Identification of Variants in the 4q35 Gene FAT1 in Patients with a Facioscapulohumeral Dystrophy-Like Phenotype

To cite this version:
<hal-01662841>

HAL Id: hal-01662841
https://hal-amu.archives-ouvertes.fr/hal-01662841
Submitted on 13 Dec 2017

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

L’archive ouverte pluridisciplinaire HAL, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d’enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.
Identification of Variants in the 4q35 Gene FAT1 in Patients with a Facioscapulohumeral Dystrophy-Like Phenotype

Francesca Puppo,1,2† Eugenie Dionnet,1,2‡ Marie-Cécile Gaillard,1,2 Pascaleine Gaillard,3 Christel Castro,1,2 Catherine Vovan,4 Karine Bertaux,4 Rafaelle Bernard,4 Shahram Attarian,1,2,5 Kanako Goto,6 Ichizo Nishino,6 Yukiko Hayashi,7 Frédérique Magdinier,1,2 Martin Krahn,1,2,4 Françoise Helmbacher,9 Marc Bartoli,1,2,4+ and Nicolas Lévy1,2,4†

1Aix Marseille Université, GMGF, Marseille, France; 2Inserm, UMR S 910, Marseille, France; 3Institute for Research and Innovation in Biomedicine (IRIB), Inserm, UMR 1079, University of Rouen, Rouen, France; 4Département de Généétique Médicale et de Biologie Cellulaire, AP-HM, Hôpital d’Enfants de la Timone, Marseille, France; 5Department of Neurology and Neuromuscular Diseases, CHU La Timone, Marseille, France; 6NCNP, National Institute of Neuroscience, Tokyo, Japan; 7Department of Neurophysiology, Tokyo Medical University, Tokyo, Japan; 8Aix Marseille Université, CNRS, IBDM, UMR 7288, Marseille, France

Communicated by Mireille Claustres
Received 16 June 2014; accepted revised manuscript 11 January 2015.
Published online 23 January 2015 in Wiley Online Library (www.wiley.com/humanmutation). DOI: 10.1002/humu.22760

ABSTRACT: Facioscapulohumeral muscular dystrophy (FSHD) is linked to copy-number reduction (N < 10) of the 4q D4Z4 subtelomeric array, in association with DUX4-permissive haplotypes. This main form is indicated as FSHD1. FSHD-like phenotypes may also appear in the absence of D4Z4 copy-number reduction. Variants of the SMCHD1 gene have been reported to associate with D4Z4 hypomethylation in DUX4-compatible haplotypes, thus defining FSHD2. Recently, mice carrying a musclespecific knock-out of the protocadherin gene Fat1 or its constitutive hypomorphic allele were shown to develop muscular and nonmuscular defects mimicking human FSHD. Here, we report FAT1 variants in a group of patients presenting with neuromuscular symptoms reminiscent of FSHD. The patients do not carry D4Z4 copy-number reduction, 4q hypomethylation, or SMCHD1 variants. However, abnormal splicing of the FAT1 transcript is predicted for all identified variants. To determine their pathogenicity, we elaborated a minigene approach coupled to an antisense oligonucleotide (AON) assay. In vitro, four out of five selected variants induced partial or complete alteration of splicing by creating new splice sites or modifying splicing regulators. AONs confirmed these effects. Altered transcripts may affect FAT1 protein interactions or stability. Altogether, our data suggest that defective FAT1 is associated with an FSHD-like phenotype.


KEY WORDS: neuromuscular pathology; facioscapulohumeral dystrophy; FAT1-protocadherin

Introduction

Facioscapulohumeral muscular dystrophy (FSHD; MIM #158900) [Justin Besancon et al., 1964; Balatsouras et al., 2007] is the third most common human neuromuscular disease. Symptoms mainly appear during the second decade of life and are characterized by progressive weakness of the facial, scapular, and humeral muscles, later spreading toward the lower limb muscles. Nonmuscular tissues are also frequently affected in FSHD; hearing loss manifests in 75% of patients, and retinal telangiectasia occurs in 60% of patients [Wohlgemuth et al., 2004]. Learning difficulties and epilepsy may also occur as atypical symptoms in subsets of severely affected children [Saito et al., 2007].

Genetically, FSHD is linked to the presence of a reduced copy number (1–10 copies) of a tandemly repeated 3.3-kb segment (D4Z4 macrosatellite) on chromosome 4q35, whereas in healthy individuals, this repeat varies from 11 to 100 copies [Wijmenga et al., 1990; van Deutekom et al., 1993]. This main form of FSHD, referred to as FSHD1, accounts for approximately 95% of cases [Richards et al., 2012]. Such a heterozygous D4Z4-array reduction is associated with chromatin relaxation on specific permissive haplotypes [Lemmens et al., 2002], which might result in stabilization of the DUX4 transcript, encoded by the retrogene located in D4Z4, and expression of this transcription factor [Gabrieli et al., 1999; Dixit et al., 2007; Lemmers et al., 2010; Spurlock et al., 2010; Vanderplank et al., 2011]. A form of FSHD not linked to D4Z4 contraction accounts for 5% of patients (contraction-independent FSHD) [van Overveld et al., 2003]. In a majority of these patients described at present, now referred to as FSHD2, variants of SMCHD1 (MIM #614982; GenBank NM_015295.2) [Lemmens et al., 2012; Larsen et al., 2014; Lemmers et al., 2015], a gene known to play a role in chromatin relaxation and maintenance of X chromosome inactivation [Blewitt et al., 2008; Gendrel et al., 2012], have been reported to associate with D4Z4 hypomethylation in DUX4-compatible 4q haplotypes, thereby leading to DUX4 overexpression as well.

