Modifying the modifier: discovering mechanisms of SMCHD1 mediated chromatin repression

Goossens, R.

Remko Goossens[1*], Marlinde L van den Boogaard[1*], Richard JLF Lemmers[1], Judit Balog[1], Patrick J van der Vliet[1], Iris M Willemsen[1], Julie Schouten[2], Ignazio Maggio[3,4], Nienke van der Stoep[5], Rob C Hoeben[3], Stephen J Tapscott[6], Niels Geijsen[2], Manuel AFV Gonçalves[3], Sabrina Sacconi[7,8], Rabi Tawil[9] & Silvère M van der Maarel[1]

1 Department of Human Genetics, Leiden University Medical Center, Leiden, The Netherlands

2 KNAW-Hubrecht Institute, Utrecht, the Netherlands

3 Department of Cell and Chemical Biology, Leiden University Medical Center, Leiden, The Netherlands

4 Department of Pediatrics, Leiden University Medical Center, Leiden, The Netherlands

5 Department of Clinical Genetics, Leiden University Medical Center, Leiden, The Netherlands

6 Human Biology Division, Fred Hutchinson Cancer Research Center, Seattle, USA

7 Neuromuscular Diseases Centre, University Hospital of Nice, Nice, France

8 CNRS UMR7277, IBV, Faculty of Medicine, University of Nice, Nice, France

9 Department of Neurology, University of Rochester Medical Center, Rochester, USA

*: Authors RG and MLvdB contributed equally to this work

CHAPTER 3

Intronic *SMCHD1* variants in FSHD: testing the potential for CRISPR-Cas9 genome editing

Journal of Medical Genetics 2019; 56(12): 828-837

Abstract

Background
Facioscapulohumeral dystrophy (FSHD) is associated with partial chromatin relaxation of the *DUX4* retrogene containing D4Z4 macrosatellite repeats on chromosome 4, and transcriptional derepression of *DUX4* in skeletal muscle. The common form of FSHD, FSHD1, is caused by a D4Z4 repeat array contraction. The less common form FSHD2 is generally caused by heterozygous variants in *SMCHD1*.

Methods
We employed whole exome sequencing combined with Sanger sequencing to screen uncharacterized FSHD2 patients for extra-exonic *SMCHD1* mutations. We also used CRISPR-Cas9 genome editing to repair pathogenic intronic *SMCHD1* variants from patient myoblasts.

Results
We identified intronic *SMCHD1* variants in two FSHD families. In the first family an intronic variant resulted in partial intron retention and inclusion of the distal 14 nucleotides of intron 13 into the transcript. In the second family a deep intronic variant in intron 34 resulted in exonisation of 53 nucleotides of intron 34. In both families the aberrant transcripts are predicted to be non-functional. Deleting the pseudo-exon by CRISPR-Cas9 mediated genome editing in primary and immortalized myoblasts from the index case of the second family restored wild-type SMCHD1 expression to a level that resulted in efficient suppression of *DUX4*.

Conclusions
The estimated intronic mutation frequency of almost 2% in FSHD2, as exemplified by the two novel intronic *SMCHD1* variants identified here, emphasizes the importance of screening for intronic variants in *SMCHD1*. Furthermore, the efficient suppression of *DUX4* after restoring SMCHD1 levels by genome editing of the mutant allele provides further guidance for therapeutic strategies.

Introduction

Facioscapulohumeral dystrophy (FSHD, [FSHD1; OMIM 158900 and FSHD2; 158901]) is a common muscular dystrophy (prevalence ~1:8.500) mainly characterized by progressive weakness and wasting of the facial, shoulder girdle, trunk and upper arm muscles.(1, 2) With a disease onset typically in the second decade of life, there is a large variability in onset and progression.(3) Two genetic forms have been identified, FSHD1 and FSHD2, which are clinically indistinguishable,(4) and seem to represent a disease continuum.(5, 6) Both forms are associated with partial chromatin relaxation of the D4Z4 macrosatellite repeat on chromosome 4 in somatic tissue, characterized by reduced CpG methylation and loss of repressive histone marks, as well as changes in other chromatin factors that result in a more relaxed chromatin structure.(7-10) This chromatin relaxation results in transcriptional derepression of the D4Z4-encoded *DUX4* [MIM 606009] retrogene in skeletal muscle. (11) The DUX4 transcription factor is normally expressed in the germ line and in cleavage stage embryos, while being suppressed in most somatic tissues.(11-14) DUX4 causes cell death when overexpressed in somatic cell lines or when endogenously expressed in FSHD myotubes.(15, 16) D4Z4 chromatin relaxation must occur on a so-called permissive chromosome 4 (4qA haplotype), which contains a polymorphic *DUX4* polyadenylation signal (PAS) distal to the D4Z4 repeat, to cause FSHD.(17, 18) This PAS is required for the production of stable *DUX4* mRNA in somatic cells. Consequently, chromatin relaxation of the homologous D4Z4 repeats on non-permissive 4qB or 10q chromosomes does not cause FSHD since these chromosomal backgrounds lack a somatic *DUX4* PAS.(17, 19)

FSHD1, accounting for >95% of cases, is caused by contraction of the D4Z4 repeat to 1-10 units on a 4qA chromosome.(20, 21) FSHD2 is most often caused by heterozygous variants in *structural maintenance of chromosomes flexible hinge domain containing 1* (*SMCHD1*) [MIM 614982] in combination with a D4Z4 repeat of 8-20 units on a 4qA chromosome. (5, 6, 22) *SMCHD1* is an atypical member of the SMC gene superfamily and was originally identified as a regulator of epigenetic silencing.(23, 24) SMCHD1 binds to the D4Z4 repeat, thereby repressing DUX4 in somatic cells by yet largely unknown mechanisms.(22) FSHD2 patients with a pathogenic *SMCHD1* variant show reduced SMCHD1 binding to the D4Z4 repeat, resulting in D4Z4 chromatin relaxation and *DUX4* (mis)expression in skeletal muscle. (22) SMCHD1 is also a disease modifier for FSHD1 since pathogenic *SMCHD1* variants have been identified in some severely affected members of FSHD1 families.(25) The *SMCHD1* mutation spectrum in FSHD2 patients includes locus-wide missense, nonsense, and splice site variants, as well as insertions and deletions.(5, 22, 25-31) Some FSHD2 patients with D4Z4 hypomethylation cannot be explained by (exonic) *SMCHD1* variants. In some of these patients D4Z4 hypomethylation can be explained by *SMCHD1* hemizygosity,(30) or by heterozygous variants in *DNA methyltransferase 3B* (*DNMT3B*)[MIM 602900].(32) Intriguingly, missense variants in the ATPase domain of SMCHD1 can also cause Bosma arhinia microphthalmia syndrome (BAMS), an unrelated severe developmental disorder.(33, 34) There is currently no comprehensive explanation for this discordant clinical outcome of missense variants in the ATPase domain of SMCHD1, although recent biochemical and modelling studies have pointed towards differences in the mutation spectrum and ATPase activity between the two conditions.(35-37)

Currently, >180 FSHD causing *SMCHD1* variants have been described.(37) In this study we describe two independent intronic *SMCHD1* variants which result in aberrant *SMCHD1* transcripts. One variant alters splicing by partial intron retention. The other deep intronic

variant leads to exonisation of 53 nucleotides. We designed a genome editing strategy to delete this deep intronic variant with the objective to restore wild-type SMCHD1 expression and DUX4 repression in myoblasts from a patient carrying this variant.

Material and Methods

All material and methods are available as supplementary data.

Results

Clinical and genetic characterization of Rf744 and Rf1034 individuals

Index case Rf744.1 was suspected of FSHD based on physical examination with a Clinical Severity Score (CSS)(38) of 9 at age 66. 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 size and haplotype analysis showed that the shortest permissive D4Z4 allele of Rf744.1 contains 14 units (Figure 1A). D4Z4 methylation analysis in Rf744.1 revealed a FseI methylation level of 19% (Delta1 value -27%), which is well within the FSHD2 range.(5) The unaffected sister of the proband (Rf744.4) also shows D4Z4 hypomethylation but she does not have a permissive allele. The daughter of the proband (Rf744.3) does not show D4Z4 hypomethylation and is unaffected (Figure 1A).

Index case Rf1034.5 was suspected of FSHD based on physical examination with a CSS of 3 at age 19.(38) 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 analysis showed that Rf1034.5 has a 7 unit D4Z4 repeat on a permissive chromosome and D4Z4 hypomethylation (FseI: 10%, Delta1 score: -29%), suggestive for both FSHD1 and FSHD2 (Figure 1B). The father (Rf1034.1) of the proband carries a disease permissive 7-unit D4Z4 repeat (Figure 1B), and has pectus excavatum, which is frequently observed in FSHD.(39) He does not have muscle weakness. The unaffected mother (Rf1034.2) of the proband shows D4Z4 hypomethylation and she has two permissive 4qA alleles of 44 and 74 units. The two sisters (Rf1034.3 and Rf1034.4) both have the 7-unit D4Z4 repeat and D4Z4 hypomethylation, and they are also affected. Physical examination of Rf1034.3 showed a combination of weakness of the scapular stabilizers and weakness of the right arm. Physical examination of Rf1034.4 showed weakness of the facial muscles. This family information strengthened the suggestion that there is a combination of FSHD1 and FSHD2 in this family.

Identification of an intronic variant in *SMCHD1* in Rf744

SMCHD1 variant analysis of coding exons and splice regions identified an intronic

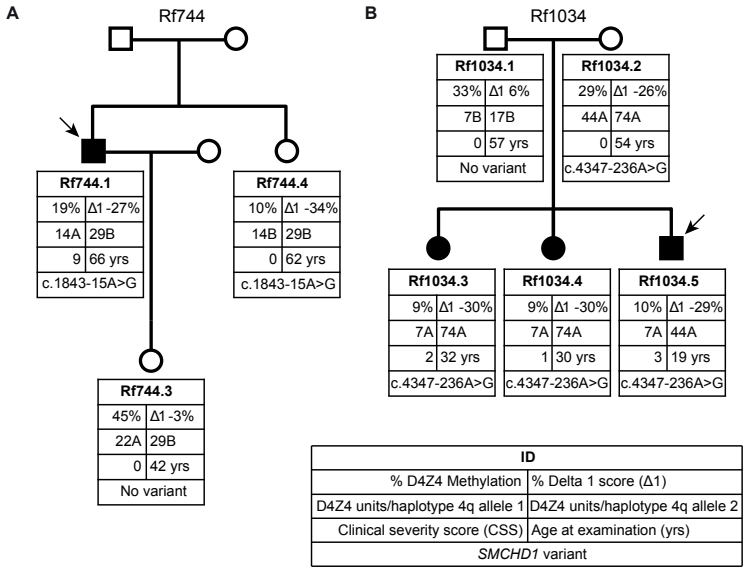


Figure 1: Pedigrees of families Rf744 (A) and Rf1034 (B). Clinically affected individuals are indicated in black, the index cases are marked by an arrow. The following information is provided: the family identifier, D4Z4 methylation, Delta1 score, the size and type (A permissive, B non-permissive) of 4q-linked D4Z4 repeats, the clinical severity score, the age at examination and the *SMCHD1* variant. Key is shown on the bottom right. N/A, not available.

SMCHD1 variant in peripheral blood-derived RNA from patient Rf744.1. This variant (NG_031972.1(SMCHD1):c.1843-15A>G, g.2705677A>G) is located 15 base pairs proximal to exon 14 and various splicing prediction tools suggest that this variant creates a 3' splice site and has not been reported in public variant databases (Figure 2A, online supplementary table S1).

The variant was also identified in Rf744.4, who also shows D4Z4 hypomethylation, but not in Rf744.3, who does not present D4Z4 hypomethylation (online supplementary figure S1A). To investigate whether this variant leads to an altered transcript, an RT-PCR targeting *SMCHD1* exon 12 through exon 14 was performed. Besides the normal PCR product of expected size, a longer PCR product was also detected (Figure 2B). Sanger sequencing of individual clones derived from PCR products of the target region shows that they contained the altered transcript sequence from c.1843-14 to c.1843-1 confirming that c.1843-15A>G creates a 3' splice site (Figure 2C and online supplementary figures S1B). The inclusion of these 14 nucleotides is predicted to disrupt the open reading frame and to result in a premature stop codon in exon 14. Sanger sequencing also confirmed the wild-type transcript sequence in some clones, consistent with the heterozygous expression. No RNA was available from Rf744.3 and Rf744.4. No further material was available from index case Rf744.1 for additional functional testing of the *SMCHD1* variant, but previous studies support the possibility for the development of FSHD from *SMCHD1* haploinsufficiency.(6, 40) This is further supported by the negative Delta1 methylation scores observed exclusively in carriers of the *SMCHD1* variant (Rf744.1 and Rf744.4, figure 1A), which is typical for reduced *SMCHD1* activity at D4Z4.(5)

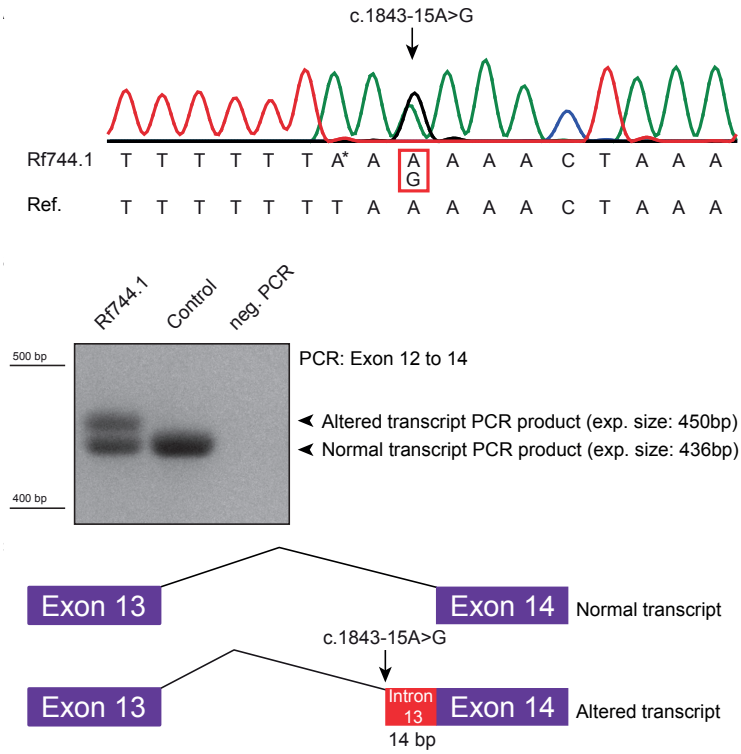


Figure 2: Identification of an intronic SMCHD1 variant in Rf744. A) Sanger sequence track from Rf744.1 showing the intronic variant in *SMCHD1* at position c.1843-15, highlighted with a rectangle. * indicates common SNP rs8090988 (T/A, ancestral T, minor allele frequency 0.33 (A)). B) RT-PCR analysis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

