. (Scale bar, 20 $\mu$m.)

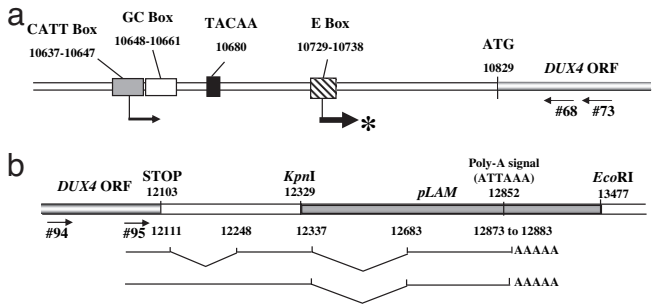


**Fig. 4.** Determination of the DUX4 mRNA 5′ and 3′ends. (*a*) Schematic representation of the *DUX4* promoter with the putative CATT, GC, TACAA, and E boxes, the ATG translation start codon, and the primers used in 5′-RACE (arrows, nos. 68 and 73). The two 5′ ends found previously in total RNAs of C2C12 cells transfected with *pGEM42* are indicated. One of these is identical to the single transcription start site determined here on total RNAs of FSHD and control primary myoblasts (*). (*b Upper*) Scheme of the EcoRI genomic fragment end as cloned in *pGEM42* with the stop codon of the *DUX4* ORF and the *pLAM* region. The KpnI and EcoRI restriction sites, the polyA addition signal (ATTAAA), and the primers used in 3′-RACE (arrows, nos. 94 and 95) are indicated. (*Lower*) Mapping of the ends obtained by 3′-RACE, with introns A and B.

in luciferase activity ($P = 0.047$), whereas *DUX1* showed no effect as expected.

To evaluate direct and specific binding between the DUX4 protein and its putative binding site in the *Pitx1* promoter region, a 30-bp oligonucleotide probe 5′-CGGATGCTGTCTTCTAATT-AGTTTGGACCC-3′ containing the putative binding site was tested in EMSA. When the probe was incubated with nuclear extracts of C2C12 cells transfected with *pClneo-DUX4 or DUX4c*, complexes were formed (Fig. 2*b*). Addition of the 9A12 mAb directed against DUX4 and DUX4c caused further retardation of the complexes and confirmed the presence of either protein. A probe with the two mismatches could still form the specific complexes (data not shown). However, a probe with six mutated base pairs abolished the binding of both DUX4 and DUX4c. No binding or supershift was observed with extracts containing DUX1.

Many homeodomain proteins bind to the TAAT core motif *in vitro* with similar affinities, whereas their specificity is regulated *in vivo* through interaction with different protein partners that help discriminate TAAT sites in different sequence contexts (20–24). To show that DUX4 and DUX4c did not interact with just any TAAT motif, we used another 30-bp oligonucleotide probe containing the other TAAT site in the same evolutionary conserved region (SI Fig. 6) and indeed observed no complex formation with DUX4, DUX4c, or DUX1 (data not shown).

To evaluate whether DUX4 could also activate the endogenous *Pitx1* gene, we transfected C2C12 cells with *pClneo-DUX4* and performed a coimmunofluorescence staining of DUX4 and Pitx1. Both proteins were detected in the same nuclei 24 h after transfection, whereas only background signals were observed in cells transfected with insertless *pClneo* (Fig. 3). These experiments were performed in the presence of a proteasome inhibitor, because we found that Pitx1 contained a PEST motif, which is a sequence enriched in proline (P), glutamic acid (E), serine (S), and threonine (T). Such a motif targets proteins for rapid degradation by the proteasome (25). Indeed, in the absence of a proteasome inhibitor, Pitx1 was detected at a later time (36 h) (data not shown).

