HHS Public Access

Author manuscript

Nat Genet. Author manuscript; available in PMC 2013 June 04.

Published in final edited form as:

Nat Genet. 2012 December ; 44(12): 1370–1374. doi:10.1038/ng.2454.

Digenic inheritance of an *SMCHD1* mutation and an FSHD-permissive D4Z4 allele causes facioscapulohumeral muscular dystrophy type 2

Richard J.L.F. Lemmers[1,*], **Rabi Tawil**[2,*], **Lisa M. Petek**[3], **Judit Balog**[1], **Gregory J. Block**[3], **Gijs W.E. Santen**[4], **Amanda M. Amell**[3], **Patrick J. van der Vliet**[1], **Rowida Almomani**[4], **Kirsten R. Straasheijm**[1], **Yvonne D. Krom**[1], **Rinse Klooster**[1], **Yu Sun**[1], **Johan T. den Dunnen**[1,4], **Quinta Helmer**[5], **Colleen M. Donlin-Smith**[2], **George W. Padberg**[6], **Baziel G.M. van Engelen**[6], **Jessica C. de Greef**[1,$], **Annemieke M. Aartsma-Rus**[1], **Rune R. Frants**[1], **Marianne de Visser**[7], **Claude Desnuelle**[8], **Sabrina Sacconi**[8], **Galina N. Filippova**[9], **Bert Bakker**[4], **Michael J. Bamshad**[3], **Stephen J. Tapscott**[9,#], **Daniel G. Miller**[3,#], and **Silvère M. van der Maarel**[1,#]

[1]Department of Human Genetics, Leiden University Medical Center, Leiden, Netherlands [2]Neuromuscular Disease Unit, Department of Neurology, University of Rochester Medical Center, Rochester, NY, USA [3]Department of Pediatrics, University of Washington, Seattle, Washington, USA. Seattle Children's Hospital, Seattle, WA, USA [4]Department of Clinical Genetics, Leiden University Medical Center, Leiden, Netherlands [5]Department of Medical Statistics and Bioinformatics, Leiden University Medical Center, Leiden, Netherlands [6]Neuromuscular Centre Nijmegen, Department of Neurology, Radboud University Nijmegen Medical Centre, Nijmegen, Netherlands [7]Department of Neurology, Academic Medical Center, Amsterdam, Netherlands [8]Centre de référence des Maladies neuromusculaires and CNRS UMR6543, Nice University Hospital, Nice, France [9]Division of Human Biology, Fred Hutchinson Cancer Research Center, Seattle, WA, USA

Abstract

Facioscapulohumeral dystrophy (FSHD) is characterized by chromatin relaxation of the D4Z4 macrosatellite array on chromosome 4 and expression of the D4Z4-encoded *DUX4* gene in skeletal muscle. The more common form, autosomal dominant FSHD1, is caused by a contraction

[#]Corresponding authors: Silvère M. van der Maarel, Ph.D. Leiden University Medical Center, Department of Human Genetics, Albinusdreef 2, 2333 ZA Leiden, The Netherlands, Phone: +31 71 526 9480, Fax: +31 71 526 8285, Maarel@lumc.nl. Stephen J. Tapscott, M.D., Ph.D. Division of Human Biology, Fred Hutchinson Cancer Research Center, 1100 Fairview Avenue North, Seattle, WA 98109, Ph: 206-667-4499, FAX: 206-667-6524, stapscot@fhcrc.org. Daniel G. Miller, M.D., Ph.D. Department of Pediatrics, Division of Genetic Medicine, University of Washington, Box 358056, 850 Republican Street, N416, Seattle, WA 98195-8056, Phone: 206 685 3882, Fax : 206 685 1357, dgmiller@uw.edu.

[*]These authors contributed equally to this study

[$]Current address: Howard Hughes Medical Institute, Carver College of Medicine, University of Iowa, Iowa City, Iowa, USA

Author contributions

R.J.L.F.L, R.T, M.J.B, S.J.T, D.G.M. R.R.F., B.B., A.A.R. and S.M. conceived of and designed the study. D.G.M. and S.M. directed the study. G.W.E.S., Y.S, Q.H. and D.G.M. performed the bioinformatics data analysis. R.J.L.F.L., D.G.M., L.M.P., J.B., G.J.B., A.M.A., P.J.V., R.A., K.R.S., Y.D.K, R.K. and J.C.G. performed experiments. R.T., J.T.D., C.M.D.S, G.W.P., B.G.M.E., G.N.F., M.V., C.D., and S.S. contributed samples, reagents, data and comments on the manuscript. R.J.L.F.L, S.J.T., D.G.M. and S.M. analyzed and interpreted data and wrote the manuscript with the assistance and final approval from all authors.

of the D4Z4 array, whereas the genetic determinants and inheritance of D4Z4 array contraction-independent FSHD2 are unclear. Here we show that mutations in *SMCHD1 (structural maintenance of chromosomes flexible hinge domain containing 1*) on chromosome 18 reduce SMCHD1 protein levels and segregate with genome-wide D4Z4 CpG hypomethylation in human kindreds. FSHD2 occurs in individuals who inherited both the *SMCHD1* mutation and a normal-sized D4Z4 array on a chromosome 4 haplotype permissive for *DUX4* expression. Reducing SMCHD1 levels in skeletal muscle results in contraction-independent DUX4 expression. Our study identifies *SMCHD1* as an epigenetic modifier of the D4Z4 metastable epiallele and as a causal genetic determinant of FSHD2 and possibly other human diseases subject to epigenetic regulation.

FSHD [MIM158900] is clinically characterized by the initial onset of facial and upper-extremity muscle weakness that is often asymmetric and progresses to involve both upper and lower extremities[1]. FSHD1 and FSHD2 are phenotypically indistinguishable and both are associated with DNA hypomethylation and decreased repressive heterochromatin of the D4Z4 array, which we will collectively refer to as chromatin relaxation[2–8] (Supplementary Fig. 1). Each D4Z4 unit contains a copy of the *DUX4* (double homeobox 4) retrogene[9–13], a transcription factor expressed in the germline and epigenetically repressed in somatic tissues. The D4Z4 chromatin relaxation in FSHD results in inefficient epigenetic repression of *DUX4* and a variegated pattern of DUX4 protein expression in a subset of skeletal muscle nuclei[14] (Supplementary Fig. 1). Ectopic expression of DUX4 in skeletal muscle activates the expression of stem cell and germline genes[15] and when over-expressed in somatic cells DUX4 can ultimately lead to cell death[12,16–20]. Chromatin relaxation in FSHD1 is associated with a contraction of the array to 1–10 D4Z4 repeat units and therefore has a dominant inheritance pattern linked to the contracted array. In FSHD2, chromatin relaxation is independent of the size of the D4Z4 array and occurs on both chromosome-4 D4Z4 arrays and also on the highly homologous arrays on chromosome 10[2,7,8,21,22] (Supplementary Fig. 1).