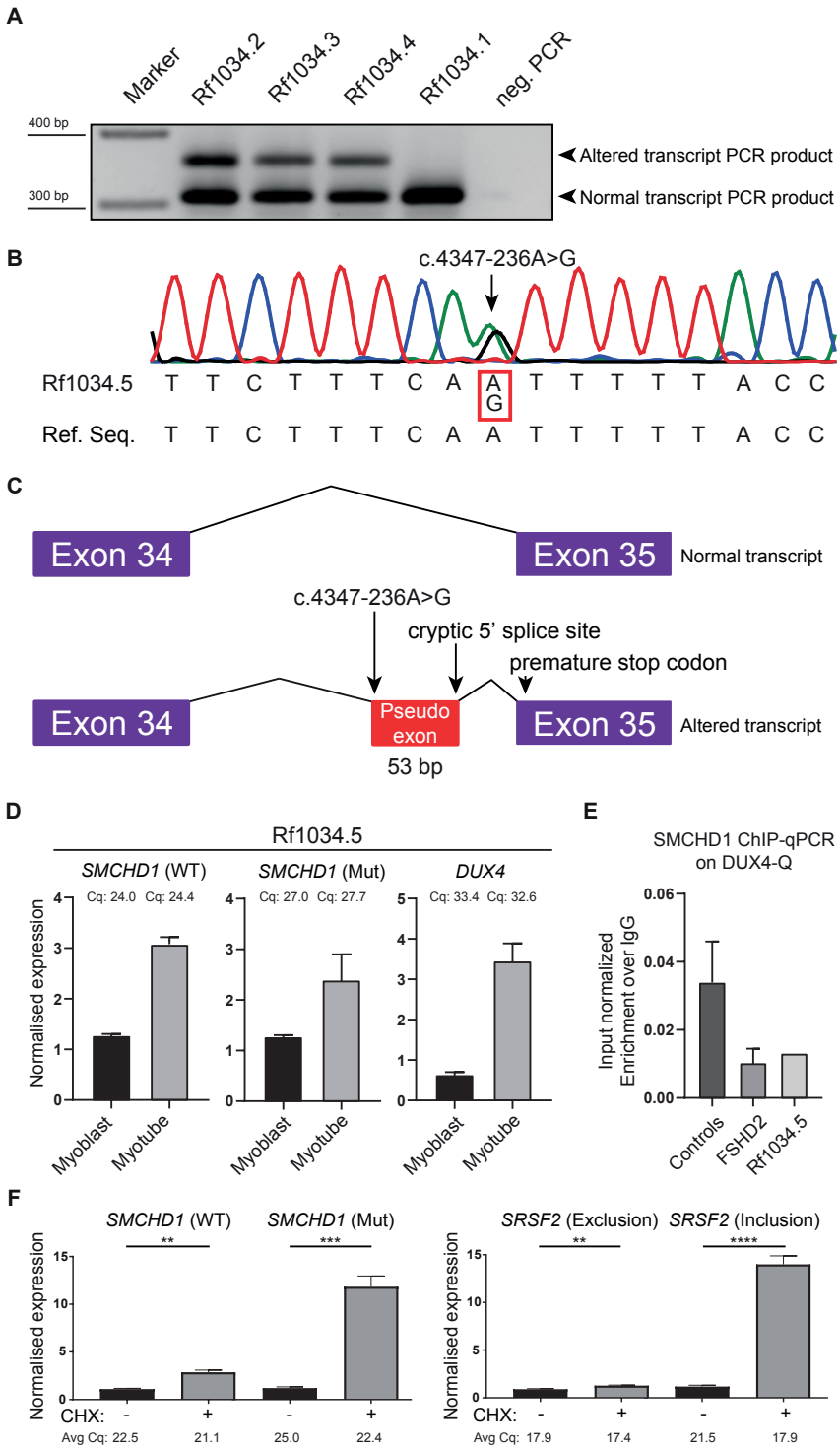
Identification of a deep intronic variant in SMCHD1 in Rf1034

SMCHD1 variant analysis of all *SMCHD1* exons and splice regions or whole exome sequencing (WES) in the proband did not identify any putative pathogenic *SMCHD1* variant or pathogenic variants elsewhere in the genome.(5) *SMCHD1* transcript analysis using partially overlapping amplicons identified a fragment of increased size suggestive for aberrant splicing. An RT-PCR targeting *SMCHD1* exon 32 through exon 35 revealed two PCR products for Rf1034.3, i.e., a product of expected size and a longer PCR product (Figure 3A). This larger PCR product was also identified with an RT-PCR performed on RNA isolated from blood of Rf1034.2 and Rf1034.4 (Figure 3A), and RNA from myoblasts of Rf1034.5 (not shown), while it was absent in Rf1034.1 (Figure 3A). Sequencing of the larger PCR product revealed the presence of a sequence corresponding to 53 nucleotides of intron 34, from c.-235 to c.-183 proximal to exon 35 (online supplementary figure S2A). These 53 nucleotides are included in the transcript as a pseudo-exon and are predicted to disrupt the open reading frame and to lead to a premature stop codon in exon 35 (online supplementary figure S2A). To identify the variant responsible for this pseudo-exon, we used an intronic PCR followed by Sanger sequencing. A heterozygous deep intronic variant (NG_031972.1(SMCHD1):c.4347-236A>G, g.2760414A>G) in *SMCHD1*, not reported in public databases, was identified in subjects Rf1034.2-5, which was absent in Rf1034.1 (Figure 3B, online supplementary figure S2B). Splicing prediction tools suggest 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 (online supplementary table S1, Figure 3C). In this family the deep intronic *SMCHD1* variant segregates with D4Z4 hypomethylation. We further characterized RNA from primary muscle cell cultures from patient Rf1034.5 using RT-qPCR for *DUX4* and for the wild-type and mutant forms of *SMCHD1* (Figure 3D). The inclusion of the pseudo-exon in the mutant mRNA allowed us to use primers targeting this exon for specific amplification of the mutant transcript. This analysis supported the diagnosis of FSHD by the apparent expression of *DUX4,* while also showing that the mutant form of *SMCHD1* is readily detectible on mRNA level (Figure 3D). The observed increased expression of *SMCHD1* mRNA upon myogenic differentiation is consistent with a previous study.(41) Whether mutant *SMCHD1* mRNA is stable and leads to translation of a truncated SMCHD1 protein is unknown, but ChIP-qPCR analysis of SMCHD1 occupancy on D4Z4 in Rf1034.5 myoblasts compared to controls and unrelated FSHD2 myoblast samples suggests that the inclusion of this pseudo-exon creates SMCHD1 haplo-insufficiency with consequent partial decompaction of the D4Z4 chromatin structure in myonuclei (Figure 3E). To determine whether the mutant transcript is a target for nonsense mediated decay (NMD), we inhibited NMD using cycloheximide (CHX) in Rf1034.5 myotube cultures (Figure 3F). RT-qPCR analysis after CHX treatment showed a modest (~2-fold) increase in *SMCHD1* wildtype mRNA, but a ~10-fold increase in mutant transcript. This response of the mutant *SMCHD1* transcript to inhibition of NMD is similar to the known endogenously produced NMD-sensitive isoform of *SRSF2* (Figure 3F, *SRSF2* intron inclusion)(42). This indicates that the *SMCHD1* transcript retaining the pseudo-exon is indeed degraded by NMD, leading to SMCHD1 haplo-insufficiency. We compared expression of total *SMCHD1* RNA in Rf1034.5 myotubes to other FSHD and control myotube cultures (Online supplementary figure S2C). Due to endogenous variability in *SMCHD1* transcript levels, it is not possible to distinguish between control and FSHD2 samples. Only hemizygous expression of *SMCHD1* significantly alters total *SMCHD1* RNA levels, as reported previously.(40) Attempts to detect a truncated SMCHD1 protein by western blotting using an N-terminal targeting antibody (HPA039441) did not yield any detectable specific signal of lower molecular weight when compared to unrelated samples (data not shown), consistent with a haploinsufficiency situation and NMD-mediated degradation of the transcript for Rf1034.5.

Genome editing designed to remove the pathologic intronic variant in Rf1034

In an attempt to suppress *DUX4* in primary muscle cell cultures from individual Rf1034.5 by restoring the wild-type *SMCHD1* open reading frame at the expense of the mutant version, we aimed to delete the *SMCHD1* pseudo-exon from the genome. We performed CRISPR-Cas9-mediated genome editing using two gRNA constructs targeting sequences upstream and downstream of the pseudo-exon. To test whether deletion of the intronic target region would not impair SMCHD1 protein expression from a wild-type allele, we transfected HeLa cells with plasmids encoding the two gRNAs (U/D3) and sp.Cas9-2A-GFP. Monoclonal cultures of GFP-positive HeLa cells were genotyped to screen for clones harbouring a homozygous or heterozygous deletion of the targeted region (online supplementary figure S3, bottom panel). We analysed SMCHD1 protein levels in all clones and compared these with those observed in monoclonal cultures transfected with a plasmid encoding the control gRNA X50 and sp.Cas9-2A-GFP. Although some variation in protein levels can be observed, none of the edited clones showed loss of SMCHD1, even when homozygously edited (online


A
Marker
Rf1034.2
Rf1034.3
Rf1034.4
Rf1034.1
neg. PCR
400 bp
300 bp
Altered transcript PCR product
Normal transcript PCR product
B
c.4347-236A>G
Rf1034.5 T T C T T T C A A T T T T T A C C
G
Ref. Seq. T T C T T T C A A T T T T T A C C
C
Exon 34
Exon 35
Normal transcript
c.4347-236A>G
cryptic 5' splice site
premature stop codon
Exon 34
Pseudo exon
Exon 35
Altered transcript
53 bp
D
Rf1034.5
SMCHD1 (WT)
SMCHD1 (Mut)
DUX4
Cq: 24.0 Cq: 24.4
Cq: 27.0 Cq: 27.7
Cq: 33.4 Cq: 32.6
Normalised expression
Myoblast
Myotube
E
SMCHD1 ChIP-qPCR on DUX4-Q
Input normalized Enrichment over IgG
Controls
FSHD2
Rf1034.5
F
SMCHD1 (WT)
SMCHD1 (Mut)
SRSF2 (Exclusion)
SRSF2 (Inclusion)
Normalised expression
CHX:
Avg Cq: 22.5 21.1 25.0 22.4
Avg Cq: 17.9 17.4 21.5 17.9

supplementary figure S3, top panels). This indicates that deletion of the intronic region corresponding to the location of the pseudo-exon in Rf1034 by means of CRISPR-Cas9 genome editing does not impair SMCHD1 expression in HeLa cells. Thus, deleting the same region in Rf1034 seems a feasible approach to restore SMCHD1 protein levels.

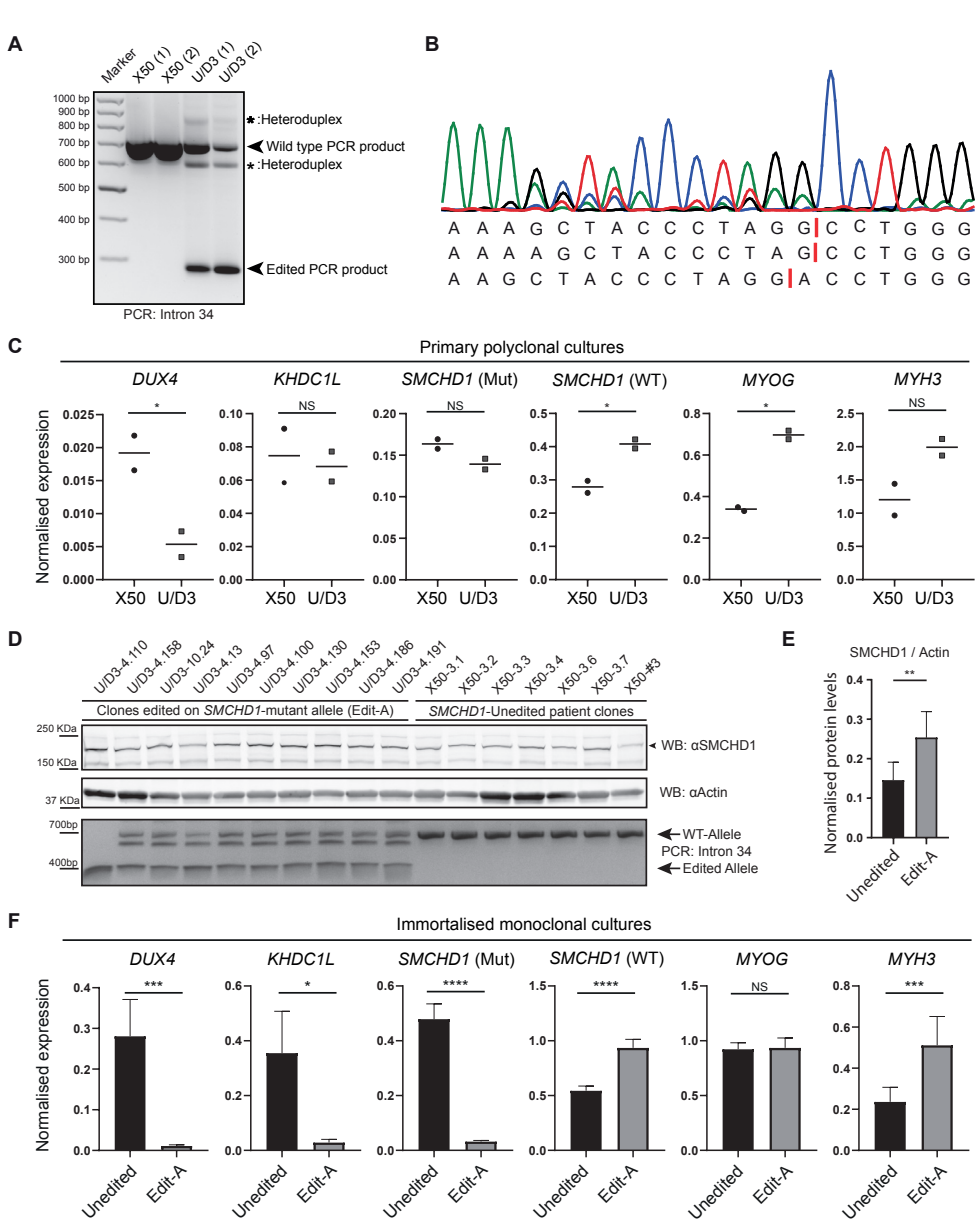
For genome editing of primary myoblast cultures of Rf1034.5, we first employed an approach in which gRNAs were first delivered by lentiviral transduction, and Cas9 protein was subsequently delivered by iTOP. To achieve enrichment of targeted myoblasts, a gRNA targeting *B2M* was co-introduced to allow for FACS sorting of MHC class I-negative cells. In a biological replicate experiment, PCR analysis identified genomic deletions in Rf1034.5 myoblasts treated with gRNAs U/D3, but not with control gRNA X50 (Figure 4A). Sanger sequencing of the smaller PCR product of the edited genomic DNA showed that there is a deletion of 407 +/-1 bp confirming that the deep intronic variant is absent in this product (Figure 4B). Additional RT-PCR analysis indicated that there are no extra products after treatment with Cas9 and gRNAs U/D3 besides the wild-type and mutant products (online supplementary figure S4A-B). The aforementioned myoblasts were allowed to differentiate to myotubes, and subsequently *SMCHD1* expression levels from the wild-type and mutant alleles were determined by using RT-qPCR. A significantly higher expression of the wild-type *SMCHD1* transcript (p value: 0.0286, Mann-Whitney test) was observed in the myotube samples treated with the gRNAs flanking the pseudo-exon, although we could not detect a significant change in mutant *SMCHD1* transcript expression (p value: 0.8857, Mann-Whitney test). The increase of wild-type *SMCHD1* transcript was concomitant with reduced expression of *DUX4* but not with the DUX4 target genes *KHDC1L* (Figure 4C) and *ZSCAN4* (online supplementary figure S4E) (p values: 0.0286, 0.8857, 0.1143, respectively (Mann-Whitney test)) levels suggesting that the gain of *SMCHD1* restores D4Z4 chromatin repression (Figure 4C). Expression levels of *MYOG* (p value: 0.0286, Mann-Whitney test) but not *MYH3* (p value: 0.2000, Mann-Whitney test) were significantly higher in Cas9-treated cells expressing gRNAs U/D3 when compared to those expressing control gRNA X50 (Figure 4C). This outcome suggests that the myogenic differentiation process is not strongly impaired due to genomic editing of *SMCHD1*.