**Characterization of the *DUX4* Gene.** To characterize the 5′ end of the transcript, we performed 5′-RACE by using nested PCR with primers nos. 68 and 73 (Fig. 4*a*; SI Table 2), which yielded a 400-bp fragment from both control and FSHD samples. After cloning the amplicons into the *pCR4* vector, 9 FSHD and 17 control inserts were sequenced and found to be identical to the *DUX4* but not the *DUX4c* genomic sequence. All 5′-RACE products corresponded to a single start site mapped 48 bp downstream from the TACAA box (position 10,732 in GenBank no. AF117653) within a typical initiator sequence (GCACCTG) overlapping with a putative E box upstream of the ORF start codon (Fig. 4*a*). This site was previously detected in C2C12 cells transfected with a genomic fragment containing one D4Z4 unit (26).

To characterize the 3′ end of the *DUX4* mRNA, we first set up 3′-RACE conditions on total RNA extracted from C2C12 cells transfected with *pGEM42* that contains the 13.5-kb genomic fragment of a patient with two D4Z4 repeats (12). This approach avoided the transcription background of hundreds of human *DUX* genes not linked to FSHD. We used an oligo dT containing primer for the RT step and obtained two 3′-RACE products of ≈490 and 350 bp that were cloned in *pCR4*. Their sequences demonstrated alternative splicing of a 136-bp intron between the stop codon and the end of the D4Z4 unit (Fig. 4*b*). In addition, all of the products had spliced out a 345-bp intron located in the *pLAM* region that flanks the telomeric side of the D4Z4 array. Both splicing variants have heterogeneous 3′ ends between positions 12873 and 12883 (in GenBank accession no. AF117653), that most likely resulted from the use of a polyadenylation signal in the *pLAM* sequence (ATTAAA, 12852–12858 in GenBank accession no. AF117653). This signal was predicted by the FGENESH gene-finding program (27). By using the 3′-RACE, both transcripts were then detected in total RNAs extracted from FSHD myoblast primary cultures but not controls.

In conclusion, the *DUX4* gene showed alternatively spliced introns in its 3′ untranslated region. Its heterogeneous mRNA 3′ ends mapped downstream of the ORF, outside of the D4Z4 unit, in the *pLAM* region that contributed an additional intron to the *DUX4* gene in the most distal D4Z4 unit. In addition, our data suggest the *DUX4* gene in the most distal D4Z4 element is likely to be the only one that could be transcribed into a polyadenylated RNA *in vivo* and yield a protein product.

**Expression of *DUX4* mRNAs in Primary Myoblasts of Patients with FSHD.** To evaluate expression of full-length *DUX4* mRNA, total RNA was extracted from FSHD (5 and 7 D4Z4 units) and control myoblast cultures, reverse-transcribed with a primer (no. 407) located in *pLAM*, 3′ of both introns, and amplified by PCR with an additional primer (no. 222; SI Table 2) located upstream of the ORF start codon (Fig. 5*a*). A strong band of 2,000 and a weaker one of 1,700 bp were observed in cells transfected with *pGEM42* (Fig. 5*b*). No fragment was obtained either upon omission of reverse transcriptase or in RNA extracted from cells transfected with the

