Expanding the mutation spectrum in FSHD and ICF syndrome

Boogaard, T.L. van den

Citation

Boogaard, T. L. van den. (2018, February 13). *Expanding the mutation spectrum in FSHD and ICF syndrome*. Retrieved from https://hdl.handle.net/1887/60938

Author: Boogaard, T.L. van den
Title: Expanding the mutation spectrum in FSHD and ICF syndrome
Issue Date: 2018-02-13

Marlinde L van den Boogaard[1], Richard JFL Lemmers[1], Judit Balog[1], Patrick J van der Vliet[1], Nienke van der Stoep[2], Stephen J Tapscott[3], Sabrina Sacconi[4,5], Rabi Tawil[6], Silvère van der Maarel[1]

1 Department of Human Genetics, Leiden University Medical Center, Leiden, The Netherlands
2 Department of Clinical Genetics, Leiden University Medical Center, Leiden, The Netherlands
3 Division of Human Biology, Fred Hutchinson Cancer Research Center, Seattle, WA, USA
4 Neuromuscular Diseases Centre, University Hospital of Nice, Nice, France
5 CNRS UMR7277, IBV, Faculty of Medicine, University of Nice, Nice, France
6 Department of Neurology, University of Rochester Medical Center, Rochester, USA

4

INTRONIC *SMCHD1* VARIANTS IN FSHD

In preparation

ABSTRACT

Fascioscapulohumeral dystrophy is associated with partial chromatin relaxation of the *DUX4* encoding D4Z4 macrosatellite repeat located on chromosome 4, and transcriptional derepression of *DUX4* in skeletal muscle. The most common form, FSHD1, is caused by a D4Z4 repeat array contraction to 1-10 units (normal range 8-100 units). The less common form, FSHD2, is most often caused by heterozygous variants in *SMCHD1*, which encodes a chromatin modifier which binds to D4Z4 to maintain a repressed chromatin state. In this study we identified intronic variants in *SMCHD1* in two FSHD families. In the first family we identified a variant 15 nucleotides proximal from the 3' splice site of exon 14. This *SMCHD1* variant creates a 3' splice site, which results in partial intron retention with inclusion of the distal 14 nucleotides of intron 13 into the transcript. In the second family we identified a deep intronic variant in intron 34. This *SMCHD1* variant creates a 3' splice site in intron 34, which results in exonisation of 53 nucleotides of intron 34. In this family the deep intronic variant acts as a modifier of disease severity. In both families the aberrant transcripts are predicted to lead to a premature stop codon. The identification of these intronic variants further expands the *SMCHD1* mutation spectrum in FSHD2 and emphasizes the importance of screening for intronic variants in *SMCHD1*.

INTRODUCTION

Fascioscapulohumeral dystrophy (FSHD, [OMIM 158900 and 158901]) is a common muscular dystrophy in adults (prevalence ~1:8.000) and is clinically mainly characterized by progressive weakness and wasting of the facial, shoulder girdle, trunk and upper arm muscles[1; 2]. Most often the onset of the disease occurs during the second decade of life. However, both within and between families there is a large variability in disease onset and progression[3]. Two genetic forms of FSHD have been identified, FSHD1 and FSHD2, which are clinically almost indistinguishable[4], but seem to represent opposite extremes of a disease spectrum[5]. Both forms are associated with partial chromatin relaxation of the D4Z4 macrosatellite repeat array on chromosome 4 in somatic tissue, characterized by reduced CpG methylation and loss of repressive histone marks[6-8]. This chromatin relaxation results in transcriptional derepression of the D4Z4 encoded *DUX4* gene in skeletal muscle[9]. DUX4 is a transcription factor normally expressed in the germ line and cleavage stage embryos, that is normally suppressed in most other somatic tissues[9-12]. DUX4 causes cell death when overexpressed in somatic cell lines or endogenously but inappropriately expressed in FSHD myotubes[13; 14]. The D4Z4 chromatin relaxation in FSHD must occur on a permissive chromosome 4 (4qA haplotype), which contains a polymorphic *DUX4* polyadenylation signal distal to the D4Z4 repeat array[15]. This polyadenylation signal is required for the production of stable DUX4 mRNA in somatic cells. Chromatin relaxation on the homologous D4Z4 repeats on non-permissive 4qB or 10q chromosomes do not cause FSHD since these chromosomal backgrounds lack a somatic *DUX4* polyadenylation signal[15].

FSHD1, the most common form of FSHD (>95%), is caused by contraction of the D4Z4 repeat array to 1-10 units on a 4qA chromosome[16]. FSHD2 is most often caused by heterozygous variants in *structural maintenance of chromosomes flexible hinge domain containing 1* (*SMCHD1*) in combination with a smaller, but normal-sized permissive D4Z4 repeat array (8-20) on a 4qA chromosome[5; 17]. SMCHD1 is an atypical member of the SMC gene superfamily and originally identified as regulator of epigenetic silencing in an ENU mutagenesis screen in mice[18; 19]. SMCHD1 is normally binding to the D4Z4 repeat, thereby repressing DUX4 expression in somatic cells[17]. FSHD2 patients with an *SMCHD1* variant show reduced binding of SMCHD1 to the D4Z4 repeat, which leads to D4Z4 chromatin relaxation and *DUX4* expression in skeletal muscle[17]. Additionally, SMCHD1 is also a modifier of disease severity and progression in FSHD1 since *SMCHD1* variants have been identified in some unusually severely affected members of FSHD1 families who carry both a contracted repeat array and an *SMCHD1* variant[20]. For some FSHD2 patients, however, we found D4Z4 hypomethylation but could not identify an (exonic) *SMCHD1* variant. Two of these families are now explained by *SMCHD1* hemizygosity[21]. In two other unexplained FSHD2 families heterozygous variants in *DNA* *methyltransferase 3B* (*DNMT3B*) have recently been identified[22]. DNMT3B is one of the *de novo* DNA methyltransferases[23], and is likely important for establishing a repressed D4Z4 chromatin structure in somatic cells.

