Supplementary information

Supplemental Materials and Methods

Embryoid body formation.

Prior to cell collection, cells are transiently treated with Thiazovivin (Rock inhibitor, 6μg/ml; StemCell. Cat. N°. 72252) to increase cell viability. HiPSCs colonies are then enzymatically disrupted with Dispase (StemCell. Cat. N°. 07923), rinsed in cell culture medium and transferred to EB medium (50% of mTeSR™1 and 50% of hESs cell culture medium (78% KO DMEM ; 20% KSR 20% ; 1% Glutamax ; 1% non-essential amino acids ; 2.5 µL β-Mercaptoethanol) in the presence of Thiazovivin (6μg/ml). Cells are grown in suspension in untreated petri dishes. Embryoid bodies are maintained between 15 to 30 days with medium replacement every two days. Expression of pluripotency and differentiation markers is quantified at different time points (Day 5, 10, 15 or 30).

Teratoma formation.

We also tested the capacity of the different clones to differentiate *in vivo*. To this aim, approximately 10 x 10^6 cells were resuspended in ice-cold Matrigel and subcutaneously injected in 6-8 weeks-old female non-obese diabetic/severely combined immunodeficient mice (NOD/SCID) (Janvier Laboratory). One to two months after injection, teratomas are collected, fixed in paraformaldehyde, paraffin embedded and stained with hematoxylin-eosin.

Western blot.

Proteins were resolved on a 4-12% MOPS-NuPAGE gel (Life Technologies) and transferred on a PVDF membrane (Milipore). The SMCHD1 protein was detected using two different commercial antibodies targeting either the N-terminal (Sigma, HPA 039441) or the C-terminal part of the protein (Abcam, ab31865). Anti-Lamin B2 (Abcam, ab151735) antibodies were used as loading control. After 3 washes in PBS-T, anti-mouse IgG secondary antibody coupled to HRP (ThermoFisher) was incubated for 60 min (1/20 000). Proteins were revealed

by chemiluminescence (ECL, Milipore, WBKLS0100) and visualized using a Quantity One Bio-rad camera.

Immunofluorescence and 3D FISH.

Cells were grown on 4-well slides (Millipore), fixed in 4% paraformaldehyde and treated as described(11) to maintain the 3D structure of the cells. Analysis of active (Xa) and inactive (Xi) chromosomes was done using a Texas red whole Chromosome X painting probe (Aquarius LPP0X R). All probes were denatured at 80±1°C for 5 minutes before hybridization. Nuclei were counterstained with DAPI (Sigma) diluted to 1 mg/ml in PBS and mounted in Vectashield (Vector Laboratories).

Images were acquired using a confocal scanning laser system from Zeiss (Germany). A 63x Plan-APOCHROMAT, oil immersion, NA 1.40 objective (Zeiss) was used to record optical sections at intervals of 0.48µm. The pinhole was set the closest to 1 Airy with optical slices in all wavelengths with identical thickness (0.8µm). Generated .lsm files with a voxel size of 0.1µm x 0.1µm x 0.48µm were processed using the Imaris software (Bitplane AG). After 3D analysis, we compared the mean volume ratio of nuclei and of the X signals. Values were compared using a non-parametric Kruskal-Wallis test.

Supplementary table 1: List of primary fibroblasts and hiPSCS clones.

BAMS-Case 1; BAMS-Case 2 and BAMS-Case 9 were described in (12).