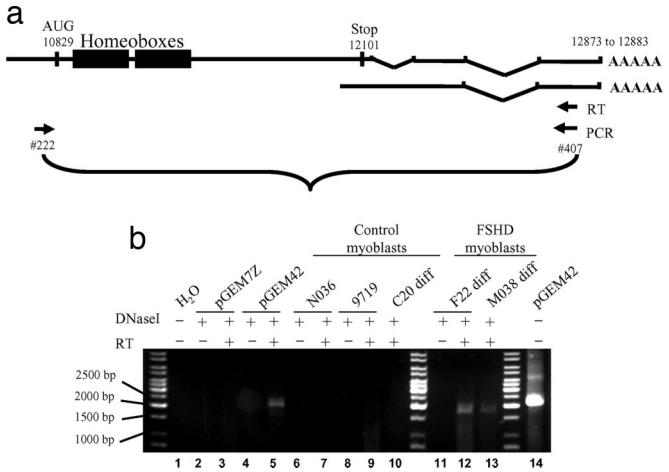


**Fig. 5.** Detection of the *DUX4* mRNAs in FSHD primary myoblasts. (*a*) Schematic representation of the *DUX4* RNAs with alternative 3′ ends and the primers used for RT (no. 407) and PCR (nos. 222 and 407). (*b*) For controls, RT-PCR was performed on 3–4 $\mu$g of total RNA extracted from C2C12 cells transfected with the *pGEM7Z* vector containing either no insert (lanes 2 and 3) or the 13.5-kb genomic fragment of a patient with 2 D4Z4 units (*pGEM42*, lanes 4 and 5). RT-PCR was similarly performed on total RNA of control (N036, 9719, and C20) and FSHD primary myoblasts (F22, 5 D4Z4 units; M038, 7 units) either in proliferation (lanes 6–9) or in differentiation (diff) (lanes 10–13). RNA samples were incubated (+) or not (−) with DNase I and RT, as indicated. As a positive control (lane 14), PCR was performed on the *pGEM42* vector present in a control sample (as in lanes 4 and 5) not treated with DNase I.

insertless vector. The 1,700-bp fragment and a weak signal at ≈2 kb were observed in differentiating FSHD myoblasts but not controls. The products were cloned in *pCR4*, and only the sequences corresponding to the 1,700-bp fragment were found to be identical to *DUX4*; they had all spliced out the intron located in the *pLAM* region. Identical sequences were also obtained on clones derived from two additional FSHD myoblast lines. No sequence identical to DUX4 was found in any control myoblast sample. The observed intron splicing clearly showed our amplified products derived from *bona fide* mRNAs and not from contaminating DNA.

**Expression of the DUX4 Protein in Primary Myoblasts of Patients with FSHD.** Conceptual translation of the *DUX4* ORF (GenBank accession no. AF117653) provided a 424-residue protein, a size larger than initially reported following correction of sequencing errors in very GC-rich D4Z4 (12). We raised a mAb (9A12 mAb) against the 253 last residues of DUX4 expressed as a recombinant protein in *E. coli* and found that it cross reacted with DUX4c but no other DUX protein.

The antibody specificity was shown on extracts of human rhabdomyosarcoma TE671 cells transfected with *pCIneo-DUX4* or *-DUX4c* (SI Fig. 7*a*). The 52-kDa DUX4 and 47-kDa DUX4c bands disappeared upon competition with a 10-fold excess of the recombinant protein. A band with the same apparent molecular weight as DUX4 but not DUX4c could be observed in the FSHD but not in the control sample (SI Fig. 7*b*). To evaluate mAb 9A12 sensitivity, we performed a similar Western blot with decreasing amounts of FSHD myoblast extract (SI Fig. 7*c Top*). DUX4 was noticed in all lanes but not in 30-$\mu$g control myoblast extract. The blot was stripped and incubated with a rabbit serum raised against a DUX4-specific peptide. As found in a previous study, this serum detected DUX4 only in extracts of cells transfected with *pCIneo-DUX4* but not in myoblasts (SI Fig. 7*c Middle*) (26). We performed additional Western blots with mAb 9A12 and again detected DUX4 in 40-$\mu$g extracts of four FSHD but not one additional control myoblast lines (SI Fig. 7*d*). We detected DUX4 protein expression in four additional FSHD lines but not in two other control myoblast lines (data not shown). Densitometric quantitation of DUX4 expression levels failed to show a correlation with D4Z4 copy number in these samples.

## Discussion

FSHD is believed to be a disorder of gene regulation. Although the genetic defect linked to the disease has been known for 15 years, the molecular pathways leading to muscular dystrophy and other unique clinical features are far from clear (3). By comparing with other muscular dystrophies, one may filter out shared secondary molecular pathways such as degeneration/regeneration, inflammatory infiltration, and fibrosis. Therefore, the genes that show disease-specific changes in patients with FSHD are more likely to be involved in the early stages of the disease progression. Indeed, we identified *PITX1* specifically up-regulated in the patients with FSHD. *PITX1* is a member of a paired family of homeoproteins and is involved in specification of hind limb identity. *Pitx1*-deleted mice exhibit striking abnormalities in morphogenesis and growth of the hind limb, resulting in structural changes in both muscle and bone to more closely resemble the corresponding forelimb structures (28, 29). The gene shows a similar expression pattern in chick, and the targeted ectopic expression of chick *Pitx1* in the developing wing bud causes several morphological changes that affect the appearance, digital patterning, and muscle organization in a way that more resembles the hind limb (30). In addition to affecting anterior/posterior limb development, *Pitx1* has also been shown to affect left–right symmetry (28, 31–33). FSHD differs from all other types of muscular dystrophy in both the pattern of muscle involvement (head and shoulder girdle) and the marked right/left asymmetry seen in many patients. Our finding of aberrant regulation of the *PITX1* gene provides a potential molecular basis for the enigmatic presentation and progression of FSHD. Recent studies reporting that *Pitx1* regulates expression of the *IFN*$\alpha$ gene in inflammatory cells and is a suppressor of RAS and tumorigenicity suggest the gene could be contributing to other clinical aspects of FSHD (34–36), including the involvement of atypical inflammatory responses and defects in myoblasts reported previously (37–39).