Figure 3: Identification of deep intronic SMCHD1 variant in Rf1034. A) RT-PCR analysis of *SMCHD1* transcripts spanning exon 32 through 35 in four members of family Rf1034. A negative control PCR (no DNA) was performed in parallel. B) Sanger sequence track showing the deep intronic variant in *SMCHD1* at position c.4347-236 in Rf1034.5, highlighted with a rectangle C) Schematic representation of splicing of the normal transcript and the altered transcript containing the deep intronic variant. The altered transcript shows exonisation of the 53 base pair pseudo-exon. D) RT-qPCR analysis of primary Rf1034.5 myoblast and myotube samples. Expression of *SMCHD1* (wild-type and mutant) and *DUX4* are shown, normalized to *GUSB* expression. Error bars indicate SEM. E) ChIP-qPCR analysis of SMCHD1 occupancy on the DUX4-Q region in D4Z4 on chromosome 4 in myoblasts of three control individuals, three unrelated FSHD2 patients, and Rf1034.5. Input normalized enrichment is shown, subtracted for IgG values of the corresponding sample, error bars indicate SD. F) Inhibition of NMD by cycloheximide treatment in Rf1034.5 myotubes. RT-qPCR analysis showing that CHX treatment results in a ~10-fold increase of *SMCHD1* mutant transcript, with a smaller (~2-fold) increase in WT transcript. This CHX mediated increase is similar to a known NMD target, an isoform of *SRSF2* including an intron (*SRSF2* Inclusion). This increase of transcript levels is not seen for an *SRSF2* transcript excluding this intron. Expression was normalized to *RPL13.* CHX-: N=3, CHX+: N=4 (**: P value <0.01, ***: P value <0.001, ****: P value <0.0001 – Student's t-test)

While editing of primary Rf1034.5 myoblasts successfully increased *SMCHD1* expression levels and decreased expression of *DUX4,* we could not rule out that incomplete editing of *SMCHD1* in the polyclonal cultures obscured a more robust phenotypical change. Therefore, we immortalized Rf1034.5 myoblasts (Rf1034.5-iMB) to allow generation of monoclonal cultures after genome editing. After iTOP mediated editing and expansion of U/D3 transduced cells, we confirmed genomic editing on the mutated allele, and thus restoration of the *SMCHD1* open reading frame, in 10 independent myocyte clones, of which one clone (U/D3-4.110) was edited on both alleles (Figure 4D, lower panel). For comparison, we isolated material from seven clones transduced with the control gRNA X50 (RNA, DNA and protein analysis shown), as well as 18 unedited U/D3 transduced clones (RNA analysis shown) (collectively called *SMCHD1*-unedited patient clones (N=25)), which still have the pseudo-exon and contain no genomic aberrations at the U/D3 gRNA target sites, as confirmed by PCR and Sanger sequencing (data not shown). Additional RT-PCR analysis of X50 and U/D3 treated samples showed no mis-splicing of the *SMCHD1* mRNA (online supplementary figure S4C-D). SMCHD1 western blot analysis of a representative set of edited (N=10) and X50 *SMCHD1*-unedited myotube clones (N=7) (Figure 4D, upper panel) and subsequent quantification of SMCHD1 protein levels normalized to the housekeeping protein Actin showed a significant increase (p value: 0.0020, Mann-Whitney test) in cellular SMCHD1 protein after genome editing (Figure 4E). Protein quantification data corresponding to the monoclonal myotube cultures were in agreement with the RT-qPCR data for the wild-type and variant *SMCHD1* mRNA forms, in that they were significantly increased (p value: <0.0001, Mann-Whitney test) and decreased (p value: <0.0001, Mann-Whitney test), respectively (Figure 4F). Again, we observed a significant decrease in *DUX4* expression (p value: 0.0001, Mann-Whitney test). However, now, the decrease in *DUX4* expression also correlated with decreased levels in the amounts of the DUX4 targets *KHDC1L* (p value: 0.0342, Mann-Whitney test), and *ZSCAN4* (p value: 0.0039, Mann-Whitney test) (Figure 4F and online supplementary figure S4F). This data indicates a more robust phenotypic rescue in edited myotube clones when compared to the primary polyclonal myotube cultures (Figure 4C). Expression of myogenic differentiation marker *MYH3* (p value: 0.0008, Mann-Whitney test), but not *MYOG* (p value: 0.7877) was significantly higher in gRNA U/D3 edited cultures, which in combination with

➔ **Figure 4: Genomic deletion of the pseudo-exon in SMCHD1 in myocytes of Rf1034 by CRISPR-Cas9-based editing.** A) Gel electrophoresis of genomic PCR on primary Rf1034.5 myoblasts treated with Cas9 and control gRNA X50 (targeted against *AAVS1*) or treated with Cas9 and gRNAs U/D3, which cleave upstream and downstream of the pseudo-exon, respectively. Gel electrophoresis image showing the wild-type PCR product, a PCR product with a genomic deletion in intron 34 (edited PCR product), and a heteroduplex formed by hybridization of the wild-type and edited PCR product. Samples from a biological replicate experiment are shown. B) Sanger sequencing track of the PCR product with the deletion in intron 34 (including the deep intronic variant) after genomic editing. The track shows that position chr18:2,760,070 is mainly repaired to chr18:2,760,478 (upper line). The other repaired products include position chr18:2,760,069 to chr18:2,760,478 (middle line) and position chr18:2,760,070 to chr18:2,760,477 (lower line). The vertical lines indicate the breakpoints. C) Expression of wild-type *SMCHD1*, mutant *SMCHD1* (i.e. pseudo-exon containing), *MYOG, MYH3, DUX4* and DUX4 target gene *KHDC1L* by RT-qPCR in Rf1034.5 myotubes treated with Cas9 and gRNA X50 or Cas9 and gRNAs U/D3. Two biological replicates are shown as independent data points, each containing two technical replicate cultures. Expression was normalized to *RPL13* and *GUSB* expression, horizontal bars indicate mean log transformed expression. D) Western blot for SMCHD1 and Actin (top panels) and genomic analysis of intron 34 (bottom panel) of monoclonal Rf1034.5-iMB cultures, edited in intron 34 with the U/D3 gRNA combination (Edit-A) or a representative set of *SMCHD1*-unedited patient clones (X50). E) Quantification of SMCHD1 levels of the Western blot data presented in panel 4D, normalized to Actin. Error bars: SD. F) RT-qPCR analysis of monoclonal *SMCHD1*-edited (N=10) or *SMCHD1*-unedited (N=25) Rf1034.5-iMB myotube clones. Error bars: SEM. (*: P value <0.05, **: P value <0.01, ***: P value <0.001, ****: P value <0.0001, NS: Not-Significant – Mann-Whitney Test)

typical morphological changes (i.e. formation of aligned multinucleated myotubes) observed in differentiating myotube cultures (representative examples online supplementary figure S4G) suggests that restoration of *SMCHD1* by genomic editing to physiological levels does not negatively influence myogenic differentiation *in vitro.* Previously published work has shown that *DUX4* expression in FSHD patient-derived myogenic cultures correlates with high expression of myogenic markers such as *MYH3.*(41) The negative correlation we



observe here (i.e. higher *MYH3* with lower *DUX4* expression in *SMCHD1*-edited compared to *SMCHD1*-unedited patient cells), strengthens our conclusion that restoration of SMCHD1 levels supresses expression of *DUX4.*

Loss of SMCHD1 in FSHD2 leads to decreased methylation levels of the DR1 region in D4Z4.(43) We therefore analysed a set of four edited U/D3 clones and four *SMCHD1*-unedited patient clones by bisulphite Sanger sequencing, and found no evidence that increased SMCHD1 levels lead to increased CpG methylation at DR1 in these clones (online supplementary figure S5).

Discussion

In this study, we identified two intronic *SMCHD1* variants: one likely acting as a modifier of disease severity in an FSHD1 family and another likely acting as a disease-causing variant in an FSHD2 family.

In family Rf744, an intronic variant located 15 base pairs proximal to exon 14 creates a 3' splice site. This variant results in the inclusion of the distal 14 nucleotides of intron 13 into the transcript, which is predicted to disrupt the open reading frame to result in the presence of a premature stop codon in exon 14. The intronic variant and the D4Z4 hypomethylation status 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, no disease-causing variants were identified by exonic *SMCHD1* variant analysis by Sanger sequencing or elsewhere in the exome using WES.(5) However, in this study, a deep intronic variant was identified, which segregates with D4Z4 hypomethylation. This *SMCHD1* variant creates a 3' splice site in intron 34 resulting 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 of disease severity. The proband and his two sisters all carry a permissive D4Z4 repeat of 7 units and the deep intronic variant in *SMCHD1*, and are affected. 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 D4Z4 repeat size on chromosome 4 in controls is 23 units. The length of the D4Z4 repeats in the mother is much longer than the median length of the shortest permissive allele in FSHD2 patients, which is only 13 units.(44) Probably, the permissive alleles of the mother contain too many repeat units to become severely de-repressed by SMCHD1 loss associated with this *SMCHD1* variant, explaining her FSHD2-free status. This has also been shown in other FSHD2 families, in which *SMCHD1* variant carriers are typically only affected when they also carry a permissive D4Z4 repeat of 11-20 units.(5) The father carries an FSHD1-sized allele of 7 units and is unaffected. However, pectus excavatum was noticed in him, a condition often observed in FSHD.(39) Non-penetrance and mild phenotypes are often seen in carriers FSHD1-sized alleles of 7-10 units.(45) Indeed, in 1-3% of the control population D4Z4 repeats of 7–10 units on disease permissive chromosomes are found, indicating the reduced penetrance of these alleles.(46, 47) Thus, in Rf1034 it is likely that only the combination of a permissive D4Z4 repeat of 7 units with the deep

intronic variant in *SMCHD1* causes FSHD. This modifying role of *SMCHD1* variants has been described in multiple FSHD1 families with upper-sized FSHD1 D4Z4 repeats, which provides an explanation for the clinical variability observed in these families.(6, 25, 26)

In order to restore the *SMCHD1* open reading frame in primary myoblasts of Rf1034, we aimed to remove the splice site created by the deep intronic variant by CRISPR-Cas9-mediated genome editing. The simultaneous treatment with these gRNAs and Cas9 was expected to create a genomic deletion in intron 34 of 407 (+/- 1) base pairs on both mutant and wild-type alleles. Since the deletion would be intronic, it was expected to disrupt inclusion of the pseudo-exon without affecting wild-type splicing. This predicted lack of consequences for the wild-type allele was supported by the experiments in HeLa cells in which we did not observe a loss of SMCHD1 protein after deletion of the target sequence in intron 34.

We performed genome editing experiments using plasmid transfection, lentiviral vector gRNA transduction combined with iTOP delivery of Cas9, or complete delivery of the gRNA-Cas9 complex by iTOP. The iTOP transduction was previously shown to be an efficient strategy to deliver proteins to a variety of primary cell types.(48) In this study, we show that this strategy can also be applied in primary and immortalized myoblasts.

Expression analysis showed a significant increase in levels of wild-type *SMCHD1* transcript upon CRISPR-Cas9-mediated pseudo-exon excision. In turn, this allowed us to detect a consistent reduction of *DUX4* expression in the edited primary and immortalized myogenic cells. Several factors might have affected wild-type *SMCHD1* levels including the efficiency of the genome editing procedure at wild-type and variant alleles and dynamic changes in *SMCHD1* and *DUX4* expression during muscle cell differentiation.(41) In Rf1034, the mutation is likely to cause *SMCHD1* haploinsufficiency, as supported by the sensitivity of the mutant transcript to NMD, and the reduced binding of SMCHD1 to D4Z4 observed in ChIP-qPCR experiments. Our results suggest that restoring wild-type *SMCHD1* expression levels to near-normal, bi-allelic levels is sufficient to effectively repress *DUX4*. Previously we have shown that mild SMCHD1 overexpression by lentiviral transduction can also repress *DUX4* in FSHD primary muscle cell cultures.(41) Combined with this study, our data suggest that therapeutic strategies aiming at SMCHD1 upregulation to normal or close-to-normal levels in FSHD2 skeletal muscle cells results in efficient suppression of *DUX4*. This suppression of *DUX4* is not dependent on D4Z4 methylation as we did not observe an increase in D4Z4 methylation in edited clones. This seems consistent with a recent report showing that SMCHD1 is important for *de novo* methylation at the pluripotent stage, but dispensable for methylation maintenance in somatic cells.(49)

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.(5) In total, we have identified >180 variants in *SMCHD1*, which affect gene function.(5, 22, 25, 29, 30, 37) This includes 3 intronic variants outside the splice consensus sequence that introduce a 3' splice site, of which two cases are further described in the current work, and Rf1352 was published previously.(22) This indicates that intronic variants in *SMCHD1* that introduce a new splice site are present in approximately 2% of the FSHD2 patient population. This type of variants might explain FSHD phenotypes in patients in whom no variant has yet been identified in the exonic regions of the *SMCHD1* gene or its splice site consensus sequences.