	Diagnosis	Genotype	*SMCHD1* status	Age	Gender	F	iPS	ESC
GENEA002	Healthy	>10	No mutation	-	Male			✓
GENEA015	Healthy	>10	No mutation	-	Male			✓
GENEA019	Healthy	>10	No mutation	-	Female			✓
GENEA049	FSHD1	6RU	No mutation	-	Female			✓
GENEA050	FSHD1	6RU	No mutation	-	Male			✓
GENEA096	FSHD1	7RU	No mutation	-	Female			✓
AG04148	Healthy	>10	No mutation	56	Female	✓	✓	
AG08498	Healthy	>10	No mutation	1	Male	✓	✓	
CTRL3	Healthy	>10	No mutation	8	Male	✓	✓	
HFF	Healthy	>10	No mutation	1	Male	✓	✓	
12566	FSHD1	3RU	No mutation	64	Female	✓	✓	
12759	FSHD1	7RU	No mutation	51	Female	✓	✓	
TalF	FSHD1	2RU	No mutation	0	Male		✓	
17706	FSHD1 Mosaic	2RU (25%)	No mutation	56	Female	✓	✓	
14586	FSHD2	>10	c.573A>C ; p.Q193P	67	Male	✓	✓	
11440	FSHD2	>10	c.2338+4A>G; p.S754*	37	Male	✓	✓	
10480	FSHD2	>10	c.2260+6T>C	53	Female	✓		
11491	FSHD2	>10	c.5476+3 A>G; r.5477_5547del; p.V1826Gfs*19	66	Female	✓	✓	
15166	FSHD2	>10	r.4614_4615insTATAATA; p.A1539Yfs*4	67	Female	✓	✓	
BAMS-1	BAMS	N/A	c.407A>G p.E136G	5	Male	✓	✓	
BAMS-2	BAMS	N/A	c.403A>T p.S135C	28	Female	✓	✓	
BAMS-9	BAMS	N/A	c.1259A>T p.D420V	3	Male	✓	✓	

Supplementary Table 2: D4Z4 mean methylation levels in peripheral blood cells (PBLs), primary fibroblasts, induced pluripotent stem cells (hiPSCs) and cancer cell lines. Results in blood DNA from healthy individuals are presented in Roche et al. submitted.

		DR1	5P	MID	3P
Blood	Control	61.4% ± 4.5	70.8% ± 4.0	70.9% ± 1	58.5% ± 0.8
	18p-Dup	63.3% ± 12.3	65.6% ± 10.3	65.7% ± 2.9	58.4% ± 18.3
	18p-Del	22.7% ± 9.6	51.7% ± 5.4	64.2% ± 2.6	55.6% ± 1.5
hESC	Control	82.2% ± 8.7	84.6% ± 9.3	76.2% ± 6.1	58.4% ± 12.4
	FSHD1	78% ± 4.7	80% ± 2.9	84.7% ± 0.9	69.8% ± 2.5
Fibroblasts	Control	71.4% ± 8.8	73.1% ± 13.1	45.3% ± 26.9	56.9% ± 14.6
	FSHD1	39.9% ± 4.4	40.5% ± 3	46% ± 4.7	45.7% ± 8.6
	Mosaic	38.6%± 2.4	47.7%± 3.1	54.1%± 1.8	41.2%± 1.7
	FSHD2	15.7% ± 14.1	44.8% ± 21.1	49.4% ± 8.9	42.4% ± 20.6
	BOSMA	21.3% ± 11.7	40.6% ± 4.6	58.1% ± 9.2	43.6% ± 5
hiPSC	Control	79.8% ± 16.6	84.1% ± 7.4	53.1% ± 23	75.9% ± 7
	FSHD1	67.8% ± 23.4	69.7% ± 11.1	66.2% ± 8.9	56.7% ± 5.2
	Mosaic-long	84.1%± 4.87	82.8%± 2.52	74.8%± 0.03	66.9%± 0.5
	Mosaic-short	76.8%± 4.75	76%± 4.14	72.3%± 1.26	62.6%± 0.04
	FSHD2	28% ± 12.7	59% ± 8.2	54.9% ± 15.2	55.0% ± 7.4
	BOSMA	46.2% ± 6.3	65.5% ± 6.3	71.0% ± 3.1	57.8% ± 3.8
HEK	Control	74.9% ±1.8	67.7% ± 0.5	45.5% ± 7.9	42.1% ± 3.3
	SMCHD1$^{-/-}$	62.1% ±5.2	63.0% ± 1.6	36.7% ± 4.3	28.8% ± 8.6
HCT116	WT	74.8%± 18	80.2%± 11	69.4%± 1	59.5%± 3.5
	DNMT3b$^{-/-}$	75.0%± 11.5	79.5%± 7.5	65.6%± 1	54.3%± 2
	DNMT3b$^{-/-}$ DNMT1$^{-/-}$	11.3%± 5.5	28.5%± 17.5	23.0%± 11	33.9%± 14.5
	SMCHD1$^{-/-}$	73.1%± 11.5	75.9%± 7.5	55.7%± 10	56.3%± 0.5

