nature genetics

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

Richard J L F Lemmers[1,13], Rabi Tawil[2,13], 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[1,12], Annemieke M Aartsma-Rus[1], Rune R Frants[1], Marianne de Visser[7], Claude Desnuelle[8,9], Sabrina Sacconi[8,9], Galina N Filippova[10], Bert Bakker[4], Michael J Bamshad[3,11], Stephen J Tapscott[10], Daniel G Miller[3,11] & Silvère M van der Maarel[1]

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 contraction 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* (encoding 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 D4Z4 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 (MIM 158900) 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 at the D4Z4 macrosatellite array, which we collectively refer to as chromatin relaxation[2–8] (**Supplementary Fig. 1**). Each D4Z4 unit encodes a copy of the *DUX4* retrogene (encoding double homeobox 4)[9–13], a transcription factor expressed in the germ line that is epigenetically repressed in somatic tissues. 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 overexpressed in somatic cells, DUX4 can ultimately lead to cell death[12,16–20]. Chromatin relaxation in FSHD1 is associated with contraction of the array to 1–10 D4Z4 repeat units and 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 at both chromosome 4 D4Z4 arrays and at the highly homologous arrays on chromosome 10 (**Supplementary Fig. 1**)[2,7,8,21,22].

D4Z4 chromatin relaxation must occur on a specific chromosome 4 haplotype to cause FSHD1 and FSHD2. This haplotype contains a polyadenylation signal to stabilize *DUX4* mRNA in skeletal muscle[13,23–27]. Chromosomes 4 and 10 that lack this polyadenylation signal do not produce DUX4 protein; consequently, D4Z4 chromatin relaxation and transcriptional derepression on these nonpermissive haplotypes does not lead to disease. Because chromatin relaxation occurs at D4Z4 repeats on both chromosomes 4 and 10 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.

[1]Department of Human Genetics, Leiden University Medical Center, Leiden, The Netherlands. [2]Neuromuscular Disease Unit, Department of Neurology, University of Rochester Medical Center, Rochester, New York, USA. [3]Department of Pediatrics, University of Washington, Seattle, Washington, USA. [4]Department of Clinical Genetics, Leiden University Medical Center, Leiden, The Netherlands. [5]Department of Medical Statistics and Bioinformatics, Leiden University Medical Center, Leiden, The Netherlands. [6]Neuromuscular Centre Nijmegen, Department of Neurology, Radboud University Nijmegen Medical Centre, Nijmegen, The Netherlands. [7]Department of Neurology, Academic Medical Center, Amsterdam, The Netherlands. [8]Centre de Référence des Maladies Neuromusculaires, Nice University Hospital, Nice, France. [9]Centre National de la Recherche Scientifique (CNRS) Unité Mixte de Recherche (UMR) 7277, Nice University Hospital, Nice, France. [10]Division of Human Biology, Fred Hutchinson Cancer Research Center, Seattle, Washington, USA. [11]Department of Genome Sciences, University of Washington, Seattle, Washington, USA. [12]Present addresses: Carver College of Medicine, University of Iowa, Iowa City, Iowa, USA, and Howard Hughes Medical Institute, Iowa City, Iowa, USA. [13]These authors contributed equally to this work. Correspondence should be addressed to S.M.v.d.M. (maarel@lumc.nl), D.G.M. (dgmiller@uw.edu) or S.J.T. (stapscot@fhcrc.org).

Received 24 May; accepted 4 October; published online 11 November 2012; doi:10.1038/ng.2454