Nevertheless, large-scale studies performed on healthy individuals and FSHD patients have shown that the permissive haplotype (4qA161) on FSHD-sized repeats, although rare, is present in the healthy population, thus constituting a rare uncommon polymorphism [van Overveld et al., 2000; Lemmers et al., 2007; Sciot et al., 2012a, 2012b] and indicating that DUX4 protein expression is not in itself sufficient to trigger FSHD symptoms. Furthermore, DUX4 RNA and protein have been detected, although at low levels, in muscle biopsies of healthy individuals carrying DUX4-compatible haplotypes in the absence of 4q contraction [Jones et al., 2012; Brocqsaui et al., 2013]. This finding indicates that although necessary, DUX4 activation is not sufficient on its own to trigger the appearance.
of symptoms [Jones et al., 2012], implying the existence of modifier genes/mutations that contribute to this complex syndrome. As such, SMCHD1 variants were recently shown to constitute modifier alleles of FSHD1 [Sacconi et al., 2013; Larsen et al., 2014]. Therefore, evidence of cosegregation for multiple genetic factors began to accumulate, contributing to the genotype–phenotype correlation in a proportion of FSHD families. However, other cases of contraction-independent FSHD are not explained by SMCHD1/DUX4 synergy, leaving room for the identification of other genes involved in FSHD pathogenesis.

Recently, mouse models carrying a muscle-specific knock-out of the protocadherin gene Fat1 or its constitutive hypomorphic allele have been shown to develop defects in the shape and position of specific groups of muscles in the shoulder and face [Caruso et al., 2013]. The human FAT1 gene (MIM #600976; GenBank NM_005245.3) is composed of 27 exons spanning a region of 139 kb and encodes a 506-kDa transmembrane protein with 34 cadherin domains followed by a laminin and five EGF motifs in its extracellular part. Studies on the protocadherin Fat1 demonstrated its influence on smooth muscle cell motility [Hou and Sibinga, 2009], actin accumulation at neuronal synapses [Moeller et al., 2004], the Hippo signaling pathway [Bennett and Harvey, 2006; Cho et al., 2006], and cell polarity [Skouloudaki et al., 2009]. While constitutive loss of FAT1 leads to a significant degree of perinatal lethality in mice [Ciani et al., 2003; Caruso et al., 2013], incomplete alterations of the FAT1 gene have been identified in rare developmental 4q syndromes characterized by mild facial dysmorphisms, abnormalities in skeletal and cardiac development, and growth and mental retardation [Ockey et al., 1967; Bendavid et al., 2007; Kitisou-Tzeli et al., 2008]. Genetic links between FAT1 gene polymorphisms and susceptibility to bipolar disorder or schizophrenia have also been established [Blair et al., 2006; Abou Jamra et al., 2008; Jung and Jun 2013]. Finally, somatic mutations inducing high expression of FAT1 or its variant isoforms have been associated with both acute lymphoblastic leukemia [de Bock et al., 2012] and breast cancer progression [Lee et al., 2012], whereas other mutations (and CNV [copy number variation]) causing FAT1 protein loss of function are associated with glioblastoma, colorectal, or head and neck cancers [Morris et al., 2013a].

For the first time, in a study by Caruso et al. (2013), FAT1 was shown to control muscle patterning by modulating the polarity of myoblast migration during embryonic development, to be involved in regionalized muscle wasting, and to have a role in adult muscle fiber functions. Mice also exhibit extramuscular defects, such as retinal vasculopathy and abnormal inner ear patterning, which possibly represents a hearing impairment [Caruso et al., 2013]. Furthermore, misregulation of FAT1 expression was observed in fetal FSHD1 muscle tissue, and a CNV mapping in a putative regulatory enhancer of FAT1 preferentially segregated with FSHD-like patients who presented no D4Z4 contraction [Caruso et al., 2013]. Although FSHD-like symptoms only represent a subset of FAT loss-of-function phenotypes, the map of muscles that show developmental impairment in FAT1 mutants strongly resembles one affected by human FSHD, raising the provocative hypothesis that FAT1 acts as a modifier gene to the disease.

Here, we sought to identify FAT1 variants in 49 Japanese cases affected by neuromuscular disease that closely resembles FSHD, according to the diagnostic criteria defined by the European Expert Group on FSHD [Padberg et al., 1991] and presenting no D4Z4 copy-number reduction in either 4q35 or 10q26, no reciprocal rearrangements [Yamanaka et al., 2004], no hypomethylation at D4Z4, and no SMCHD1 mutations. In 10 out of 49 cases analyzed, we identified 10 different variants of the FAT1 gene. To determine their pathogenic effect on FAT1 splicing, we elaborated a minigene approach, coupled to an antisense oligonucleotide (AON) assay. The in vitro results showed a partial or complete alteration of splicing either by creating a new acceptor splice site or by modifying splicing regulators in FAT1 variants. AONs confirmed the effect of these nucleotide substitutions. Altogether, our data suggest that a defective FAT1 protein may be associated with an FSHD-like phenotype.

Methods

Bioinformatic predictions

Amino acid substitutions were predicted for their impact on protein function by three online algorithms: polymorphism phenotyping v2 (Polyphen-2), SIFT, and phyphyl (see Web links for references and Supp. Table S1 for results). PolyPhen-2 version 2.2.2 (http://genetics.bwh.harvard.edu/pph2) calculates the impact of amino acid substitutions using protein sequences from the UniProt database, structural data from PDB/DSSP, and comparative considerations from UCSC multiple alignments of 45 vertebrate genomes. For a mutation, PolyPhen-2 calculates the probability that this mutation is damaging and estimates a false-positive rate (the chance that the mutation is classified as damaging when it is in fact nondamaging). Based on the threshold of this rate, a mutation is qualitatively defined as benign, possibly damaging, or probably damaging. SIFT (v1.03, http://sift.jcvi.org) predictions are based on the degree of conservation of amino acid residues in sequence alignments derived from closely related sequences. The score ranges from 0 to 1. The amino acid substitution is predicted to be damaging if the score is ≤0.05. The impact of variants on FAT1 splicing was predicted by Human Splicing Finder (HSF; version 2.4.1) (cut-off for donor or acceptor sites is 55%) [Desmet et al., 2009] and Alamut Visual 2.2, released in June 2012. This software integrates four different algorithms for splicing prediction: SpliceSiteFinder (http://www.umd.be/HSF/; cut-off for donor and acceptor sites 70%), MaxEntScan (no cut-off provided), NN Splice (score for donor and acceptor ranges between 0 and 1, cut-off is 0.4), and GeneSplicer (no cut-off provided). For HSF, as well as for Alamut, the threshold for ESEFinder is the following: SF2/ASF, 1.956; SC35, 2.383; SRp40, 2.67; and SRp55, 2.676. Predictions were considered strong enough to be tested in a minigene reporter assay in the following cases: nucleotide substitutions that create or break regulatory ESE or SSE splicing sites or determine the creation of a new splicing site stronger than the natural one. The inclusion criteria for experimental analysis of variants also accounted for available data from dbSNP137 and included a minor allele frequency lower than 0.005 (Table 1; Supp. Table S1).