Since *SMCHD1* is expressed in blood, RNA-seq or targeted SMCHD1 RNA analysis approaches might be considered to identify these intronic variants, although care must be taken to avoid false-negative results due to potential NMD of the mutant transcript.

In summary, this report expands the *SMCHD1* mutation spectrum in FSHD2 by characterizing two additional intronic variants in *SMCHD1*. Both variants lead to aberrant splicing with the altered *SMCHD1* transcripts leading to frameshifts generating premature stop codons. Our study also highlights the importance of, whenever warranted, performing additional variant screening in FSHD2 patients that are negative for exonic *SMCHD1* variants.

Acknowledgements

We thank all families for participating in our studies. This study was supported by grants from the US National Institutes of Health (NIH) (National Institute of Neurological Disorders and Stroke (NINDS) P01NS069539, and National Institute of Arthritis and Musculoskeletal and Skin Diseases (NIAMS) R01AR045203 and R01AR066248, the Prinses Beatrix Spierfonds (W.OP14-01; W.OR11–18, W.OR14-04), the European Union Framework Programme 7 (agreement 2012-305121, NEUROMICS) and Spieren voor Spieren. The authors are members of the European Reference Network for Rare Neuromuscular Diseases [ERN EURO-NMD]

Conflict of interest statement

N. Geijsen is co-founder of NTrans Technologies, a company developing gene editing therapies to treat monogenetic disease.

Funding

This study was supported by grants from the US National Institutes of Health (NIH) (National Institute of Neurological Disorders and Stroke (NINDS) P01NS069539, and National Institute of Arthritis and Musculoskeletal and Skin Diseases (NIAMS) R01AR045203 and R01AR066248, the Prinses Beatrix Spierfonds (W.OP14-01; W.OR11–18), the European Union Framework Programme 7 (agreement 2012-305121, NEUROMICS) and Spieren voor Spieren.

References

1. Deenen JC, Arnts H, van der Maarel SM, Padberg GW, Verschuuren JJ, Bakker E, Weinreich SS, Verbeek AL, van Engelen BG. Population-based incidence and prevalence of facioscapulohumeral dystrophy. Neurology. 2014;83(12):1056-9.

2. Mul K, Lassche S, Voermans NC, Padberg GW, Horlings CG, van Engelen BG. What's in a name? The clinical features of facioscapulohumeral muscular dystrophy. Practical neurology. 2016;16(3):201-7.

3. Statland JM, Tawil R. Facioscapulohumeral muscular dystrophy: molecular pathological advances and future directions. Current opinion in neurology. 2011;24(5):423-8.

4. de Greef JC, Lemmers RJ, Camano P, Day JW, Sacconi S, Dunand M, van Engelen BG, Kiuru-Enari S, Padberg GW, Rosa AL, Desnuelle C, Spuler S, Tarnopolsky M, Venance SL, Frants RR, van der Maarel SM, Tawil R. Clinical features of facioscapulohumeral muscular dystrophy 2. Neurology. 2010;75(17):1548-54.

5. Lemmers RJ, Goeman JJ, van der Vliet PJ, van Nieuwenhuizen MP, Balog J, Vos-Versteeg M, Camano P, Ramos Arroyo MA, Jerico I, Rogers MT, Miller DG, Upadhyaya M, Verschuuren JJ, Lopez de Munain Arregui A, van Engelen BG, Padberg GW, Sacconi S, Tawil R, Tapscott SJ, Bakker B, van der Maarel SM. Inter-individual differences in CpG methylation at D4Z4 correlate with clinical variability in FSHD1 and FSHD2. Human molecular genetics. 2014.

6. Sacconi S, Briand-Suleau A, Gros M, Baudoin C, Lemmers R, Rondeau S, Lagha N, Nigumann P, Cambieri C, Puma A, Chapon F, Stojkovic T, Vial C, Bouhour F, Cao M, Pegoraro E, Petiot P, Behin A, Marc B, Eymard B, Echaniz-Laguna A, Laforet P, Salviati L, Jeanpierre M, Cristofari G, van der Maarel SM. FSHD1 and FSHD2 form a disease continuum. Neurology. 2019.

7. Balog J, Thijssen PE, de Greef JC, Shah B, van Engelen BG, Yokomori K, Tapscott SJ, Tawil R, van der Maarel SM. Correlation analysis of clinical parameters with epigenetic modifications in the DUX4 promoter in FSHD. Epigenetics : official journal of the DNA Methylation Society. 2012;7(6):579-84.

8. van Overveld PG, Lemmers RJ, Sandkuijl LA, Enthoven L, Winokur ST, Bakels F, Padberg GW, van Ommen GJ, Frants RR, van der Maarel SM. Hypomethylation of D4Z4 in 4q-linked and non-4q-linked facioscapulohumeral muscular dystrophy. Nat Genet. 2003;35(4):315-7.

9. Zeng W, de Greef JC, Chen YY, Chien R, Kong X, Gregson HC, Winokur ST, Pyle A, Robertson KD, Schmiesing JA, Kimonis VE, Balog J, Frants RR, Ball AR, Jr., Lock LF, Donovan PJ, van der Maarel SM, Yokomori K. Specific loss of histone H3 lysine 9 trimethylation and HP1gamma/cohesin binding at D4Z4 repeats is associated with facioscapulohumeral dystrophy (FSHD). PLoS genetics. 2009;5(7):e1000559.

10. Haynes P, Bomsztyk K, Miller DG. Sporadic DUX4 expression in FSHD myocytes is associated with incomplete repression by the PRC2 complex and gain of H3K9 acetylation on the contracted D4Z4 allele. Epigenetics Chromatin. 2018;11(1):47.

11. Snider L, Geng LN, Lemmers RJ, Kyba M, Ware CB, Nelson AM, Tawil R, Filippova GN, van der Maarel SM, Tapscott SJ, Miller DG. Facioscapulohumeral dystrophy: incomplete suppression of a retrotransposed gene. PLoS genetics. 2010;6(10):e1001181.

12. Hendrickson PG, Dorais JA, Grow EJ, Whiddon JL, Lim JW, Wike CL, Weaver BD, Pflueger C, Emery BR, Wilcox AL, Nix DA, Peterson CM, Tapscott SJ, Carrell DT, Cairns BR. Conserved roles of mouse DUX and human DUX4 in activating cleavage-stage genes and MERVL/HERVL retrotransposons. Nat Genet. 2017.

13. Whiddon JL, Langford AT, Wong CJ, Zhong JW, Tapscott SJ. Conservation and innovation in the DUX4-family gene network. Nat Genet. 2017.

14. De Iaco A, Planet E, Coluccio A, Verp S, Duc J, Trono D. DUX-family transcription factors regulate zygotic genome activation in placental mammals. Nat Genet. 2017.

15. Kowaljow V, Marcowycz A, Ansseau E, Conde CB, Sauvage S, Matteotti C, Arias C, Corona ED, Nunez NG, Leo O, Wattiez R, Figlewicz D, Laoudj-Chenivesse D, Belayew A, Coppee F, Rosa AL. The DUX4 gene at the FSHD1A locus encodes a pro-apoptotic protein. Neuromuscul Disord. 2007;17(8):611-23.

16. Rickard AM, Petek LM, Miller DG. Endogenous DUX4 expression in FSHD myotubes is sufficient to cause cell death and disrupts RNA splicing and cell migration pathways. Hum Mol Genet. 2015;24(20):5901-14.

17. Lemmers RJ, van der Vliet PJ, Klooster R, Sacconi S, Camano P, Dauwerse JG, Snider L, Straasheijm KR, van Ommen GJ, Padberg GW, Miller DG, Tapscott SJ, Tawil R, Frants RR, van der Maarel SM. A unifying genetic model for facioscapulohumeral muscular dystrophy. Science (New York, NY). 2010;329(5999):1650-3.

18. Dixit M, Ansseau E, Tassin A, Winokur S, Shi R, Qian H, Sauvage S, Mattéotti C, van Acker AM, Leo O, Figlewicz D, Barro M, Laoudj-Chenivesse D, Belayew A, Coppée F, YW C. DUX4, a candidate gene of facioscapulohumeral muscular dystrophy, encodes a transcriptional activator of PITX1. PNAS. 2007;104(46):18157-62.

19. Lemmers RJ, Wohlgemuth M, Frants RR, Padberg GW, Morava E, van der Maarel SM. Contractions of D4Z4 on 4qB subtelomeres do not cause facioscapulohumeral muscular dystrophy. Am J Hum Genet. 2004;75(6):1124-30.

20. Wijmenga C, Hewitt JE, Sandkuijl LA, Clark LN, Wright TJ, Dauwerse HG, Gruter AM, Hofker MH, Moerer P, Williamson R, van Ommen GJ, Padberg GW, Frants R. Chromosome 4q DNA rearrangements associated with facioscapulohumeral muscular dystrophy. Nat Genet. 1992;2(1):26-30.

21. Lemmers RJ, de Kievit P, Sandkuijl L, Padberg GW, van Ommen GJ, Frants RR, van der Maarel SM. Facioscapulohumeral muscular dystrophy is uniquely associated with one of the two variants of the 4q subtelomere. Nat Genet. 2002;32(2):235-6.

22. Lemmers RJ, Tawil R, Petek LM, Balog J, Block GJ, Santen GW, Amell AM, van der Vliet PJ, Almomani R, Straasheijm KR, Krom YD, Klooster R, Sun Y, den Dunnen JT, Helmer Q, Donlin-Smith CM, Padberg GW, van Engelen BG, de Greef JC, Aartsma-Rus AM, Frants RR, de Visser M, Desnuelle C, Sacconi S, Filippova GN, Bakker B, Bamshad MJ, Tapscott SJ, Miller DG, van der Maarel SM. Digenic inheritance of an SMCHD1 mutation and an FSHD-permissive D4Z4 allele causes facioscapulohumeral muscular dystrophy type 2. Nat Genet. 2012;44(12):1370-4.

23. Hirano T. SMC proteins and chromosome mechanics: from bacteria to humans. Philosophical transactions of the Royal Society of London Series B, Biological sciences. 2005;360(1455):507-14.

24. Blewitt ME, Vickaryous NK, Hemley SJ, Ashe A, Bruxner TJ, Preis JI, Arkell R, Whitelaw E. An N-ethyl-N-nitrosourea screen for genes involved in variegation in the mouse. Proc Natl Acad Sci U S A. 2005;102(21):7629-34.

25. Sacconi S, Lemmers RJ, Balog J, van der Vliet PJ, Lahaut P, van Nieuwenhuizen MP, Straasheijm KR, Debipersad RD, Vos-Versteeg M, Salviati L, Casarin A, Pegoraro E, Tawil R, Bakker E, Tapscott SJ, Desnuelle C, van der Maarel SM. The FSHD2 gene SMCHD1 is a modifier of disease severity in families affected by FSHD1. Am J Hum Genet. 2013;93(4):744-51.

26. Larsen M, Rost S, El Hajj N, Ferbert A, Deschauer M, Walter MC, Schoser B, Tacik P, Kress W, Muller CR. Diagnostic approach for FSHD revisited: SMCHD1 mutations cause FSHD2 and act as modifiers of disease severity in FSHD1. Eur J Hum Genet. 2014.

27. Mitsuhashi S, Boyden SE, Estrella EA, Jones TI, Rahimov F, Yu TW, Darras BT, Amato AA, Folkerth RD, Jones PL, Kunkel LM, Kang PB. Exome sequencing identifies a novel SMCHD1 mutation in facioscapulohumeral muscular dystrophy 2. Neuromuscul Disord. 2013;23(12):975-80.

28. Winston J, Duerden L, Mort M, Frayling IM, Rogers MT, Upadhyaya M. Identification of two novel SMCHD1 sequence variants in families with FSHD-like muscular dystrophy. Eur J Hum Genet. 2014.

29. van den Boogaard ML, Lemmers R, Camano P, van der Vliet PJ, Voermans N, van Engelen BG, Lopez de Munain A, Tapscott SJ, van der Stoep N, Tawil R, van der Maarel SM. 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. 2016;24(1):78-85.

30. Lemmers RJ, van den Boogaard ML, van der Vliet PJ, Donlin-Smith CM, Nations SP, Ruivenkamp CA, Heard P, Bakker B, Tapscott S, Cody JD, Tawil R, van der Maarel SM. Hemizygosity for SMCHD1 in Facioscapulohumeral Muscular Dystrophy Type 2: Consequences for 18p Deletion Syndrome. Hum Mutat. 2015;36(7):679-83.

31. Hamanaka K, Goto K, Arai M, Nagao K, Obuse C, Noguchi S, Hayashi YK, Mitsuhashi S, Nishino I. Clinical, muscle pathological, and genetic features of Japanese facioscapulohumeral muscular dystrophy 2 (FSHD2) patients with SMCHD1 mutations. Neuromuscul Disord. 2016;26(4-5):300-8.

32. van den Boogaard ML, Lemmers RJ, Balog J, Wohlgemuth M, Auranen M, Mitsuhashi S, van der Vliet PJ, Straasheijm KR, van den Akker RF, Kriek M, Laurense-Bik ME, Raz V, van Ostaijen-Ten Dam MM, Hansson KB, van der Kooi EL, Kiuru-Enari S, Udd B, van Tol MJ, Nishino I, Tawil R, Tapscott SJ, van Engelen BG, van der Maarel SM. Mutations in DNMT3B Modify Epigenetic Repression of the D4Z4 Repeat and the Penetrance of Facioscapulohumeral Dystrophy. Am J Hum Genet. 2016;98(5):1020-9.

33. Gordon CT, Xue S, Yigit G, Filali H, Chen K, Rosin N, Yoshiura KI, Oufadem M, Beck TJ, McGowan R, Magee AC, Altmuller J, Dion C, Thiele H, Gurzau AD, Nurnberg P, Meschede D, Muhlbauer W, Okamoto N, Varghese V, Irving R, Sigaudy S, Williams D, Ahmed SF, Bonnard C, Kong MK, Ratbi I, Fejjal N, Fikri M, Elalaoui SC, Reigstad H, Bole-Feysot C, Nitschke P, Ragge N, Levy N, Tuncbilek G, Teo AS, Cunningham ML, Sefiani A, Kayserili H, Murphy JM, Chatdokmaiprai C, Hillmer AM, Wattanasirichaigoon D, Lyonnet S, Magdinier F, Javed A, Blewitt ME, Amiel J, Wollnik B, Reversade B. De novo mutations in SMCHD1 cause Bosma arhinia microphthalmia syndrome and abrogate nasal development. Nat Genet. 2017.