D4Z4 chromatin relaxation must occur on a specific chromosome-4 haplotype in order to cause FSHD1 and FSHD2. This haplotype contains a polyadenylation (pA) signal to stabilize *DUX4* mRNA in skeletal muscle[13,23–27]. Chromosomes 4 and 10 that lack this pA signal fail to produce DUX4 protein; consequently, D4Z4 chromatin relaxation and transcriptional derepression on these non-permissive haplotypes does not lead to disease. Because chromatin relaxation occurs at both chromosome 4 and chromosome 10 D4Z4 repeats in FSHD2, we sought to determine whether an inherited defect in a modifier of D4Z4 repeat-mediated epigenetic repression might cause FSHD2 when combined with an FSHD-permissive *DUX4* allele.

To measure D4Z4 chromatin relaxation, we quantified the percentage of CpG methylation based on cleavage by the methylation sensitive *Fse*I endonuclease, an assay that averages the percentage D4Z4 methylation on both alleles of chromosomes 4 and 10 in a cohort of 72 controls, 93 FSHD1 patients and 53 FSHD2 patients. In FSHD2 affected individuals D4Z4 methylation was at least 2SD below the average levels in the general population (44+/−10% for the general population and 11+/−5% for FSHD2, **see** Fig. 1a, Supplementary Note and

Supplementary Fig. 2). Using a stringent methylation threshold of <25%, we discovered that in some kindreds identified by an FSHD2 proband D4Z4 hypomethylation segregated in a pattern consistent with autosomal dominant inheritance that was not linked to the chromosome-4 or -10 D4Z4 array haplotypes (Fig. 1b). In these kindreds, FSHD2 individuals inherited both the hypomethylation trait and the FSHD-permissive chromosome-4 haplotype with the *DUX4* pA signal, suggesting that two independently segregating loci cause and determine the penetrance of FSHD2.

In order to identify the locus controlling the D4Z4 hypomethylation trait, we performed whole exome sequencing[28] of twelve individuals in seven unrelated FSHD2 families: five with dominant segregation of the hypomethylation trait and two with sporadic hypomethylation and FSHD2). Detailed genetic analysis of the repeat lengths and haplotypes did not reveal evidence for non-paternity in these families (Fig. 1b). Families were stratified according to the criteria listed in Supplementary Table 1 and described in Supplementary Information. We identified rare and potentially pathogenic mutations in the *SMCHD1* (*Structural maintenance of chromosomes flexible hinge domain-containing 1)* gene in all individuals with D4Z4 hypomethylation with the exception of members of one family (Rf854: Table 1). These mutations were not present in public (dbSNP132 and the 1000 Genomes Project) or internal databases or in family members with normal D4Z4 methylation levels.

We confirmed the presence of these mutations by Sanger sequencing and included 12 additional unrelated FSHD2 families for which DNA or RNA was available. We identified heterozygous out-of-frame deletions, splice-site mutations, and heterozygous missense mutations in *SMCHD1* in 15/19 (79%) families (Table 1 and Fig. 1b). We also confirmed that the splice-site mutations altered the normal *SMCHD1* mRNA by exclusion of exons and cryptic splice site usage (Supplementary Fig. 4a,b).

Because heterozygous *SMCHD1* mutations co-segregated with D4Z4 hypomethylation in FSHD2 families or occurred *de novo* in sporadic hypomethylation/FSHD2 individuals (Fig. 1b), we considered SMCHD1 haploinsufficiency as a candidate disease mechanism, particularly since many of the mutations were predicted to affect the production of the full protein. Indeed, fibroblasts from FSHD2 patients with non-synonymous or splice-site mutations in *SMCHD1* had substantially reduced SMCHD1 protein levels (Fig. 2a). We found normal levels of SMCHD1 protein in the hypomethylated FSHD2 individual in family Rf854 that did not have an *SMCHD1* mutation (Fig. 2a), suggesting that FSHD2 in this family has a genetic cause other than SMCHD1 haploinsufficiency. Finally, chromatin immunoprecipitation (ChIP) demonstrated the presence of SMCHD1 on the D4Z4 array and reduced levels of this association in FSHD2 individuals with *SMCHD1* mutations (Fig. 2b). Together, these results support haploinsufficiency of SMCHD1 as a cause of D4Z4 hypomethylation in unrelated FSHD2 kindreds.

FSHD is characterized by low-level variegated expression of DUX4 in skeletal muscle. Therefore, we assessed *DUX4* expression in skeletal muscle cells from control individuals after decreasing SMCHD1 by RNA interference (Figs. 3a,b). We detected no *DUX4* mRNA in primary myotubes from an unaffected individual with a normal-sized and methylated

D4Z4 array on the FSHD-permissive *DUX4* pA haplotype. In contrast, *DUX4* was transcriptionally activated in these myotubes (Fig. 3c) when SMCHD1 transcripts and protein were suppressed to <50% of normal levels. We observed a variegated pattern of DUX4 protein in myotubes in all samples with adequate SMCHD1 knockdown (Fig. 3d); this pattern is similar to that seen in myotubes from FSHD2 patients. Cells expressing a scrambled or ineffective shRNA did not express DUX4 (Fig. 3, cont. and 4059).

In order to demonstrate that the *SMCHD1* splice mutations identified in FSHD2 patients result in *DUX4* expression, we manipulated *SMCHD1* pre-mRNA splicing in skeletal muscle cells using antisense oligonucleotides (AONs) directed to exon 29 or 36. These AONs induced skipping of *SMCHD1* exon 29 or 36 at rates comparable to those detected in some FSHD2 patients and resulted in transcription of *DUX4* (Fig. 3e,f). Thus, SMCHD1 activity is necessary for the somatic repression of *DUX4*, and reduction of this activity produces D4Z4 arrays that express DUX4 when an FSHD-permissive *DUX4* haplotype is present, with a pattern of variegated expression similar to that observed in FSHD1 and FSHD2 myotube cultures.

SMCHD1 belongs to the SMC gene superfamily that regulates chromatin repression of loci in many different organisms, including silencing mating loci in yeast[29], dosage compensation in *C. elegans*[30,31], position-effect variegation in *D. melanogaster*[32], and RNA-directed DNA methylation in *Arabidopsis*[33]. SMCHD1 was first identified in a mouse mutagenesis screen for modifiers of the variegated expression of a multi-copy transgene[34]. Gene targeting confirmed that Smchd1 was necessary for hypermethylation of (a subset of) CpG islands associated with X-inactivation, and continued association of the Smchd1 protein with the inactive X suggested its continuous requirement in maintaining X inactivation[35,36]. Our observations paint a strikingly similar picture for SMCHD1 and the D4Z4 arrays: SMCHD1 is necessary for D4Z4 hypermethylation, SMCHD1 remains associated with the D4Z4 array in skeletal muscle cells, and its continuous expression is required to maintain array silencing. It will be interesting to examine individuals with *SMCHD1* mutations for subclinical abnormalities of X-inactivation.