Methylation analysis

For sodium bisulfite sequencing, 2 μg of genomic DNA was denatured for 30 min at 37°C in 0.4 M NaOH and incubated overnight in a solution of 3 M sodium bisulfite, pH 5, and 10 mM hydroquinone using a previously described protocol [Ehrlich et al., 1982]. Converted DNA was then purified using the Wizard DNA CleanUp kit (Promega, Marseille, France) following manufacturer’s recommendation and precipitated with ethanol. Modified DNA was amplified using the forward 5′-AAATATGTTAGGGAAGGTGTAAGTT-3′ and
Table 1. Molecular and Clinical Details for Selected Mutations

<table>
<thead>
<tr>
<th>Exon</th>
<th>East Asian allele count/allele number</th>
<th>General allele count/allele number</th>
</tr>
</thead>
<tbody>
<tr>
<td>2</td>
<td>1/8892</td>
<td>2/124670</td>
</tr>
<tr>
<td>3</td>
<td>12/8894</td>
<td>22/124752</td>
</tr>
<tr>
<td>5</td>
<td>15/8724</td>
<td>195/120308</td>
</tr>
<tr>
<td>10</td>
<td>9/8878</td>
<td>9/124606</td>
</tr>
<tr>
<td>11</td>
<td>141/8888</td>
<td>155/124660</td>
</tr>
<tr>
<td>17</td>
<td>0</td>
<td>20/124668</td>
</tr>
<tr>
<td>22</td>
<td>0</td>
<td>25/124654</td>
</tr>
<tr>
<td>27</td>
<td>2/8894</td>
<td>2/124730</td>
</tr>
</tbody>
</table>

X indicates the presence of mutation relative to cDNA sequence and relative to GenBank reference NM_005245.3. Nucleotide numbering uses +1 as the A of the ATG translation initiation codon in the reference sequence, with the initiation codon as codon 1. Black shadow indicates validated effect on splicing, whereas gray one potentially damaging amino acid changes. Last two columns represent minor allele count in East Asian and general healthy population, respectively, as reported by Version0.2 ExAC (Exome Aggregation Consortium) database.

Reverse 5'-GGAGAGAGGGTTGTTATATTAAG-3' primer set designed with the MethPrimer software [Bird 2002] to amplify 21 CpGs (275 bp) within the D4Z4 proximal region [Gaillard et al., 2014]. We avoided CpGs in the primer sequence in order to amplify methylated and unmethylated DNA with the same efficiency. Amplification was conducted using the High Fidelity Taq polymerase (Roche, Paris, France) according to manufacturer's instructions. After initial denaturation, the amplification conditions were the following: 94°C for 20 sec, 54°C for 30 sec, 72°C for 1 min for 10 cycles, and then the elongation was increased by 30 sec at each subsequent cycle for 25 cycles. The PCR products were then purified using the Wizard SV gel and PCR Purification System (Promega) and cloned using the pGEM®-T Easy Vector cloning kit (Promega). Randomly selected colonies were grown overnight at 37°C with ampicillin selection and PCR amplified directly using T7 and SP6 primers. At least 10 independent colonies were sequenced according to the Sanger method by Eurofins MWG Operon (Ebersberg, Germany) with either T7 or SP6 primers.

Sequences were analyzed using BiQ Analyzer software, and the average methylation score was calculated as the number of methylated CpGs for the total number of CpGs in the reference sequence (Table 2).

**Cell culture**

HEK293 cells (human embryonic kidney cell line) were grown in Dulbecco’s modified Eagle medium (Gibco, Saint Aubin, France) supplemented with 15% fetal calf serum and 1% PSA (penicillin streptomycin actinomycin) in a 5% CO₂ atmosphere at 37°C.

**Generation of constructs for splicing minigene reporter assay**

The splicing minigene assay has been previously described [Gaillard et al., 2010]. DNA fragments corresponding to FAT1 wild-type and mutant exons surrounded by upstream and downstream intronic sequences were amplified from genomic DNA of patients heterozygous for the nucleotide substitution by using exon–specific forward and reverse primers (Supp. Table S2) carrying 5' tails with BamHI (or BglIII) and MluI restriction sites, respectively (underlined in Supp. Table S3). After digestion with BamHI (or BglIII) and MluI, the PCR products were inserted into the BamHI and MluI cloning sites in the intron of the pCAS2 vector containing a two-exon splicing reporter minigene. The insert was then sequenced to identify the minigene constructs containing the FAT1 exon carrying the wild-type or mutant nucleotide and to ensure that no extra substitutions were added during amplification or cloning.