Supplementary Table 3: Blood samples (PBL) from patients with mutation, deletion or duplication of the *SMCHD1* gene.

	Diagnosis	*SMCHD1* status	Age	Gender
18pDel-3586	18pDel	18p Deletion	7	Male
18pDel-23617	18pDel	18p Deletion	21 Mo	Female
Del-SMCHD1-3	18pDel	18p Deletion	6	Female
Del-SMCHD1-4	18pDel	18p Deletion	5	Male
Dup-SMCHD1-5	18pDup	18p Duplication	38	Male
Dup-SMCHD1-6	18pDup	18p Duplication	10	Female

Supplementary Table 4: Mean methylation levels of SINE, LINE and others repeats.

		SINE/LINE		Macrosatellite	
		AluI	LINE-1	RS447	TAR1
Blood	Control	58.1% ± 2.6	74.9% ± 0.8	84.4% ± 0.9	48.9% ± 9.8
	18p-Dup	58.8% ± 4.5	57.2% ± 18.3	NA	NA ±
	18p-Del	59.4% ± 1.7	74.3% ± 1.5	84.7% ± 1.5	44.0% ± 2.5
hESC	Control	47.1% ± 3.4	55.5% ± 3.9	79.1% ± 2	78.4% ± 2
	FSHD1	60% ± 6.7	60.9% ± 4.3	79.7% ± 5.9	51.7% ± 25
Fibroblasts	Control	52% ± 6.2	64.1% ± 4.9	73.5% ± 9.7	67.1% ± 11.2
	FSHD1	49.1% ± 3.9	53.5% ± 3.3	49.2% ± 7.1	46.7% ± 27.2
	FSHD2	52.9% ± 7.3	60.1% ± 6.1	64.7% ± 8.7	57.1% ± 18.1
	BOSMA	53% ± 7.5	52.9% ± 10.7	64.5% ± 8.9	46% ± 24.7
hiPSC	Control	53.3% ± 5.7	63.8% ± 3.5	82.7% ± 0.1	80.9% ± 5
	FSHD1	60% ± 6.7	60.9% ± 4.3	79.7% ± 5.9	51.7% ± 25
	FSHD2	54.3% ± 6.4	60.0% ± 6.4	84.1% ± 1.1	72.7% ± 21.7
	BOSMA	50.8% ± 9.2	63.1% ± 0.5	85% ± 0.6	62.1% ± 13
HEK	WT	50.0% ± 5.8	43.8% ± 3.0		
	SMCHD1$^{-/-}$	39.1% ± 11.2	30.8% ± 9.0		
HCT116	WT	58%±1.2	70.4%±0.4	78.8%±4.6	47.4%±3.5
	DNMT3b$^{-/-}$	51.3%±3.7	68.2%±2	73.9%±4.8	50.1%±14.3
	DNMT3b$^{-/-}$ DNMT1$^{-/-}$	45.6%±5.9	45.9%±7.9	46.3%±8.6	17.1%±8
	SMCHD1$^{-/-}$	52.6%±5.7	70.9%±1.1	79.4%±2.7	51.4%±15

Supplementary Table 5: Mean methylation levels of DXZ4 and *H19*.