The *Smchd1* mutation was originally called the *Momme D1* (*Modifiers of Murine Metastable Epialleles D1*) locus[34]. The term metastable epiallele has been applied to genes that show variable expression because of probabilistic determinants of epigenetic repression[37]. An example of a metastable epiallele in mice is the *agouti viable yellow* (A^{vy}) locus; coat colors of isogenic mice can vary based on the epigenetic state of a retrotransposon integrated near the *agouti* promoter[38]. SmcHD1 is a modifier of metastable epialleles because SmcHD1 haploinsufficiency increased the penetrance of *agouti* expression[34]. In the case of FSHD, decreased levels of SMCHD1 resulted in decreased D4Z4 CpG methylation and variegated expression of DUX4 in myonuclei. In both FSHD1 and FSHD2, the penetrance is incomplete, and the presentation is often asymmetric. Out of the 26 hypomethylated individuals with a *SMCHD1* mutation and carrying a permissive D4Z4 haplotype, five of them are asymptomatic (19%) (Supplementary Table 2). This proportion of clinically unaffected carriers is remarkably similar to FSHD1[39] although a recent publication corroborates an earlier observation that non-penetrance may be much more frequent[40,41]. Thus, both features are consistent with FSHD as a metastable epiallele disease. Our

demonstration that independently variable modifiers of D4Z4 chromatin relaxation (repeat size for FSHD1 and SMCHD1 activity for FSHD2) modulate the variegated expression of DUX4, suggests that *DUX4* should be regarded as a metastable epiallele causing phenotypic variation in humans.

The disease mechanisms of FSHD1 and FSHD2 converge at the level of D4Z4 chromatin relaxation and the variegated expression of *DUX4*[14,15]. Both FSHD1 and FSHD2 require inheritance of two independent genetic variations: a version of the *DUX4* gene with a polyadenylation signal and a second genetic variant that results in D4Z4 chromatin relaxation. For FSHD1 the genetic variant associated with chromatin relaxation is contraction of the D4Z4 array and is therefore transmitted as a dominant trait. For FSHD2, mutations in *SMCHD1*, which is on chromosome 18, segregate independently of the FSHD-permissive *DUX4* allele on chromosome 4 and result in a digenic inheritance pattern in affected kindreds. Considering the variable clinical severity and asymmetric disease presentation, as well as the FSHD2 families without *SMCHD1* mutations, it is likely that other modifier loci will be identified that affect the chromatin structure of D4Z4. *SMCHD1* mutations could also modify the penetrance of FSHD1. Finally, many other human diseases show variable penetrance that might be related to epigenetic control. Our findings establish the possibility that *SMCHD1* mutations modify the epigenetic repression of other genomic regions and the penetrance of other human diseases.

URLs

SAMtools: http://samtools.sourceforge.net/;

Picard: http://picard.sourceforge.net/;

SeattleSeq Annotation: http://snp.gs.washington.edu/SeattleSeqAnnotation131/;

1000 Genomes: http://www.1000genomes.org/;

Alamut: http://www.interactive-biosoftware.com/;

Mutalyzer 2.0.beta-21: https://mutalyzer.nl/;

NCBI: http://www.ncbi.nlm.nih.gov/

Genecards: http://www.genecards.org/

Ensemble: http://www.ensembl.org/Homo_sapiens/Gene/

FSHD genotyping and methylation analysis protocols: http://www.urmc.rochester.edu/fields-center/

Referenced accessions numbers

NCBI NM_015295.2 (mRNA SMCHD1) and NP_056110.2 (protein SMCHD1)

Materials and Methods

Cases and samples

Forty-one FSHD2 patients were selected based on published clinical and molecular criteria[5,7,43,44] and D4Z4 methylation levels <25% as described in the previous section (Supplementary Table 1). Assessment of the FSHD2 phenotype was determined by experienced neurologists (RT, BGME, GWP, SS, CD, MV). Initial testing was performed using Pulsed Field Gel electrophoresis and hybridization of Southern blots with P13E-11, "A" and "B" probes, and SSLP length determined using an ABI Prism 3100 Genetic analyzer[41,45,46] according to protocols at the Fields Center for FSHD Research website. Forty of them had D4Z4 array sizes >10 units on both chromosomes 4 ruling out FSHD1. One patient had 2 contracted alleles on chromosome 10 possibly explaining the low D4Z4 methylation and was therefore excluded from further studies. Of the 39 remaining families, of 13 we had sufficient family information suggesting dominant inheritance of the D4Z4 hypomethylation and in 7 cases the hypomethylation appeared to have occurred *de novo* (Fig. 1b). For exome sequencing, we selected 5 families with a dominant inheritance pattern and 2 with *de novo* hypomethylation in the patient. In total 14 individuals from these families were analyzed by exome sequencing. All participants provided written consent, and the Institutional Review Boards of participating institutes approved all studies.

D4Z4 methylation analysis

Genomic DNA was double digested with *Eco*RI and *Bgl*II overnight at 37°C and cleaved DNA was purified using PCR extraction columns (supplementary note). Purified *Eco*RI/*Bgl*II digested DNA was digested with *Fse*I for 4 hours, separated by size on 0.8% agarose gels, transferred to a nylon membrane (Hybond XL, Amersham) by Southern blotting and probed using the p13E-11 radiolabeled probe[22]. Probe signals were quantified using a phosphorimager and Image Quant software. The signal from the total amount of hybridizing fragments 4061 bp (methylated fragments) and 3387 bp (unmethylated fraction) was divided by the signal quantity from the 4061 bp fragment to give the percentage of methylated *Fse*I sites within the most proximal D4Z4 unit (see supplementary note).

Exome definition, array design and target masking

We targeted all protein-coding regions as defined by RefSeq 36.3. Entries were filtered for the following: (i) CDS as the feature type, (ii) transcript name starting with "NM_" or "-", (iii) reference as the group_label, (iv) not being on an unplaced contig (for example, 17|NT_113931.1). Overlapping coordinates were collapsed for a total of 31,922,798 bases over 186,040 discontiguous regions. A single custom array (Agilent, 1M features, aCGH format) was designed to have probes over these coordinates as previously described, except here, the maximum melting temperature (T_m) was raised to 73 °C. The mappable exome was also determined as previously described using this RefSeq exome definition instead. After masking for 'unmappable' regions, 30,923,460 bases were left as the mappable target.

Targeted capture and massive parallel sequencing

Genomic DNA was extracted from peripheral blood lymphocytes using standard protocols. Five micrograms of DNA from each of the eight individuals was used for construction of a shotgun sequencing library as described previously using paired-end adaptors for sequencing on an Illumina Genome Analyzer II (GAII). Each shotgun library was hybridized to an array for target enrichment; this was then followed by washing, elution and additional amplification. Enriched libraries were then sequenced on a GAII to get either single-end or paired-end reads.