34. Shaw ND, Brand H, Kupchinsky ZA, Bengani H, Plummer L, Jones TI, Erdin S, Williamson KA, Rainger J, Stortchevoi A, Samocha K, Currall BB, Dunican DS, Collins RL, Willer JR, Lek A, Lek M, Nassan M, Pereira S, Kammin T, Lucente D, Silva A, Seabra CM, Chiang C, An Y, Ansari M, Rainger JK, Joss S, Smith JC, Lippincott MF, Singh SS, Patel N, Jing JW, Law JR, Ferraro N, Verloes A, Rauch A, Steindl K, Zweier M, Scheer I, Sato D, Okamoto N, Jacobsen C, Tryggestad J, Chernausek S, Schimmenti LA, Brasseur B, Cesaretti C, Garcia-Ortiz JE, Buitrago TP, Silva OP, Hoffman JD, Muhlbauer W, Ruprecht KW, Loeys BL, Shino M, Kaindl AM, Cho CH, Morton CC, Meehan RR, van Heyningen V, Liao EC, Balasubramanian R, Hall JE, Seminara SB, Macarthur D, Moore SA, Yoshiura KI, Gusella JF, Marsh JA, Graham JM, Jr., Lin AE, Katsanis N, Jones PL, Crowley WF, Jr., Davis EE, FitzPatrick DR, Talkowski ME. SMCHD1 mutations associated with a rare muscular dystrophy can also cause isolated arhinia and Bosma arhinia microphthalmia syndrome. Nat Genet. 2017.

35. Mul K, Lemmers R, Kriek M, van der Vliet PJ, van den Boogaard ML, Badrising UA, Graham JM, Jr., Lin AE, Brand H, Moore SA, Johnson K, Evangelista T, Topf A, Straub V, Kapetanovic Garcia S, Sacconi S, Tawil R, Tapscott SJ, Voermans NC, van Engelen BGM, Horlings CGC, Shaw ND, van der Maarel SM. FSHD type 2 and Bosma arhinia microphthalmia syndrome: Two faces of the same mutation. Neurology. 2018.

36. Gurzau AD, Chen K, Xue S, Dai W, Lucet IS, Ly TTN, Reversade B, Blewitt ME, Murphy JM. FSHD2- and BAMS-associated mutations confer opposing effects on SMCHD1 function. The Journal of biological chemistry. 2018.

37. Lemmers RJLF, van der Stoep N, Vliet PJvd, Moore SA, San Leon Granado D, Johnson K, Topf A, Straub V, Evangelista T, Mozaffar T, Kimonis V, Selvatici R, Ferlini A, Voermans N, van Engelen B, Sacconi S, Tawil R, Lamers M, van der Maarel SM. SMCHD1 mutation spectrum for facioscapulohumeral muscular dystrophy type 2 (FSHD2) and Bosma arhinia microphthalmia syndrome (BAMS) reveals disease-specific

localisation of variants in the ATPase domain. Journal of medical genetics. 2019.

38. Ricci E, Galluzzi, G., Deidda, G., Cacurri, S., Colantoni, L., Merico, B., Piazzo, N., Servidei, S., Vigneti, E., Pasceri, V. 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. 1999;45:751-7.

39. van der Maarel SM, Miller DG, Tawil R, Filippova GN, Tapscott SJ. Facioscapulohumeral muscular dystrophy: consequences of chromatin relaxation. Current opinion in neurology. 2012;25(5):614-20.

40. Balog J, Goossens R, Lemmers R, Straasheijm KR, van der Vliet PJ, Heuvel AVD, Cambieri C, Capet N, Feasson L, Manel V, Contet J, Kriek M, Donlin-Smith CM, Ruivenkamp CAL, Heard P, Tapscott SJ, Cody JD, Tawil R, Sacconi S, van der Maarel SM. Monosomy 18p is a risk factor for facioscapulohumeral dystrophy. Journal of medical genetics. 2018.

41. Balog J, Thijssen PE, Shadle S, Straasheijm KR, van der Vliet PJ, Krom YD, van den Boogaard ML, de Jong A, RJ FL, Tawil R, Tapscott SJ, van der Maarel SM. Increased DUX4 expression during muscle differentiation correlates with decreased SMCHD1 protein levels at D4Z4. Epigenetics : official journal of the DNA Methylation Society. 2015;10(12):1133-42.

42. Lareau LF, Inada M, Green RE, Wengrod JC, Brenner SE. Unproductive splicing of SR genes associated with highly conserved and ultraconserved DNA elements. Nature. 2007;446:926.

43. Hartweck LM, Anderson LJ, Lemmers RJ, Dandapat A, Toso EA, Dalton JC, Tawil R, Day JW, van der Maarel SM, Kyba M. A focal domain of extreme demethylation within D4Z4 in FSHD2. Neurology. 2013;80(4):392-9.

44. Lemmers RJLF, van der Vliet PJ, Vreijling JP, Henderson D, van der Stoep N, Voermans N, van Engelen B, Baas F, Sacconi S, Tawil R, van der Maarel SM. Cis D4Z4 repeat duplications associated with facioscapulohumeral muscular dystrophy type 2. Human Molecular Genetics. 2018:ddy236-ddy.

45. Statland JM, Donlin-Smith CM, Tapscott SJ, Lemmers RJ, van der Maarel SM, Tawil R. Milder phenotype in facioscapulohumeral dystrophy with 7-10 residual D4Z4 repeats. Neurology. 2015;85(24):2147-50.

46. Lemmers RJ, Wohlgemuth M, van der Gaag KJ, van der Vliet PJ, van Teijlingen CM, de Knijff P, Padberg GW, Frants RR, van der Maarel SM. Specific sequence variations within the 4q35 region are associated with facioscapulohumeral muscular dystrophy. Am J Hum Genet. 2007;81(5):884-94.

47. Scionti I, Fabbri G, Fiorillo C, Ricci G, Greco F, D'Amico R, Termanini A, Vercelli L, Tomelleri G, Cao M, Santoro L, Percesepe A, Tupler R. Facioscapulohumeral muscular dystrophy: new insights from compound heterozygotes and implication for prenatal genetic counselling. Journal of medical genetics. 2012;49(3):171-8.

48. D'Astolfo DS, Pagliero RJ, Pras A, Karthaus WR, Clevers H, Prasad V, Lebbink RJ, Rehmann H, Geijsen N. Efficient intracellular delivery of native proteins. Cell. 2015;161(3):674-90.

49. Dion C, Roche S, Laberthonniere C, Broucqsault N, Mariot V, Xue S, Gurzau AD, Nowak A, Gordon CT, Gaillard MC, El-Yazidi C, Thomas M, Schlupp-Robaglia A, Missirian C, Malan V, Ratbi L, Sefiani A, Wollnik B, Binetruy B, Salort Campana E, Attarian S, Bernard R, Nguyen K, Amiel J, Dumonceaux J, Murphy JM, Dejardin J, Blewitt ME, Reversade B, Robin JD, Magdinier F. SMCHD1 is involved in de novo methylation of the DUX4-encoding D4Z4 macrosatellite. Nucleic acids research. 2019.

Full text and figures also available at:

Supplementary information

Material and methods

Subjects

A French family (Rf744, Figure 1A) and an American family (Rf1034, Figure 1B) were studied after informed consent and after the study protocol had been 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). (1)

D4Z4 repeat sizing, haplotype analysis and methylation analysis

For genotyping, high molecular-weight 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.(2) 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.(2, 3) D4Z4 methylation analysis in genomic DNA derived from peripheral blood mononuclear cells was assessed using the FseI restriction site located in the most proximal unit of the D4Z4 arrays on chromosomes 4 and 10 as previously published.(4, 5) The Delta1 value, representing repeat size-corrected D4Z4 methylation, was calculated as described in Lemmers et al. 2014. In immortalized myoblast clones, DNA methylation analysis was performed by bisulfite sequencing of the DR1 region in D4Z4 as previously described.(6) After PCR amplification, PCR products were cloned into the pCR™4-TOPO vector (Invitrogen, Life Technologies). From each culture, a minimum of 10 independent clones were sequenced by Sanger sequencing. CpG methylation analysis was performed using BiQ Analyzer (V2.02). (7)

Genomic *SMCHD1* variant analysis

To identify possible *SMCHD1* variants in the index cases, the sequences of all coding exons and splice regions were analysed after PCR amplification. The previously published *SMCHD1* intron-specific primers target a sequence positioned at least 50 nucleotides from the splice donor or acceptor sites.(8) 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 mixture (30 μl/reaction) was composed of 0.5 units DreamTaq DNA polymerase (5U/μl Thermo Fisher Scientific), 1× DreamTaq buffer (Thermo Fisher Scientific), 3 μl of dNTPs (2mM of each nucleotide) and 25 pmol of each primer. The following PCR protocol was followed: denaturation for 5 min at 95°C, followed by 35 cycles of: 95°C for 30 sec, 60°C for 30 sec, 72°C for 45 sec, with a final elongation step of 10 min at 72°C. The PCR products were 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). Information about all variants mentioned in this manuscript are publicly available at http://www.lovd.nl/SMCHD1.

Blood 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 random hexamer primers and the RevertAid First Strand cDNA Synthesis Kit (Thermo Fisher Scientific) according to the manufacturer's recommendations. A reverse transcription polymerase chain reaction (RT-PCR) screen on *SMCHD1* mRNA was performed with partially overlapping primer sets to identify potential mRNA changes. Additional RT-PCR for *SMCHD1* exon 32 to 35, exon 32 to 36 and exon 30 to 37 were performed using primers 4098F/4406R, 3233F/Ex36R and 4062F/Ex37R, respectively (online supplementary table S2). RT-PCRs were performed in 30 µl reactions using 0.5 units DreamTaq DNA polymerase (5U/µl Thermo Fisher Scientific), 1× DreamTaq buffer (Thermo Fisher Scientific), 3 µl of dNTPs (2mM of each nucleotide) and 25 pmol of each primer. The RT-PCR protocol was as follows: 95°C for 5 min, 35 cycles of: 95°C for 30 sec, 60°C for 30 sec, 72°C for 30 sec, then 72°C for 10 min. RT-PCR products were separated by size on 2% agarose gels after which PCR products were purified from gel (NucleoSpin® Gel and PCR Clean-up, Machery Nagel). Purified PCR products were cloned into a pCR™4-TOPO vector and transformed in DH5α heat-shock competent cells (Subcloning Efficiency DH5α Competent Cells, Invitrogen, Life Technologies). The inserts of multiple clones were analysed by Sanger sequencing.

Cell culture

The human primary myoblasts from Rf1034.5 originated from the University of Rochester bio repository (http://www.urmc.rochester.edu/fields-center/). Muscle samples were obtained after subjects consent under a protocol approved by the institutional review board at the University of Rochester (Rochester, USA). The biopsy for patient Rf1034.5 was taken from the vastus lateralis (quadriceps) muscle, and had a pathologic score of 2 (on a 0-12 scale). (9) Myoblasts were cultured in F-10 Nut Mix (1x) + GlutaMax™-I (Gibco by Life Technologies, Bleiswijk, The Netherlands) supplemented with 20% heat-inactivated fetal bovine serum (FBS, Gibco by Life Technologies), 1% penicillin/streptomycin (Gibco by Life Technologies), 10 ng/ml rhFGF (Promega, Leiden, The Netherlands) and 1 mM dexamethasone (Sigma-Aldrich, Zwijndrecht, The Netherlands). Myoblast fusion was induced by exposing cultures at approximately 80% confluency to DMEM (1x) + GlutaMax™-I media (Gibco by Life Technologies) supplemented with 2% KnockOut serum replacement formulation (Gibco by Life Technologies) for 48 h.

HeLa cells were a kind gift of Prof. A.C. Vertegaal (Leiden University Medical Center, Department of Chemical and Cell Biology) and were maintained in high glucose DMEM Glutamax supplemented with 10% heat inactivated FBS and 1% penicillin/streptomycin. The identity of the HeLa cell line was verified by in-house STR genotyping.

Rf1034.5 primary myoblasts were immortalized by retroviral transduction with viral particles encoding hTERT-Hygromycin and CDK4-Neomycin.(10) Cells were selected and maintained in media containing 400µg/ml G418 (Thermo Fischer Scientific) and 200µg/ml hygromycin B (Thermo Fischer Scientific).

Genome editing

Two plasmids expressing gRNAs targeting *SMCHD1* intron 34, D3 and U3 (collectively called U/D3 from hereon) (online supplementary table S3), were assembled by cloning the annealed oligonucleotide pairs D3+/D3- and U3+/U3 into the BveI-digested lentiviral vector constructs AO58_pLKO.1-Blast::EGFP.U6.BveI-stuffer and AA19_pLKO1.puro.U6.sgRNA.BveI-stuffer.(11) respectively. Specificity of designed guides was checked by use of the Cas-OFFinder algorithm (online supplementary table S4).(12) The gRNA sequences in U/D3 were confirmed by Sanger sequencing. A previously described lentiviral vector construct (X50, online supplementary table S3) encoding a gRNA targeting the human *AAVS1* locus was taken as a control.(11) Lentiviral particles encoding the different gRNAs were produced by transient transfection of HEK293T cells. In brief, transfection mixtures consisting of 7.5 µg pCMV-VSVG, 11.4 µg pMDLg-RRE (gag/pol), 5.4 µg pRSV-REV and 13.7 µg of each of the gRNA-expressing lentiviral plasmid were diluted in 1 ml 150 mM NaCl and were mixed with 114 µg PEI diluted in 1 ml 150 mM NaCl. After 10 minutes incubation at room temperature, the transfection mixtures were added to a T175 flask containing 70% confluent HEK293T cells (in 20 ml medium). One day after transfection, the medium was refreshed. The lentiviral vector particles were harvested 48 and 72 hours after transfection by collecting producer-cell supernatants. The supernatants were centrifuged (700 ×g, 10 min) and subsequently filtered through a 0.45µm filter (Acrodisc Syringe filter, HT Tuffryn membrane, Pall Corporation). Lentiviral vector titers were determined with a p24 Antigen ELISA (ZeptoMetrixCorporation). Next, the human primary myoblasts of Rf1034.5 were seeded at a density of 300,000 cells in a 3.5 cm^2 dish. Five hours after seeding the myoblasts were transduced with gRNA-expressing lentiviral vectors D3 and U3 or control gRNA-expressing lentiviral vector X50 for 3 days at a vector particle dose of 45 ng p24 per 3.5 cm^2 dish (Greiner Bio-One). After 3 days incubation at 37°C the inocula were removed and cells transduced with the lentiviral vectors encoding for gRNA D3 and U3 were kept in growth medium supplemented with blasticidin (20 µg/ml) and puromycin (0.75 µg/ml) for 10 days. The parallel myoblast cultures transduced with lentivirus encoding for gRNA X50 were only exposed to blasticidin (20 µg/ml) for 10 days.