		DXZ4	H19
Fibroblasts	Control	59.9% ±22.4	27.9% ± 19.2
	FSHD1	52.3% ± 12.7	60.9% ± 10.4
	FSHD2	38.8% ± 5.5	54.0% ± 33.2
	BAMS	65.2% ± 23.1	61.3% ± 8.3
hiPSC	Control	88.0% ± 1.8	9.4% ± 14.5
	FSHD1	39.6% ± 18.2	64.8% ± 10.4
	FSHD2	57.1% ±16.3	63.7%
	BAMS	77.2% ± 16.1	56.9% ± 38.6

Supplementary Table 6: Sequence of the primers used for Sodium bisulfite PCR.

Sequence	Name	Primers	Temperature	Size
D4Z4	5P	AAATATGTAGGGAAGGGTGTAAGTT CTTAAATATACCAAACCCTCTCTCC	56°C	341
D4Z4	MID	ATTTATGAAGGGGTGGAGTTT ATAACCTAAACCAACCRTTCTCTA	56°C	419
D4Z4	3P	GTTTTGTTGGAGGAGTTTTAGGA CTAAACCTAAAAAACAAAAATCCCA	56°C	237
D4Z4	DR1	GAAGGTAGGGAGGAAAAG ACTCAACCTAAAAATATACAATCT	56°C	254
RS447	5P	TGTGATTTTTTTTTGAATTGAG CTCACCTCCCAAATAAAATAAA	56°C	318
LINE	5UTR	GTAAGGGGTTAGGGAGTTTTTT TTATCTATACCCTACCCCCAAA	56°C	432
ALU	ALU1	GGATTATTTGAGGTTAGGAGAT TCCCRAATAACTAAAACTACAA	56°C	250
Telomere	TAR1	GGAGTAGAGTTTTTTTTAGGTTAGATT AAACAAACAATACCCCCAAC	56°C	125
DXZ4	DXZ4	GATGGTAGTATTGTTTTAGAAATGTT ATCCCATCTACCAAAAACA	56°C	172
H19	H19-Ext	TTTTTGGTAGGTATAGAGTT AAACCATAACACTAAAACCC	50°C	343
H19	H19-Int	TGTATAGTATATGGGTATTTTTGGAGG TTT TCCCATAAATATCCTATTCCCAAATAAC C	52°C	231

Supplementary table 7: Sequence of the primers used for RT-qPCR.

Primers for DUX4 and DUX4 targets were described in (8)

Gene	Forward Primer	Reverse Primer
DUX4-1A	GAGCTCCTGGCGAGCCCGGAGTTTCTG	
DUX4-14A	CCCCGAGCCAAAGCGAGGCCCTGCGAGCCT	
DUX4-15A	CGGCCCTGGCCCGGGAGACGCGGCCCGC	
DUX4-175		TCTAATCCAGGTTTGCCTAGACAGC
DUX4-174		GTAACTCTAATCCAGGTTTGCCTAGA
DUX4-182	CACTCCCCTGCGGCCTGCTGCTGGATGA	
DUX4-183		CCAGGAGATGTAACTCTAATCCAGGTTTGC
DUX4-184		GTAACTCTAATCCAGGTTTGCCTAGACAGC
MDB3L2	CAGGAGTGGGGTCAGCAGGAGGA	TTTGAGCTTCCCCAGAACAGGCAGG
ZSCAN4	TGCCTCCTGGATTCAAACA	TGTTCTATACCATCACTGGTCCTG
TRIM 43	ACCCATCACTGGACTGGTGT	CACATCCTCAAAGAGCCTGA
MBDL3	CGTTCACCTCTTTTCCAAGC	AGTCTCATGGGGAGAGCAGA
TRIM 28	ATGGTGCAGACAGCACTGG	GCAGTACACGCTCACATTTCC
HPRT	TGATAGATCCATTCCTATGACTGTAGA	CAAGACATTCTTTCCAGTTAAAGTTG
GAPDH	AGCCACATCGCTCAGACAC	GCCCAATACGACCAAATCC
PPIA	ATGCTGGACCCAACACAAAT	TCTTTCACTTTGCCAAACACC
36B4	TCTACAACCCTGAAGTGCTTGAT	CAATCTCGAGACAGACACTGG
NANOG	TTTGGAAGCTGCTGGGGAAG	GATGGGAGGAGGGGAGAGGA
OCT4	CTTGCTGCAGAAGTGGGTGGAGGAA	CTGCAGTGTGGGTTTCGGGCA
SOX2	TCAGGAGTTGTCAAGGCAGAGAAG	GCCGCCGCCGATGATTGTTATTAT
PAX6	CCGGTCAAGAAACAGAAGACCAGA	CCATTGCTATTCTTCGGCCAGTTG
TBXT	CACCTGCAAATCCTCATCCTCAGT	TGTCATGGGATTGCAGCATGGA
FOXA2	GCATTCCCAATCTTGACACGGTGA	GCCCTTGCAGCCAGAATACACAT