Read mapping and variant analysis

Reads were mapped and processed largely as previously described. In brief, reads were quality recalibrated using Eland and then aligned to the reference human genome (hg19) using Maq. When reads with the same start site and orientation were filtered, paired-end reads were treated like separate single-end reads; this method is overly conservative and hence the actual coverage of the exomes is higher than reported here. Sequence calls were performed using Maq and these calls were filtered to coordinates with ≥8× coverage and consensus quality ≥20.

Indels affecting coding sequences were identified as previously described, but we used phaster instead of cross_match and Maq. Specifically, unmapped reads from Maq were aligned to the reference sequence using phaster (version 1.100122a) with the parameters -max_ins:21 -max_del:21 -gapextend_ins:-1 -gapextend_del:-1 -match_report_type:1. Reads were then filtered for those with at most two substitutions and one indel. Reads that mapped to the negative strand were reverse complemented and, together with the other filtered reads, were remapped using the same parameters to reduce ambiguity in the called indel positions. These reads were then filtered for (i) having a single indel more than 3 bp from the ends and (ii) having no other substitutions in the read. Putative indels were then called per individual if they were supported by at least two filtered reads that started from different positions. An ‘indel reference’ was generated as previously described, and all the reads from each individual were mapped back to this reference using phaster with default settings and -match_report_type:1. Indel genotypes were called as previously described.

To determine the novelty of the variants, sequence calls were compared against 1200 individuals for whom we had previously reported exome data, and to the 1000 genomes database dbSNP. Annotations of variants were based on NCBI and UCSC databases using an in-house server (SeattleSeqAnnotation). Loss-of-function variants were defined as nonsense mutations (premature stop) or frame-shifting indels. For each variant, we also generated constraint scores as implemented in GERP.

Post hoc ranking of candidate genes

Candidate genes were ranked by summation of variant scores calculated by counting the total number of nonsense and nonsynonymous variants across the five FSHD2 exomes.

Mutation validation

Sanger sequencing of PCR amplicons (LGTC, Leiden, Netherlands) from genomic DNA was used to confirm the presence and identity of mutations in *SMCHD1* via exome sequencing and to screen the mutation in affected and unaffected family members of FSHD2 families.

Cells and culture conditions

Primary human myoblasts were obtained through the Fields Center at the University of Rochester (http://www.urmc.rochester.edu/fields-center/protocols/myoblast-cell-cultures.cfm). Biopsies were obtained after full consent with an IRB-approved protocol. Consents included the possibility of exome sequencing and sharing of samples with other investigators. Normal human myoblasts were grown on dishes coated with .01% Calf skin collagen (Sigma Aldrich, St. Louis, MI) in F10 medium (Invitrogen) supplemented with 20% FBS, 100U/ml penicillin and 100μg/ml streptomycin, 4μg/ml bFGF (Invitrogen), and 1 μM dexamethasone (Sigma Aldrich), in a humidified atmosphere containing 5% CO2 at 37°C[13]. Transduction of human myoblasts with retroviral vectors was accomplished by seeding cells at 5×10^4 cells/cm^2 density on day −1. On Day 0 the medium is changed and cells are incubated with vector preparations and polybrene (4 μg/ml, Sigma Aldrich). 2–4 hours later the medium is replaced with a fresh sample and cells are cultured and split at ~75% confluence to prevent differentiation. Human myoblasts transduced with pGIPZ shRNA expression vectors were selected with puromycin (0.5 μg/ml). Differentiation was induced using F10 medium supplemented with 1% horse serum and ITS supplement (insulin 0.1%, 0.000067% sodium selenite, 0.055% transferrin; Invitrogen).

Fibroblast obtained from FSHD2 patients and family members were cultured in DMEM/F-12 media supplemented with 20% heat inactivated fetal bovine serum, 1% penicillin/streptomycin, 10mM HEPES, 1mM Sodium Pyruvate (all Invitrogen).

RNA extraction and cDNA synthesis

Total RNA was extracted using the Qiagen miRNeasy mini isolation kit with DnaseI treatment. The RNA concentration was determined on a ND-1000 spectrophotometer (Thermo Scientific, Wilmington, USA) and the quality was analyzed with a RNA 6000 Nanochip Labchip on an Agilent 2100 BioAnalyzer (Agilent Technologies Netherlands BV, Amstelveen, The Netherlands). cDNA was synthesized from 2 μg of total RNA using random hexamer primers (Fermentas, St Leon-Rot, Germany) and the RevertAid H Minus M-MuLV First Strand Kit (Fermentas Life Sciences, Burlington, ON, Canada) according to the manufacturer's instructions. After the cDNA reaction 30 μL of water was added to an end volume of 50 μL.

Semiquantitative RNA analysis and sequencing of *SMCHD1* mutations

Splicing alterations were analyzed by RT-PCR using different primer sets covering the exons surrounding the possible splice site mutation. Subsequently, PCR fragments obtained from *SMCHD1* heterozygotes and control samples were analyzed on 1,5-2% agarose gels. Fragments were isolated from gel and analyzed by Sanger sequencing (LGTC).

Allelic expression analysis of missense mutations (wild type versus mutant allele) was done by Sanger sequencing (LGTC) by comparison of nucleotide peak heights of wild type and mutant alleles.

DUX4 mRNA levels were analyzed in duplo by real-time RT-PCR using SYBR Green QPCR master mix kit (Stratagene) on a MyiQ (Biorad Laboratories, Veenendaal, The Netherlands) running an initial denaturation step at 95°C for 6 min, followed by 40 cycles of 10 s at 95°C and 30 s at 60°C (35 cycles for *DUX4* RT-PCR samples shown in Fig. 3e,f). All PCR products were analyzed on a 2% agarose gel. Expression levels were corrected for *GAPDH* and *GUS* as constitutively expressed standard for cDNA input, and the relative steady-state RNA levels of the genes of interest were calculated by the method of Pfaffl[47]. All primers were designed using Primer 3 software and sequences are shown in Supplementary Table 3

Chromatin immunoprecipitation

Chromatin was prepared from myoblast cells lines fixed with 1% formaldehyde according to a published protocol[48]. Control and FSHD2 myoblast carried a comparable total number D4Z4 repeat units on permissive and nonpermissive chromosomes. 60ug chromatin was incubated with the different antibodies. Every sample was independently studied twice. Antibodies against SMCHD1 (ab31865) and H3 (ab1791) were purchased from Abcam (Cambridge, MA, USA). Normal rabbit serum was used to measure unspecific binding of proteins to beads. Immunopurified DNA was quantified with D4Z4 Q-PCR primer pair[8] and quantitative PCR measurements were done with CFX96TM real time system using iQTM SYBRR Green Supermix. Relative enrichment values were calculated by subtracting the IgG ChIP values representing background from the ChIP values with the SMCHD1 and H3 antibody and SMCHD1 values were divided by H3 enrichment values for D4Z4 copy number correction.