The Cas9 protein was delivered using induced transduction by osmocytosis and propanebetaine (iTOP).(13) For iTOP treatment Rf1034.5 myoblasts, containing the aforementioned gRNAs, were seeded at a density of 20,000 cells per well of a 96 well plate coated with collagen. Six hours after seeding 50 µl transduction mix, containing 25 µl 2× CRISPR supplement (online supplementary table S4, previously described in extended experimental procedures in D'Astolfo et al. 2015), 10 µl recombinant Cas9 protein (75 µM) (Utrecht University Gene Editing Facility), 50 ng of a gRNA (target sequence: GAAGTTGACTTACTGAAGAA) targeting the β2 *microglobulin* (*B2M*) gene coding for a component of MHC class I molecules was added to each well, and incubated for 40 min at 37°C. The gRNA targeting *B2M* was added to achieve selection of targeted myoblasts by FACS sorting for B2M negative cells. Five days after iTOP treatment the cells were prepared for FACS sorting for B2M negative cells (ARIA,

BD Biosciences). Myoblasts were trypsinized and resuspended in medium, washed with FACS buffer (2% FBS, 0.1% NaN3, 5 mM EDTA in 1x PBS) and incubated for 20 min at 4°C with 25 µl of primary antibody anti-human HLA ABC (MHC1)–FITC (#311404, Biolegend) diluted 1:150 in FACS buffer. Next, Rf1034.5 myoblasts were washed with FACS buffer without NaN3 (2% FBS, 5 mM EDTA in 1x PBS) and incubated for 20 min at 4°C with 200 µl of secondary antibody sheep anti-mouse Atto647N (Cy5 labeled) (Atto-Tec) diluted 1:2500 in FACS buffer without NaN3. Rf1034.5 myoblasts were washed twice with FACS buffer without NaN3 and pipetted through a cell strainer before FACS sorting. Since the myoblasts are $eGFP^+$ due to the integration of the gRNA-expressing lentiviral vector, stained cells were FACS sorted for the cells negative for Cy5 (indicating negative for B2M) and positive for eGFP in a 48 well plate. Fourteen days after iTOP treatment, Rf1034.5 myoblasts were harvested for (genomic) DNA extraction and seeded for fusion (differentiation) in wells of a 6 well plate (Greiner Bio-One). Three days after this seeding (17 days after iTOP treatment), the Rf1034.5 myoblasts were exposed to fusion medium. After an additional two days of incubation (19 days after iTOP treatment) differentiated myocytes were harvested in Qiazol (Qiagen) for RNA extraction.

For genomic DNA extraction, cells were trypsinized, resuspended in medium and centrifuged at 220×g for 5 min. The cell pellets were resuspended in 1 ml of nucleus lysis buffer (10 mM Tris-HCl (pH 8.2), 2 mM EDTA, 0.4 M NaCl) supplemented with 50 µl 20% SDS and 33.3 µl pronase (20 mg/ml in 10 mM NaCl/10 mM Tris-HCl (pH7.5)), and were subsequently incubated at 37°C for 2 days. The proteins were precipitated by adding 400 µl of 5M NaCl and subsequent centrifugation. The genomic DNA was precipitated from the supernatant by adding 800 µl of isopropanol and, after centrifugation, was washed with 70% EtOH. The recovered genomic DNA was suspended in 50 µl TE and the DNA was used to assess targeted deletions in intron 34 using primers SMCHD1_ex35-633fwd (5'-CTGCATTGAGCCGGTATTTT-3') and SMCHD1 c.4406rev (5'-TCCATCATAAAACCAAACTGGA-3'). PCRs were performed in 30 µl reactions using 0.5 units DreamTaq DNA polymerase (5U/µl Thermo Fisher Scientific), 1× DreamTaq buffer (Thermo Fisher Scientific), 3 µl of dNTPs (2mM of each nucleotide) and 25 pmol of each primer. The following PCR protocol was followed: denaturation for 5 min at 95°C, followed by 35 cycles of: 95°C for 30 sec, 60°C for 30 sec, 72°C for 45 sec, with a final elongation step of 10 min at 72°C. PCR products were analysed on a 2% agarose gel containing 0.001% Ethidium Bromide.(14, 15)

Single cell gene editing of immortalized Rf1034.5 myoblasts (Rf1034.5-iMB) was performed as described above for primary cell pools, except for the following modifications. The guide RNA (gRNA) oligos (U3, D3, X50 and B2M) were sub-cloned into the PX458 plasmid (Addgene #48138, Teddington, UK).(16) The generated plasmids were used as template in PCR reactions with primers containing a T7 RNA polymerase promoter sequence (online supplementary table S3). The purified PCR products were subsequently used to generate *in vitro* transcribed gRNA using T7 RNA polymerase. The transcribed gRNAs were purified using the MEGAclear Transcription Clean-Up kit (Ambion) according to the manufacturer's instructions. Per condition, two wells of a 96 well plate of iMB-Rf1034.5 were transduced using iTOP as described earlier (25µl CRISPR supplement, 10µl Cas9 protein, 10µl gRNA (10µg total gRNA) and 5µl H_2O), and were allowed to recover before single cell seeding using an AriaIII flowcytometer (100µM nozzle, 20psi) (BD-Bioscience). For detection of genome editing outcomes, genomic DNA from single cell derived colonies was extracted with 50µl Single Worm Lysis Buffer (SWLB) (50 mM KCl, 10 mM Tris-HCl pH 8.3, 2.5 mM $MgCl_2$, 0.45% NP40, 0.45% Tween-20 and 0.01% gelatin), supplemented with 0.1mg/ml proteinase K). The lysates were digested overnight at 55°C and were heat inactivated for 10 minutes at

95°C.(17) Next, isolated genomic DNA was subjected to PCR using primers ex35-633fwd and c.4406rev as described earlier. Unedited and heterozygously edited clones were genotyped by Sanger sequencing of the PCR product to identify the edited allele. Parallel cell cultures were subsequently expanded to allow differentiation of myoblasts and harvesting of DNA, RNA and protein from myotubes.

For gene editing of HeLa cells, plasmids containing wild-type sp.Cas9 PX458(16) and the gRNAs targeting the positions described earlier as U/D3 or X50 were used. Cells were transfected using Polyethylenimine (PEI) (1mg/ml, pH7.4 – Ratio: 3µl PEI / 1µg DNA) (Polysciences), and 72 hours post-transfection, GFP positive cells were single cell sorted using an AriaIII flowcytometer (BD-Bioscience). Single cells capable of forming a colony were transferred to two 96-well plates. One plate of cells was lysed in SWLB and analyzed by PCR. Clones showing a pattern of heterozygous or homozygous editing were cultured further on the duplicate plate until sufficient material was available for protein analysis and secondary verification of genotype by PCR and Sanger sequencing.

Expression analysis

RNA was isolated using the Direct-zol RNA miniprep kit (Zymo Research), according to the manufacturer's instructions, including the DNase I treatment. cDNA synthesis and qPCR analysis were performed as described earlier(18) qPCR and RT-PCR primers are listed in (online supplementary tables S2 and S6). *GUSB* and *RPL13* transcripts were used as internal controls for every individual sample. Statistical analysis was performed using Graphpad Prism Version 8.

Inhibition of NMD

Primary myoblasts from patient Rf1034.5 were seeded in 6 wells plates, and were differentiated into myotubes for 48 hours upon reaching confluency. After 42 hours of fusion, Cycloheximide (Sigma-Aldrich, C4859-1ML) was added to cells as indicated for 6 hours. RNA was harvested and transcripts were measured using RT-qPCR as described. *SRSF2*, a transcript known to produce isoforms targeted by NMD was used a positive control for treatment (19, 20). Normalization was performed on *RPL13* only, as *GUSB* responds positively to CHX treatment.

Protein analysis

Levels of cellular SMCHD1 protein were analysed by Western blotting. Cells grown in 6 well plates were lysed in Laemmli lysis buffer (2% SDS, 10% glycerol, 60 mM Tris pH6.8) and boiled for 10 minutes at 95°C, after which total protein concentration was determined by BCA protein quantification kit (Thermo Fischer Scientific). Equal amounts of protein were supplemented with β-mercaptoethanol and bromophenol blue (end concentration 2% and 0.01%, respectively), boiled for 10 minutes at 95°C and loaded on Criterion TGX 4-20%

protein gels (Bio-Rad). After electrophoresis, gels were transferred to Immobilon-FL (Merck-Millipore) PVDF membranes, which were afterwards blocked in Odyssey blocking buffer (Li-Cor). Blocked membranes were subsequently probed with Rabbit-α-SMCHD1 (Abcam 176731) (1:1000), Mouse-α-Tubulin (Sigma-Aldrich T6199) (1:5000) or Goat-α-Actin (Santa-Cruz SC-1616) (1:500) in Takara Immunobooster (Takara)(SMCHD1) or 4% Skim milk (Sigma-Aldrich)(Tubulin and Actin). For secondary antibodies, Donkey-α-Rabbit800CW, Donkey-α-Mouse680RD and Donkey-α-Goat800CW IRDyes (Li-Cor) were used. Immunocomplexes were visualized and quantified using an Odyssey Classic Infrared scanner and the accompanying software package (V3.0) (Li-Cor). Statistical analysis was performed using Graphpad Prism Version 8.

Chromatin immunoprecipitation

Chromatin immunoprecipitation qPCR (ChIP-qPCR) for SMCHD1 was performed as described previously.(4, 18) 30 µg sheared crosslinked chromatin was used per ChIP, using 5µg SMCHD1 antibody (Abcam ab31865) or 5µg control rabbit IgG (Abcam ab37415). ChIP-qPCR primers used are listed in supp. Table S6.

Supplementary references:

1. Ricci E, Galluzzi, G., Deidda, G., Cacurri, S., Colantoni, L., Merico, B., Piazzo, N., Servidei, S., Vigneti, E., Pasceri, V. 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. 1999;45:751-7.
2. Lemmers RJ, van der Vliet PJ, Klooster R, Sacconi S, Camano P, Dauwerse JG, Snider L, Straasheijm KR, van Ommen GJ, Padberg GW, Miller DG, Tapscott SJ, Tawil R, Frants RR, van der Maarel SM. A unifying genetic model for facioscapulohumeral muscular dystrophy. Science (New York, NY). 2010;329(5999):1650-3.
3. Lemmers RJ, de Kievit P, Sandkuijl L, Padberg GW, van Ommen GJ, Frants RR, van der Maarel SM. Facioscapulohumeral muscular dystrophy is uniquely associated with one of the two variants of the 4q subtelomere. Nat Genet. 2002;32(2):235-6.
4. Lemmers RJ, Tawil R, Petek LM, Balog J, Block GJ, Santen GW, Amell AM, van der Vliet PJ, Almomani R, Straasheijm KR, Krom YD, Klooster R, Sun Y, den Dunnen JT, Helmer Q, Donlin-Smith CM, Padberg GW, van Engelen BG, de Greef JC, Aartsma-Rus AM, Frants RR, de Visser M, Desnuelle C, Sacconi S, Filippova GN, Bakker B, Bamshad MJ, Tapscott SJ, Miller DG, van der Maarel SM. Digenic inheritance of an SMCHD1 mutation and an FSHD-permissive D4Z4 allele causes facioscapulohumeral muscular dystrophy type 2. Nat Genet. 2012;44(12):1370-4.
5. Deak KL, Lemmers RJ, Stajich JM, Klooster R, Tawil R, Frants RR, Speer MC, van der Maarel SM, Gilbert JR. Genotype-phenotype study in an FSHD family with a proximal deletion encompassing p13E-11 and D4Z4. Neurology. 2007;68(8):578-82.
6. Hartweck LM, Anderson LJ, Lemmers RJ, Dandapat A, Toso EA, Dalton JC, Tawil R, Day JW, van der Maarel SM, Kyba M. A focal domain of extreme demethylation within D4Z4 in FSHD2. Neurology. 2013;80(4):392-9.
7. Bock C, Reither S, Mikeska T, Paulsen M, Walter J, Lengauer T. BiQ Analyzer: visualization and quality control for DNA methylation data from bisulfite sequencing. Bioinformatics (Oxford, England). 2005;21(21):4067-8.
8. Lemmers RJ, Goeman JJ, van der Vliet PJ, van Nieuwenhuizen MP, Balog J, Vos-Versteeg M, Camano P, Ramos Arroyo MA, Jerico I, Rogers MT, Miller DG, Upadhyaya M, Verschuuren JJ, Lopez de Munain Arregui A, van Engelen BG, Padberg GW, Sacconi S, Tawil R, Tapscott SJ, Bakker B, van der Maarel SM. Inter-individual differences in CpG methylation at D4Z4 correlate with clinical variability in FSHD1 and FSHD2. Human molecular genetics. 2014.
9. Statland JM, Shah B, Henderson D, Van Der Maarel S, Tapscott SJ, Tawil R. Muscle pathology grade for

facioscapulohumeral muscular dystrophy biopsies. Muscle Nerve. 2015;52(4):521-6.

10. Stadler G, Chen JC, Wagner K, Robin JD, Shay JW, Emerson CP, Jr., Wright WE. Establishment of clonal myogenic cell lines from severely affected dystrophic muscles - CDK4 maintains the myogenic population. Skelet Muscle. 2011;1(1):12.

11. Maggio I, Stefanucci L, Janssen JM, Liu J, Chen X, Mouly V, Goncalves MA. Selection-free gene repair after adenoviral vector transduction of designer nucleases: rescue of dystrophin synthesis in DMD muscle cell populations. Nucleic acids research. 2016;44(3):1449-70.

12. Bae S, Park J, Kim JS. Cas-OFFinder: a fast and versatile algorithm that searches for potential off-target sites of Cas9 RNA-guided endonucleases. Bioinformatics (Oxford, England). 2014;30(10):1473-5.

13. D'Astolfo DS, Pagliero RJ, Pras A, Karthaus WR, Clevers H, Prasad V, Lebbink RJ, Rehmann H, Geijsen N. Efficient intracellular delivery of native proteins. Cell. 2015;161(3):674-90.

14. Sharp PA, Sugden B, Sambrook J. Detection of two restriction endonuclease activities in Haemophilus parainfluenzae using analytical agarose--ethidium bromide electrophoresis. Biochemistry. 1973;12(16):3055-63.

15. Aaij C, Borst P. The gel electrophoresis of DNA. Biochim Biophys Acta. 1972;269(2):192-200.

16. Ran FA, Hsu PD, Wright J, Agarwala V, Scott DA, Zhang F. Genome engineering using the CRISPR-Cas9 system. Nature protocols. 2013;8(11):2281-308.