Supplementary table 8: Sequence of primers used for ChIP

Gene	Forward Primer	Reverse Primer
D4Z4-DR1	CCCGCCTCCGGGAAAAC	GGGATGTGCGGTCTGTGA A
D4Z4-5'	ACGACGGAGGCGTGATTT	AGTGTGGCCGGTTTGGAA
D4Z4-Mid	TCATGAAGGGGTGGAGCCTG	TCCAAACGAGTCTCCGTCGC
D4Z4-3'	CTCAGCGAGGAAGAATACCG	ACCGGGCCTAGACCTAGAAG
CHR5	GGAGTTGGGGAAGCTAGGAA	GATCATCCGTGGCTTGAGAT

Legend to the supplementary figures

Supplementary figure 1. Expression of pluripotency transcription factors for validation of hiPSC clones.

Reverse transcription and quantification by Real-Time PCR of *NANOG*, *OCT4* and *SOX2* expression in representative clones is provided. Histograms display the mean fold-change of expression compared to the H9 hESCs. Values are normalized to the *36B4* standard gene. Error bars represent SD from three independent reactions. The AG04 and AG08 control clones were previously described (13). For validation of pluripotency, expression of the Keratin Sulfate antigens Tra1-60 and Tra1-81 and the glycolipid antigens *SSEA3* and *SSEA4* is verified by flow cytometry (data not shown). The presence of alkaline phosphate-positive colonies is verified by colorimetric staining (data not shown).

Supplementary figure 2. Validation of *in vitro* and *in vivo* differentiation.

One of the main features of pluripotent cells is the ability to differentiate toward the three embryonic lineages, *i.e.* ectoderm, endoderm and mesoderm. To further validate the different clones we tested whether these cells are able to differentiate and form embryoid bodies when grown in suspension and form teratomas after subcutaneous injection in NOD/SCID immunodeficient mice. **A.** Teratomas were collected and processed 1 to 2 months post-injection, fixed in paraformaldehyde, paraffin embedded and stained with hematoxylin-eosin. Representative histological sections containing tissue derived from the neuroectoderm layer (NE), mesoderm (M) or endoderm (E) are presented. **B.** Differentiation capacity was tested *in vitro* after formation of embryoid bodies (EB). Reverse transcription and quantification by Real-Time PCR of the *NANOG*, *OCT4* and *SOX2* pluripotency markers in the different clones during embryoid bodies formation (day 1 to day 14, Panel B) C42 corresponds to HFF (control), C9 to FSHD1-12566; C3 to FSHD1-12759 and C31 to FSHD1-TalF. Histogram display the mean fold-change of expression compared to the H9 hESCs. Values are

normalized to the *36B4* standard gene. Expression of the pluripotency markers is significantly decreased at day 7 of differentiation in the different clones. **c.** Following the previous validation, for expression of pluripotency markers, EB were collected at day 14 for the different clones. Expression of the *FOXOA2* endoderm marker, *PAX6* neuroectoderm marker and *Brachyury* mesoderm marker was analyzed by RT-qPCR. Values are normalized to the *36B4* standard gene. Error bars represent SD from three independent reactions.