Antisense-mediated exon skipping

Antisense oligonucleotides (AONs) for *SMCHD1* exons 29 (29AON5) and 36 (36AON1) were designed based on the guidelines for Duchenne Muscular Dystrophy (DMD) exons (Supplementary Table 3)[49]. All AONs target exon-internal sequences and consist of 2′-O-methyl RNA with a full-length phosphorothioate backbone and were manufactured by Eurogentec (Seraing, Belgium). Human control myoblasts were seeded in 6 wells plates or 6 cm dishes at a cell density of approximately $1*10^4$ cells per cm^2 and cultured for 2 days. Myotubes were obtained by growing 70% confluence myoblasts for 4 days on differentiation media (DMEM (+glucose, +L-glutamin, +pyruvate), 2% horse serum). Four hours after the differentiation medium was added AONs were transfected at a 250nM concentration, using 2.5µl polyethyleneimine (MBI-Fermentas, Leon-Rot, Germany) per µg AON according to the manufacturer's instructions. A FAM-labeled AON targeting exon 50 of the *DMD* gene was used to confirm the efficiency of transfection and exon skipping. Primers flanking the targeted exons were used to study splicing of the *SMCHD1* or *DMD* gene.

Knock down of SMCHD1 mRNA's in normal human myoblasts

SMCHD1 transcripts were targeted for degradation using lentiviral vectors expressing short hairpin RNA's from a CMV promoter and linked to a puromycin selection cassette by an internal ribosome entry site (IRES). Five different pGIPZ (Open Biosystems, Huntsville, AL) vectors were purchased and each was tested in normal human myoblasts for the effect on SMCHD1 transcripts by quantitative PCR, immunofluorescence signal intensity, and western blot.

Antibodies, immunofluorescence and western blotting

Immunofluorescence for human DUX4 was performed using a rabbit monoclonal C-terminal specific antibody (Epitomics E5-5) as previously described[15]. Immunoreactivity was detected with a mouse anti-rabbit secondary antibody conjugated to Alexa Fluor 594 (Molecular Probes, 1:1000 dilution).

For western blotting, fibroblast or myoblast lysates were run on a 7.5% SDS-PAGE and transferred to PVDF membrane. SMCHD1 protein was detected using a commercially available rabbit polyclonal antibody (Sigma, HPA039441 (1:250 dilution)), and as reference protein Tubulin was detected with a commercially available mouse monoclonal antibody (Sigma, T6199 (1:2000)). Bound antibodies were detected with an HRP-conjugated donkey anti-rabbit (Pierce, 31458 (1:5000)) and an IRDye 800CW-conjugated goat anti-mouse antibody (Westburg, 926-32210 (1:5000)), respectively.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Acknowledgments

The authors thank all patients and family members for their participation. We thank Dr. Debbie Nickerson and Dr. Jay Shendure for excellent assistance, and Dr. Barbara Trask for helpful discussions and critical reading. This work was supported by grants from the NIH (NINDS P01NS069539; CTSA UL1RR024160; NIAMS R01AR045203; NHGRI HG005608 and HG006493), NGI Horizon Valorization Project Grant (Nr 93515504), The University of Washington Center for Mendelian Genomics, the MDA (217596), the Fields Center for FSHD Research, the Geraldi Norton and Eklund family foundation, the FSH Society, The Friends of FSH Research, EU FP7 framework program agreements 223026 (NMD-chip) and 223143 (TechGene), and the Stichting FSHD. Yu Sun is supported by China Scholarship Council.

References

1. Statland JM, Tawil R. Facioscapulohumeral muscular dystrophy: molecular pathological advances and future directions. Curr Opin Neurol. 2011; 24:423–428. [PubMed: 21734574]
2. Balog J, et al. Correlation analysis of clinical parameters with epigenetic modifications in the DUX4 promoter in FSHD. Epigenetics. 2012; 7:579–584. [PubMed: 22522912]
3. Bodega B, et al. Remodeling of the chromatin structure of the facioscapulohumeral muscular dystrophy (FSHD) locus and upregulation of FSHD-related gene 1 (FRG1) expression during human myogenic differentiation. BMC Biol. 2009; 7:41. [PubMed: 19607661]
4. Cabianca DS, et al. A Long ncRNA Links Copy Number Variation to a Polycomb/Trithorax Epigenetic Switch in FSHD Muscular Dystrophy. Cell. 2012; 149:819–831. [PubMed: 22541069]
5. de Greef JC, et al. Common epigenetic changes of D4Z4 in contraction-dependent and contraction-independent FSHD. Hum Mutat. 2009; 30:1449–1459. [PubMed: 19728363]