This study reports introns in the *DUX4* gene and identifies a direct transcription target of the DUX4 protein. Detection of *DUX4* expression has been unsuccessful in the past (10, 15, 40–43). SI Table 3 presents a step-by-step comparison of the unsuccessful protocols with ours. In our study, a single initiation site at the 5′ end of the gene and alternative spliced isoforms of DUX4 transcripts were identified in FSHD myoblasts. These transcripts likely use a polyadenylation site in the pLAM region. Although we detected the 5′ end of *DUX4* mRNAs in both FSHD and control myoblasts, the 3′ end sequence specific to the most distal unit of DUX4 was detected only in patients' samples. This observation suggested that, whereas other *DUX4* copies might be transcribed, only the gene in the most distal D4Z4 unit could extend to a polyadenylation signal present in the flanking pLAM region. Therefore, these polyA+ *DUX4* mRNAs are likely to be the only ones with an extended half life to be translated into a protein product *in vivo*. This could explain why at least one or part of a D4Z4 unit with one full copy of the *DUX4* gene is required on the FSHD allele to cause the disease. Interestingly, some of the differences between the 4qA and 4qB alleles are found in the region just distal to the D4Z4 array, which may affect both the splicing and polyadenylation signals (44, 45). Contractions of the D4Z4 repeat array are associated with the FSHD phenotype only when they occur on the 4qA allele (45, 46). It will be interesting to investigate whether the distal D4Z4 unit on a contracted 4qB allele is able to produce a protein product. The 10q26 region presents high sequence similarity to 4q35, including a D4Z4-like repeat array with *DUX10* genes encoding a protein identical to DUX4 (47). Contractions of this repeat array are not associated with FSHD (44). Interestingly, none of our 3′ RACE or RT-PCR products contained the BlnI restriction site characteristic

of the D4Z4-like array in 10q26, suggesting no stable DUX transcript could derive from chromosome 10.

We showed by transient cotransfection with a *Pitx1-LUC* reporter and *DUX* expression vectors that DUX4 could activate transcription about twice as much as DUX4c, although both proteins could similarly bind to the *Pitx1* promoter element in EMSA. Similar DNA binding is expected, because these proteins are identical over their first 342 residues, including the double homeodomain. However, in contrast to DUX4c, DUX4 presented a strong transcriptional activity in the yeast one-hybrid system that was mediated by its unique C-terminal (tail) region (ref. 47 and unpublished data). This difference in transcriptional activity was much stronger in the yeast system than was observed here in mammalian muscle cells and probably reflects the nature of cofactors within the transcriptional complex (20–24). Low expression of a transcription factor is an event frequently occurring as an initial trigger leading to differentiation during embryonic development. Similarly, low levels of a potent trans-factor such as DUX4 could activate expression of a gene encoding another trans-factor such as *PITX1* that would be produced at higher levels and could itself activate other genes. Such transcription factor cascades are well known for homeoproteins and could contribute to the mechanism of FSHD.

The lack of *Pitx1-LUC* activation by DUX1 is correlated with its lack of interaction with the *Pitx1* promoter site as confirmed by EMSA. DUX1 is a transcription factor with unknown function that binds to a palindromic TAAT/ATTA site instead of a single TAATTA as found here for DUX4/DUX4c. This difference in binding specificity could be explained by sequence divergence in their DNA-binding domains. The first and second homeodomains present 11 and 2 mismatches in DUX4/DUX4c vs. DUX1, respectively.