Since the discovery of *SMCHD1* as the most common FSHD2 gene, disease causing variants in *SMCHD1* have been identified in approximately 80 FSHD2 families. The mutation spectrum of *SMCHD1* in FSHD2 includes missense, nonsense, and splice site variants, insertions and deletions[5; 17; 20; 21; 24-28]. In this study we describe two families with an intronic variant in *SMCHD1* which results in aberrant SMCHD1 transcripts. In the first family an intronic variant in *SMCHD1* was identified which alters splicing and results in partial intron retention. In the second family a deep intronic variant in *SMCHD1* was identified, resulting in exonisation of 53 nucleotides of intron 34. These variants further expand the *SMCHD1* mutation spectrum in FSHD2.

MATERIAL METHODS

Subjects

A French family (Rf744, fig. 1A) and an American family (Rf1034, fig. 1B) were studied after informed consent and the study protocol was approved by the relevant institutional review boards. Clinical assessment of disease severity was performed using the 11 point (0: unaffected – 10: wheelchair bound) standardized Clinical Severity Score (CSS)[29].

D4Z4 repeat sizing, haplotype analysis and methylation analysis

For genotyping high quality genomic DNA was isolated from peripheral blood mononuclear cells (PBMCs). The sizing of the D4Z4 repeats on chromosomes 4 and 10 was done by pulsed field gel electrophoresis (PFGE) as described previously[15]. Haplotype analysis was done by hybridization of PFGE blots with probes specific for the 4qA and 4qB haplotype in combination with PCR-based SSLP analysis according to previously described protocols[15]. D4Z4 methylation analysis was measured using the FseI restriction site in the most proximal unit of the D4Z4 arrays on chromosomes 4 and 10 as published previously[17]. The Delta1 value of D4Z4 methylation was calculated as described in Lemmers et al. 2014[5].

Genomic SMCHD1 variant analysis

For the index cases *SMCHD1* variant analysis of all coding exons and splice regions was performed by Sanger sequencing after PCR amplification. The intronic primers were located at a position of at least 50 nucleotides from the splice donor or acceptor site and were previously published [5]. For Rf1034 a PCR was performed in intron 34 to identify a deep intronic variant using primers intron_34fwd (5′-TTGAAATACAAAACTGTCGCTTAGA-3′) and intron_34rev (5′-AGGGGGAAGGAATTCAAAGA-3′). The PCR product was analysed by Sanger sequencing.

The *SMCHD1* genomic sequence was obtained from Ensembl human assembly GRCh37 [GRCh37:18:2655286:2805615] (Genomic Refseq: NG_031972.1, Transcript Refseq: NM_015295.2), exons were numbered like in NG_031972.1. The functional consequences of *SMCHD1* variants were predicted using Alamut Visual version 2.6 (Interactive Biosoftware, Rouen, France).

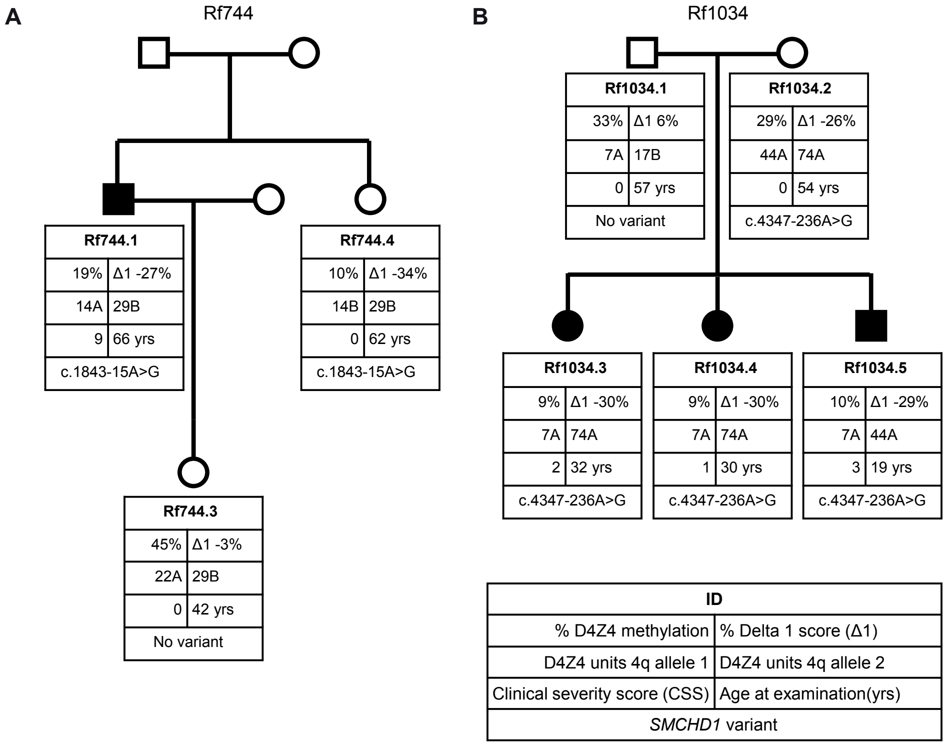


FIGURE 1. Pedigrees of families Rf744 (A) and Rf1034 (B). Clinically affected individuals are indicated in black. The following information is provided: the family identifier, D4Z4 methylation, Delta1 score, the size and type (A permissive, B non-permissive) of 4q-linked D4Z4 repeats, the clinical severity score, the age of examination and the *SMCHD1* variant. Key is shown below.