**Transfection and analysis of RT-PCR products**

The wild-type and mutant minigene constructs were transiently transfected into HEK293 cells using the promofectene transfection reagent, according to manufacturer’s instructions (PromoKine, Heidelberg, Germany). Cells were collected 24 hr post-transfection. Total RNA was extracted using the PureLink® RNA Mini Kit (Ambion; Life Technologies, Saint Aubin, France), according to the manufacturer’s instructions, followed by a DNase treatment with the DNA-free Kit (Ambion; Life Technologies). The RT-PCR reactions were performed using the SuperScript OneStep RT-PCR with Platinum Taq kit (Life Technologies), according to the manufacturer’s instructions, with 500 ng RNA as a template in a 50-μl reaction volume. Reactions were performed using 300 nM of forward pCAS-KO1F (5’-TACGCTGGCGCCCATCACTAC-3') and reverse pCAS2R primers (5’-ATTGGGTAGTGAAGTTGGTGTC-3') [Gaillard et al., 2010]. The reverse transcription program had one cycle: 50°C for 30 min. The PCR program had 35 cycles of amplification of 96°C for 45 sec, 50°C for 45 sec, and 72°C for 1 min. RT-PCR products were separated by electrophoresis on 2% agarose 1000 (Invitrogen, Saint Aubin, France) gels containing ethidium bromide and visualized by exposure to nonsaturating ultraviolet light.

**AON design and transfection**

AONs with full-length 2'-O-methyl-substituted ribose moieties and phosphorothioate internucleotide linkages were purchased from Eurogentec (Angers, France). AONs were designed according to several criteria: sequences targeting exon 5 (5’-U-U-C-U-U-C-C-G-C-A-G-G-G-A-U-G-U-U-C-U-G-3') and exon 17 (5’-G-A-C-A-C-U-U-G-A-G-U-U-C-U-C-C-U-G-G-A-3') were designed complementarily to the wild-type and mutant nucleotide, respectively. In parallel, the AON sequence for exon 11
<table>
<thead>
<tr>
<th>5'-D4Z4 methylation percentage</th>
<th>Patients</th>
<th>5'-D4Z4 Methylation Percentage and SMCHD1 (GenBank NM_015295) Screening Results</th>
</tr>
</thead>
<tbody>
<tr>
<td>J2</td>
<td>84%</td>
<td>HTZ rs2430853, MAF gp = 46%, MAF ip (C) = 29%</td>
</tr>
<tr>
<td>J6</td>
<td>65%</td>
<td>HTZ rs2430853, MAF gp = 46%, MAF ip (C) = 29%</td>
</tr>
<tr>
<td>J7</td>
<td>71%</td>
<td>HTZ rs2430853, MAF gp = 46%, MAF ip (C) = 29%</td>
</tr>
<tr>
<td>J15</td>
<td>80%</td>
<td>HTZ rs2430853, MAF gp = 46%, MAF ip (C) = 29%</td>
</tr>
<tr>
<td>J16</td>
<td>57%</td>
<td>HTZ rs2430853, MAF gp = 46%, MAF ip (C) = 29%</td>
</tr>
<tr>
<td>J21</td>
<td>65%</td>
<td>HTZ rs2430853, MAF gp = 46%, MAF ip (C) = 29%</td>
</tr>
<tr>
<td>J29</td>
<td>72%</td>
<td>HTZ rs2430853, MAF gp = 46%, MAF ip (C) = 29%</td>
</tr>
<tr>
<td>J51</td>
<td>72%</td>
<td>HTZ rs2430853, MAF gp = 46%, MAF ip (C) = 29%</td>
</tr>
</tbody>
</table>

SMCHD1 exons and SNPs identified:

1. HTZ rs2430853, MAF gp = 46%, MAF ip (C) = 29%

14. HTZ rs8090967, MAF gp = 37%, MAF ip = 37%
   HTZ rs377473058, MAF gp = 37%, MAF ip = 37%
   HTZ rs635132, MAF gp = 37%, MAF ip = 37%

18/19 HTZ rs28446296, MAF gp = 31%, MAF ip = 22%
   HZT rs8090967, MAF gp = 31%, MAF ip = 22%

21. HTZ rs635422, MAF gp = 31%, MAF ip = 31%
    HZT rs635422, MAF gp = 31%, MAF ip = 31%

22. HTZ rs1693516, MAF gp = 31%, MAF ip = 31%
    HZT rs1693516, MAF gp = 31%, MAF ip = 31%

26. HZT rs1693516, MAF gp = 31%, MAF ip = 31%
    HZT rs1693516, MAF gp = 31%, MAF ip = 31%

28. PCR not available

(Continued)
<table>
<thead>
<tr>
<th>Patients</th>
<th>J2</th>
<th>J6</th>
<th>J7</th>
<th>J15</th>
<th>J16</th>
<th>J21</th>
<th>J29</th>
<th>J51</th>
</tr>
</thead>
<tbody>
<tr>
<td>HTZ rs2019793</td>
<td>31/32</td>
<td>MAF gp = 18%</td>
<td>HTZ rs2019793</td>
<td>3%</td>
<td>MAF gp = 3%</td>
<td>HTZ rs2019793</td>
<td>1%</td>
<td>MAF gp = 1%</td>
</tr>
<tr>
<td>HMZ rs2019793</td>
<td>31/32</td>
<td>MAF gp = 3%</td>
<td>HMZ rs2019793</td>
<td>5%</td>
<td>MAF gp = 5%</td>
<td>HMZ rs2019793</td>
<td>1%</td>
<td>MAF gp = 1%</td>
</tr>
<tr>
<td>HTZ rs3214732</td>
<td>48</td>
<td>MAF gp = 33%</td>
<td>MGZ rs300291</td>
<td>3%</td>
<td>MAF gp = 3%</td>
<td>MGZ rs300291</td>
<td>1%</td>
<td>MAF gp = 1%</td>
</tr>
<tr>
<td>MGZ rs300291</td>
<td>48</td>
<td>MAF gp = 33%</td>
<td>MGZ rs300291</td>
<td>3%</td>
<td>MAF gp = 3%</td>
<td>MGZ rs300291</td>
<td>1%</td>
<td>MAF gp = 1%</td>
</tr>
</tbody>
</table>

In the upper part of the table, black sections of pie charts represent percentage of proximal D4Z4 methylated CpG/sequence in eight out of 10 patients carrying FAT1 variants. In the lower part of the table, Minor Allele Frequencies from general population (gp) and from Asian population (jp) are expressed as MAF for the SNPs detected each of the eight patients analyzed. No variants were depicted in exons not showed in this table.
(5′-A′U-C′A′G′U-U′G′C′U-G′U′G′A′U′A′G′U′A′A′-3′) was designed complementarily to the variant sequence.

