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#], 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

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

* These authors contributed equally to this study

\# Corresponding authors

CONTENTS

Supplementary Note

Genomic DNA isolated from peripheral blood lymphocytes from a large panel of controls, sporadic patients with FSHD and FSHD families were included in this study after obtaining informed consent. The clinical diagnosis of FSHD was based on a standardized clinical form made available through the Fields Center: http://www.urmc.rochester.edu/fields-center/). For all individuals we performed a detailed genotyping, including D4Z4 repeat array length and chromosomal background analysis of chromosomes 4q and 10q.

The observation that in FSHD1 patients D4Z4 hypomethylation is restricted to the disease allele while in FSHD2 patients the repeats on all four chromosomes are affected provides a unique opportunity to develop a more sensitive and specific diagnostic test for FSHD2. Rather than separating the chromosome 4-derived fragments from the chromosome 10-derived fragments by using restriction enzyme *Bln*I, as done before[1, 2], a collective measurement of D4Z4 methylation on both chromosomes 4 and 10 should yield a more sensitive and specific diagnostic test for FSHD2. From our previous tests involving three methylation-sensitive restriction enzymes, *Fse*I was shown to be the most informative enzyme[1, 2]. Therefore, we redesigned the *Fse*I D4Z4 methylation test so that it interrogates all four alleles simultaneously by omitting *Bln*I from the digestion (Supplementary Fig. 1). Previously, we showed that the *Fse*I methylation value of the first D4Z4 unit in controls is ~50% on both chromosomes 4q[1, 2]. The average *Fse*I methylation level of the first unit in pathogenic chromosomes 4 in FSHD1 patients (n=21) was shown to be 20%[3], while in FSHD2 patients we found for both chromosomes 4 on average a value of 13% (n=32)[4]. While in controls and FSHD1 patients we would expect (near-) normal methylation values (as in FSHD1 the hypomethylation signal from the disease allele would be diluted 3x by the normal methylation levels of the normal chromosome 4 and chromosomes 10), in FSHD2 patients we would expect to see profound hypomethylation. As the activity of restriction enzymes is sensitive to salt or protein impurities in the gDNA we introduced an extra DNA clean-up step preceding digestion with *Fse*I (Supplementary Fig. 1a). This extraction column-based purification step can also be applied to gDNA embedded in agarose plugs and to samples with low gDNA concentrations.

Upon digesting with *Bgl*II a 4061 bp fragment is released (M in Supplementary Fig. 1c) while digesting with *Fse*I yields a fragment of 3387 bp when the restriction site is unmethylated (U in Supplementary Fig. 1c). The previously used enzyme *Bln*I to separate chromosomes 4 (white) from chromosomes 10 (black) is also shown.

To validate the modified methylation test, we re-analyzed the same gDNA samples from a previous study[4]. While we obtained nearly identical average methylation levels in all three populations analyzed, the modified methylation test clearly improves discrimination between FSHD1 and FSHD2 by reducing the error bars particularly in FSHD1 patients (Supplementary Fig. 1d). Supplementary Fig. 1b shows a typical example of the D4Z4 methylation analysis on a *de novo* FSHD2 patient and his unaffected family members. The FSHD2 patient has comparable methylation levels (%) to her unaffected mother who carries a non-permissive alleles (NP) only. The unaffected father has significant lower methylation levels than mother and daughter as quantified by fragment intensities.

Criteria	Number of Families
*Fse*I Methylation <25%	41
Both chromosome 4q D4Z4 arrays > 10 units	40
Not more than one chromosome 10q D4Z4 array <11 units	39
Inheritance Pattern:	
Dominant inheritance	13
de novo D4Z4 hypomethylation	7
Unknown (not informative)	19

Supplementary Table 1:

Selection criteria used to prioritize FSHD2 families for whole exome sequencing.

Maximum D4Z4 methylation at *Fse*I site in patients was set at 25%. We excluded families in which one of the individuals with D4Z4 methylation <25% had a D4Z4 repeat array of <10 units on a permissive allele or more than one array of <10 units. Families were further categorized according to the inheritance pattern of D4Z4 hypomethylation.