In summary, we identified *PITX1* as showing aberrant transcriptional regulation in FSHD compared with 11 other neuromuscular disorders. We further showed that the DUX4 protein encoded by a gene mapped within the D4Z4 repeat, interacted *in vitro* with a cis-element in the *Pitx1* promoter region and activated transient expression of both a *Pitx1-LUC* vector and the endogenous *Pitx1* gene in C2C12 cells. Our characterization of the *DUX4* gene structure suggested that polyadenylated *DUX4* mRNAs could be produced only from the most distal D4Z4 unit at the FSHD locus. Our data provide a direct link between the genetic defect in 4q35 and the specific up-regulation of a gene in FSHD muscles. Moreover the transcription factor properties of both *DUX4* and *PITX1* would explain the large transcriptome dysregulation observed in FSHD myoblasts and muscles and the unique clinical characteristics in patients (9, 10, 48). In conclusion, our data identified a candidate mechanism involving interplay between *DUX4* and *PITX1*, which may help to explain the long-elusive biology of FSHD.

## Materials and Methods

**Muscle Biopsies and Expression Profiling.** The muscle biopsies of nine patients with FSHD described in a previous study were used here for expression profiling and quantitative RT-PCR (10). Nine affected muscles from patients with FSHD and five clinically unaffected muscles from five of those patients were expression-profiled by using the Affymetrix Human Genome U133A microarrays. Normal sex- and age-matched adult controls were obtained from six healthy adult volunteers. Expression profiles of 11 neuromuscular disorders were used as disease controls, as described (49). The procedures and quality control were done as described in our previous publications (50–52). For details, see *SI Text*.

**Primary Myoblasts.** Muscle biopsies were performed according to a procedure approved either by the University of Rochester Research Subjects Review Board or by current ethical and legislative rules of France, as described (53). Primary myoblast cultures were derived from biopsies as described (39). Myoblasts were grown in DMEM with 10% FCS and 1% Ultroser (Cyphergen). Differentiation to myotubes was induced by changing confluent myoblasts to DMEM containing 2% horse serum (differentiation medium). Cells were harvested either during the exponential proliferation phase or 4–6 days after differentiation was induced.

**Luciferase Reporter Assay.** C2C12 cells were cotransfected with the *Pitx1-LUC* wild-type or mutant promoter construct and either *pCIneo* vector by using FuGENE 6 (Roche Applied Sciences) and 0.5 $\mu$g of each plasmid (for details, see *SI Text*). Quadruplicates were performed. After 24 h, transfected cells were lysed directly in the dish with 100 $\mu$l of reporter lysis buffer, and luciferase activity was evaluated by using Luciferase Assay System (Promega) in Centro LB 960 Luminometer (Berthold Technologies).

**EMSA.** EMSA was performed with a 30-bp oligonucleotide containing the putative DUX4-binding site as a probe (for details, see *SI Text*). Nuclear protein extract was prepared from C2C12 cells 16 h posttransfected with *pCIneo-DUX4*, *DUX4c*, and *DUX1* expression vectors by using Cell Lytic NuCLEAR Extraction Kit (Sigma). To obtain supershift of protein–DNA complexes, 1 $\mu$g of the anti-DUX4/DUX4c mAb was added to the mixtures. DNA–protein complexes were resolved on a 6% Novex DNA retardation polyacrylamide gel (Invitrogen), then transferred onto positively charged nylon membranes by electroblotting. For signal detection, the chemiluminescent substrate CSPD (10 mg/ml) was used in 1:100 dilutions, and the emitted light was recorded on an x-ray film.

**5′- and 3′-RACE.** Primary myoblast total RNA was extracted with the Aurum Total RNA Mini kit (Bio-Rad) or NucleoSpin RNAII (Macherey-Nagel). DNase-treated total RNA were submitted to 5′- and 3′-RACE by using the RLM-RACE kit (Ambion). The products were cloned and sequenced (for details, see *SI Text*.)