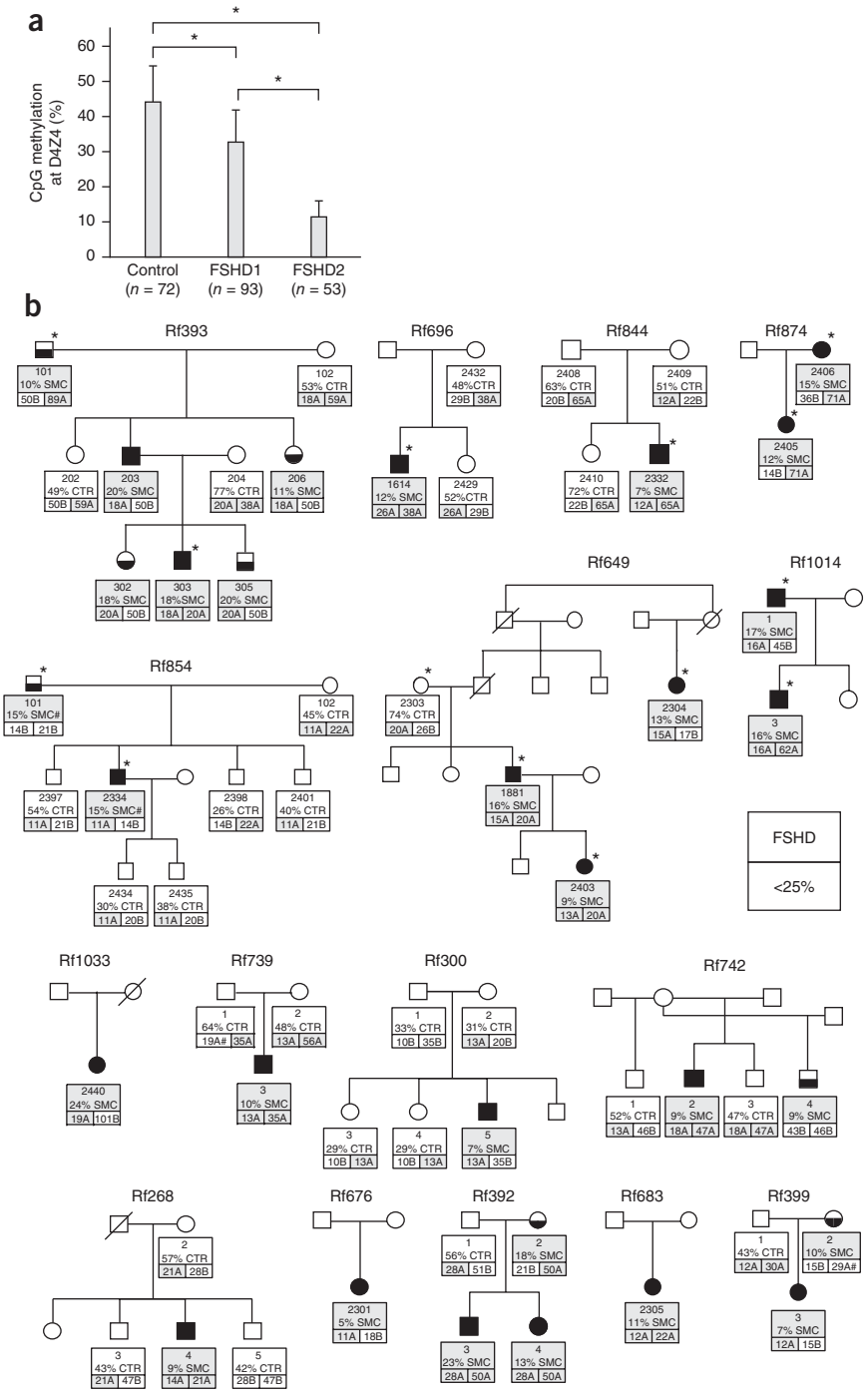

Figure 1 D4Z4 methylation test and families with FSHD2. (**a**) FseI methylation levels in 72 control, 93 FSHD1 and 53 FSHD2 genomic DNA samples. Error bars, s.d. $*P < 0.005$. (**b**) Pedigrees of families with FSHD2. For each individual, the subject ID, FseI-detected methylation level (%) and presence (SMC, gray) or absence (CTR, white) of a *SMCHD1* mutation is indicated (upper box). Also indicated in the lower two boxes are the lengths of both D4Z4 arrays on chromosomes 4 in units. Permissive alleles, typically A alleles defined on the basis of a polymorphism distal to the repeat[24], are indicated by gray boxes. B alleles, which are nonpermissive[42], are indicated by white boxes. Some less common subtypes of the A allele are considered to be nonpermissive[41]; these are marked with # 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, whereas individuals with D4Z4 hypomethylation carrying nonpermissive alleles were unaffected by FSHD. Individuals selected for whole-exome sequencing (upper seven pedigrees) are indicated by asterisks. SMC# indicates a coding synonymous SNP identified in Rf854. A key for the symbols is included.

To quantify D4Z4 chromatin relaxation, we determined the percentage of CpG methylation on the basis of measurements following cleavage with the methylation-sensitive FseI endonuclease, in an assay that averaged the percentage of D4Z4 methylation on both alleles of chromosomes 4 and 10 in a cohort of 72 controls, 93 individuals with FSHD1 and 53 individuals with FSHD2. In FSHD2-affected individuals, D4Z4 methylation was at least 2 s.d. below the average levels in the general population (44% ± 10% for the general population and 11 ± 5% for individuals with FSHD2; **Fig. 1a**, **Supplementary Fig. 2** and **Supplementary Note**). Using a stringent methylation threshold of <25%, we discovered that, in some kindreds identified by a proband with FSHD2, D4Z4 hypomethylation segregated in a pattern consistent with autosomal dominant inheritance that was not linked to the chromosome 4 or 10 D4Z4 array haplotype (**Fig. 1b**). In these kindreds, individuals with FSHD2 inherited both the hypomethylation trait and the FSHD-permissive chromosome 4 haplotype with the *DUX4* polyadenylation signal, suggesting that two independently segregating loci cause and determine the penetrance of FSHD2.

To identify the locus controlling the D4Z4 hypomethylation trait, we performed whole-exome sequencing[28] of 14 individuals in 7 unrelated families with FSHD2: 5 with dominant segregation of the hypomethylation trait and 2 with sporadic hypomethylation and FSHD2. Detailed genetic analysis of the repeat lengths and haplotypes did not provide evidence of non-paternity in these families (**Fig. 1b**). Families were stratified according to the criterion that the D4Z4 methylation level had to be <25%, not as a result of contracted repeats on chromosomes 4 and 10 (**Supplementary Table 1** and **Supplementary Note**). We identified rare and potentially pathogenic mutations in the *SMCHD1* gene (encoding structural maintenance of chromosomes flexible hinge domain containing 1) 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 in house 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 families with FSHD2 from whom DNA or RNA was available. We identified heterozygous

Table 1 SMCHD1 mutations identified

Family	Inheritance	Mutation type	Nr[a]	Position[b]	Chromosome position[c]	Transcript position[d]	Protein position[e]	RNA analysis
Rf742	Maternal	Missense	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 transcript[g]
Rf739	*De novo*	Missense	M2	Exon 11	g.2700630G>C	c.1436G>C	p.Arg479Pro	WT + mutant transcript[g]
Rf300	*De novo*	Missense	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[g]
Rf696	Unknown	5′ splice site	S1	Intron12	g.2701019A>G	c.1647+103A>G		WT + skipped exon 12[g] + cryptic splicing of exon12[g]
Rf399	Maternal	Missense	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 + skipped exon 25 + cryptic splicing of exon 25
Rf874	Maternal	5′ splice site	S3	Intron 25	g.2732488_2732492del	c.3274_3276+2del		–
Rf854	Paternal	Synonymous	CS	Exon 27	g.2739448T>A[f]	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 + skipped exon 36
Rf392	Maternal	5′ splice site	S5	Exon 36	g.2762234G>A	c.4566G>A	p.Thr1522Thr	WT + skipped exon 36 + cryptic splicing of exon 36
Rf683	Unknown	Missense	M5	Exon 37	g.2763729T>C	c.4661T>C	p.Phe1554Ser	WT + mutant transcript