RNA analysis

RNA was isolated from PAXgene Blood RNA Tubes using the PAXgene Blood RNA Kit (PreAnalytiX, a Qiagen/BD company). cDNA was synthesized with 800 ng to 2000 ng of RNA using the RevertAid First Strand cDNA Synthesis Kit (Thermo Fisher Scientific) using random hexamer primers. A reverse transcriptase polymerase chain reaction (RT-PCR) for *SMCHD1* exon 12 to 16 was performed using primers 1482F (5′-TCCTAAGAAGAGAGGGCTTGC-3′) and 2105R (5′-TCATCTCCTTCAGGCCAAGT-3′). A RT-PCR for *SMCHD1* exon 32 to 35 was performed using primers 4098F (5′-AAAACCCGTTCGTCTCAATG-3′) and 4406R (5′-TCCATCATAAAACCAAACTGGA-3′). RT-PCRs were performed in 30 µl reactions using 0.5 units DreamTaq DNA polymerase (5U/µl Thermo Fisher Scientific), 1x DreamTaq buffer (Thermo Fisher Scientific), 3 µl of dNTPs (2mM of each nucleotide) and 25 pmol of each primer. The following RT-PCR protocol was used: 95°C 5 min, 35 cycles of: 95°C 30 sec, 60°C 30 sec, 72°C 30 sec, then 72°C 10 min. RT-PCR products were separated by size on 2% agarose gels after which PCR products were gel purified (NucleoSpin® Gel and PCR Clean-up, Machery Nagel). Purified PCR products were cloned into a pCR™4-TOPO vector (Invitrogen, Life Technologies) and transformed in DH5α heat-shock competent cells (Subcloning Efficiency DH5α Competent Cells, Invitrogen, Life Technologies). Multiple clones were analysed by sequencing their insert to find the sequence of the altered transcript.

RESULTS

D4Z4 length and methylation analysis in Rf744 and Rf1034 individuals

Index case Rf744.1 was suspected of FSHD based on physical examination with a CSS of 9 at age 66. His physical examination showed asymmetric scapular winging, right foot drop, asymmetric distribution of facial weakness, symmetric weakness of fixator shoulder girdle muscles, weakness of the pelvic girdle muscles, humeral weakness involving both biceps and triceps brachii, abdominal weakness with positive Beevor's sign and tibialis anterior weakness. Rf744.1 also has a benign myelodysplastic syndrome. D4Z4 repeat length and haplotype analysis showed that the shortest permissive D4Z4 allele of Rf744.1 consists of 14 units (Fig. 1A). D4Z4 methylation was measured and the Delta1 value was calculated. The Delta 1 value ranges between -42% and -22% (5th and 95th percentile, respectively) in carriers of an *SMCHD1* variant which affects functions[5]. D4Z4 methylation analysis in Rf744.1 revealed a FseI methylation level of 19% (Delta1 value -27%), indicative of FSHD2. The unaffected sister of the proband (Rf744.4) also shows D4Z4 hypomethylation but she does not carry a permissive allele. The daughter of the proband (Rf744.3) does not show D4Z4 hypomethylation and she is unaffected (Fig. 1A).

Index case Rf1034.5 was suspected of FSHD based on physical examination with a CSS of 3 at age 19 [29]. His physical examination showed a combination of pectus excavatum, progressive weakness of the right arm, bilateral scapular winging, facial weakness, and Beevor's sign. D4Z4 repeat length and haplotype analysis showed that Rf1034.5 carries a 7 units D4Z4 repeat on a permissive chromosome and D4Z4 hypomethylation (Delta1 score -29%), suggestive for both FSHD1 and FSHD2 (Fig. 1B). Additional family-member material was obtained and D4Z4 repeat sizes, haplotypes and D4Z4 methylation levels were determined and a physical examination was performed for the four additional family members of Rf1034[22] (Fig. 1B). The father (Rf1034.1) of the proband carries a 7 unit D4Z4 repeat array, but he is unaffected. The unaffected mother (Rf1034.2) of the proband shows D4Z4 hypomethylation and she carries two permissive 4qA alleles of 44 and 74 units. The two sisters (Rf1034.3 and Rf1034.4) of the proband both carry the 7 unit D4Z4 repeat array as well as D4Z4 hypomethylation and they are also affected. The physical examination of Rf1034.3 showed a combination of weakness of the scapular stabilizers and weakness of the right arm. The physical examination of Rf1034.4 showed only weakness of the facial muscles. This family information strengthened the suggestion that there is a combination of FSHD1 and FSHD2 in this family.

TABLE 1. Splice site predictions in *SMCHD1*

	3'splice site c.4347-236A>G	**5'splice site c.4347-183**	**3' splice site c.1843-15A>G**
SpliceSiteFinder-like (0-100)	87.4	94.7	89.9
MaxEntScan (0-16)	8.9	10.8	7.4
NNSPLICE (0-1)	0.9	1	1
GeneSplicer (0-15)	5.7	0.54	5.1
Human Splicing Finder (0-100)	89.4	97.7	86.1

Identification of an intronic variant in *SMCHD1* in Rf744

SMCHD1 variant analysis of all coding exons and splice regions in Rf744 identified an intronic *SMCHD1* variant in Rf744.1. This variant (c.1843-15A>G) is located 15 base pairs proximal to exon 14 and various splicing prediction tools predict that this variant creates a 3' splice site (Fig. 2A, Table 1). The variant was also identified in Rf744.4, which also shows D4Z4 hypomethylation, but not in Rf744.3 without D4Z4 hypomethylation (Fig.

S1A). To investigate whether this variant alters the transcript, an RT-PCR from *SMCHD1* exon 12 to 16 was performed and analysed by gel electrophoresis. Besides the normal PCR product of the expected size, two longer PCR products were identified (Fig. 2B). Sanger sequencing of TOPO clones of those two additional PCR products identified that the altered transcript contains the sequence from c.1843-14 to c.1843-1, confirming that c.1843-15A>G creates a 3' splice site (Fig. 2C, S1B). The inclusion of these 14 nucleotides is predicted to disrupt the open reading frame with a premature stop codon in exon 14. No other sequences were identified, suggesting that the highest band in the gel is a heteroduplex of the normal and altered transcript. No RNA was available from Rf744.3 and Rf744.4.