6. Jiang G, et al. Testing the position-effect variegation hypothesis for facioscapulohumeral muscular dystrophy by analysis of histone modification and gene expression in subtelomeric 4q. Hum Mol Genet. 2003; 12:2909–2921. [PubMed: 14506132]
7. van Overveld PG, et al. Hypomethylation of D4Z4 in 4q-linked and non-4q-linked facioscapulohumeral muscular dystrophy. Nat Genet. 2003; 35:315–317. [PubMed: 14634647]
8. Zeng W, et al. Specific loss of histone H3 lysine 9 trimethylation and HP1gamma/cohesin binding at D4Z4 repeats is associated with facioscapulohumeral dystrophy (FSHD). PLoS Genet. 2009; 5:e1000559. [PubMed: 19593370]
9. Gabriels J, et al. Nucleotide sequence of the partially deleted D4Z4 locus in a patient with FSHD identifies a putative gene within each 3.3 kb element. Gene. 1999; 236:25–32. [PubMed: 10433963]
10. Hewitt JE, et al. Analysis of the tandem repeat locus D4Z4 associated with facioscapulohumeral muscular dystrophy. Hum Mol Genet. 1994; 3:1287–1295. [PubMed: 7987304]
11. Lyle R, Wright TJ, Clark LN, Hewitt JE. The FSHD-associated repeat, D4Z4, is a member of a dispersed family of homeobox-containing repeats, subsets of which are clustered on the short arms of the acrocentric chromosomes. Genomics. 1995; 28:389–397. [PubMed: 7490072]
12. Snider L, et al. RNA transcripts, miRNA-sized fragments and proteins produced from D4Z4 units: new candidates for the pathophysiology of facioscapulohumeral dystrophy. Hum Mol Genet. 2009; 18:2414–2430. [PubMed: 19359275]
13. Snider L, et al. Facioscapulohumeral Dystrophy: Incomplete Suppression of a Retrotransposed Gene. PLoS Genet. 2010; 6:e1001181. [PubMed: 21060811]
14. van der Maarel SM, Tawil R, Tapscott SJ. Facioscapulohumeral muscular dystrophy and DUX4: breaking the silence. Trends Mol Med. 2011; 17:252–258. [PubMed: 21288772]
15. Geng LN, et al. DUX4 activates germline genes, retroelements, and immune mediators: implications for facioscapulohumeral dystrophy. Dev Cell. 2012; 22:38–51. [PubMed: 22209328]
16. Bosnakovski D, et al. An isogenetic myoblast expression screen identifies DUX4-mediated FSHD-associated molecular pathologies. EMBO J. 2008; 27:2766–2779. [PubMed: 18833193]
17. Kowaljow V, et al. The DUX4 gene at the FSHD1A locus encodes a pro-apoptotic protein. Neuromuscul Disord. 2007; 17:611–623. [PubMed: 17588759]
18. Vanderplanck C, et al. The FSHD atrophic myotube phenotype is caused by DUX4 expression. PLoS One. 2011; 6:e26820. [PubMed: 22053214]
19. Wallace LM, et al. DUX4, a candidate gene for facioscapulohumeral muscular dystrophy, causes p53-dependent myopathy in vivo. Ann Neurol. 2010; 69:540–552. [PubMed: 21446026]
20. Wuebbles RD, Long SW, Hanel ML, Jones PL. Testing the effects of FSHD candidate gene expression in vertebrate muscle development. Int J Clin Exp Pathol. 2010; 3:386–400. [PubMed: 20490329]
21. van Deutekom JC, et al. FSHD associated DNA rearrangements are due to deletions of integral copies of a 3.2 kb tandemly repeated unit. Hum Mol Genet. 1993; 2:2037–2042. [PubMed: 8111371]
22. Wijmenga C, et al. Chromosome 4q DNA rearrangements associated with facioscapulohumeral muscular dystrophy. Nat Genet. 1992; 2:26–30. [PubMed: 1363881]
23. Dixit M, et al. DUX4, a candidate gene of facioscapulohumeral muscular dystrophy, encodes a transcriptional activator of PITX1. Proc Natl Acad Sci U S A. 2007; 104:18157–18162. [PubMed: 17984056]
24. Lemmers RJ, et al. Facioscapulohumeral muscular dystrophy is uniquely associated with one of the two variants of the 4q subtelomere. Nat Genet. 2002; 32:235–236. [PubMed: 12355084]
25. Lemmers RJ, et al. A Unifying Genetic Model for Facioscapulohumeral Muscular Dystrophy. Science. 2010; 329:1650–1653. [PubMed: 20724583]
26. Spurlock G, Jim HP, Upadhyaya M. Confirmation that the specific SSLP microsatellite allele 4qA161 segregates with fascioscapulohumeral muscular dystrophy (FSHD) in a cohort of multiplex and simplex FSHD families. Muscle Nerve. 2010; 42:820–821. [PubMed: 20928905]
27. Thomas NS, et al. A large patient study confirming that facioscapulohumeral muscular dystrophy (FSHD) disease expression is almost exclusively associated with an FSHD locus located on a 4qA-defined 4qter subtelomere. J Med Genet. 2007; 44:215–218. [PubMed: 16987949]

28. Bamshad MJ, et al. Exome sequencing as a tool for Mendelian disease gene discovery. Nat Rev Genet. 2011; 12:745–755. [PubMed: 21946919]
29. Bhalla N, Biggins S, Murray AW. Mutation of YCS4, a budding yeast condensin subunit, affects mitotic and nonmitotic chromosome behavior. Mol Biol Cell. 2002; 13:632–645. [PubMed: 11854418]
30. Lieb JD, Capowski EE, Meneely P, Meyer BJ. DPY-26, a link between dosage compensation and meiotic chromosome segregation in the nematode. Science. 1996; 274:1732–1736. [PubMed: 8939869]
31. Chuang PT, Albertson DG, Meyer BJ. DPY-27:a chromosome condensation protein homolog that regulates C. elegans dosage compensation through association with the X chromosome. Cell. 1994; 79:459–474. [PubMed: 7954812]
32. Dej KJ, Ahn C, Orr-Weaver TL. Mutations in the Drosophila condensin subunit dCAP-G: defining the role of condensin for chromosome condensation in mitosis and gene expression in interphase. Genetics. 2004; 168:895–906. [PubMed: 15514062]
33. Kanno T, et al. A structural-maintenance-of-chromosomes hinge domain-containing protein is required for RNA-directed DNA methylation. Nat Genet. 2008; 40:670–675. [PubMed: 18425128]
34. Blewitt ME, et al. An N-ethyl-N-nitrosourea screen for genes involved in variegation in the mouse. Proc Natl Acad Sci U S A. 2005; 102:7629–7634. [PubMed: 15890782]
35. Blewitt ME, et al. SmcHD1, containing a structural-maintenance-of-chromosomes hinge domain, has a critical role in X inactivation. Nat Genet. 2008; 40:663–669. [PubMed: 18425126]
36. Gendrel AV, et al. Smchd1-Dependent and -Independent Pathways Determine Developmental Dynamics of CpG Island Methylation on the Inactive X Chromosome. Dev Cell. 201210.1016/j.devcel.2012.06.011
37. Rakyan VK, Blewitt ME, Druker R, Preis JI, Whitelaw E. Metastable epialleles in mammals. Trends Genet. 2002; 18:348–351. [PubMed: 12127774]
38. Duhl DM, Vrieling H, Miller KA, Wolff GL, Barsh GS. Neomorphic agouti mutations in obese yellow mice. Nat Genet. 1994; 8:59–65. [PubMed: 7987393]
39. van der Maarel SM, Frants RR, Padberg GW. Facioscapulohumeral muscular dystrophy. Biochim Biophys Acta. 2007; 1772:186–194. [PubMed: 16837171]
40. Scionti I, et al. Facioscapulohumeral muscular dystrophy: new insights from compound heterozygotes and implication for prenatal genetic counselling. J Med Genet. 2012; 49:171–178. [PubMed: 22217918]
41. Lemmers RJ, et al. Specific sequence variations within the 4q35 region are associated with facioscapulohumeral muscular dystrophy. Am J Hum Genet. 2007; 81:884–894. [PubMed: 17924332]
42. Lemmers RJ, et al. Contractions of D4Z4 on 4qB subtelomeres do not cause facioscapulohumeral muscular dystrophy. Am J Hum Genet. 2004; 75:1124–1130. [PubMed: 15467981]
43. van Overveld PG, et al. Variable hypomethylation of D4Z4 in facioscapulohumeral muscular dystrophy. Ann Neurol. 2005; 58:569–576. [PubMed: 16178028]
44. de Greef JC, et al. Clinical features of facioscapulohumeral muscular dystrophy 2. Neurology. 2010; 75:1548–1554. [PubMed: 20975055]
45. Lemmers RJL, et al. Complete allele information in the diagnosis of facioscapulohumeral muscular dystrophy by triple DNA analysis. Ann Neurol. 2001; 50:816–9. [PubMed: 11761483]
46. Lemmers RJLF, et al. Worldwide population analysis of the 4q and 10q subtelomeres identifies only four discrete duplication events in human evolution. Am J Hum Genet. 2010; 86:364–377. [PubMed: 20206332]
47. Pfaffl MW. A new mathematical model for relative quantification in real-time RT-PCR. Nucleic Acids Res. 2012; 2001:e45.
48. Nelson JD, Denisenko O, Bomsztyk K. Protocol for the fast chromatin immunoprecipitation (ChIP) method. Nat Protoc. 2006; 1:179–185. [PubMed: 17406230]
49. Aartsma-Rus A. Overview on AON design. Methods Mol Biol. 2012; 867:117–129. [PubMed: 22454058]