–, no RNA available; WT, wild type.
[a]The position of each mutation is shown according to mutation number (Nr) in **Supplementary Figure 3**. [b]Exon number is based on Ensembl transcript ENST00000320876. [c]Genomic position is based on hg19. [d]Transcript position is based on NM_015295.2. [e]Protein position is based on NP_056110.2. [f]Present at frequency of 0.0055 in 1000 Genomes Project data. [g]Disrupts ORF.

out-of-frame deletions, heterozygous splice-site mutations and heterozygous missense mutations in *SMCHD1* in 15 out of 19 families (79%; **Fig. 1b**, **Table 1** and **Supplementary Fig. 3**). We also confirmed that the splice-site mutations altered normal *SMCHD1* mRNA by excluding exons and introducing the usage of cryptic splice sites (**Supplementary Fig. 4a**,**b**).

Because heterozygous *SMCHD1* mutations cosegregated with D4Z4 hypomethylation in families with FSHD2 or occurred *de novo* in individuals with sporadic hypomethylation and FSHD2 (**Fig. 1b**), we considered *SMCHD1* haploinsufficiency to be a candidate disease mechanism, particularly because many of the mutations were predicted to affect production of the full-length protein. Indeed, fibroblasts from individuals with FSHD2 who had nonsynonymous or splice-site mutations in *SMCHD1* expressed substantially lower amounts of SMCHD1 protein relative to control individuals (**Fig. 2a**). We found normal levels of SMCHD1 protein in the individual with FSHD2 with hypomethylated D4Z4 in family Rf854 who 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) showed the presence of SMCHD1 on the D4Z4 array and detected lower levels of this association in individuals with FSHD2 who had *SMCHD1* mutations (**Fig. 2b**). Taken together, these results support haploinsufficiency of *SMCHD1* as a cause of D4Z4 hypomethylation in unrelated kindreds with FSHD2.

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* levels by RNA interference (**Fig. 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* polyadenlated haplotype. In contrast, *DUX4* was transcriptionally activated in these myotubes (**Fig. 3c**) when SMCHD1 transcripts and protein amounts were reduced to <50% of normal. We observed a variegated pattern of DUX4 protein expression in myotubes in all samples with adequate *SMCHD1* knockdown (**Fig. 3d**); this pattern was similar to that seen in myotubes from individuals with FSHD2. Cells expressing a scrambled or ineffective short hairpin RNA (shRNA) did not express DUX4 (**Fig. 3**, control and 4059).

To show that the *SMCHD1* splice-site mutations identified in individuals with FSHD2 result in *DUX4* expression, we manipulated *SMCHD1* pre-mRNA splicing in skeletal muscle cells using antisense oligonucleotides directed against exon 29 or 36. These antisense oligonucleotides caused skipping of *SMCHD1* exon 29 or 36 at rates comparable to those detected in some individuals with FSHD2 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 results in 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 in many different organisms, mediating the silencing of mating loci in yeast[29], dosage compensation in *Caenorhabditis elegans*[30,31], position effect variegation in *Drosophila melanogaster*[32]

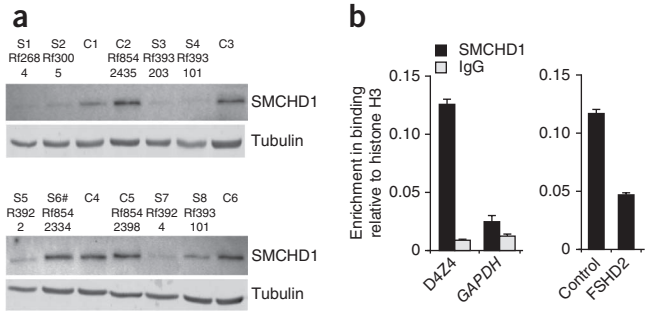

Figure 2 Families with FSHD2 with *SMCHD1* mutations. (**a**) Protein blot analysis of fibroblast cultures from six controls (C) and eight individuals carrying a *SMCHD1* mutation (S). Sample identifiers refer to the pedigrees in **Figure 1b**, and S6 denotes the individual with FSHD2 with only a synonymous coding SNP. (**b**) The results of ChIP analysis showing binding of SMCHD1 to D4Z4 arrays but not to *GAPDH* (left) and weaker binding of SMCHD1 to D4Z4 arrays in individual 2305 with FSHD2 from family Rf683 (right). Error bars, ± 1 s.d. from duplicate experiments. IgG, immunoglubulin G.