Identification of a deep intronic variant in *SMCHD1* in Rf1034

SMCHD1 variant analysis in all *SMCHD1* exons and splice regions in the proband did not identify any putative *SMCHD1* variants that affect function[5]. Therefore, whole exome sequencing (WES) was performed in Rf1034 but this did not identify a causative variant. By serendipity, an RT-PCR targeting *SMCHD1* exon 32 to 35 followed by agarose gel electrophoresis revealed two PCR products for Rf1034.3, the normal PCR product with the expected size and a PCR product that was larger than expected (Fig. 3A). This larger PCR product was also identified with an RT-PCR for Rf1034.2, Rf1034.4 (Fig. 3A) and Rf1034.5 (not shown), while it was absent in Rf1034.1 (Fig. 3A). This additional PCR product contained a sequence corresponding to 53 nucleotides of intron 34, from c.-235 to c.-183 proximal to exon 35 (Fig S2A). These 53 nucleotides are included in the transcript as a new exon and are predicted to disrupt the open reading frame and lead to a premature stop codon in exon 35 (Fig. S2A). Subsequently, an intronic PCR was performed, followed by Sanger sequencing, to identify the variant which is responsible for this new exon. A heterozygous deep intronic variant (c.4347-236A>G, g.2760414A>G) in *SMCHD1* was identified in individuals Rf1034.2, Rf1034.3, Rf1034.4, and Rf1034.5, which was absent in Rf1034.1 (Fig S2B). Various splicing prediction tools predict that this variant creates a 3' splice site, while a cryptic 5' splice site is already predicted in the reference sequence at position c.4347-183 (Table 1). In this family this deep intronic variant in *SMCHD1* segregates with D4Z4 hypomethylation and modifies disease severity.

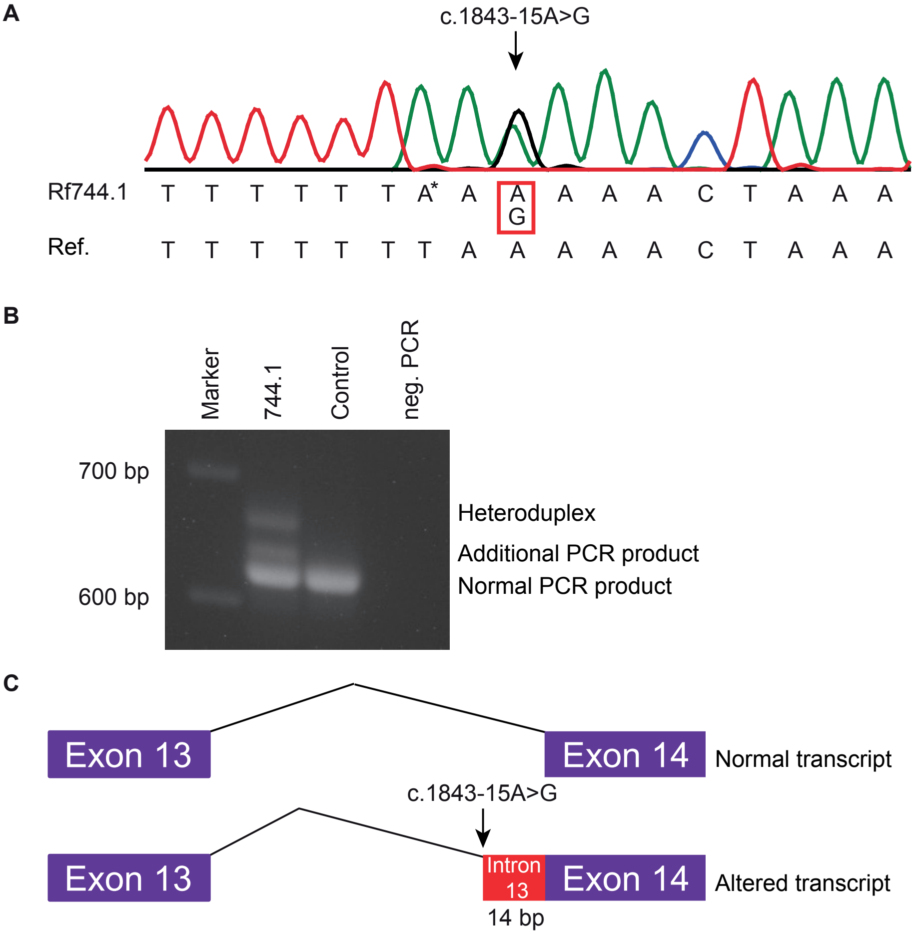


FIGURE 2. Identification of intronic variant in *SMCHD1* in Rf744. A) Sanger sequence track from Rf744.1 showing the intronic variant in *SMCHD1* at position c.1843-15, highlighted with a red rectangle. * indicates common SNP rs8090988 (T/A, ancestral T, minor allele frequency 0.33 (A)) B) Gel of RT-PCR of *SMCHD1* exon 12 to 16 in Rf744.1, a control and a negative PCR (no DNA). C) Schematic representation of splicing of the normal transcript and the altered transcript containing the intronic variant