Family	Individual	Gender	age	FseI	Units
Rf392	102	F	54	17	50U
Rf393	101	M	75	11	89U
Rf393	206	F	42	11	18U
Rf393	302	F	27	19	20U
Rf393	305	M	34	21	20U

Supplementary Table 2.

Information on unaffected *SMCHD1* heterozygotes with a permissive D4Z4 haplotype.

Indicated are family ID, individual ID, gender, age, *Fse*I methylation level and D4Z4 array size in units (U) of smallest permissive allele.

Primers for analysis SMCHD1 splicing at exons 12, 25, 29 and 36

Name	Sequence (5' to 3')	position SMCHD1
exon 10F	5'-TGA TCC ATG CTT TCC ATC AA-3'	NM015295_1512F
exon 13R	5'-CCT TCA GCC ACA AAG CAA AT-3'	NM015295_1882R
exon 24F	5'-TCT GGA ACC AGT ATT TTA ACA GGA-3'	NM015295_3151F
exon 26R	5'-TTG CAC ATC AGG AAG CAG AC-3'	NM015295_3518R
exon 28F	5'-CTG GGG TTG GAC TTG ATA GC-3'	NM015295_3779F
exon 30R	5'-GGT GCT GGA TTA TCC CAC TG-3'	NM015295_4070R
exon 35F	5'-TCC AGT TTG GTT TTA TGA TGG A-3'	NM015295_4574F
exon 37R	5'-TTC ACG AAG GGG AAT TCA AG-3'	NM015295_4889R

qPCR primers

Name	Sequence (5' to 3')	position SMCHD1
SMCHD1_F (exon 47F)	5'- CGA CAG ATT GTC CAG TTC CTC-3'	NM015295_6125F
SMCHD1_R (exon 48R)	5'- CCA ATG GCC TCT TCT CTC TG-3'	NM015295_6225R
DUX4RT-F2	5'-CCC AGG TAC CAG CAG ACC-3'	
DUX4-pLAMR4	5'-TCC AGG AGA TGT AAC TCT AAT CCA-3'	
hGAPDHFw	5'-AGC ACA TCG CTC AGA CAC-3'	
hGAPDHRev	5'-GCC CAA TAC GAC CAA ATC C-3'	
qPCR GUS fw	5'-CTC ATT TGG AAT TTT GCC GAT T-3'	
qPCR GUS rev	5'-CCG AGT GAA GAT CCC CTT TTT A-3'	