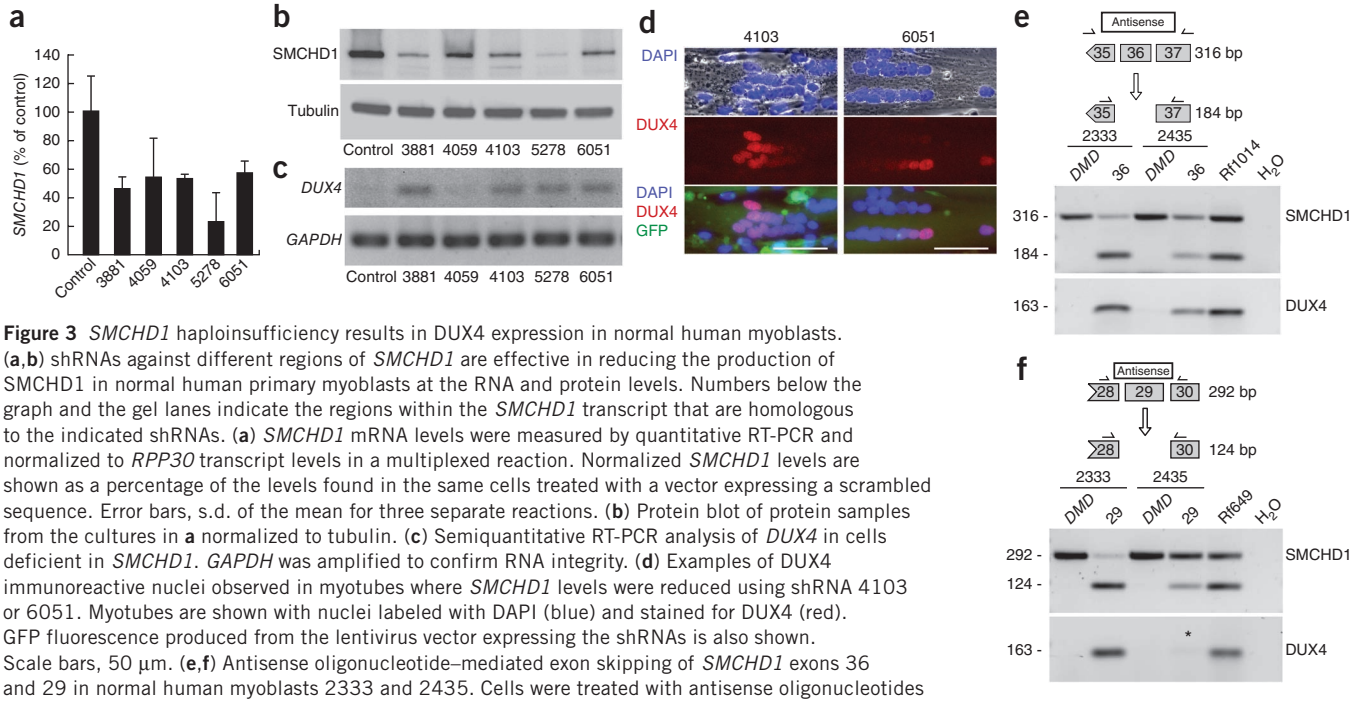

Figure 3 *SMCHD1* haploinsufficiency results in DUX4 expression in normal human myoblasts. (**a**,**b**) shRNAs against different regions of *SMCHD1* are effective in reducing the production of SMCHD1 in normal human primary myoblasts at the RNA and protein levels. Numbers below the graph and the gel lanes indicate the regions within the *SMCHD1* transcript that are homologous to the indicated shRNAs. (**a**) *SMCHD1* mRNA levels were measured by quantitative RT-PCR and normalized to *RPP30* transcript levels 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, s.d. of the mean for three separate reactions. (**b**) Protein blot of protein samples from the cultures in **a** normalized to tubulin. (**c**) Semiquantitative RT-PCR analysis of *DUX4* in cells deficient in *SMCHD1*. *GAPDH* was amplified to confirm RNA integrity. (**d**) Examples of DUX4 immunoreactive nuclei observed in myotubes where *SMCHD1* levels were reduced using shRNA 4103 or 6051. Myotubes are shown with nuclei labeled with DAPI (blue) and stained for DUX4 (red). GFP fluorescence produced from the lentivirus vector expressing the shRNAs is also shown. Scale bars, 50 μm. (**e**,**f**) Antisense oligonucleotide–mediated exon skipping of *SMCHD1* exons 36 and 29 in normal human myoblasts 2333 and 2435. Cells were treated with antisense oligonucleotides designed to reproduce this skipping, and primers homologous to flanking exons (shown above each gel) were used to evaluate the proportion of transcripts with skipped exons. *DUX4* expression from the same cells is shown below each panel of *SMCHD1* exon analysis. Results are also shown for myotube RNA from affected individuals in both families with the mutations. (**e**) An 184-bp fragment is produced when exon 36 is skipped. (**f**) An 124-bp fragment is produced when exon 29 is skipped. *, low *DUX4* expression levels consistent with inefficient *SMCHD1* exon skipping. An antisense oligonucleotide targeting exon 50 of the *DMD* gene (encoding dystrophin) was used as a negative control.

and RNA-directed DNA methylation in *Arabidopsis thaliana*[33]. SMCHD1 was first identified in a mouse mutagenesis screen for modifiers of the variegated expression of a multicopy transgene[34]. Gene targeting confirmed that *Smchd1* was necessary for hypermethylation of a subset of CpG islands associated with X-chromosome inactivation, and continued association of the Smchd1 protein with the inactive X chromosome suggested its continuous requirement in maintaining X-chromosome inactivation[35,36]. Our observations paint a similar picture of the role of 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 in X-chromosome inactivation.