17. Williams BD, Schrank B, Huynh C, Shownkeen R, Waterston RH. A genetic mapping system in Caenorhabditis elegans based on polymorphic sequence-tagged sites. Genetics. 1992;131(3):609-24.

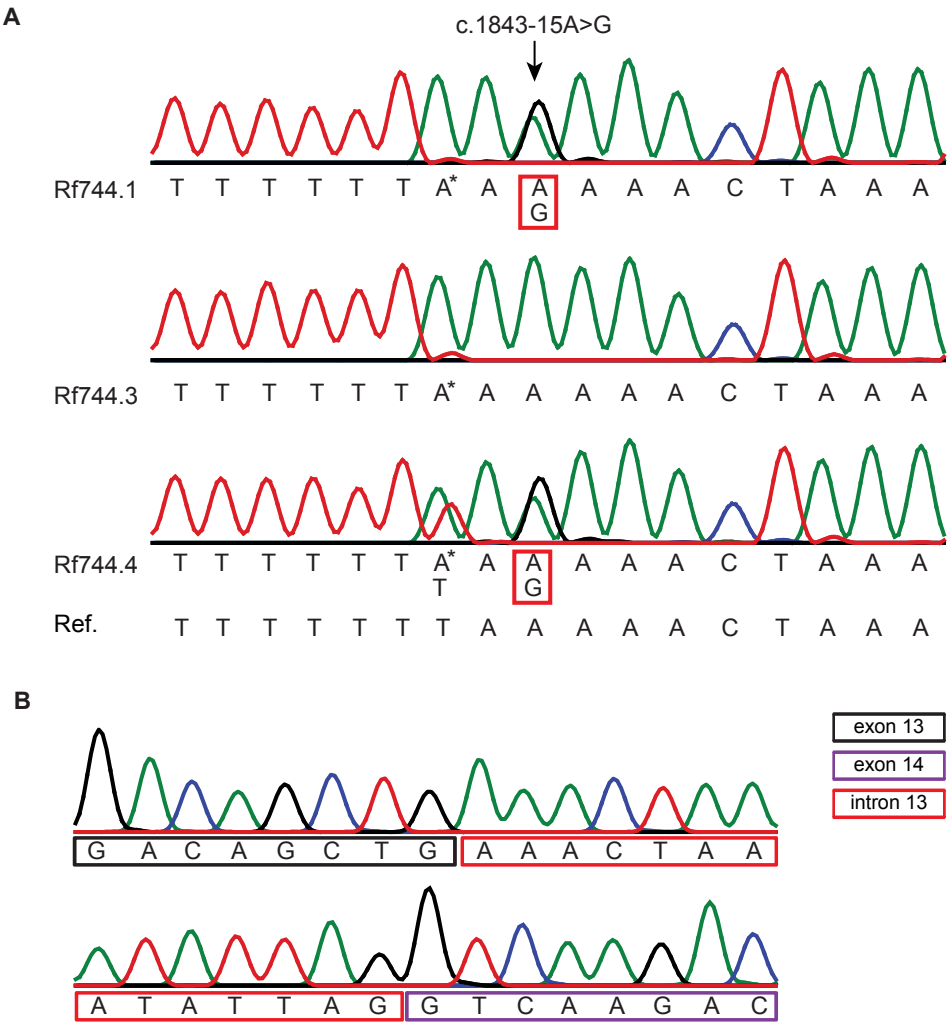
18. Balog J, Thijssen PE, Shadle S, Straasheijm KR, van der Vliet PJ, Krom YD, van den Boogaard ML, de Jong A, RJ FL, Tawil R, Tapscott SJ, van der Maarel SM. Increased DUX4 expression during muscle differentiation correlates with decreased SMCHD1 protein levels at D4Z4. Epigenetics : official journal of the DNA Methylation Society. 2015;10(12):1133-42.

19. Lareau LF, Inada M, Green RE, Wengrod JC, Brenner SE. Unproductive splicing of SR genes associated with highly conserved and ultraconserved DNA elements. Nature. 2007;446:926.

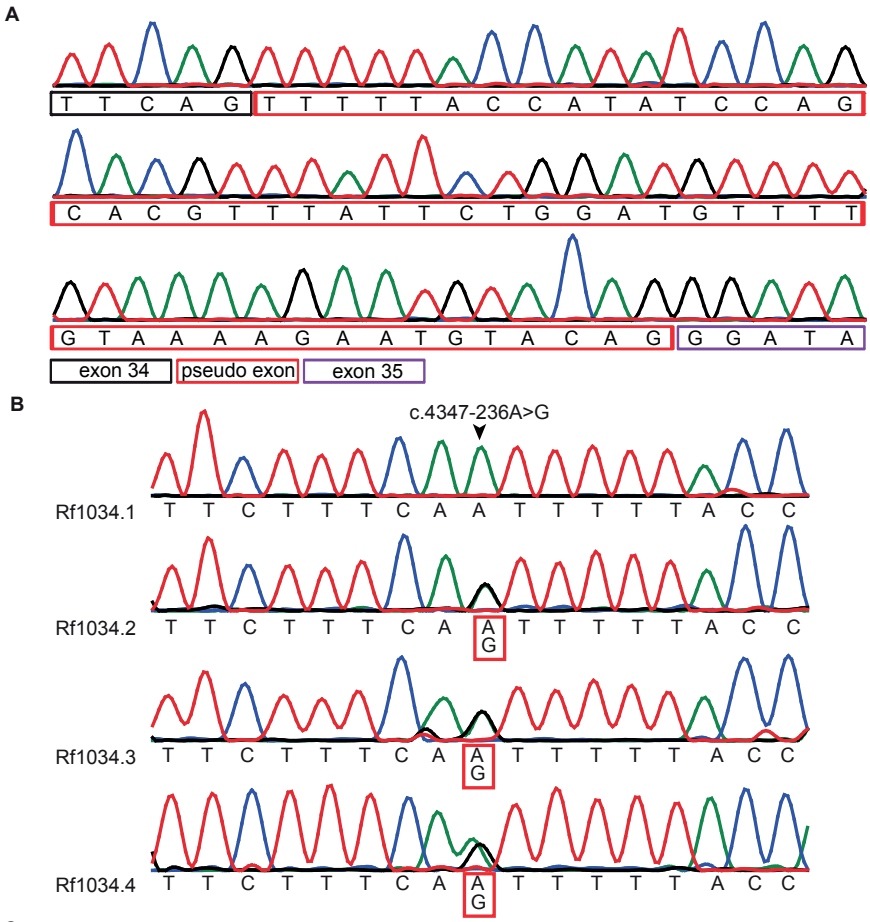
20. Feng Q, Snider L, Jagannathan S, Tawil R, van der Maarel SM, Tapscott SJ, Bradley RK. A feedback loop between nonsense-mediated decay and the retrogene DUX4 in facioscapulohumeral muscular dystrophy. eLife. 2015;4.

21. Balog J, Goossens R, Lemmers R, Straasheijm KR, van der Vliet PJ, Heuvel AVD, Cambieri C, Capet N, Feasson L, Manel V, Contet J, Kriek M, Donlin-Smith CM, Ruivenkamp CAL, Heard P, Tapscott SJ, Cody JD, Tawil R, Sacconi S, van der Maarel SM. Monosomy 18p is a risk factor for facioscapulohumeral dystrophy. Journal of medical genetics. 2018.

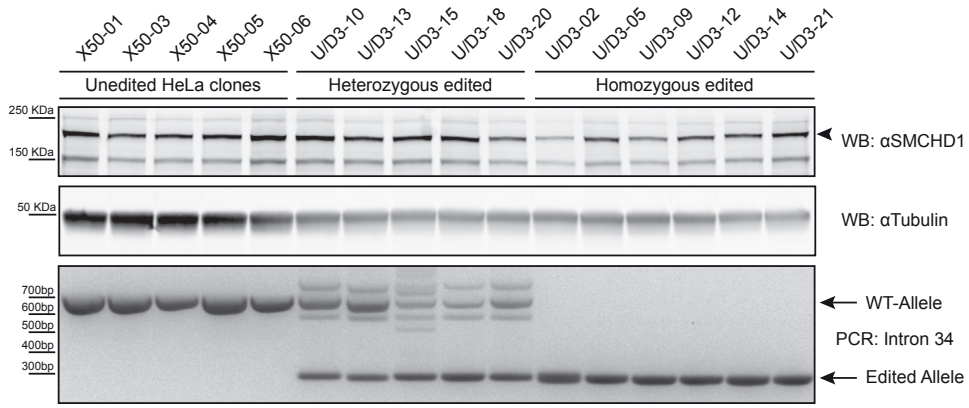
22. Lemmers RJ, van der Vliet PJ, Balog J, Goeman JJ, Arindrarto W, Krom YD, Straasheijm KR, Debipersad RD, Ozel G, Sowden J, Snider L, Mul K, Sacconi S, van Engelen B, Tapscott SJ, Tawil R, van der Maarel SM. Deep characterization of a common D4Z4 variant identifies biallelic DUX4 expression as a modifier for disease penetrance in FSHD2. Eur J Hum Genet. 2018;26(1):94-106.



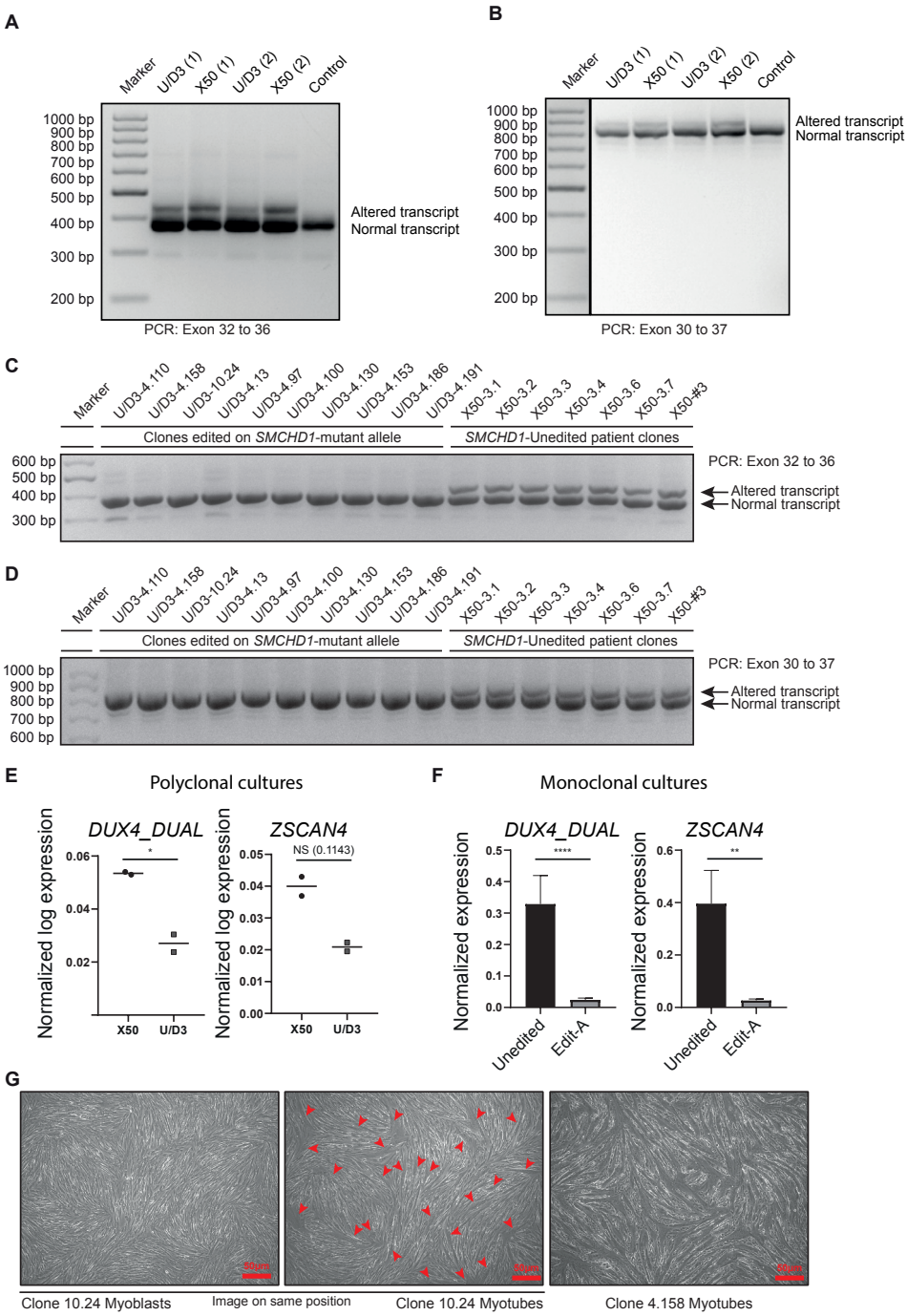
Supplementary figure S1: Identification of the *SMCHD1* intronic variant in Rf744. A) Sanger sequence track from Rf744 showing the intronic variant in *SMCHD1* at position c.1843-15 in Rf744.1 and Rf744.4, highlighted with a red rectangle. * indicates common SNP rs8090988 (T/A, ancestral T, minor allele frequency 0.33 (A)) B) Sanger sequence track of a TOPO clone corresponding to a RT-PCR product containing the *SMCHD1* sequence from c.1843-14 to c.1843-1.