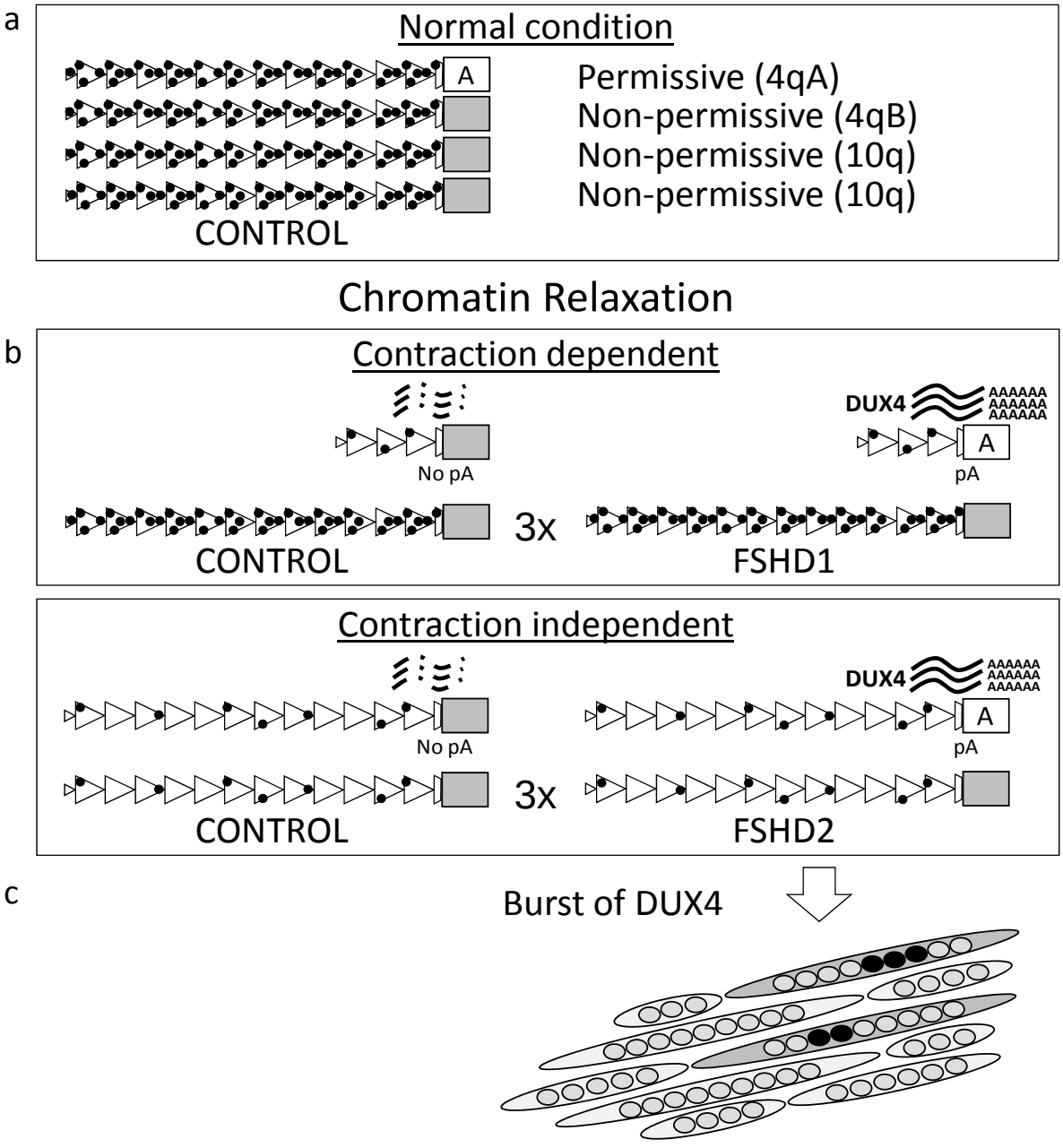
ChIP primers

Name	Sequence (5' to 3')
DUX4 ChIP F	5'-CCG CGTC CGT CCG TGA AA-3'
DUX4 ChIP R	5'-TCC GTC GCC GTC CTC GTC-3'
GAPDH ChIP F	5'-CTG AGC AGT CCG GTG TCA CTA C-3'
GAPDH ChIP R	5'-GAG GAC TTT GGG AAC GAC TGA G-3'