The *Smchd1* mutation was originally called the *Momme D1* locus (encoding modifiers of murine metastable epialleles D1)[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 A^{vy} locus (encoding agouti viable yellow); coat colors of isogenic mice vary on the basis of the epigenetic state of a retrotransposon integrated near the *A* promoter[38]. SmcHD1 is a modifier of metastable epialleles, as *Smchd1* haploinsufficiency results in higher penetrance of agouti expression[34]. In the case of FSHD, lower levels of SMCHD1 resulted in lower 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 individuals with hypomethylation at D4Z4 with a *SMCHD1* mutation and carrying a permissive D4Z4 haplotype, 5 (19%) are asymptomatic (**Supplementary Table 2**). This proportion of clinically unaffected carriers is notably similar to that observed in FSHD1 (ref. 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 being a metastable epiallele–linked disease. Our demonstration that independently variable modifiers of D4Z4 chromatin relaxation (repeat size in FSHD1 and SMCHD1 activity in 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* (refs. 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 involves 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 from 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 families with FSHD2 without *SMCHD1* mutations, it is likely that other modifier loci will be identified that affect D4Z4 chromatin structure. *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 may 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 Project, 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/; Ensembl, http://www.ensembl.org/Homo_sapiens/Info/Index; FSHD genotyping and methylation analysis protocols, http://www.urmc.rochester.edu/fields-center/.

METHODS

Methods and any associated references are available in the online version of the paper.

Note: Supplementary information is available in the online version of the paper.

ACKNOWLEDGMENTS

The authors thank all subjects and family members for their participation. We thank D. Nickerson and J. Shendure for excellent assistance and B. Trask for helpful discussions and critical reading of the manuscript. This work was supported by grants from the US National Institutes of Health (NIH) (National Institute of Neurological Disorders and Stroke (NINDS) P01NS069539, Clinical & Translational Science Award (CTSA) UL1RR024160, National Institute of Arthritis and Musculoskeletal and Skin Diseases (NIAMS) R01AR045203 and National Human Genome Research Institute (NHGRI) HG005608 and HG006493), a Netherlands Genomics Initiative (NGI) Horizon Valorization Project Grant (93515504), The University of Washington Center for Mendelian Genomics, the Muscular Dystrophy Association (MDA; 217596), the Fields Center for FSHD Research, the Geraldi Norton and Eklund family foundation, the FSH Society, The Friends of FSH Research, European Union Framework Programme 7 agreements 223026 (NMD-chip), 223143 (TechGene) and 2012-305121 (NEUROMICS) and the Stichting FSHD. Y.S. is supported by the China Scholarship Council.

AUTHOR CONTRIBUTIONS

R.J.L.F.L., R.T., M.J.B., S.J.T., D.G.M., R.R.F., B.B., A.M.A.-R. and S.M.v.d.M. conceived of and designed the study. D.G.M., S.J.T. and S.M.v.d.M. directed the study. G.W.E.S., Y.S., Q.H. and D.G.M. performed the bioinformatics data analysis. R.T., B.G.M.v.E., G.W.P., S.S., C.D. and M.d.V. performed assessments of the FSHD2 phenotype. R.J.L.F.L., D.G.M., L.M.P., J.B., G.J.B., A.M.A., P.J.v.d.V., R.A., K.R.S., Y.D.K., R.K. and J.C.d.G. performed experiments. R.T., J.T.d.D., C.M.D.-S., G.W.P., B.G.M.v.E., G.N.F., M.d.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.v.d.M. analyzed and interpreted data, and wrote the manuscript with the assistance and final approval of all authors.

COMPETING FINANCIAL INTERESTS

The authors declare no competing financial interests.

Published online at http://www.nature.com/doifinder/10.1038/ng.2454.
Reprints and permissions information is available online at http://www.nature.com/reprints/index.html.