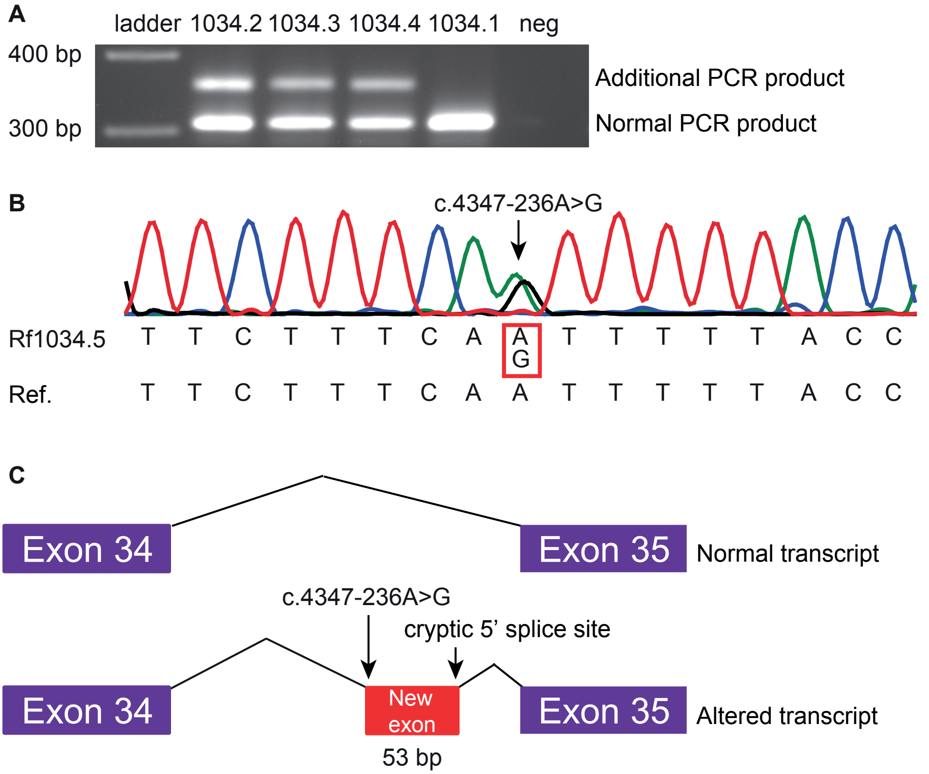


FIGURE 3. Identification of deep intronic variant in *SMCHD1* in Rf1034. A) Gel of RT-PCR of *SMCHD1* exon 32 to 35 in four members of family Rf1034 and a negative PCR (no DNA), B) Sanger sequence track showing the deep intronic variant in *SMCHD1* at position c.4347-236 in Rf1034.5, highlighted with a red rectangle C) Schematic representation of splicing of the normal transcript and the altered transcript containing the deep intronic variant and showing the exonisation of 53 basepairs in red.

DISCUSSION

In this study we identified a deep intronic variant in *SMCHD1* to act as a modifier for disease severity in an FSHD1 family. Furthermore, we identified an intronic variant in *SMCHD1* in an FSHD2 family.

In family Rf744 we identified an intronic variant located at 15 base pairs proximal to exon 14, which creates a 3' splice site. This *SMCHD1* variant results in the inclusion of the distal 14 nucleotides of intron 13 (c.1843-14 to c.1843-1) into the transcript, which is predicted to disrupt the open reading frame with a premature stop codon in exon 14.

The intronic variant and D4Z4 hypomethylation were also detected in the unaffected sister of the proband. She carries two non-permissive alleles, which explains why she remained unaffected. The unaffected daughter of the proband does not carry the variant and shows no D4Z4 hypomethylation.

In family Rf1034, exonic *SMCHD1* variant analysis by *SMCHD1* Sanger sequencing and WES did not identify any variants that affect function in *SMCHD1* or elsewhere in the genome[5]. However, in this study, by serendipity, a deep intronic variant was identified, which segregates with D4Z4 hypomethylation. This *SMCHD1* variant creates a 3' splice site in intron 34 and this results in exonisation of 53 nucleotides of intron 34. Inclusion of these 53 nucleotides in the transcript is predicted to disrupt the open reading frame and to result in a premature stop codon in exon 35. In family Rf1034 this *SMCHD1* variant acts as a modifier for disease severity. The proband and his two sisters all carry both a permissive D4Z4 repeat array of 7 units and the deep intronic variant in *SMCHD1* and present an FSHD phenotype. The proband is more severely affected than his sisters, indicating clinical variability, which is common in FSHD[5]. The mother (Rf1034.2) carries the deep intronic variant in *SMCHD1* and two permissive 4qA alleles of 44 and 74 units, while the median repeat size in controls is 23 units. The length of the D4Z4 repeats of the mother is much longer than the median length of the shortest permissive allele in FSHD2 patients, which is only 13 units[5]. Probably, the permissive allele of the mother is too long to develop FSHD2. This has also been shown in other FSHD2 families, where carriers of an *SMCHD1* variant are most often only affected with FSHD when they also carry a relatively short but normal sized permissive D4Z4 repeat of 11-20 units[5]. The father (Rf1034.1) carries an FSHD1 sized allele of 7 units and is unaffected. Non-penetrance and mild phenotypes are seen more often in carriers of a 7-10 units FSHD1 size allele[30]. Additionally, in 1-3% of the control population D4Z4 repeats of 7–10 units on disease permissive 4qA chromosomes are found, indicating the reduced penetrance of these alleles[31; 32]. In conclusion, in Rf1034 only the combination of a permissive D4Z4 repeat array of 7 units with the deep intronic variant in *SMCHD1* causes an FSHD phenotype, illustrating this *SMCHD1* variant modifies disease severity. This modifying role of *SMCHD1* variants has been described in multiple FSHD1 families with upper sized FSHD1 repeat arrays, which explains clinical variability in these families[20; 24].

The variants identified in this study affect splicing by introducing new 3' splice sites in *SMCHD1* outside the consensus sequence. Previously, an intronic *SMCHD1* variant with a similar effect as the variant in Rf744 was identified in another FSHD2 patient (Rf1352 in Lemmers et al. 2014)[5]. The variant c.3634-19A>G creates a 3' splice site, which results in the inclusion of the distal 18 nucleotides of intron 28 into the transcript and introduces a premature stop codon immediately proximal to exon 29[5]. In total, we have identified approximately 100 variants in *SMCHD1* which affect function (ref. 5,17,20,21,22 and unpublished results), including 3 intronic variants outside the consensus sequence that introduce a 3' splice site. This indicates that the frequency of intronic variants in *SMCHD1* that introduce a new splice site is approximately 3% and that this type of variants might explain FSHD in patients in which no variant was identified in the exonic SMCHD1 region or in the splice site consensus.

Therefore, it would be useful to perform whole genome sequencing in FSHD2 patients without exonic *SMCHD1* variants, in combination with RT-PCR to identify alternative splicing. However, the products of such variants might be masked by efficient nonsense mediate decay (NMD), which would make it difficult to study the splicing effect of intronic variants. One way to address this issue would be to culture cells from blood of FSHD patients and controls in the presence of cycloheximide to block NMD. Alternatively, intronic variants might influence expression levels of *SMCHD1*. Furthermore, variants in the promoter or regulatory regions of *SMCHD1* might cause FSHD2. The functional consequences of these types of variants will be difficult to predict since the information on regulatory regions of *SMCHD1* is limited. Recently, in two FSHD families a variant in a putative regulatory region of *SMCHD1* was identified, however segregation with D4Z4 hypomethylation was inconclusive[33]. The functional effects of variants in regulatory regions could be studied with reporter assays in combination with segregation analysis of D4Z4 hypomethylation.