Antisense oligo nucleotide

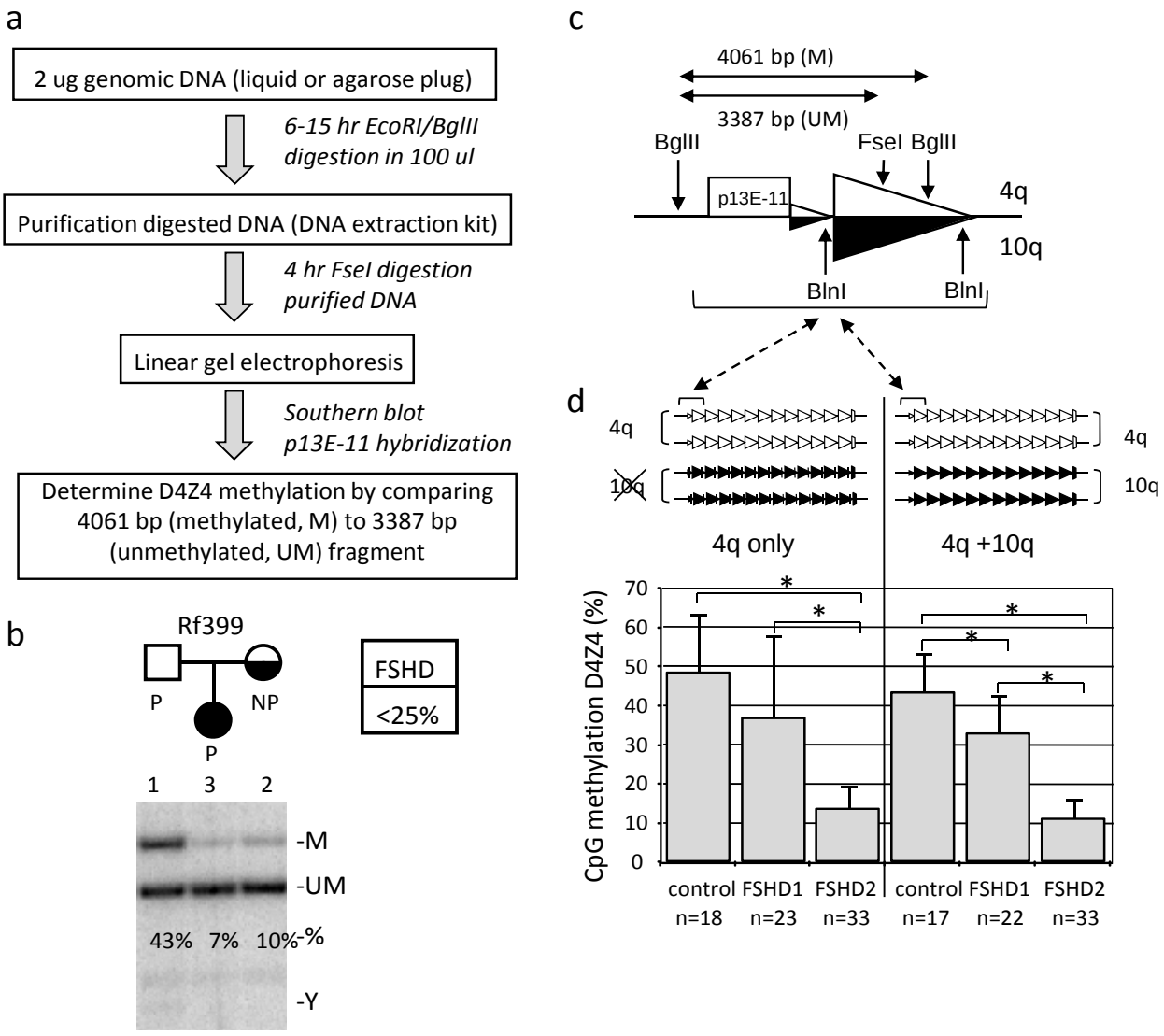
Name	Sequence (5' to 3')	position
29AON5	5'-GUC CAG AAA UUA GUU GCA CUC-3'	exon 29 SMCHD1
36AON1	5'-GAU UAG GCA GGA CUU CAA CU-3'	exon 36 SMCHD1
h50AON2	5'-(6FAM)-GGC UGC UUU GCC CUC-3'	exon 50 DMD

Supplementary Table 3. **Sequences primers and antisense oligo nucleotides used in this study**


a
Normal condition
A
Permissive (4qA)
Non-permissive (4qB)
Non-permissive (10q)
Non-permissive (10q)
CONTROL
Chromatin Relaxation
b
Contraction dependent
No pA
DUX4
AAAAAA
AAAAAA
AAAAAA
A
pA
3x
CONTROL
FSHD1
Contraction independent
No pA
DUX4
AAAAAA
AAAAAA
AAAAAA
A
pA
3x
CONTROL
FSHD2
c
Burst of DUX4