1. Statland, J.M. & Tawil, R. Facioscapulohumeral muscular dystrophy: molecular pathological advances and future directions. *Curr. Opin. Neurol.* **24**, 423–428 (2011).
2. Balog, J. *et al.* Correlation analysis of clinical parameters with epigenetic modifications in the *DUX4* promoter in FSHD. *Epigenetics* **7**, 579–584 (2012).
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.* **7**, 41 (2009).
4. Cabianca, D.S. *et al.* A long ncRNA links copy number variation to a polycomb/trithorax epigenetic switch in FSHD muscular dystrophy. *Cell* **149**, 819–831 (2012).
5. de Greef, J.C. *et al.* Common epigenetic changes of D4Z4 in contraction-dependent and contraction-independent FSHD. *Hum. Mutat.* **30**, 1449–1459 (2009).
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.* **12**, 2909–2921 (2003).
7. van Overveld, P.G. *et al.* Hypomethylation of D4Z4 in 4q-linked and non-4q–linked facioscapulohumeral muscular dystrophy. *Nat. Genet.* **35**, 315–317 (2003).
8. Zeng, W. *et al.* Specific loss of histone H3 lysine 9 trimethylation and HP1γ/cohesin binding at D4Z4 repeats is associated with facioscapulohumeral dystrophy (FSHD). *PLoS Genet.* **5**, e1000559 (2009).
9. Gabriëls, 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* **236**, 25–32 (1999).
10. Hewitt, J.E. *et al.* Analysis of the tandem repeat locus D4Z4 associated with facioscapulohumeral muscular dystrophy. *Hum. Mol. Genet.* **3**, 1287–1295 (1994).
11. Lyle, R., Wright, T.J., Clark, L.N. & Hewitt, J.E. 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* **28**, 389–397 (1995).
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.* **18**, 2414–2430 (2009).
13. Snider, L. *et al.* Facioscapulohumeral dystrophy: incomplete suppression of a retrotransposed gene. *PLoS Genet.* **6**, e1001181 (2010).
14. van der Maarel, S.M., Tawil, R. & Tapscott, S.J. Facioscapulohumeral muscular dystrophy and DUX4: breaking the silence. *Trends Mol. Med.* **17**, 252–258 (2011).
15. Geng, L.N. *et al.* DUX4 activates germline genes, retroelements, and immune mediators: implications for facioscapulohumeral dystrophy. *Dev. Cell* **22**, 38–51 (2012).
16. Bosnakovski, D. *et al.* An isogenetic myoblast expression screen identifies DUX4-mediated FSHD-associated molecular pathologies. *EMBO J.* **27**, 2766–2779 (2008).
17. Kowaljow, V. *et al.* The *DUX4* gene at the FSHD1A locus encodes a pro-apoptotic protein. *Neuromuscul. Disord.* **17**, 611–623 (2007).
18. Vanderplanck, C. *et al.* The FSHD atrophic myotube phenotype is caused by DUX4 expression. *PLoS ONE* **6**, e26820 (2011).
19. Wallace, L.M. *et al. DUX4*, a candidate gene for facioscapulohumeral muscular dystrophy, causes p53-dependent myopathy *in vivo*. *Ann. Neurol.* **69**, 540–552 (2011).
20. Wuebbles, R.D., Long, S.W., Hanel, M.L. & Jones, P.L. Testing the effects of FSHD candidate gene expression in vertebrate muscle development. *Int. J. Clin. Exp. Pathol.* **3**, 386–400 (2010).
21. van Deutekom, J.C. *et al.* FSHD associated DNA rearrangements are due to deletions of integral copies of a 3.2 kb tandemly repeated unit. *Hum. Mol. Genet.* **2**, 2037–2042 (1993).
22. Wijmenga, C. *et al.* Chromosome 4q DNA rearrangements associated with facioscapulohumeral muscular dystrophy. *Nat. Genet.* **2**, 26–30 (1992).
23. Dixit, M. *et al. DUX4*, a candidate gene of facioscapulohumeral muscular dystrophy, encodes a transcriptional activator of *PITX1*. *Proc. Natl. Acad. Sci. USA* **104**, 18157–18162 (2007).
24. Lemmers, R.J. *et al.* Facioscapulohumeral muscular dystrophy is uniquely associated with one of the two variants of the 4q subtelomere. *Nat. Genet.* **32**, 235–236 (2002).
25. Lemmers, R.J. *et al.* A unifying genetic model for facioscapulohumeral muscular dystrophy. *Science* **329**, 1650–1653 (2010).
26. Spurlock, G., Jim, H.P. & 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* **42**, 820–821 (2010).
27. Thomas, N.S. *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.* **44**, 215–218 (2007).
28. Bamshad, M.J. *et al.* Exome sequencing as a tool for Mendelian disease gene discovery. *Nat. Rev. Genet.* **12**, 745–755 (2011).
29. Bhalla, N., Biggins, S. & Murray, A.W. Mutation of *YCS4*, a budding yeast condensin subunit, affects mitotic and nonmitotic chromosome behavior. *Mol. Biol. Cell* **13**, 632–645 (2002).
30. Lieb, J.D., Capowski, E.E., Meneely, P. & Meyer, B.J. DPY-26, a link between dosage compensation and meiotic chromosome segregation in the nematode. *Science* **274**, 1732–1736 (1996).
31. Chuang, P.T., Albertson, D.G. & Meyer, B.J. DPY-27:a chromosome condensation protein homolog that regulates *C. elegans* dosage compensation through association with the X chromosome. *Cell* **79**, 459–474 (1994).
32. Dej, K.J., Ahn, C. & Orr-Weaver, T.L. Mutations in the *Drosophila* condensin subunit dCAP-G: defining the role of condensin for chromosome condensation in mitosis and gene expression in interphase. *Genetics* **168**, 895–906 (2004).
33. Kanno, T. *et al.* A structural-maintenance-of-chromosomes hinge domain–containing protein is required for RNA-directed DNA methylation. *Nat. Genet.* **40**, 670–675 (2008).
34. Blewitt, M.E. *et al.* An N-ethyl-N-nitrosourea screen for genes involved in variegation in the mouse. *Proc. Natl. Acad. Sci. USA* **102**, 7629–7634 (2005).
35. Blewitt, M.E. *et al.* SmcHD1, containing a structural-maintenance-of-chromosomes hinge domain, has a critical role in X inactivation. *Nat. Genet.* **40**, 663–669 (2008).
36. Gendrel, A.V. *et al.* Smchd1-dependent and -independent pathways determine developmental dynamics of CpG island methylation on the inactive X chromosome. *Dev. Cell* **23**, 265–279 (2012).
37. Rakyan, V.K., Blewitt, M.E., Druker, R., Preis, J.I. & Whitelaw, E. Metastable epialleles in mammals. *Trends Genet.* **18**, 348–351 (2002).
38. Duhl, D.M., Vrieling, H., Miller, K.A., Wolff, G.L. & Barsh, G.S. Neomorphic agouti mutations in obese yellow mice. *Nat. Genet.* **8**, 59–65 (1994).
39. van der Maarel, S.M., Frants, R.R. & Padberg, G.W. Facioscapulohumeral muscular dystrophy. *Biochim. Biophys. Acta* **1772**, 186–194 (2007).
40. Scionti, I. *et al.* Facioscapulohumeral muscular dystrophy: new insights from compound heterozygotes and implication for prenatal genetic counselling. *J. Med. Genet.* **49**, 171–178 (2012).
41. Lemmers, R.J. *et al.* Specific sequence variations within the 4q35 region are associated with facioscapulohumeral muscular dystrophy. *Am. J. Hum. Genet.* **81**, 884–894 (2007).
42. Lemmers, R.J. *et al.* Contractions of D4Z4 on 4qB subtelomeres do not cause facioscapulohumeral muscular dystrophy. *Am. J. Hum. Genet.* **75**, 1124–1130 (2004).