HEK293 cells (4 × 10^5 cells) grown in 12-well plates to 80% confluence were transfected with 1.1 μg of AON and 0.5 μg of pCAS2 constructions using Lipofectamine 2000 reagent according to the manufacturer’s instructions (Life Technologies). Total RNA was isolated from cultured cells 24 hr after transfection, and RT-PCR analysis was performed as described above.

**Results**

In this work, 49 Japanese cases affected by neuromuscular disease were selected based on the homogeneity of muscles presenting clinical signs (Table 3) and closely resembling FSHD according to the diagnostic criteria defined by the European Expert Group on FSHD [Padberg et al., 1991]. No patients presented with D4Z4 copy-number reduction [Yamanaka et al., 2004]. The age distribution of patients with available clinical information was 60.8% of patients under the age of 30 at the time of disease presentation. The facial, scapular, and humeral muscles were simultaneously affected in 45.8% of patients, and 89.6% of them had at least two out of the three muscle groups affected. Progression of the disease toward the lower limbs was present in 72.9% of patients, with most affected at the proximal limb muscles. Other myopathies presenting with similar phenotypic appearances but specific histological defects were ruled out in 47 out of 49 individuals by muscular biopsy examination. Specifically, muscle from the J21 case showed scattered fibers with rimmed vacuoles. Protein analyses showed normal expression of dystrophin, sarcoglycans, dysglycans, merosin, collagen 6, dysferlin, caveolin-3, calpain 3, and emerin. A sequence analysis of the dystrophin, sarcoglycans, dystroglycans, merosin, collagen 6, dysferlin, caveolin-3, calpain 3, and emerin. A sequence analysis of the dystrophin, sarcoglycans, dystroglycans, merosin, collagen 6, dysferlin, caveolin-3, calpain 3, and emerin. A sequence analysis of the dystrophin, sarcoglycans, dystroglycans, merosin, collagen 6, dysferlin, caveolin-3, calpain 3, and emerin. A sequence analysis of the dystrophin, sarcoglycans, dystroglycans, merosin, collagen 6, dysferlin, caveolin-3, calpain 3, and emerin. A sequence analysis of the dystrophin, sarcoglycans, dystroglycans, merosin, collagen 6, dysferlin, caveolin-3, calpain 3, and emerin. A sequence analysis of the dystrophin, sarcoglycans, dystroglycans, merosin, collagen 6, dysferlin, caveolin-3, calpain 3, and emerin. A sequence analysis of the dystrophin, sarcoglycans, dystroglycans, merosin, collagen 6, dysferlin, caveolin-3, calpain 3, and emerin. A sequence analysis of the dystrophin, sarcoglycans, dystroglycans, merosin, collagen 6, dysferlin, caveolin-3, calpain 3, and emerin. A sequence analysis of the dystrophin, sarcoglycans, dystroglycans, merosin, collagen 6, dysferlin, caveolin-3, calpain 3, and emerin. A sequence analysis of the dystrophin, sarcoglycans, dystroglycans, merosin, collagen 6, dysferlin, caveolin-3, calpain 3, and emerin. A sequence analysis of the dystrophin, sarcoglycans, dystroglycans, merosin, collagen 6, dysferlin, caveolin-3, calpain 3, and emerin. A sequence analysis of the dystrophin, sarcoglycans, dystroglycans, merosin, collagen 6, dysferlin, caveolin-3, calpain 3, and emerin. A sequence analysis of the dystrophin, sarcoglycans, dystroglycans, merosin, collagen 6, dysferlin, caveolin-3, calpain 3, and emerin. A sequence analysis of the dystrophin, sarcoglycans, dystroglycans, merosin, collagen 6, dysferlin, caveolin-3, calpain 3, and emerin. A sequence analysis of the dystrophin, sarcoglycans, dystroglycans, merosin, collagen 6, dysferlin, caveolin-3, calpain 3, and emerin. A sequence analysis of the dystrophin, sarcoglycans, dystroglycans, merosin, collagen 6, dysferlin, caveolin-3, calpain 3, and emerin. A sequence analysis of the dystrophin, sarcoglycans, dystroglycans, merosin, collagen 6, dysferlin, caveolin-3, calpain 3, and emerin. A sequence analysis of the dystrophin, sarcoglycans, dystroglycans, merosin, collagen 6, dysferlin, caveolin-3, calpain 3, and emerin. A sequence analysis of the dystrophin, sarcoglycans, dystroglycans, merosin, collagen 6, dysferlin, caveolin-3, calpain 3, and emerin. A sequence analysis of the dystrophin, sarcoglycans, dystroglycans, merosin, collagen 6, dysferlin, caveolin-3, calpain 3, and emerin. A sequence analysis of the dystrophin, sarcoglycans, dystroglycans, merosin, collagen 6, dysferlin, caveolin-3, calpain 3, and emerin. A sequence analysis of the dystrophin, sarcoglycans, dystroglycans, merosin, collagen 6, dysferlin, caveolin-3, calpain 3, and emerin. A sequence analysis of the dystrophin, sarcoglycans, dystroglycans, merosin, collagen 6, dysferlin, caveolin-3, calpain 3, and emerin. A sequence analysis of the dystrophin, sarcoglycans, dystroglycans, merosin, collagen 6, dysferlin, caveolin-3, calpain 3, and emerin. A sequence analysis of the dystrophin, sarcoglycans, dystroglycans, merosin, collagen 6, dysferlin, caveolin-3, calpain 3, and emerin. A sequence analysis of the dystrophin, sarcoglycans, dystroglycans, merosin, collagen 6, dysferlin, caveolin-3, calpain 3, and emerin. A sequence analysis of the dystrophin, sarcoglycans, dystroglycans, merosin, collagen 6, dysferlin, caveolin-3, calpain 3, and emerin. A sequence analysis of the dystrophin, sarcoglycans, dystroglycans, merosin, collagen 6, dysferlin, caveolin-3, calpain 3, and emerin. A sequence analysis of the dystrophin, sarcoglycans, dystroglycans, merosin, collagen 