Supplementary figure 3. Similar D4Z4 remethylation in hiPSCs clones carrying a short or a long D4Z4 array derived from a mosaic patient.

A. For each sample, the number of sequenced molecule was plotted in different classes defined by their mean methylation level. Histogram bars correspond to the different classes of molecules from low (0-10%) to high (90-100%) of methylated CpG. The global distribution does not follow a normal distribution as samples are distributed in two to three groups visualized using the "Mixtools" bootstrap algorithm in R. In all graphs, the red curves corresponds to molecules with a low methylation level (< 35%). For fibroblasts, the percentage of molecules with a medium methylation level is calculated as the area under the green curve (mean methylation 50%); the percentage of molecules with a high methylation level (>50%) is calculated as the area under the blue curve. For DR1 in hiPSCs, the green curves corresponds to molecules with a high methylation level. For 5P in fibroblasts, the green curve corresponds to molecules with a mean methylation level of 50%; the blue curve to molecules with a high methylation level (>50%). We report here the multimodal distribution for the DR1 and 5P differentially methylated regions in primary fibroblasts from a mosaic patient affected with FSHD and carrying a short 2 units D4Z4 array in 25% of cells and a long 11 units array in 75% of cells (upper panel) and individual hiPSCs clones. For each clone, the size of the chromosome 4 D4Z4 array was determined by Southern blotting using the p13E11 probe. **B**. Boxplot showing the percentage of molecules with a low (red curve) or high (green curve) level of methylation for the DR1 sequence in the different samples and percentage of molecules with a low (red curve) or high (blue curve) level of methylation for

the 5P sequence. We observe a decrease in the number of molecules with a low level of methylation and a subsequent increase of the number of molecules with a high methylation level for the DR1 and 5P in hiPSCs clones compared to fibroblasts with no significant difference between the short and long array clones. These results confirm that D4Z4 is remethylated upon reprogramming regardless of the number of D4Z4 unit.

Supplementary figure 4. Methylation level in control and FSHD1 human embryonic stem cells. A. Histograms display the mean methylation level for the different regions analyzed by bisulfite sequencing across D4Z4 in hiPSCs (black) and hESCs (white) in control and FSHD1 samples. Error bars represent SD from three independent assays. Methylation level is not significantly different in the two groups of samples. **B.** We compared the methylation profile of AluY, LINE-1, RS447 and TAR1 repetitive elements in three different human embryonic stem cells (hESC) from healthy donors or donors carrying a short D4Z4 array (FSHD1). GENEA049 and GENEA050 are siblings. Histograms represent the mean methylation level in each group of samples. Histogram bars correspond to the mean percentage of methylated (black) or unmethylated (white) CpG for each position. Position of primers used for PCR amplification is indicated together with the different CpGs (white dots). **C.** Histograms display the mean methylation level for the different repetitive DNA elements in hiPSCs (black) and hESCs (white) in control and FSHD1 samples. Error bars represent SD from three independent reactions. Methylation level is not significantly different in the two groups of samples

Supplementary figure 5. Patients carrying 18p deletion display marked D4Z4 hypomethylation while methylation is identical to controls in patients with duplication of the SMCHD1 locus. A. The 18p deletion syndrome (18pDel syndrome, OMIM #146390) is a rare disease characterized by a number of typical features (cognitive impairment, growth retardation and short stature, minor craniofacial dysmorphism, ptosis and strabismus)(14) but absence of muscular dystrophy(15). The different patients carry an 18p deletion