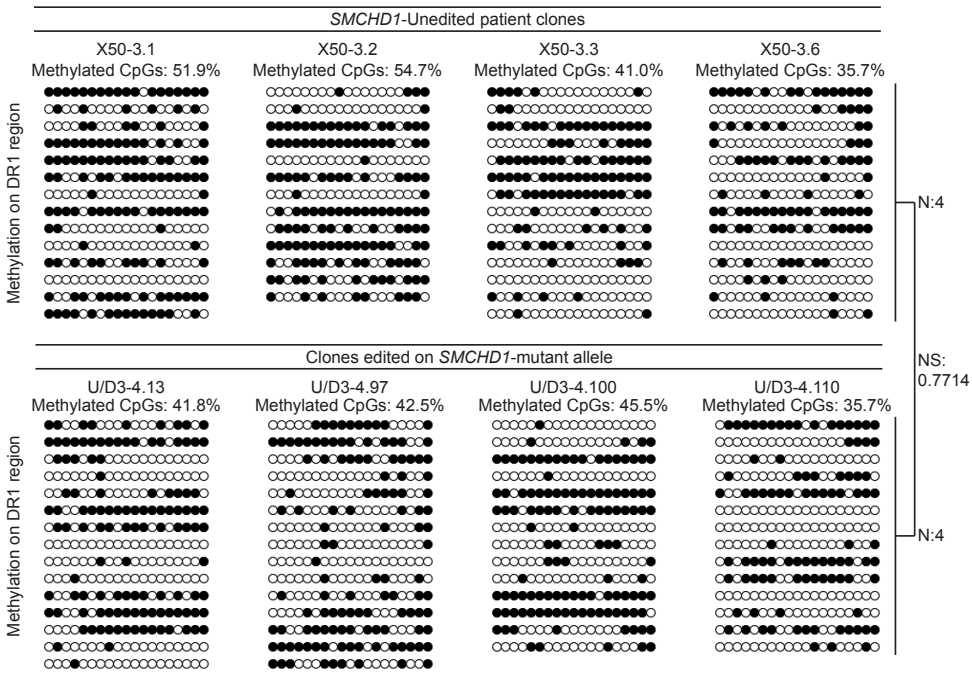
Supplementary figure S2: Identification of intronic variant in *SMCHD1* in Rf1034. A) Sanger sequence track of a TOPO clone corresponding to a RT-PCR product containing the sequence of the pseudo-exon. B) Sanger sequence track showing the deep intronic variant in *SMCHD1* at position c.4347-236 in Rf1034.2, Rf1034.3, Rf1034.4, and Rf1034.5. C) RT-qPCR of *SMCHD1* normalized to *GUSB* in myotube cultures of control individuals (Controls, N=6), FSHD1 patients (FSHD1, N=7), FSHD2 patients (FSHD2, N=7) and patients hemizygous for *SMCHD1* (*SMCHD1* -/+, N=5). A myotube sample from Rf1034.5 is included in the FSHD2 group and is plotted as an asterisk. Analysis of control and *SMCHD1* -/+ RNA samples (but not including samples in the FSHD1 and FSHD2 groups) was published previously by Balog et al.(21) (*: P value <0.05, **: P value <0.01 – One-way ANOVA).



Supplementary figure S3: Genomic editing of *SMCHD1* intron 34 in HeLa cells. HeLa cells transfected with plasmids encoding sp.Cas9 and gRNA X50 or gRNA combination U/D3 were grown as monoclonal cultures and genotyped for editing on intron 34 by PCR (bottom panel). Unedited (N=5), heterozygous edited (N=5) and homozygous edited (N=6) clones were analysed by western blot for SMCHD1 and the housekeeping protein Tubulin (top panels). No Loss of SMCHD1 upon editing of *SMCHD1* intron 34 was observed.


A
Marker
U/D3 (1)
X50 (1)
U/D3 (2)
X50 (2)
Control
1000 bp
900 bp
800 bp
700 bp
600 bp
500 bp
400 bp
300 bp
200 bp
Altered transcript
Normal transcript
PCR: Exon 32 to 36
B
Marker
U/D3 (1)
X50 (1)
U/D3 (2)
X50 (2)
Control
1000 bp
900 bp
800 bp
700 bp
600 bp
500 bp
400 bp
300 bp
200 bp
Altered transcript
Normal transcript
PCR: Exon 30 to 37
C
Marker
U/D3-4.110
U/D3-4.158
U/D3-10.24
U/D3-4.13
U/D3-4.97
U/D3-4.100
U/D3-4.130
U/D3-4.153
U/D3-4.186
U/D3-4.191
X50-3.1
X50-3.2
X50-3.3
X50-3.4
X50-3.6
X50-3.7
X50-#3
Clones edited on SMCHD1-mutant allele
SMCHD1-Unedited patient clones
600 bp
500 bp
400 bp
300 bp
PCR: Exon 32 to 36
Altered transcript
Normal transcript
D
Marker
U/D3-4.110
U/D3-4.158
U/D3-10.24
U/D3-4.13
U/D3-4.97
U/D3-4.100
U/D3-4.130
U/D3-4.153
U/D3-4.186
U/D3-4.191
X50-3.1
X50-3.2
X50-3.3
X50-3.4
X50-3.6
X50-3.7
X50-#3
Clones edited on SMCHD1-mutant allele
SMCHD1-Unedited patient clones
1000 bp
900 bp
800 bp
700 bp
600 bp
PCR: Exon 30 to 37
Altered transcript
Normal transcript
E
Polyclonal cultures
DUX4_DUAL
ZSCAN4
NS (0.1143)
Normalized log expression
0.06
0.04
0.02
0.05
0.04
0.03
0.02
0.01
0.00
X50
U/D3
F
Monoclonal cultures
DUX4_DUAL
ZSCAN4
Normalized expression
0.5
0.4
0.3
0.2
0.1
0.0
0.6
Unedited
Edit-A
G
Clone 10.24 Myoblasts
Image on same position
Clone 10.24 Myotubes
Clone 4.158 Myotubes

Supplementary figure S4: Additional validation and measurements of Rf1034.5 cells edited by CRISPR-Cas9. A) RT-PCR analysis of *SMCHD1* transcripts spanning exon 32 through 36 in Rf1034.5 polyclonal myoblasts treated with Cas9 and control gRNA X50 (targeted against *AAVS1*) or treated with Cas9 and gRNAs U/D3, which cut upstream and downstream of the pseudo-exon, respectively. Image acquired after agarose gel electrophoresis shows the wild-type RT-PCR product and the altered RT-PCR product including the pseudo exon. B) RT-PCR analysis of *SMCHD1* transcripts spanning exon 30 through 37 in Rf1034.5 myoblasts treated with Cas9 and control gRNA X50 or treated with Cas9 and gRNAs U/D3. Image acquired after agarose gel electrophoresis shows the wild-type RT-PCR product and the RT-PCR product including the pseudo exon. C) RT-PCR analysis performed as in the experiments presented in Fig. S4A, for a representative sample of monoclonal Rf1034.5-iMB cultures used in Fig. 4., either edited or unedited on *SMCHD1* intron 34. D) RT-PCR analysis performed as in the experiments presented in Fig. S4B, for a representative sample of monoclonal Rf1034.5-iMB cultures used in Fig. 4., either edited or unedited on *SMCHD1* intron 34. E) Gene expression analysis on the polyclonal edited primary Rf1034.5 myoblasts. Depicted are *DUX4,* which could originate from the long or short 4A161 allele (DUAL) (22), and DUX4 target *ZSCAN4.* F) Gene expression analysis on the monoclonal cultures of *SMCHD1*-edited and *SMCHD1*-unedit Rf1034.5-iMB myoblasts. As in S4E, *DUX4*_DUAL and DUX4 target *ZSCAN4* are shown. G) Representative images (100× magnification, scale bar 50 µm) of proliferating myoblasts cultures and differentiated myotubes. The left two panels depict the same location of clone 10.24 before (left) and after differentiation (middle), arrowheads depict a few examples of multinucleated myotubes. Error bars: SEM. (*: P value <0.05, **: P value <0.01, ***: P value <0.001, ****: P value <0.0001, NS: Not-Significant – Mann-Whitney Test)



Supplementary figure S5: Bisulphite sequencing of the DR1 region in the D4Z4 sequence in *SMCHD1*-unedited Rf1034.5-iMB clones and Rf1034.5-iMB clones edited on the mutant *SMCHD1* allele. More than 10 independent clones were sequenced to assess changes in CpG methylation. No changes were observed when comparing two groups of four clones. (NS: Not-Significant – Mann-Whitney Test)

Supplementary table S1
Splice site predictions in *SMCHD1* using Alamut Visual, version 2.6. The different prediction methods output range is indicated between brackets, a higher score indicates a higher chance of splicing aberrations.

Alamut Visual version 2.6: **Prediction method (output)**	3'splice site c.4347-236A>G (Rf1034) Variant	5'splice site c.4347-183 (Rf1034) Cryptic site naturally present	3' splice site c.1843-15A>G (Rf744) Variant
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

Supplementary table S2
RT-PCR primers used in this study.

Primer name	Sequence (5' → 3')
SMCHD1 Ex12_1671F	TCCTAAGAAGAGAGGGCTTGC
SMCHD1 Ex16_2105R	TCATCTCCTTCAGGCCAAGT
SMCHD1 Ex14_2106R	TTTATGGCGATCATGATGGA
SMCHD1 4098F	AAAACCCGTTCGTCTCAATG
SMCHD1 4406R	TCCATCATAAAACCAAACTGGA
SMCHD1 ex3233F	GGGGTCTTTTCACTGATTTTATGA
SMCHD1-ex36R	TACTGGCAACTGAGCGAACA
SMCHD1-4062F	TCCAGCACCGGTACAACAT
SMCHD1-ex37R	TTCACGAAGGGGAATTCAAG

Supplementary table S3

Oligonucleotides used for gRNA assembly and corresponding target sites.

gRNA	Oligonucleotide code	Oligonucleotide sequence (5' → 3')	Target site sequence (5' → 3'), PAM underlined	Target site region
D3	D3 +	ACCGTGTACAGGTAAGAGGAACCT	TGTACAGGTAAGAGGAACCTGGG	*SMCHD1,* 10 nt downstream of pseudo exon
	D3 -	AAACAGGTTCCTCTTACCTGTACA		
UP3	UP3 -	ACCGATGTGGCTCATGATCCCTA	GATGTGGCTCATGATCCCTAGGG	*SMCHD1,* 345 nt upstream of pseudo exon
	UP3 +	AAACTAGGGATCATGAGCCACAT		
X50	#97	ACCGGGGCCACTAGGGACAGGAT	GGGGCCACTAGGGACAGGATTGG	*AAVS1*
	#98	AAACATCCTGTCCCTAGTGGCCC		

Supplementary table S4
Analysis output of the Cas-OFFinder algorithm for potential off-target effects of the guides targeting *SMCHD1* intron 34.

sgUP3(U3)	Chr.	Position	DNA sequence target	Dir.	Type	Mismatches
GATGTGGCTCATGATC-CCTANNN	chr18	2760065	GATGTGGCTCATGATC-CCTAGGG	-	Intronic	0 (target)
0 potential off-targets with 1 mismatch						
0 potential off-target with 2 mismatches						
4 potential off-targets with 3 mismatches:						
GATGTGGCTCATGATC-CCTANNN	chr1	162238011	GATGTGGCTCATGAT-gaCTgGGG	-	Intronic	3
GATGTGGCTCATGATC-CCTANNN	chr10	23720874	GAaGTGGCTCtTGtTC-CCTAAGG	+	Intergenic	3
GATGTGGCTCATGATC-CCTANNN	chr10	71823286	GATGTGGCTCtaGATC-CCaAAGG	+	Intronic	3
GATGTGGCTCATGATC-CCTANNN	chr9	81934773	GATGTGGCaCAgGATC-CCTgTGG	-	Intergenic	3
sgDOWN3(D3)						
TGTACAGGTAAGAG-GAACCTNNN	chr18	2760462	TGTACAGGTAAGAG-GAACCTGGG	+	Intronic	0 (target)
0 potential off-targets with 1 mismatch						
1 potential off-target with 2 mismatches						
8 potential off-targets with 3 mismatches:						
TGTACAGGTAAGAG-GAACCTNNN	chr18	66205833	TGTACAGGaAAGAG-GAAaCTAGG	+	Intergenic	2
TGTACAGGTAAGAG-GAACCTNNN	chr8	125318821	aGTACAGGgAAGgG-GAACCTAGG	+	lincRNA	3
TGTACAGGTAAGAG-GAACCTNNN	chr3	73018612	TGgAaAGGTA-AGAaGAACCTAGG	-	Intronic	3
TGTACAGGTAAGAG-GAACCTNNN	chr7	26476072	TGTACAGGTAAGAG-GAtCtcAGG	-	intronic	3
TGTACAGGTAAGAG-GAACCTNNN	chr7	97270799	TGTAaAGGgAAGAG-cAACCTTGG	-	intergenic	3
TGTACAGGTAAGAG-GAACCTNNN	chr1	241503116	TGTACtGGTcAGAG-GAAtCTGGG	-	intronic	3
TGTACAGGTAAGAG-GAACCTNNN	chr17	29092518	TGTACAGGaAAGAGG-gAgCTCGG	+	intronic	3
TGTACAGGTAAGAG-GAACCTNNN	chr6	132512990	gGTcCAGGTAAGAG-GAcCCTGGG	-	intergenic	3
TGTACAGGTAAGAG-GAACCTNNN	chr14	22326132	TGTACAGGgAAGtG-GAcCCTGGG	-	intronic	3

Supplementary table S5

Composition of 25 ml 2x CRISPR supplement

GABA	1.04 gram
5M NaCl	3.25 ml
100x Glutamine	375 µl
100x non-essential amino acids	375 µl
100x N2 supplement	375 µl
50x B27 supplement	750 µl
Optimem	Up to 25 ml
50 mg/ml EGF	2 µl/ml of above buffer
100 mg/ml bFGF	2 µl/ml of above buffer

Supplementary table S6

qPCR primers used in this study.

Target	Forward primer sequence (5' → 3')	Reverse primer sequence (5' → 3')
cDNA-GUSB	CCGAGTGAAGATCCCCTTTTTA	CTCATTTGGAATTTTGCCGATT
cDNA-RPL13	AACCTCCTCCTTTTCCAAGC	GCAGTACCTGTTTAGCCACGA
cDNA-MYOG	GCCAGACTATCCCCTTCCTC	GGGGATGCCCTCTCCTCTAA
cDNA-MYH3	CCTGCTGGAGGTGAAGTCTC	GATTGCAGGATCTGGTGGAT
cDNA-DUX4	TCCAGGAGATGTAACTCTAATCCA	CCCAGGTACCAGCAGACC
cDNA-DUX4 dual	CTTCCGTGAAATTCTGGCTGAATG	TTTTTTTTTTTTTTTTTCTATAGGATCCACAGG
cDNA-KHDC1L	TGAATCAGGTGGGAGCACAG	CAATGCAGCGAAGGTACGTG
cDNA-SMCHD1 wild type	GGGGTCTTTTCACTGATTTTATGA	CTTTATTAGGAATTACTTTATCCCTGAAAT
cDNA-SMCHD1 mutant	GGGGTCTTTTCACTGATTTTATGA	TCCAGAATAAACGTGCTGGAT
cDNA-ZSCAN4	TGGAAATCAAGTGGCAAAAA	CTGCATGTGGACGTGGAC
SRSF2 F1R1 (Intron Inclusion)	GTGTCCAAGAGGGAATCCAA	AGGAGACCGCAGCATTTTCT
SRSF2 F2R2 (Intron Exclusion)	GTGTCCAAGAGGGAATCCAA	TGCTTGCCGATACATCATTT
cDNA-SMCHD1 ex47-48	CGACAGATTGTCCAGTTCCTC	CCAATGGCCTCTTCTCTCTG
ChIP-DUX4-Q	CCGCGTCCGTCCGTGAAA	TCCGTCGCCGTCCTCGTC