6, dysferlin, caveolin-3, calpain 3, and emerin. A sequence analysis of the dystrophin, sarcoglycans, dystroglycans, merosin, collagen 6, dysferlin, caveolin-3, calpain 3, and emerin. A sequence analysis of the dystrophin, sarcoglycans, dystroglycans, merosin, collagen 6, dysferlin, caveolin-3, calpain 3, and emerin. A sequence analysis of the dystrophin, sarcoglycans, dystroglycans, merosin, collagen 6, dysferlin, caveolin-3, calpain 3, and emerin. A sequence analysis of the dystrophin, sarcoglycans, dystroglycans, merosin, collagen 6, dysferlin, caveolin-3, calpain 3, and emerin. A sequence analysis of the dystrophin, sarcoglycans, dystroglycans, merosin, collagen 6, dysferlin, caveolin-3, calpain 3, and emerin. A sequence analysis of the dystrophin, sarcoglycans, dystroglycans, merosin, collagen 6, dysferlin, caveolin-3, calpain 3, and emerin. A sequence analysis of the dystrophin, sarcoglycans, dystroglycans, merosin, collagen 6, dysferlin, caveolin-3, calpain 3, and emerin. A sequence analysis of the dystrophin, sarcoglycans, dystroglycans, merosin, collagen 6, dysferlin, caveolin-3, calpain 3, and emerin. A sequence analysis of the dystrophin, sarcoglycans, dystroglycans, merosin, collagen 6, dysferlin, caveolin-3, calpain 3, and emerin. A sequence analysis of the dystrophin, sarcoglycans, dystroglycans, merosin, collagen 6, dysferlin, caveolin-3, calpain 3, and emerin. A sequence analysis of the dystrophin, sarcoglycans, dystroglycans, merosin, collagen 6, dysferlin, caveolin-3, calpain 3, and emerin. A sequence analysis of the dystrophin, sarcoglycans, dystroglycans, merosin, collagen 6, dysferlin, caveolin-3, calpain 3, and emerin. A sequence analysis of the dystrophin, sarcoglycans, dystroglycans, merosin, collagen 6, dysferlin, caveolin-3, calpain 3, and emerin. A sequence analysis of the dystrophin, sarcoglycans, dystroglycans, merosin, collagen 6, dysferlin, caveolin-3, calpain 3, and emerin. A sequence analysis of the dystrophin, sar...
Figure 1. Abnormal splicing in FAT1-derived transcripts. A: Diagram representing FAT1 protein domains with the positions of the four variants tested by the minigene splicing assay coupled with the AON assay. Briefly, wild-type and mutant pCAS2 minigene constructs containing different FAT1 exons of interest were transfected into HEK293 cells alone or in combination with an AON specifically directed against the variant or wild-type allele. At 24 hr after transfection, the splicing patterns of the minigenes were monitored by RT-PCR. All variants are expressed as the cDNA sequence and refer to GenBank NM_005245.3. B: Transcripts amplified from constructs containing wild-type (pFAT1 Exon5 c.3770G) and variant (pFAT1 Exon5 c.3770A) alleles of FAT1 exon 5 in the presence or absence of AON c.3770G and the empty pCAS2 minigene were separated on an agarose gel. The transcript amplified from the exon 5 variant construction containing exons from the pCAS2 minigene only is indicated by a star (*). The effect of cotransfection of AON c.3770G and the wild-type exon 5 construct is presented together with a representation of the predicted effect on the protein. C: Agarose gel electrophoresis of transcripts produced from constructs containing wild-type (pFAT1 Exon11 c.8991G) and variant (pFAT1 Exon11 c.8991A) FAT1 exon 11 alone or cotransfected with AON c.8991A (mut). A star (*) indicates the transcript specifically produced from the variant exon 11 minigene and corresponding to 114 nucleotides deleted from exon 11. AON c.8991A rescued the exon 11 variant allele, allowing expression of the complete wild-type form of the exon. The potential effect on the FAT1 protein is also represented. D: Transcripts produced from constructs containing wild-type (pFAT1 Exon17 c.10331A) or variant (pFAT1 Exon17 c.10331G) FAT1 exon 17 in the presence or absence of AON (c.10331A, c.3770G) and the pCAS2 minigene. As indicated by a star (*), the transcript amplified from the exon 17 variant construct only contains exons from the pCAS2 minigene. The effect of cotransfection of AON c.10331A with the pFAT1 Exon17 c.10331G minigene is also presented together with the schema of the effect on the translated FAT1 protein. AON c.3770G from exon 5 does not have a specific splicing effect on wild-type exon 17.
splicing silencer, as predicted by HSF and Alamut (Supp. Table S1). Interestingly, cotransfection with the AON carrying the wild-type sequence (c.3770G) induced a partial skipping of the exon, suggesting that this AON masks the predicted ESE element (Fig. 1a). The aberrant transcript missing exon 5 (RNA r.3643_3972del) produced by the variant allele would result, if translated, into an in-frame deletion of 110 amino acids at the protein level (p.Val1215_Ser1324del) with partial truncation of the cadherin 10 and 11 extracellular domains.

The second nucleotide substitution with an evident effect on splicing was the c.8963A>T transversion, located in exon 11 and identified in patient J29. The minigene assay showed the production of two transcripts: a primary one corresponding to normal exon inclusion and an aberrant minor transcript (5% expression ratio versus the full-length transcript), which skips exon 11 (Fig. 1c). These results suggest that this variant alters splicing by disrupting ESE motifs, in agreement with the prediction (Supp. Table S1). If translated, the misspliced RNA (RNA r.8879_9075del) would result in a frameshift with the creation of a premature stop codon (p.Gly2960Asp).