In summary, this report expands the *SMCHD1* mutation spectrum in FSHD2 with two intronic variants in *SMCHD1*. Both variants lead to aberrant splicing and the altered SMCHD1 transcripts are predicted to lead to a premature stop codon. Our study also highlights the importance of the additional variant screening in FSHD2 patients negative for exonic *SMCHD1* variants.

REFERENCES

1. Deenen, J.C., Arnts, H., van der Maarel, S.M., Padberg, G.W., Verschuuren, J.J., Bakker, E., Weinreich, S.S., Verbeek, A.L., and van Engelen, B.G. (2014). Population-based incidence and prevalence of facioscapulohumeral dystrophy. Neurology 83, 1056-1059.
2. Mul, K., Lassche, S., Voermans, N.C., Padberg, G.W., Horlings, C.G., and van Engelen, B.G. (2016). What's in a name? The clinical features of facioscapulohumeral muscular dystrophy. Pract Neurol 16, 201-207.
3. Statland, J.M., and Tawil, R. (2011). Facioscapulohumeral muscular dystrophy: molecular pathological advances and future directions. Current opinion in neurology 24, 423-428.
4. de Greef, J.C., Lemmers, R.J., Camano, P., Day, J.W., Sacconi, S., Dunand, M., van Engelen, B.G., Kiuru-Enari, S., Padberg, G.W., Rosa, A.L., et al. (2010). Clinical features of facioscapulohumeral muscular dystrophy 2. Neurology 75, 1548-1554.
5. Lemmers, R.J., Goeman, J.J., van der Vliet, P.J., van Nieuwenhuizen, M.P., Balog, J., Vos-Versteeg, M., Camano, P., Ramos Arroyo, M.A., Jerico, I., Rogers, M.T., et al. (2014). Inter-individual differences in CpG methylation at D4Z4 correlate with clinical variability in FSHD1 and FSHD2. Human molecular genetics.
6. Balog, J., Thijssen, P.E., de Greef, J.C., Shah, B., van Engelen, B.G., Yokomori, K., Tapscott, S.J., Tawil, R., and van der Maarel, S.M. (2012). Correlation analysis of clinical parameters with epigenetic modifications in the DUX4 promoter in FSHD. Epigenetics : official journal of the DNA Methylation Society 7, 579-584.
7. van Overveld, P.G., Lemmers, R.J., Sandkuijl, L.A., Enthoven, L., Winokur, S.T., Bakels, F., Padberg, G.W., van Ommen, G.J., Frants, R.R., and van der Maarel, S.M. (2003). Hypomethylation of D4Z4 in 4q-linked and non-4q-linked facioscapulohumeral muscular dystrophy. Nature genetics 35, 315-317.
8. 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 genetics 5, e1000559.
9. 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 genetics 6, e1001181.
10. Hendrickson, P.G., Dorais, J.A., Grow, E.J., Whiddon, J.L., Lim, J.W., Wike, C.L., Weaver, B.D., Pflueger, C., Emery, B.R., Wilcox, A.L., et al. (2017). Conserved roles of mouse DUX and human DUX4 in activating cleavage-stage genes and MERVL/HERVL retrotransposons. Nature genetics.
11. Whiddon, J.L., Langford, A.T., Wong, C.J., Zhong, J.W., and Tapscott, S.J. (2017). Conservation and innovation in the DUX4-family gene network. Nature genetics.