**RT-PCR.** RT was done on 2–4 $\mu$g of DNase-treated RNA with primer no. 407 and 200 units of SuperScript III reverse transcriptase in a 20-$\mu$l final volume (Invitrogen; see SI Table 3). Eight or 12 $\mu$l of cDNA was used for PCR in a 50-$\mu$l final volume containing 2.5 units of platinum *Pfx*, 3× $PCR_X$ enhancer solution (Invitrogen), and 15 pmol of each primer (nos. 222 and 407). A step-by-step procedure is provided in *SI Text*.

**Immunofluorescence.** Cells were transfected with 1 $\mu$g of plasmid DNA. The proteasome inhibitor MG132 (25 mM; Sigma) was added 24 or 36 h after transfection, and immunostaining was performed by standard procedures (*SI Text*) 5 h later by using the 9A12 mAb and the anti-Pitx1 rabbit serum (1:50). The *pCMV-Pitx1* expression vector (MGC-13954) was purchased from American Type Culture Collection containing full-length Pitx1 cDNA.

**Western Blotting.** Transfected control TE671 cells or myoblast primary cultures were lysed on ice in 150 $\mu$l of 10 mM Tris·HCl, pH 7.4/150 mM NaCl/0.1% Triton X-100, with protease inhibitors.

Proteins were separated by electrophoresis (10% SDS/PAGE) and transferred onto a PVDF membrane (Amersham Biosciences). After blocking, the Western blot was incubated with the 9A12 mAb (dilution 1:1,000 in PBS/0.2% Tween/2% BSA) followed by secondary antibodies coupled to HRP (Amersham Biosciences). We also used a rabbit serum raised against a DUX4-specific peptide (26). HRP was detected either with LiteAblot (Euroclone) or Super Signal West Femto Maximum Sensitivity Substrate (Pierce). Chemiluminescence was detected on ECL films (Amersham Biosciences). Detection of loading controls $\alpha$-tubulin (mAb, Sigma–Aldrich) or actin (rabbit polyclonal; Sigma–Aldrich) was similarly performed after membrane stripping (*SI Text*).

We thank Dr. Eric P. Hoffman for constructive review of the manuscript. The study was partially supported by National Institutes of Health Grants

NIAMS 5R21AR048318 (to Y.-W.C. and S.W.) and R21-NS046242 (to H. J. Worman and A.B.) and Facioscapulohumeral Society Grants DR-01 (to A.B.) and MGBF-011 (to Y.-W.C.). A.B.'s group was also funded by the Association Belge Contre les Maladies Neuromusculaires (Belgium), the Asssociation Française Contre les Myopathies (AFM, France), and the Muscular Dystrophy Association (Grant 3209). Additional support was generously provided by the Fischer–Shaw Foundation (to S.W. and A.B.) and by Mr. D. Frenzel (A.B.). E.A. and A.T. held predoctoral fellowships from the Fonds pour la Formation à la Recherche dans l'Industrie et dans l'Agriculture (Belgium), and M.B. held a predoctoral fellowship from the Ministry of Research and Technology (France). The AFM provided pre- and postdoctoral fellowships to S.S. and D.L.-C., respectively.