Cotransfection of the AON specific for exon 5 (c.3770G) and a wild-type exon 17 construct did not interfere with exon 17 splicing. Thus, these results confirm that the wild-type AON specifically masks an element involved in exon inclusion. The main consequence of exon 17 skipping is likely the truncation of the cadherin 32 and 33 domains (p.Thr3403_Glu3431del) without the loss of the downstream reading frame.

Other nucleotide substitutions were also predicted to interfere with splicing, but they could not be easily tested, as they were not suitable for the conditions required for the minigene assay. In particular, patient J21 carries two different substitutions: c.2215A>G (p.Met739Val) and c.13374G>A (p.Gln4458Gln). Testing these two variants using a minigene-based assay was not possible because of their location in FAT1 exons 2 and 27, respectively. As present, the parents’ DNA is required to determine whether one of the two substitutions is a de novo substitution or if both substitutions are carried on the same allele. Both substitutions are predicted to affect splicing with equal strength, and an analysis of their respective impact on FAT1 transcription in the patient’s biological samples, such as a muscle biopsy, would be interesting.

**Discussion**

In this study, we identified 10 different variants in the FAT1 gene in 10 out of 49 Japanese patients affected by an FSHD-like neuromuscular disease. The diagnosis was based on criteria defined by the European Expert Group on FSHD [Padberg et al., 1991]; the patients presented no D4Z4 copy-number reduction in either 4q35 or 10q26 and no reciprocal rearrangements [Yamanaka et al., 2004]. To rule out myopathies presenting similar phenotypic appearance but specific histological defects, a muscular biopsy examination was performed in 47 out of 49 individuals. For J21, the muscle biopsy showed nonspecific scattered fibers with rimmed vacuoles, and the J51 diagnosis of Nemaline myopathy was based on histological examination of his muscle biopsy (data not shown). In both cases, neither mutations in known genes nor expression defects in myopathy-related proteins were observed. Thus, mutations in new genes responsible for the neuromuscular phenotype have not been ruled out. Interestingly, as the simultaneous presence of hypomethylation in D4Z4 regions in the context of a DUX4-permissive haplotype and variants in the SMCHD1 gene may cause FSHD2 and contribute to FSHD1 [Lemmers et al., 2012; Sacconi et al., 2013; Larsen et al., 2014; Lemmers et al., 2015], we measured the D4Z4 methylation level according to the methods and results recently published by our group [Gaillard et al., 2014]. In that publication, a significant reduction in the DNA methylation level at the D4Z4 proximal region was reported for individuals with clinical FSHD but not carrying a
copy-number reduction (FSHD2) by sodium bisulfite sequencing, with a global level of methylated CpG/sequence below the threshold of 50% [Gaillard et al., 2014]. In the group of patients investigated here, the average methylation level in the proximal D4Z4 region was above 50% (72%), meaning that in these patients, FSHD symptoms are not associated with epigenetic changes at the 4q35 region (Table 2). Furthermore, no mutations in SMCHD1 were detected in any of the patients carrying nucleotide variants for FAT1 (Table 2). Even if the diagnosis of FSHD is based on the clinical evaluation of symptoms, genetic testing is required to confirm FSHD1 or FSHD2. However, a small proportion of patients with FSHD-like symptoms are not associated with D4Z4 copy-number reduction or with hypomethylated 4qA alleles and SMCHD1 variants, as reported here. In this situation, while the diagnosis may be questioned and reevaluated for some, it remains likely that the specific FSHD clinical signs might result from other genetic changes that affect processes also involved in FSHD. Identifying such genetic causes of FSHD-like cases would likely teach us about the biological mechanisms of this pathology. Thus, hypomorphic Fat1 mice presenting a FSHD-like phenotype [Caruso et al., 2013] and the identification of FAT1 variants in FSHD-like patients raise the challenging idea that FAT1 might be a disease gene associated with FSHD-like symptoms.

For some of the FAT1 variants depicted here, the minor allele count in the general and Japanese populations is reported, when available, in Table 1. Thus, we propose that these variants might represent very rare mutations that were possibly identified in presymptomatic individuals. Nonetheless, incomplete penetrance may not be excluded at this point. Consistently, a CNV located in a putative regulatory enhancer of FAT1 has previously been shown to preferentially segregate with FSHD in noncontracted FSHD-like patients, hence constituting the basis of tissue-specific alterations in FAT1 expression [Caruso et al., 2013].

The variants identified here fall into two categories that are not incompatible with each other. Six out of 10 variants led to amino acid substitutions localized in different extracellular cadherin domains of FAT1 and have been predicted to have a deleterious impact on protein function by several algorithms. Moreover, for all nucleotide substitutions, the drastic loss of splicing regulator sites or the creation of new splicing sites stronger than the natural ones have been suggested to give rise to partial or complete splicing defects. These defects would lead to shorter half-lives of aberrant mRNAs or the production of deleterious forms of the translated FAT1 protein. Thus, to demonstrate their pathogenic effect on FAT1 transcript splicing, we elaborated a minigene approach coupled to an AON assay. Some of these variants showed potential splicing alterations as well as deleterious amino acid changes, suggesting that both possibilities can occur; for example, any abnormally spliced mRNA, if translated, could result in a functionally aberrant protein. Therefore, only further experimental validation will determine the true functional relevance of each process. The FAT1 protein can be considered a “model protein,” allowing us to correlate predictions to experimental findings. Moreover, the results shown here indicate that selected nucleotide substitutions have splicing effects in in vitro minigene assays. Interestingly, splicing defects might also depend on the chromosomal context or the presence of tissue-specific regulatory elements, which could only be analyzed in biological samples derived from patients. Nonetheless, our group recently showed that a minigene in vitro approach is reliable for confirming endogenous splicing defects (Kergourlay et al., 2014), supporting the idea that an actual damaging effect can be considered for variants showing an in vitro effect, even in the absence of in vivo confirmation.