ONLINE METHODS

Study subjects and samples. Forty-one individuals with FSHD2 were selected on the basis of published clinical and molecular criteria[5,7,43,44] and because they had D4Z4 methylation levels of <25% (**Supplementary Table 1**). Assessment of the FSHD phenotype was performed by experienced neurologists. Initial testing was performed using pulsed-field gel electrophoresis and hybridization of Southern blots with P13E-11, A and B probes, and SSLP length was determined using an ABI Prism 3100 Genetic Analyzer[41,45,46] according to protocols at the Fields Center at the University of Rochester for FSHD Research website (see URLs). Forty of the affected individuals had D4Z4 array sizes of >10 units on both copies of chromosome 4, ruling out diagnosis with FSHD1. One affected individual had two contracted alleles on chromosome 10, possibly explaining the low D4Z4 methylation observed for this subject, and was therefore excluded from further studies. Of the 39 remaining affected families, we had sufficient family information for 13 suggesting dominant inheritance of D4Z4 hypomethylation, whereas the hypomethylation seemed to have occurred *de novo* in 7 affected individuals (**Fig. 1b**). For exome sequencing, we selected five families with a dominant inheritance pattern and two with *de novo* hypomethylation in the affected individual. In total, 14 individuals from these families were analyzed by exome sequencing. All participants provided written consent, and the institutional review boards (IRBs) of participating institutes approved all studies.

D4Z4 methylation analysis. Genomic DNA was double digested with EcoRI and BglII overnight at 37 °C, and cleaved DNA was purified using PCR extraction columns (**Supplementary Note**). Purified DNA digested with EcoRI and BglII was digested with FseI for 4 h, 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 ImageQuant software. The signal from the 4,061-bp fragment was divided by the total amount of hybridizing fragments at 4,061 bp (methylated) and 3,387 bp (unmethylated) to give the percentage of methylated FseI sites within the most proximal D4Z4 unit (**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 (i) CDS as the feature type, (ii) transcript name starting with 'NM_' or '-', (iii) reference as the group_label and (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, 1 million features, array–comparative genomic hybridization (aCGH) format) was designed to have probes over these coordinates as previously described, except that the maximum melting temperature (T_m) was raised to 73 °C.

The mappable exome was also determined as previously described, instead using the RefSeq 36.3 exome definition. After masking for 'unmappable' regions, 30,923,460 bases remained as the mappable target.

Targeted capture and massively parallel sequencing. Genomic DNA was extracted from peripheral blood lymphocytes using standard protocols. DNA (5 µg) 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 IIx (GAIIx). Each shotgun library was hybridized to an array for target enrichment, which was followed by washing, elution and additional amplification. Enriched libraries were then sequenced on a GAIIx to generate 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 as separate single-end reads; this method is overly conservative, and, hence, the actual coverage of the exomes was higher than reported here. Sequence calls were performed using MAQ, and these calls were filtered to coordinates with ≥8× coverage and consensus quality of ≥20.

Insertions and/or deletions (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 whether the variants were novel, sequence calls were compared against our previously reported exome data for 1,200 individuals and the 1000 Genomes Project database and dbSNP. Annotations of variants were made on the basis of information in the NCBI and UCSC databases using an in house server (SeattleSeq Annotation). Loss-of-function variants were defined as nonsense mutations (introduction of a premature stop codon) or frameshift indels. For each variant, we also generated constraint scores, as implemented in genomic evolutionary rate profiling (GERP).

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 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* that were initially detected via exome sequencing and to screen each mutation in the affected and unaffected family members in families with FSHD2.

Cells and culture conditions. Primary human myoblasts were obtained through the Fields Center. Biopsies were obtained after obtaining 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 0.01% calf skin collagen (Sigma-Aldrich) in F10 medium (Invitrogen) supplemented with 20% FBS, 100 U/ml penicillin and 100 µg/ml streptomycin, 4 µg/ml human basic fibroblast growth factor (bFGF) (Invitrogen) and 1 µM dexamethasone (Sigma-Aldrich) in a humidified atmosphere containing 5% CO_2 at 37 °C[13]. Transduction of human myoblasts with retroviral vectors was accomplished by seeding cells at a density of 5×10^4 cells/cm^2 on day −1. On day 0, the medium was changed, and cells were incubated with vector preparations and polybrene (4 µg/ml; Sigma-Aldrich). After 2–4 h, the medium was replaced with fresh medium, and cells were 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).

Fibroblasts obtained from individuals with FSHD2 and their family members were cultured in DMEM/F-12 medium supplemented with 20% heat-inactivated FBS, 1% penicillin-streptomycin, 10 mM HEPES and 1 mM sodium pyruvate (all from Invitrogen).

RNA extraction and cDNA synthesis. Total RNA was extracted using the Qiagen miRNeasy Mini isolation kit with DNase I treatment. The RNA concentration was determined on an ND-1000 spectrophotometer (ThermoScientific), and RNA quality was analyzed with an RNA 6000 Nanochip Labchip on an Agilent 2100 Bioanalyzer (Agilent Technologies Netherlands). cDNA was synthesized from 2 µg of total RNA using random hexamer primers (Fermentas) and the RevertAid H Minus M-MuLV First Strand kit (Fermentas Life Sciences) according to the manufacturer's instructions. After completion of cDNA synthesis, 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 generated from control samples and individuals heterozygous for *SMCHD1* mutations were analyzed on 1.5–2% agarose gels. Fragments were separated by size on agarose gels, purified and analyzed by Sanger sequencing (LGTC).