1. van der Maarel SM, Frants RR, Padberg GW (2006) *Biochim Biophys Acta* 1772:186–194.
2. Tawil R, Figlewicz DA, Griggs RC, Weiffenbach B (1998) *Ann Neurol* 43:279–282.
3. Wijmenga C, Hewitt JE, Sandkuijl LA, Clark LN, Wright TJ, Dauwerse HG, Gruter AM, Hofker MH, Moerer P, Williamson R, *et al.* (1992) *Nat Genet* 2:26–30.
4. Tupler R, Berardinelli A, Barbierato L, Frants R, Hewitt JE, Lanzi G, Maraschio P, Tiepolo L (1996) *J Med Genet* 33:366–370.
5. Winokur ST, Bengtsson U, Feddersen J, Mathews KD, Weiffenbach B, Bailey H, Markovich RP, Murray JC, Wasmuth JJ, Altherr MR, *et al.* (1994) *Chromosome Res* 2:225–234.
6. Gabellini D, Green MR, Tupler R (2002) *Cell* 110:339–348.
7. Jiang G, Yang F, van Overveld PG, Vedanarayanan V, van der Maarel S, Ehrlich M (2003) *Hum Mol Genet* 12:2909–2921.
8. Masny PS, Bengtsson U, Chung SA, Martin JH, van Engelen B, van der Maarel SM, Winokur ST (2004) *Hum Mol Genet* 13:1857–1871.
9. van Overveld PG, Lemmers RJ, Sandkuijl LA, Enthoven L, Winokur ST, Bakels F, Padberg GW, van Ommen GJ, Frants RR, van der Maarel SM (2003) *Nat Genet* 35:315–317.
10. Winokur ST, Chen YW, Masny PS, Martin JH, Ehmsen JT, Tapscott SJ, van der Maarel SM, Hayashi Y, Flanigan KM (2003) *Hum Mol Genet* 12:2895–2907.
11. Gabellini D, D'Antona G, Moggio M, Prelle A, Zecca C, Adami R, Angeletti B, Ciscato P, Pellegrino MA, Bottinelli R, *et al.* (2006) *Nature* 439:973–977.
12. Gabriels J, Beckers MC, Ding H, De Vriese A, Plaisance S, van der Maarel SM, Padberg GW, Frants RR, Hewitt JE, Collen D, *et al.* (1999) *Gene* 236:25–32.
13. Winokur ST, Bengtsson U, Vargas JC, Wasmuth JJ, Altherr MR, Weiffenbach B, Jacobsen SJ (1996) *Hum Mol Genet* 5:1567–1575.
14. Clapp J, Mitchell LM, Bolland DJ, Fantes J, Corcoran AE, Scotting PJ, Armour JA, Hewitt JE (2007) *Am J Hum Genet* 81:264–279.
15. Lyle R, Wright TJ, Clark LN, Hewitt JE (1995) *Genomics* 28:389–397.
16. Beckers M, Gabriëls J, van der Maarel S, De Vriese A, Frants RR, Collen D, Belayew A (2001) *Gene* 264:51–57.
17. Tremblay JJ, Goodyer CG, Drouin J (2000) *Neuroendocrinology* 71:277–286.
18. Tumpel S, Cambronero F, Ferretti E, Blasi F, Wiedemann L, Krumlauf R (2007) *Dev Biol* 302:646–660.
19. Ding H, Beckers MC, Plaisance S, Marynen P, Collen D, Belayew A (1998) *Hum Mol Genet* 7:1681–1694.
20. Kalionis B, O'Farrell PH (1993) *Mech Dev* 43:57–70.
21. Kumar P, Nazarali AJ (2001) *Cell Mol Neurobiol* 21:369–388.
22. Gehring WJ, Qian YQ, Billeter M, Furukubo-Tokunaga K, Schier AF, Resendez-Perez D, Affolter M, Otting G, Wuthrich K (1994) *Cell* 78:211–223.
23. Chan SK, Ryoo HD, Gould A, Krumlauf R, Mann RS (1997) *Development (Cambridge, UK)* 124:2007–2014.
24. Ryoo HD, Mann RS (1999) *Genes Dev* 13:1704–1716.
25. Rechsteiner M, Rogers SW (1996) *Trends Biochem Sci* 21:267–271.
26. Kowaljow V, Marcowycz A, Ansseau E, Conde CB, Sauvage S, Matteotti C, Arias C, Corona ED, Nunez NG, Leo O, *et al.* (2007) *Neuromuscul Disord* 17:611–623.
27. Salamov AA, Solovyev VV (2000) *Genome Res* 10:516–522.