encompassing the *SMCHD1* gene locus (18p11.32) but did not display any sign of muscular dystrophy or craniofacial defect at the time of diagnosis. For each sample, we plotted the mean methylation level of each CpG analyzed for the 4 D4Z4 subdomains (DR1, 5P, MID, 3P). Each row of histograms represents the average methylation level in each individual patient calculated for each individual CpG within the four D4Z4 subregions after bisulfite modification, PCR amplification and deep sequencing. Black bars correspond to the percentage of methylated sites; white bars to unmethylated ones. Methylation percentages are provided in the supplementary table 3. Sodium bisulfite sequencing indicate that D4Z4 methylation is decreased in individuals with *SMCHD1* haploinsufficiency suggesting that D4Z4 methylation is sensitive to *SMCHD1* gene dosage and that D4Z4 methylation does not systematically segregate with muscle symptoms reminiscent of FSHD. **B.** For two individuals carrying an 18p duplication containing the *SMCHD1* gene locus (18p11.32), we plotted the mean methylation level of each CpG for the 4 D4Z4 subdomains (DR1, 5P, MID, 3P). The 18p trisomy syndrome is a very rare condition with a poorly defined phenotype including features such as a short stature, developmental delay, intellectual disability and cranial dysmorphism. The percentage of methylation was calculated for each individual CpG within the four D4Z4 subregions, Histograms represent the average methylation level with black bars corresponding to the percentage of methylated sites; white bars to unmethylated ones. The level of D4Z4 methylation is not significantly different from controls indicating that a higher SMCHD1 dosage does not lead to an increased D4Z4 methylation.

Supplementary figure 6. Analysis of individual clones reveald a potential role for the SMCHD1 ATPase domain in D4Z4 methylation regulation. A. Schematic representation of the SMCHD1 protein and position of mutations in BAMS (cyan) or FSHD2 patients (red). BAMS patients were described in (12). BAMS-1 (E136G) and BAMS-2 (S135C) carry missense mutation in the ATPase domain reported as gain of function. BAMS-9 carries a missense mutation in the BAH-Bromo adjacent domain (D420V). FSHD2-14586 also carries a mutation in the ATPase domain (Q193P) while FSHD2-11491 carry a truncating mutation

(p.V1826Gfs*) in the Hinge domain. **B.** DNA methylation was determined after sodium bisulfite modification for 4 regions across D4Z4 (DR1, 5P, MID, 3P) by PCR amplification and high throughput DNA sequencing. For each sequence, dots represent individual CpGs. **c.** The mean methylation level of each CpG was plotted for the 4 D4Z4 sub-domains. BAMS patients are presented individually. Assays were done with DNA from primary fibroblasts (rows 1-3) or HiPSC clones (rows 4-6). Histogram bars correspond to individual CpGs, black indicates the percentage of methylated CG, white corresponds to the percentage of unmethylated CGs for each position. D. Mean methylation level of each CpG plotted for the 4 D4Z4 sub-domains in primary fibroblasts (rows 7-8) or HiPSC clones (rows 9-10) for two FSHD2 patients. Histogram bars represent the average of methylated (black) or unmethylated (white) CpG for each position. We observed the absence of D4Z4 remethylation in hiPSC clones from FSHD2 patients and differences in the remethylation depending on the type of mutation in BAMS with a higher level of methylation in samples displaying a gain of ATPase activity (S135C ; E136G) compared to cells in which the mutation occurs in the BAH-Bromo adjacent domain (D420V) with no increase in ATPase activity.

Supplementary figure 7. Comparison of the global methylation level of different repetitive DNA sequences in DNA samples from healthy donors, FSHD1, FSHD2 and BAMS patients. Methylation profile of different repetitive DNA sequences (AluY, LINE1, RS447, TAR1). Histograms display the mean methylation level (ratio between the number of methylated and unmethylated CpG for each DNA fragment) in fibroblasts from controls (n=7), FSHD1 (n=4), FSHD2 (n=6) and BAMS (n=5) or hiPSCs from controls (n=7), FSHD1 (n=4), FSHD2 (n=4) and BAMS (n=3) DNA samples. Histogram bars correspond to individual CpGs, black indicates the percentage of methylated CG, white corresponds to the percentage of unmethylated CGs.

Supplementary figure 8. Absence of global changes in X chromosome conformation in SMCHD1-mutated cells.