Allelic expression analysis of missense mutations (wild-type versus mutant alleles) was carried out with Sanger sequencing (LGTC) by comparison of the nucleotide peak heights of the wild-type and mutant alleles.

DUX4 mRNA levels were analyzed in duplicate by RT-PCR using the SYBR Green QPCR master mix kit (Stratagene) on a MyiQ (Bio-Rad), 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 the *DUX4* RT-PCR samples shown in **Fig. 3e**,**f**). All PCR products were analyzed on a 2% agarose gel. Expression levels were corrected by those of *GAPDH* and *GUSB*, constitutively expressed standards for cDNA input, and the relative steady-state RNA levels of the genes of interest were calculated by a previously described method[47]. All primers were designed using Primer3 software, and sequences are provided in **Supplementary Table 3**.

ChIP assays. Chromatin was prepared from myoblast cell lines fixed with 1% formaldehyde according to a published protocol[48]. Control and FSHD2 myoblasts carried a comparable total number D4Z4 repeat units on permissive and nonpermissive chromosomes. We incubated 60 µg of chromatin with the different antibodies. Every sample was independently studied twice. Antibodies against SMCHD1 (ab31865) and histone H3 (ab1791) were purchased from Abcam. Normal rabbit serum was used to measure unspecific binding of proteins to beads. Immunopurified DNA was quantified with the D4Z4 Q-PCR primer pair[8], and quantitative PCR measurements were performed with the CFX96 Real-Time PCR Detection System using iQ SYBRR Green Supermix. Relative enrichment values were calculated by dividing the ChIP values obtained with the antibodies to SMCHD1 or IgG by the ChIP values obtained with the antibodies to histone H3.

Antisense-mediated exon skipping. Antisense oligonucleotides for *SMCHD1* exons 29 (29AON5) and 36 (36AON1) were designed on the basis of the guidelines for *DMD* exons (**Supplementary Table 3**)[49]. All antisense oligonucleotides target exon-intern sequences, consist of 2′-O-methyl RNA with a full-length phosphorothioate backbone and were manufactured by Eurogentec. Human control myoblasts were seeded in 6-well plates or 6-cm dishes at a density of approximately 1×10^4 cells/cm^2 and were cultured for 2 d. Myotubes were obtained by growing myoblasts at 70% confluence for 4 d in differentiation medium (DMEM (with glucose, L-glutamine and pyruvate) supplemented with 2% horse serum). Cells were transfected with 250 nM concentrations of antisense oligonucleotides 4 h after the differentiation medium was added, using 2.5 µl of polyethyleneimine (MBI-Fermentas) per microgram of antisense oligonucleotide according to the manufacturer's instructions. A FAM-labeled antisense oligonucleotide 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* and *DMD* genes.

Knockdown of *SMCHD1* mRNA in normal human myoblasts. *SMCHD1* transcripts were targeted for degradation using lentiviral vectors expressing shRNAs from a CMV promoter linked to a puromycin selection cassette controlled by an internal ribosome entry site (IRES). Five different pGIPZ vectors (Open Biosystems) were purchased, and each was tested in normal human myoblasts for the effect on *SMCHD1* transcripts by quantitative PCR, immunofluorescence signal intensity and protein blot analysis.

Antibodies, immunofluorescence and protein blotting. Immunofluorescence for human DUX4 was performed using a rabbit monoclonal antibody specific to its C terminus (Epitomics, E5-5), as previously described[15]. Immunoreactivity was detected with a mouse Alexa Fluor 594–conjugated secondary antibody to rabbit (Molecular Probes; 1:1,000 dilution).

For protein blotting, fibroblast or myoblast lysates were separated by 7.5% SDS-PAGE and transferred to PVDF membrane. SMCHD1 protein was detected using a commercially available rabbit polyclonal antibody (Sigma-Aldrich, HPA039441; 1:250 dilution), and the reference protein tubulin was detected with a commercially available mouse monoclonal antibody (Sigma, T6199; 1:2,000 dilution). Bound antibodies were detected with a horseradish peroxidase (HRP)-conjugated donkey secondary antibody to rabbit (Pierce, 31458; 1:5,000 dilution) and an IRDye 800CW–conjugated goat secondary antibody to mouse (Westburg, 926-32210; 1:5,000 dilution), respectively.

43. van Overveld, P.G. *et al.* Variable hypomethylation of D4Z4 in facioscapulohumeral muscular dystrophy. *Ann. Neurol.* **58**, 569–576 (2005).
44. de Greef, J.C. *et al.* Clinical features of facioscapulohumeral muscular dystrophy 2. *Neurology* **75**, 1548–1554 (2010).
45. Lemmers, R.J.L. *et al.* Complete allele information in the diagnosis of facioscapulohumeral muscular dystrophy by triple DNA analysis. *Ann. Neurol.* **50**, 816–819 (2001).
46. Lemmers, R.J.L.F. *et al.* Worldwide population analysis of the 4q and 10q subtelomeres identifies only four discrete duplication events in human evolution. *Am. J. Hum. Genet.* **86**, 364–377 (2010).
47. Pfaffl, M.W. A new mathematical model for relative quantification in real-time RT-PCR. *Nucleic Acids Res.* **29**, e45 (2001).
48. Nelson, J.D., Denisenko, O. & Bomsztyk, K. Protocol for the fast chromatin immunoprecipitation (ChIP) method. *Nat. Protoc.* **1**, 179–185 (2006).
49. Aartsma-Rus, A. Overview on AON design. *Methods Mol. Biol.* **867**, 117–129 (2012).

doi:10.1038/ng.2454