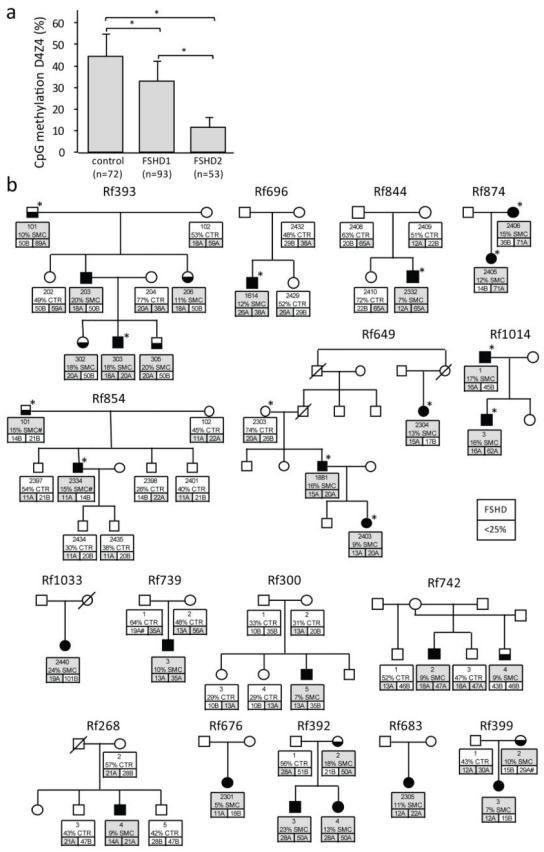

Fig. 1. D4Z4 methylation test and FSHD2 families

(a) *Fse*I methylation values of 72 control, 93 FSHD1 and 53 FSHD2 gDNA samples. Error bar represents standard deviation. FSHD2 patients are significantly hypomethylated by this test compared to controls and FSHD1 patients (*: p<0.005). **(b)** Pedigrees of FSHD2 families. For each individual in the upper box their ID, their *Fse*I methylation level (%) and whether they carry a *SMCHD1* mutation (SMC: grey) or not (CTR: white), is indicated. Also indicated in the lower two boxes are the lengths of both D4Z4 arrays on chromosomes 4 in units (U). Permissive alleles, typically A alleles based on a polymorphism distal to the repeat[24] are indicated in grey boxes. B alleles, which are non-permissive alleles[42] are indicated in white boxes. Some less common subtypes of the A allele are considered to be non-permissive[41], these are marked with an # and colored white (Rf399 and Rf739). Note the independent segregation of D4Z4 hypomethylation and FSHD-permissive alleles. Only in those individuals in whom a permissive allele combines with D4Z4 hypomethylation (<25%) was FSHD diagnosed, while D4Z4 hypomethylated individuals carrying non-permissive alleles were unaffected by FSHD. Individuals selected for whole exome sequencing (upper 7 pedigrees) are indicated by asterisks. SMC# indicates coding synonymous SNP identified in Rf854. Color key is shown in the figure.

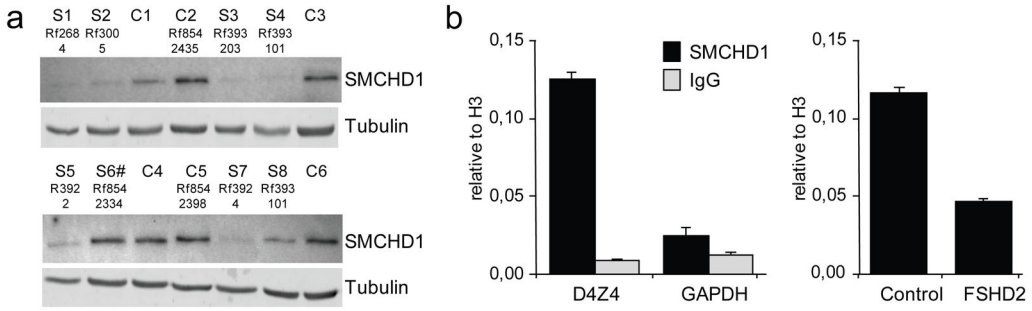

Fig. 2. FSHD2 families with *SMCHD1* mutations

(a) Western blot analysis of fibroblast cultures of 6 controls (C) and 8 individuals carrying a *SMCHD1* mutation (S). Sample identifiers refer to pedigrees in Fig. 1b and S6# denotes FSHD2 patient with only a synonymous coding SNP. **(b)** Bar diagram of ChIP analysis showing binding of SMCHD1 to D4Z4 but not to GAPDH (left panel) and reduced levels of SMCHD1 binding to D4Z4 in FSHD2 patient 2305 from family Rf683 (Fig. 1b). Error bars represent +/− 1 standard deviation of duplicate experiments.