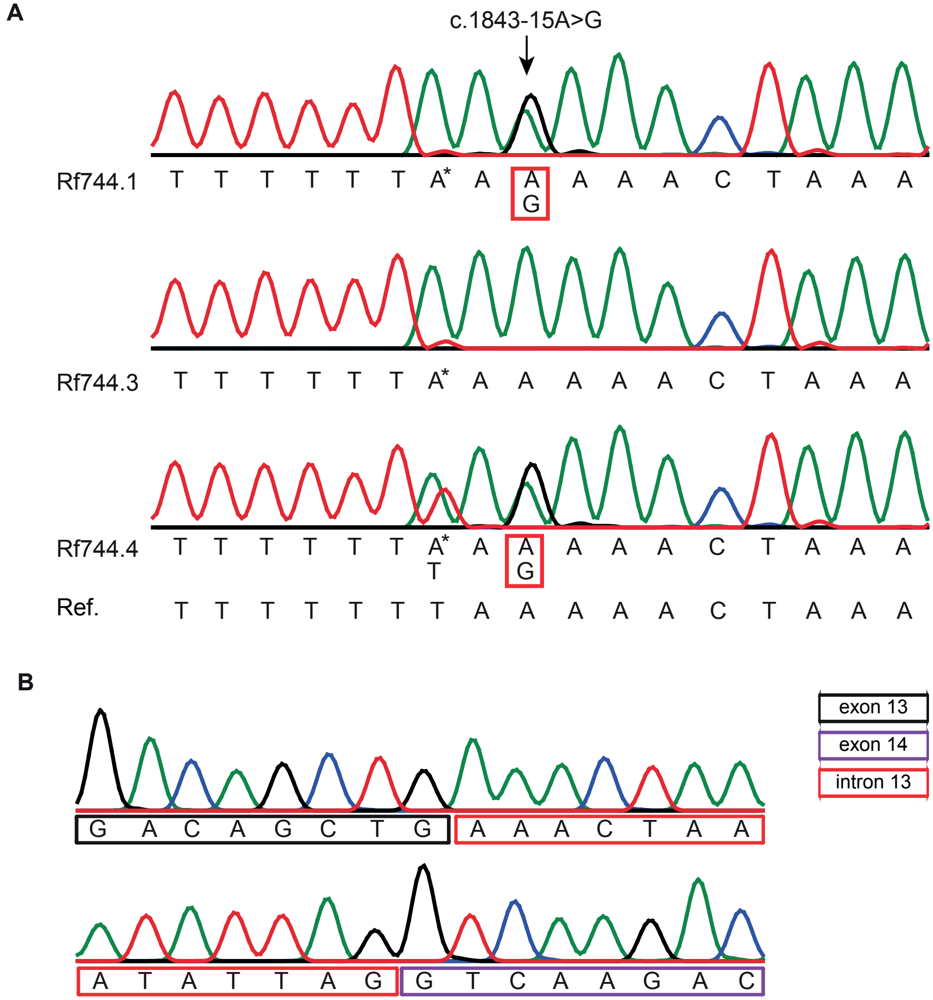
12. De Iaco, A., Planet, E., Coluccio, A., Verp, S., Duc, J., and Trono, D. (2017). DUX-family transcription factors regulate zygotic genome activation in placental mammals. Nature genetics.
13. Kowaljow, V., Marcowycz, A., Ansseau, E., Conde, C.B., Sauvage, S., Matteotti, C., Arias, C., Corona, E.D., Nunez, N.G., Leo, O., et al. (2007). The DUX4 gene at the FSHD1A locus encodes a pro-apoptotic protein. Neuromuscular disorders : NMD 17, 611-623.
14. Rickard, A.M., Petek, L.M., and Miller, D.G. (2015). Endogenous DUX4 expression in FSHD myotubes is sufficient to cause cell death and disrupts RNA splicing and cell migration pathways. Human molecular genetics 24, 5901-5914.
15. 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.
16. Wijmenga, C., Hewitt, J.E., Sandkuijl, L.A., Clark, L.N., Wright, T.J., Dauwerse, H.G., Gruter, A.M., Hofker, M.H., Moerer, P., Williamson, R., et al. (1992). Chromosome 4q DNA rearrangements associated with facioscapulohumeral muscular dystrophy. Nature genetics 2, 26-30.
17. Lemmers, R.J., Tawil, R., Petek, L.M., Balog, J., Block, G.J., Santen, G.W., Amell, A.M., van der Vliet, P.J., Almomani, R., Straasheijm, K.R., et al. (2012). Digenic inheritance of an SMCHD1 mutation and an FSHD-permissive D4Z4 allele causes facioscapulohumeral muscular dystrophy type 2. Nature genetics 44, 1370-1374.
18. Hirano, T. (2005). SMC proteins and chromosome mechanics: from bacteria to humans. Philosophical transactions of the Royal Society of London Series B, Biological sciences 360, 507-514.
19. Blewitt, M.E., Vickaryous, N.K., Hemley, S.J., Ashe, A., Bruxner, T.J., Preis, J.I., Arkell, R., and Whitelaw, E. (2005). An N-ethyl-N-nitrosourea screen for genes involved in variegation in the mouse. Proceedings of the National Academy of Sciences of the United States of America 102, 7629-7634.
20. Sacconi, S., Lemmers, R.J., Balog, J., van der Vliet, P.J., Lahaut, P., van Nieuwenhuizen, M.P., Straasheijm, K.R., Debipersad, R.D., Vos-Versteeg, M., Salviati, L., et al. (2013). The FSHD2 gene SMCHD1 is a modifier of disease severity in families affected by FSHD1. American journal of human genetics 93, 744-751.
21. Lemmers, R.J., van den Boogaard, M.L., van der Vliet, P.J., Donlin-Smith, C.M., Nations, S.P., Ruivenkamp, C.A., Heard, P., Bakker, B., Tapscott, S., Cody, J.D., et al. (2015). Hemizygosity for SMCHD1 in Facioscapulohumeral Muscular Dystrophy Type 2: Consequences for 18p Deletion Syndrome. Hum Mutat 36, 679-683.
22. van den Boogaard, M.L., Lemmers, R.J., Balog, J., Wohlgemuth, M., Auranen, M., Mitsuhashi, S., van der Vliet, P.J., Straasheijm, K.R., van den Akker, R.F., Kriek, M., et al. (2016). Mutations in DNMT3B Modify Epigenetic Repression of the D4Z4 Repeat and the Penetrance of Facioscapulohumeral Dystrophy. American journal of human genetics 98, 1020-1029.

23. Jeltsch, A., and Jurkowska, R.Z. (2016). Allosteric control of mammalian DNA methyltransferases - a new regulatory paradigm. Nucleic acids research 44, 8556-8575.
24. Larsen, M., Rost, S., El Hajj, N., Ferbert, A., Deschauer, M., Walter, M.C., Schoser, B., Tacik, P., Kress, W., and Muller, C.R. (2014). Diagnostic approach for FSHD revisited: SMCHD1 mutations cause FSHD2 and act as modifiers of disease severity in FSHD1. European journal of human genetics : EJHG.
25. Mitsuhashi, S., Boyden, S.E., Estrella, E.A., Jones, T.I., Rahimov, F., Yu, T.W., Darras, B.T., Amato, A.A., Folkerth, R.D., Jones, P.L., et al. (2013). Exome sequencing identifies a novel SMCHD1 mutation in facioscapulohumeral muscular dystrophy 2. Neuromuscular disorders : NMD 23, 975-980.
26. Winston, J., Duerden, L., Mort, M., Frayling, I.M., Rogers, M.T., and Upadhyaya, M. (2014). Identification of two novel SMCHD1 sequence variants in families with FSHD-like muscular dystrophy. European journal of human genetics : EJHG.
27. van den Boogaard, M.L., Jfl Lemmers, R., Camano, P., van der Vliet, P.J., Voermans, N., van Engelen, B.G., Lopez de Munain, A., Tapscott, S.J., van der Stoep, N., Tawil, R., et al. (2016). Double SMCHD1 variants in FSHD2: the synergistic effect of two SMCHD1 variants on D4Z4 hypomethylation and disease penetrance in FSHD2. European journal of human genetics : EJHG 24, 78-85.
28. Hamanaka, K., Goto, K., Arai, M., Nagao, K., Obuse, C., Noguchi, S., Hayashi, Y.K., Mitsuhashi, S., and Nishino, I. (2016). Clinical, muscle pathological, and genetic features of Japanese facioscapulohumeral muscular dystrophy 2 (FSHD2) patients with SMCHD1 mutations. Neuromuscular disorders : NMD 26, 300-308.
29. Ricci, E., Galluzzi, G., Deidda, G., Cacurri, S., Colantoni, L., Merico, B., Piazzo, N., Servidei, S., Vigneti, E., Pasceri, V. (1999). Progress in the molecular diagnosis of facioscapulohumeral dystrophy and correlation between the number of KpnI repeat at the 4q35 locus and clinical phenotype. Ann Neurol 45, 751-757.
30. Statland, J.M., Donlin-Smith, C.M., Tapscott, S.J., Lemmers, R.J., van der Maarel, S.M., and Tawil, R. (2015). Milder phenotype in facioscapulohumeral dystrophy with 7-10 residual D4Z4 repeats. Neurology 85, 2147-2150.
31. Lemmers, R.J., Wohlgemuth, M., van der Gaag, K.J., van der Vliet, P.J., van Teijlingen, C.M., de Knijff, P., Padberg, G.W., Frants, R.R., and van der Maarel, S.M. (2007). Specific sequence variations within the 4q35 region are associated with facioscapulohumeral muscular dystrophy. American journal of human genetics 81, 884-894.
32. Scionti, I., Fabbri, G., Fiorillo, C., Ricci, G., Greco, F., D'Amico, R., Termanini, A., Vercelli, L., Tomelleri, G., Cao, M., et al. (2012). Facioscapulohumeral muscular dystrophy: new insights from compound heterozygotes and implication for prenatal genetic counselling. Journal of medical genetics 49, 171-178.