Supplementary Fig. 1: **Schematic of the FSHD locus.**

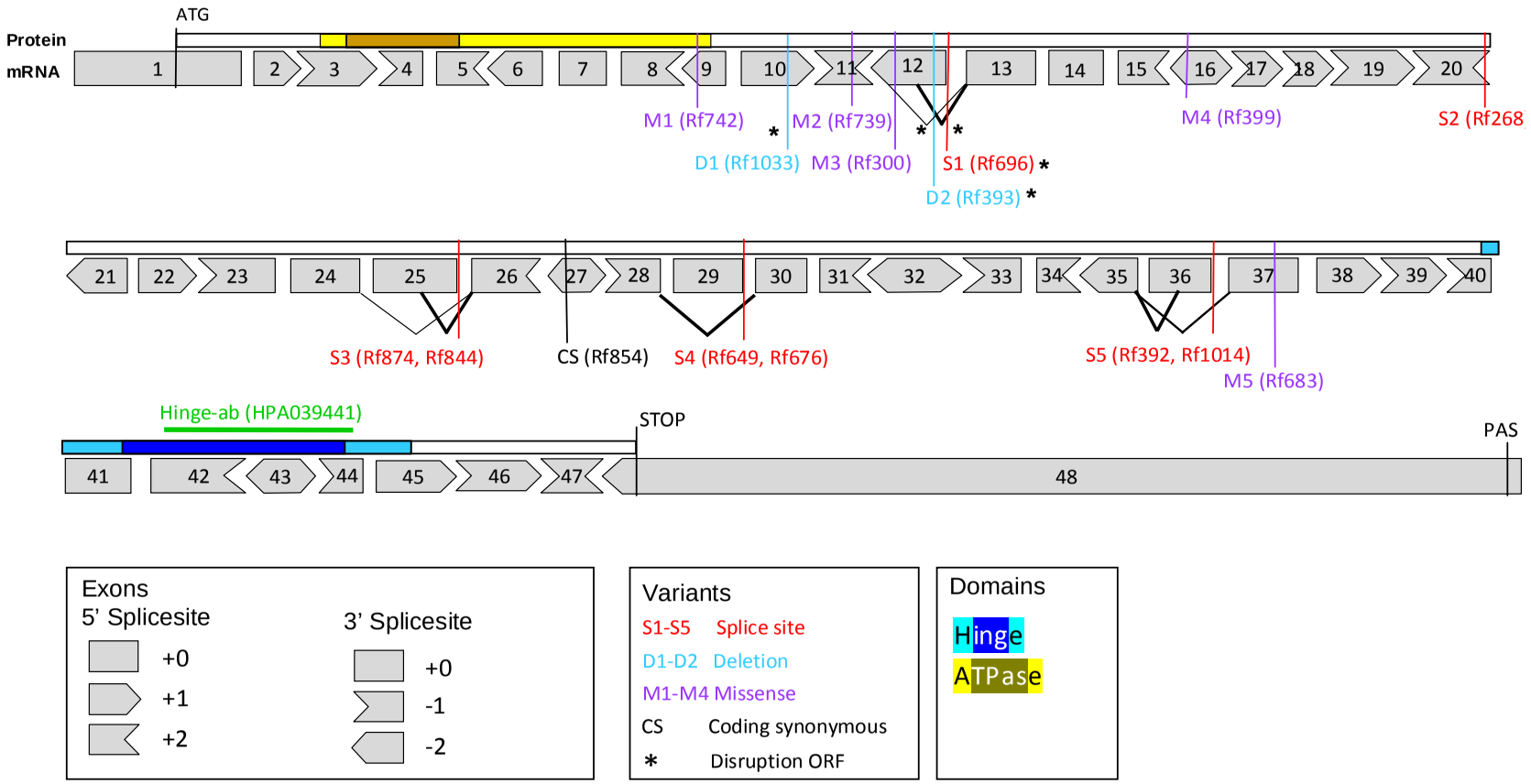
Different combinations of D4Z4 chromatin relaxation are shown with the associated chromosomal context and patient phenotype. The D4Z4 array is shown as a series of white triangles on chromosome 4. The homologous array on chromosome 10 is depicted in grey. The FSHD permissive 4qA, and FSHD non-permissive 10q and 4qB haplotypes are depicted as white and light grey boxes, respectively. **(a)** In the normal condition, D4Z4 arrays of >10 units are densely CpG methylated (black dots) on all four chromosomes. **(b)** FSHD1 is associated with D4Z4 array contraction-dependent D4Z4 hypomethylation and *DUX4* expression from the deleted chromosome having a FSHD-permissive 4qA haplotype. Permissive 4qA haplotypes have a *DUX4* polyadenylation signal (pA) distal to the last unit of the D4Z4 array. This pA signal results in stabilization of *DUX4* mRNA. Contraction-dependent chromatin relaxation on non-permissive haplotypes (4qB or 10q) do not cause disease, because they lack this *DUX4* pA signal. In FSHD1, D4Z4 hypomethylation is restricted to the contracted array. FSHD2 is caused by D4Z4 array contraction-independent chromatin relaxation of a D4Z4 locus with a permissive haplotype. In this case all four D4Z4 arrays are hypomethylated, and the hypomethylation phenotype can segregate independently of the permissive 4q haplotype within a family. Thus, family members who inherit the hypomethylation phenotype without a permissive haplotype do not develop FSHD2 (CONTROL). Chromosome 10 arrays are not depicted. **(c)** D4Z4 chromatin relaxation leads to a variegated production of the DUX4 protein in a subset of FSHD1 and FSHD2 myonuclei (black).