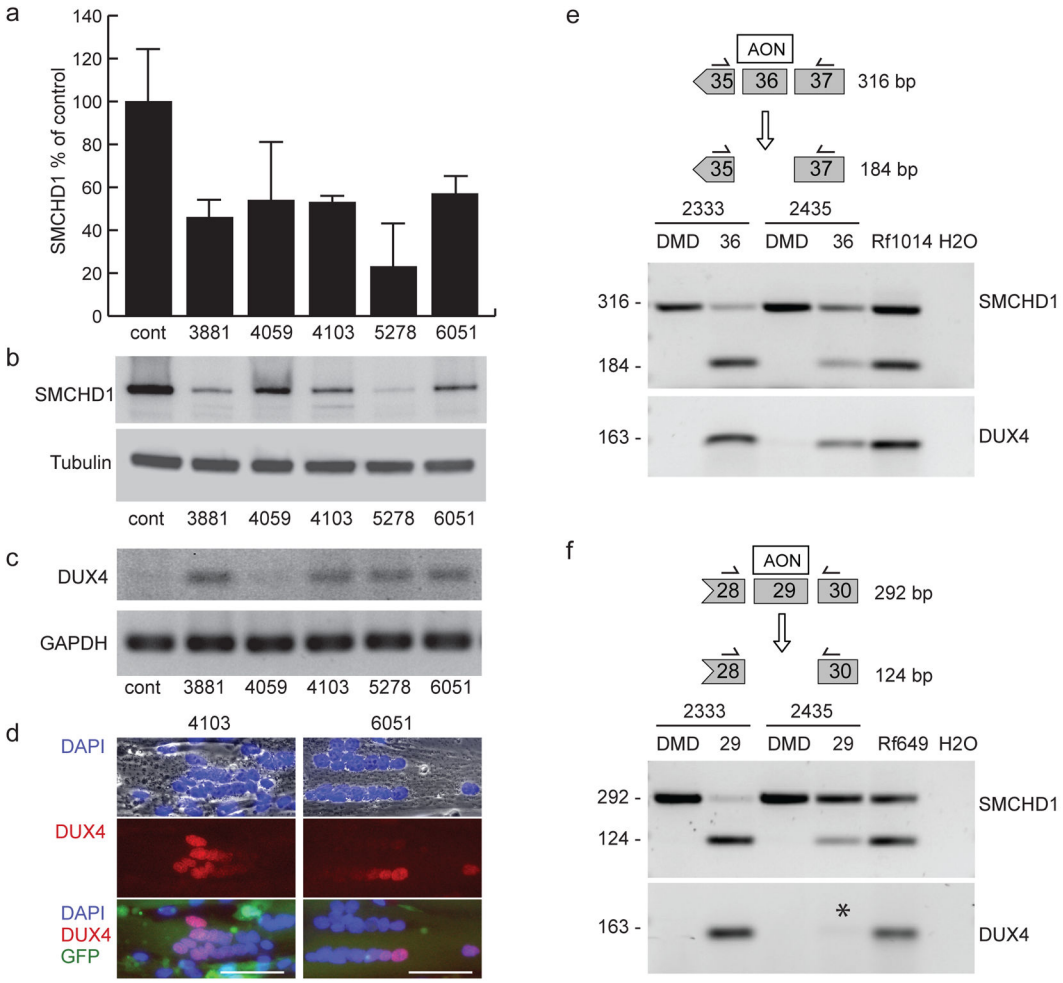

Fig. 3. SMCHD1 haploinsufficiency results in DUX4 expression in normal human myoblasts

(a) Short hairpin RNAs against different regions of *SMCHD1* are effective in reducing the production of SMCHD1 in normal human primary myoblasts on RNA and protein levels. Numbers below the graph and above the gel lanes indicate the regions within the *SMCHD1* transcript that are homologous to the indicated shRNA. *SMCHD1* mRNA levels were quantified by qRT-PCR and l normalized to *RNAse P* transcripts in a multiplexed reaction. Normalized *SMCHD1* levels are shown as a percentage of the levels found in the same cells treated with a vector expressing a scrambled sequence. Error bars show the standard deviation of the mean of three separate reactions. **(b)** Western blot of protein samples from the same cultures described in **a** normalized to tubulin. **(c)** Semi-quantitative RT-PCR analysis of *DUX4* in cells deficient for SMCHD1. *GAPDH* was amplified to demonstrate RNA integrity. **(d)** Examples of DUX4 immuno-reactive nuclei observed in tubes where SMCHD1 levels were reduced using shRNA 4103 or 6051. Myotubes are shown with nuclei labeled blue with DAPI and DUX4 (red). GFP fluorescence produced from the lentivirus vector expressing the shRNAs is also shown. Depicted scale bars are 50 μm in length. **(e)** AON-mediated exon skipping of *SMCHD1* exon 36 in normal human myoblasts 2333 and 2435. The mutation in family Rf1014 results in skipping of exon 36. Cells were treated with AONs designed to reproduce this skipping, and primers homologous to flanking exons

(shown above each gel) were used to evaluate the proportion of exon-skipped transcripts. The 184-bp fragment is produced when exon 36 is skipped. *DUX4* expression from the same cells is shown below each panel of *SMCHD1* exon analysis. Asterisk marks low *DUX4* expression levels consistent with inefficient *SMCHD1* exon skipping levels. Results are also shown for myotube RNA of affected individuals from both families. An AON targeting exon 50 of the *DMD* gene was used as a negative control. **(f)** Similar as panel e, AON-mediated exon skipping of *SMCHD1* exon 29 in normal human myoblasts 2333 and 2435. The mutation in family Rf649 results in skipping of exon 29, giving rise to the 124-bp fragment.

Table 1

Summary of *SMCHD1* mutations identified in this study showing family ID, mode of inheritance and mutation type

Column 4 shows the position of the mutation according to Supplementary Fig. 3. The position of the mutation in the *SMCHD1* gene is given with respect to chromosome, transcript and protein as well as a summary of the RNA analysis.

Family	Inheritance	Mutationtype	Nr	Position1	chromosome position2	transcript position3	protein position4	RNA analysis
Rf742	Maternal	missence	M1	exon 9	g.2697047A>G	c.1058A>G	p.Tyr353Cys	**
Rf1033	Unknown	deletion	D1	exon 10	g.2697999_2698003del	c.1302_1306del	p.Tyr434*	WT + mutant trancript*
Rf739	De novo	missence	M2	exon 11	g.2700630G>C	c.1436G>C	p.Arg479Pro	WT + mutant transcript
Rf300	De novo	missence	M3	exon 12	g.2700743T>C	c.1474T>C	p.Cys492Arg	WT + mutant transcript
Rf393	Paternal	deletion	D2	exon 12	g.2700875_2700875del	c.1608del	p.Asp537Ilefs*10	WT + mutant transcript*
Rf696	Unknown	5′ splice site	S1	intron12	g.2701019A>G	c.1647+103A>G		WT + skip exon 12* + cryptic splicing 12*
Rf399	Maternal	missence	M4	exon 16	g.2707565C>T	c.2068C>T	p.Pro690Ser	WT + mutant transcript
Rf268	Unknown	5′ splice site	S2	exon 20	g.2722661G>A	c.2603G>A	p.Ser868Asn	**
Rf844	De novo	5′ splice site	S3	intron 25	g.2732488_2732492del	c.3274_3276+2del		WT + exon 25 skip + cryptic splicing 25
Rf874	Maternal	5′ splice site	S3	intron 25	g.2732488_2732492del	c.3274_3276+2del		**
Rf854	Paternal	synonymous5	CS	exon 27	g.2739448T>A	c.3444T>A	p.Pro1148Pro	WT + mutant transcript
Rf649	Paternal	5′ splice site	S4	intron 29	g.2743927G>A	c.3801+1G>A		WT + cryptic splicing
Rf676	Unknown	5′ splice site	S4	intron 29	g.2743927G>A	c.3801+1G>A		**
Rf1014	Paternal	5′ splice site	S5	exon 36	g.2762234G>A	c.4566G>A	p.Thr1522Thr	WT + exon 36 skip
Rf392	Maternal	5′ splice site	S5	exon 36	g.2762234G>A	c.4566G>A	p.Thr1522Thr	WT + exon 36 skip + cryptic splicing 36
Rf683	Unknown	missence	M5	exon 37	g.2763729T>C	c.4661T>C	p.Phe1554Ser	WT + mutant transcript

[1] Exon number number is based on Ensembl transcript ENST00000320876

[2] Genomic position is based on hg19

[3] Transcript position is based on NM_015295.2

[4] Protein position is based on NP_056110.2

[5] Present at frequency 0.0055 in 1000 Genomes

* Disruption open reading frame

** No RNA available