In particular, the absence of an effect in c.4723G>A (p.Ala1575Thr) (Supp. Fig. S1) does not exclude the impact of this variant on the endogenous transcript. This nucleotide substitution is strongly predicted to cause a deleterious amino acid substitution, as detailed in Supp. Table S1. To further validate the reliability of prediction algorithms at the experimental level, proteomic approaches as well as functional tests focusing on cadherin domain structure are in development. In the case of the c.3770G>A (p.Arg1257Gln), c.8991G>A (pThr2997Thr), and c.10331A>G (p.Asn3444Ser) transitions with a validated impact on splicing, aberrantly spliced mRNA isoforms may have more rapid turnover and shorter half-lives, thereby affecting intracellular FAT1 concentrations. The splicing effect is also compatible with an additional issue at the protein level because these aberrant transcripts, if translated, have been predicted to contribute to deleterious amino acid substitutions as well. For one of these variants (c.3770G>A), we showed that it completely prevents the in vitro inclusion of exon 5 in the transcript, thus producing an r.3643_3972del mRNA. If translated, the corresponding protein would lack the cadherin 10 and 11 extracellular domains without loss of the downstream reading frame. Nevertheless, the effect of this variant on splicing in vivo could be less drastic, allowing the partial integration of exon 5 carrying the variant nucleotide in the transcript. Thus, it would be interesting to investigate the importance of the arginine 1257 change to glutamine on FAT1 protein function as well as the potential dominant-negative effect due to the coexistence of multiple FAT1 protein-splicing isoforms in cells. Along the same line, c.10331A>G (p.Asn3444Ser) may contribute to two coexisting mRNA transcripts: the aberrant mRNA r.10207_10351del (p.Thr3403_Glu3451del) and the normal transcript, which may be translated into a FAT1 protein with a serine replacing the aspartic acid in position 3444. Even if these variants are compatible with FAT1 protein frame conservation, extracellular cadherin domains are expected to be lost and likely affect the stability and/or protein–protein interactions. Similar consequences may be predicted for c.8991G>A (pThr2997Thr) in which the aberrant mRNA r.8879_8992del, coexisting with the normal transcript, may be translated into a FAT1 protein truncated for cadherin domain 27. Interestingly, patient J29 carries a c.8963A>T transversion (p.Lys2988Ile) that is predicted to disrupt the ESE motifs Traβ and 9G8 and to create a new hnRNP A1 site, whereas amino acid substitution is not expected to have deleterious consequences on the protein. Here, we show the partial production of a variant transcript missing exon 11. The juxtaposition of exons 10 and 12 would create a frameshift in the FAT1 transcript and a premature stop codon 23 nucleotides after the beginning of exon 12 (p.Gly2960Asp*9). This incomplete FAT1 mRNA, if translated into a truncated FAT1 protein, may exert dominant-negative activity. Alternatively, the misspliced RNA (r.8879_9075del) may be eliminated by the nonsense-mediated mRNA decay pathway, causing haploinsufficiency reminiscent of the fat1 mouse model [Caruso et al., 2013].

AONs recognize and block special sequences in the neotranslated RNA that are otherwise bound by splicing protein complexes [Wein et al., 2010]. Based on the targeted splicing sequence, AONs may lead to the partial or total skipping of flanking exons. In our case, the AONs were designed to precisely recognize either the substituted or wild-type alleles, allowing us to demonstrate the specific effect of FAT1 alterations in vitro. Next, we will apply the AONs that mimic nucleotide substitutions to FAT1-expressing cells and animal models to characterize the functional consequences of these substitutions on the translated proteins. Moreover, for c.8991G>A, the second minor transcript in which the first 114 exonic nucleotides were deleted, as observed using the minigene test, was rescued by cotransfection of AON c.8991A, which carries the variant nucleotide to mask the created splicing acceptor site. Animal models reproducing this variant are under development to evaluate the
therapeutic consequences of AON injection and splicing rescue in vivo.

Some recent algorithms, which have been developed for functional annotation of genetic variants from high-throughput sequencing data, suggest that FAT1 is a dispensable human gene based on the number of stop codons identified in the study [Wang et al., 2010]. Interestingly, our study did not identify stop codons in any of the screened individuals or in healthy controls, whereas the constitutive loss of FAT1 function leads to perinatal lethality [Ciani et al., 2003; Caruso et al., 2013]. Thus, in independent functional and genetic studies, the partial preservation of FAT1 function would still be compatible with life. Accordingly, defective forms of the protein would have tissue-specific impacts and would be exerted at low doses by a dominant-negative effect. Hence, altered FAT1 would play a pathogenic role by affecting only specific interactions with protein partners and only during specific stages of development. Nevertheless, consistent with mice carrying the tissue-specific knock-out of FAT1 or its constitutive hypomorphic allele [Caruso et al., 2013], we cannot rule out haploinsufficiency as the pathogenic mechanism, at least for some of the variants reported here.

In addition to our present results, other germline mutations in the FAT1 gene may lead to developmental defects in subtle pathologies, such as schizophrenia and bipolar disorder susceptibility [Okey et al., 1967; Blair et al., 2006; Bendavid et al., 2007; Abou Jamra et al., 2006; Blair IP, Checcutti AF, Badenhop RF, Scimone A, Moses MJ, Adams LJ, Craddock N, Green E, Kirov G, Owen MJ, Kook BJ, Donald IA, et al. 2006. Positional cloning, association analysis and expression studies provide convergent evidence that the cadherin gene FAT1 contains a bipolar disorder susceptibility allele. Mol Psychiatry 11:372–383.]

According to the evidence from the mouse model recently published and the identification of the CNV located in a transcriptional enhancer of the human FAT1 gene, which segregates in noncontracted FSHD-like patients, our genetic data further strengthen the link between the FAT1 gene and FSHD-like neuromuscular diseases.

References