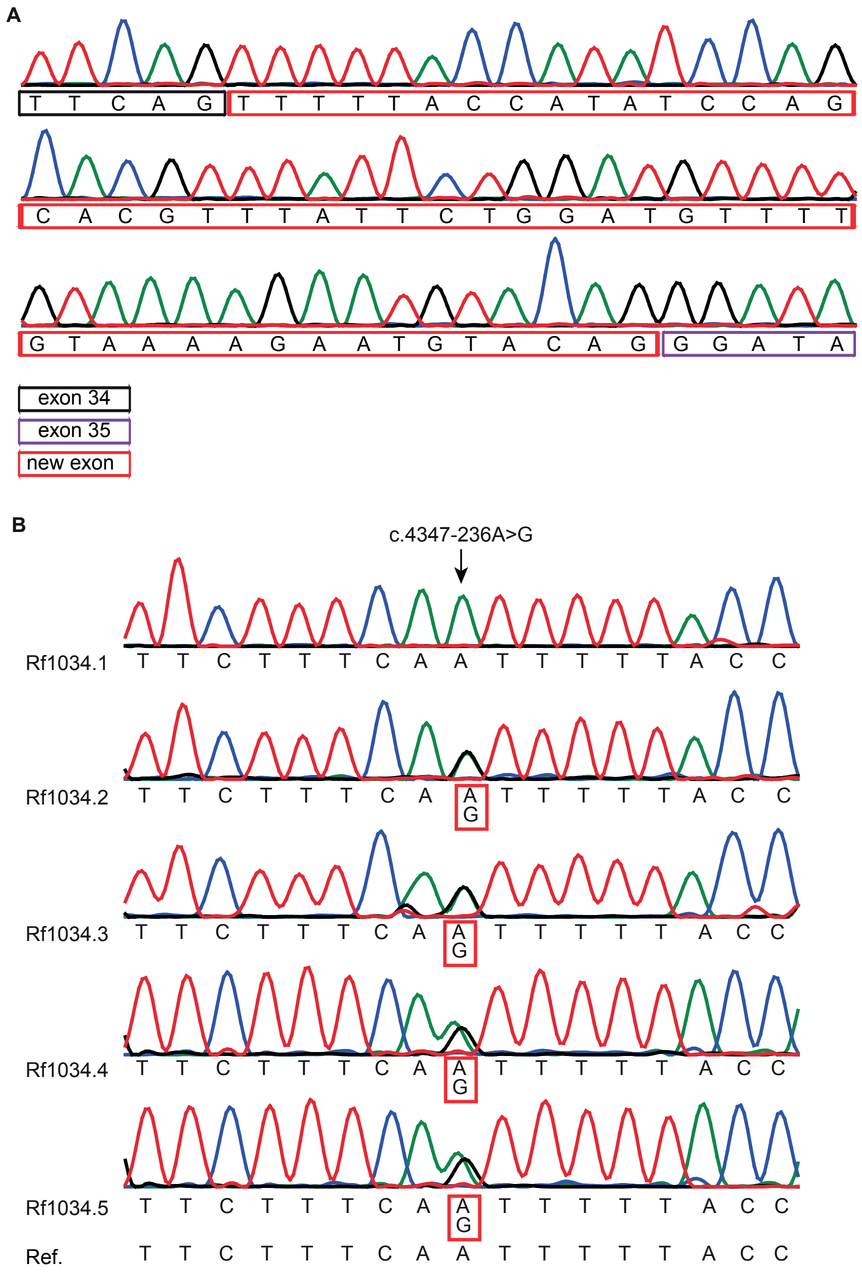
33. Mayes, M.B., Morgan, T., Winston, J., Buxton, D.S., Kamat, M.A., Smith, D., Williams, M., Martin, R.L., Kleinjan, D.A., Cooper, D.N., et al. (2015). Remotely acting SMCHD1 gene regulatory elements: in silico prediction and identification of potential regulatory variants in patients with FSHD. Hum Genomics 9, 25.

SUPPLEMENTARY INFORMATION

Supplementary figures



SUPPLEMENTARY FIGURE 1. Intronic variant in *SMCHD1* in Rf744. A) Sanger sequence track from Rf744.1, Rf744.3 and Rf744.4 showing the intronic variant in *SMCHD1* at position c.1843-15 in Rf744.1 and Rf744.4, highlighted with a red rectangle. * indicates common SNP rs8090988 (T/A, ancestral T, minor allele frequency 0.33 (A)). B) Sanger sequence track of the altered *SMCHD1* transcript in Rf744.1 shows the inclusion of the last 14 nucleotides of intron 13 between exon 13 and exon 14.



SUPPLEMENTARY FIGURE 2. Deep intronic variant in *SMCHD1* in Rf1034. A) Sanger sequence track of the altered *SMCHD1* transcript in Rf1034 shows exonisation of 53 nucleotides of intron 34 between exon 34 and exon 35. B) Sanger sequence track from family members of Rf1034, showing the deep intronic variant in *SMCHD1* at position c.4347-236 in Rf1034.2, Rf1034.3, Rf1034.3 and Rf1034.5, highlighted with a red rectangle.