28. Lanctot C, Moreau A, Chamberland M, Tremblay ML, Drouin J (1999) *Development (Cambridge, UK)* 126:1805–1810.
29. Szeto DP, Rodriguez-Esteban C, Ryan AK, O'Connell SM, Liu F, Kioussi C, Gleiberman AS, Izpisua-Belmonte JC, Rosenfeld MG (1999) *Genes Dev* 13:484–494.
30. Logan M, Tabin CJ (1999) *Science* 283:1736–1739.
31. Cole NJ, Tanaka M, Prescott A, Tickle C (2003) *Curr Biol* 13:R951–R952.
32. Shapiro MD, Marks ME, Peichel CL, Blackman BK, Nereng KS, Jonsson B, Schluter D, Kingsley DM (2004) *Nature* 428:717–723.
33. Tanaka M, Hale LA, Amores A, Yan YL, Cresko WA, Suzuki T, Postlethwait JH (2005) *Dev Biol* 281:227–239.
34. Civas A, Island ML, Genin P, Morin P, Navarro S (2002) *Biochimie* 84:643–654.
35. Island ML, Mesplede T, Darracq N, Bandu MT, Christeff N, Djian P, Drouin J, Navarro S (2002) *Mol Cell Biol* 22:7120–7133.
36. Kolfschoten IG, van Leeuwen B, Berns K, Mullenders J, Beijersbergen RL, Bernards R, Voorhoeve PM, Agami R (2005) *Cell* 121:849–858.
37. Padberg G, Adams C (2000) in *Neurogenetics*, ed Pulst S-M (Oxford Univ Press, Oxford), pp 105–116.
38. Arahata K, Ishihara T, Fukunaga H, Orimo S, Lee JH, Goto K, Nonaka I (1995) *Muscle Nerve* 2:S56–S66.
39. Winokur ST, Barrett K, Martin JH, Forrester JR, Simon M, Tawil R, Chung SA, Masny PS, Figlewicz DA (2003) *Neuromuscul Disord* 13:322–333.
40. Hewitt JE, Lyle R, Clark LN, Valleley EM, Wright TJ, Wijmenga, C., van Deutekom JC, Francis F, Sharpe PT, Hofker M, *et al.* (1994) *Hum Mol Genet* 3:1287–1295.
41. Yip DJ, Picketts DJ (2003) *FEBS Lett* 537:133–138.
42. Osborne RJ, Welle S, Venance SL, Thornton CA, Tawil R (2007) *Neurology* 68:569–577.
43. Alexiadis V, Ballestas ME, Sanchez C, Winokur S, Vedanarayanan V, Warren M, Ehrlich M (2007) *Biochim Biophys Acta* 1769:29–40.
44. van Geel M, Dickson MC, Beck AF, Bolland DJ, Frants RR, van der Maarel SM, de Jong PJ, Hewitt JE (2002) *Genomics* 79:210–217.
45. Lemmers RJ, de Kievit P, Sandkuijl L, Padberg GW, van Ommen GJ, Frants RR, van der Maarel SM (2002) *Nat Genet* 32:235–236.
46. Thomas NS, Wiseman K, Spurlock G, Macdonald M, Ustek D, Upadhyaya M (2006) *J Med Genet* 44:215–218.
47. Coppée F, Mattéotti C, Ansseau E, Sauvage S, Leclercq I, Leroy A, Marcowycz A, Figlewicz D, Ding H, Belayew A (2004) in *Facioscapulohumeral Muscular Dystrophy*: *Clinical Medicine and Mol Cell Biol*, eds Cooper D, Upadhyahya M (BIOS, Oxford), chap 9, pp 117–134.
48. Tupler R, Perini G, Pellegrino MA, Green MR (1999) *Proc Natl Acad Sci USA* 96:12650–12654.
49. Bakay M, Wang Z, Melcon G, Schiltz L, Xuan J, Zhao P, Sartorelli V, Seo J, Pegoraro E, Angelini C, *et al.* (2006) *Brain* 129:996–1013.
50. Chen YW, Hubal MJ, Hoffman EP, Thompson PD, Clarkson PM (2003) *J Appl Physiol* 95:2485–2494.
51. Chen YW, Zhao P, Borup R, Hoffman EP (2000) *J Cell Biol* 151:1321–1336.
52. Tumor Analysis Best Practices Working Group (2004) *Nat Rev Genet* 5:229–237.
53. Laoudj-Chenivesse D, Carnac G, Bisbal C, Hugon G, Bouillot S, Desnuelle C, Vassetzky Y, Fernandez A (2005) *J Mol Med* 83:216–224.