a
2 ug genomic DNA (liquid or agarose plug)
6-15 hr EcoRI/BglII digestion in 100 ul
Purification digested DNA (DNA extraction kit)
4 hr FseI digestion purified DNA
Linear gel electrophoresis
Southern blot p13E-11 hybridization
Determine D4Z4 methylation by comparing 4061 bp (methylated, M) to 3387 bp (unmethylated, UM) fragment
b
Rf399
P
NP
P
FSHD
<25%
1 3 2
-M
-UM
43% 7% 10%
-%
-Y
c
4061 bp (M)
3387 bp (UM)
BglII
FseI BglII
p13E-11
4q
10q
BlnI
BlnI
d
4q
10q
4q
10q
4q only
4q +10q
CpG methylation D4Z4 (%)
70
60
50
40
30
20
10
0
*
control FSHD1 FSHD2
n=18 n=23 n=33
control FSHD1 FSHD2
n=17 n=22 n=33

Supplementary Fig. 2.

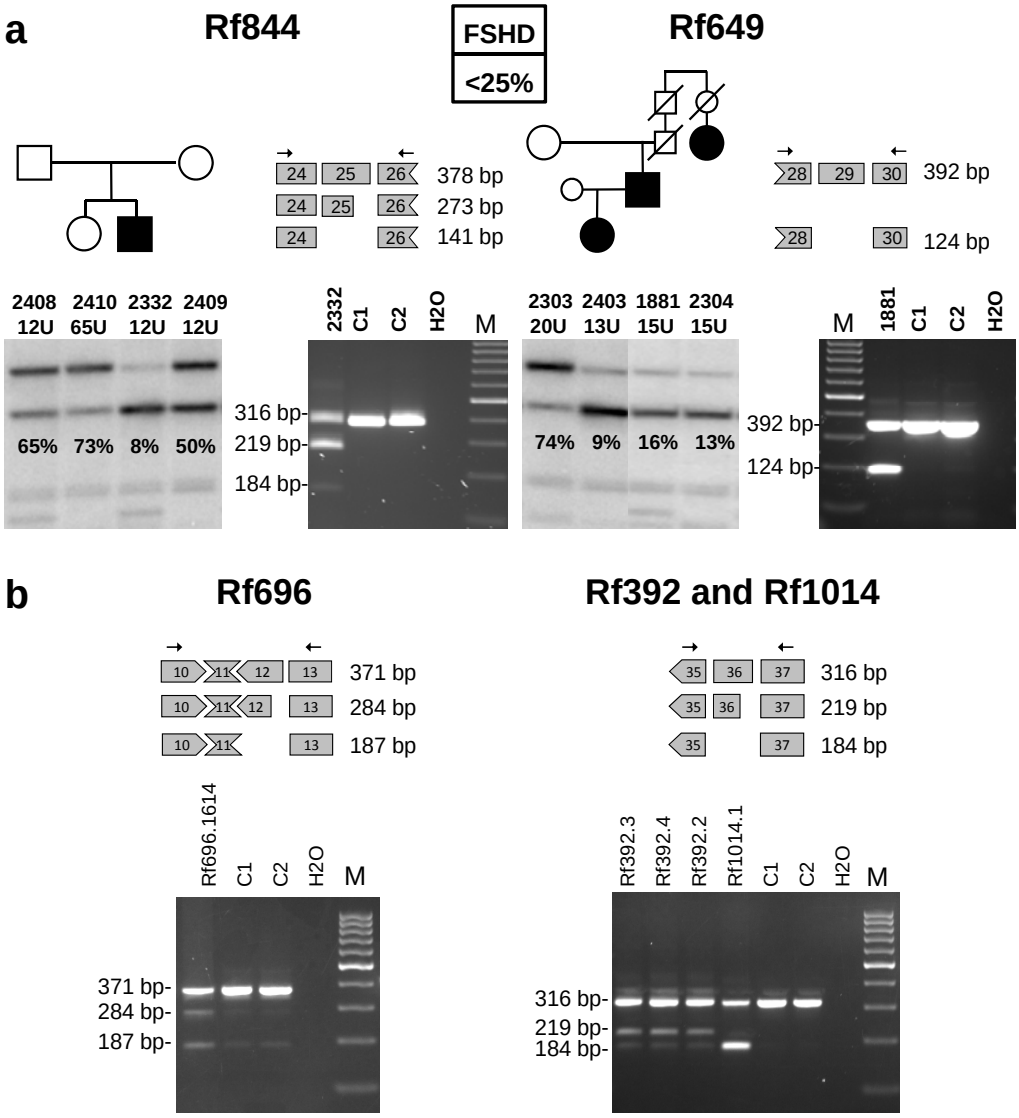
Design and results of the D4Z4 methylation test.

(a) Overview of methylation analysis method. **(b)** Example of methylation analysis in an FSHD2 family. Methylated (M) and unmethylated (UM) D4Z4 fragments are indicated. Below each lane the methylation value is indicated in %. Y indicates cross hybridizing Y fragment. The hypomethylated mother in this family is not affected in the absence of a permissive haplotype. **(c)** Schematic of methylation test showing the p13E-11 probe region at the proximal end of the D4Z4 repeat array and the expected D4Z4 fragment sizes upon digestion with restriction enzymes *Eco*RI, *Bgl*II and *Fse*I (*Eco*RI sites are not shown as they are outside the indicated area and the enzyme is only used for additional fragmentation of the gDNA). The position of the chromosome 10q-specific restriction enzyme *Bln*I (black bottom half) that was previously used for the chromosomes 4q only methylation analysis is indicated. **(d)** Schematic of *Fse*I methylation analysis for both chromosomes 4 (old method; left panel)[4] and chromosomes 4 and 10 (new method; right panel). Bar diagram of average methylation levels in controls (N=17), FSHD1 patients (N=22) and FSHD2 patients (N=33) obtained by the old method (left panel) and same samples by new method (right panel). Error bar represents standard deviation. FSHD2 patients are significantly hypomethylated by this test compared to controls and FSHD1 patients (*: $p<0.005$). Note that FSHD1 patients have methylation levels in between controls (normal methylation at all 4 alleles) and FSHD2 (hypomethylation at all 4 alleles) due to the presence of one hypomethylated allele.



Supplementary Fig. 3. Schematic of the human *SMCHD1* gene.

All exons are indicated with boxes. Information about the SMCHD1 protein domains and Hinge antibody epitope is also given. *SMCHD1* mutations identified in this study are documented with their (predicted) consequences. The position of the 5' and 3' splice sites with respect to the coding frame is also indicated. Mutations that result in a frameshift are indicated by an asterisk.



Supplementary Fig. 4.

Examples of methylation analysis and alternative splicing in *SMCHD1* heterozygotes

(a) Pedigrees of sporadic (left panel) and familial (right panel) FSHD2 kindreds. Methylation analysis of the *Fse*I site in D4Z4 shows the degree of methylation (left panels). *SMCHD1* mRNA analysis in *SMCHD1* heterozygotes and controls (C1-2) shows exon skipping or cryptic splice site usage (right panels). **(b)** RT-PCR analysis of *SMCHD1* RNA in controls (C) and individuals heterozygous for *SMCHD1* splice site mutations in families Rf696, Rf392 and Rf1014. RT-PCR products were sequence verified. Schematics of alternative splice events are shown on top and primers used to determine splicing are indicated with arrows. The splicing changes in family Rf696 can also be observed at lower frequency in the controls indicating that this variant shifts the balance (compare unspliced product with spliced products).