A. Analysis of the X chromosome volume in interphase nuclei. Upper panels, confocal images of X chromosome painting in male (left) and female (right) cells following DNA FISH with X chromosome painting probes (red). Nuclei were counterstained with DAPI. Lower panels, 3D reconstruction of nuclei for analysis of chromosome X volume. **B.** Scattergram displaying chromosome X volume (μm^3) in male control fibroblasts (n=30 nuclei counted) and hiPSCs (n=13), female control cells (fibroblasts, n=32; hiPSCs, n=58), female FSHD2 cells (11491; p.V1826Gfs*19; fibroblasts, n=32; hiPSCs, n=40) and female BAMS cells (BAMS-2, p.S135C; fibroblasts, n=39; hiPSCs, n=15). In female cells, the two X chromosomes are presented in different colors (red and black). The horizontal line corresponds to the median volume for each group of samples.

Supplementary figure 9. Expression of the DUX4-fl transcript is detectable both in cells carrying a short or a long D4Z4 array derived from a mosaic patient. A. We analyzed *DUX4* expression in hiPSCs clones derived from a mosaic patients affected with FSHD. After clone selection, size of the D4Z4 array was determined by Southern blotting. Two clones carrying a short 2 D4Z4 units repeat and two clones carrying a long 11 D4Z4 units were selected. *DUX4* expression was analyzed after nested PCR for amplification of the pathogenic DUX4-fl transcript and **B.** by RT-qPCR, values were normalized to the expression of three housekeeping genes (*GAPDH*, *HPRT* and *PPIA*). **c**. We also quantified expression of different DUX4 target genes by RT-qPCR (*ZSCAN4*, *TRIM43* and *MBD3L2*). Expression was normalized to three housekeeping genes (*GAPDH*, *HPRT* and *PPIA*).

Supplementary figure 10. 3' UTR RACE reveals that BAMS patients are permissive for the *DUX4-fl* pathogenic transcript. Different 3' UTR have been described for *DUX4* transcripts with the *DUX4* long isoform (*DUX4-fl*) associated with FSHD. Given the presence of this long isoform in both FSHD- and BAMS-derived cells, we next sought to fully

characterize the amplified transcripts and 3' UTR. We performed 3' RACE in different BAMS and FSHD samples. The 3'RACE PCR was performed using a forward primer matching Exon1 and a reverse primer recognizing the Poly(A). After amplification, amplicons were gel purified, cloned in a Topo TA cloning vector and at least ten clones were sequenced for each sample. Representative results are presented for three FSHD2 hiPSCs clones (11440; 11491 and 14586) and BAMS-2.

Supplementary figure 11. *DUX4* and DUX4 target genes are expressed in both FSHD and BAMS-derived hiPSCs and embryoid bodies.

A. Based on overexpression assays in somatic tissues, DUX4 has been shown to activate a number of genes including genes expressed in the germline. We tested expression of three different of these genes (*ZSCAN4*, *TRIM43* and *MBD3L2*) in control (dark blue; **1**: AG08; **2**: AG09), BAMS (cyan ; **3**: clone derived from BAMS 2; **4**, **5**, **6**: three clones derived from BAMS 9), **7**, **8**: two clones derived from BAMS 1), FSHD2 (red; **9** : 11491, **10**: 11440, **11**-**12**: 14586) and FSHD1 (grey ; **13**: 12759 carrying 7 D4Z4 units, **14**: TalF carrying two D4Z4 units) hiPSCs. Expression was normalized to three housekeeping genes (*GAPDH*, *HPRT* and *PPIA*). **B.** We evaluated *DUX4* expression in embryoid bodies (EB) obtained from different hiPSCs clones. EB were collected at day 14. *DUX4* transcripts were analyzed by RT-PCR using primers designed to amplify the short isoform (*DUX4-s*), detectable in all cells and the *DUX4-fl* pathogenic transcript. Controls (dark blue; **1**: AG08; **2**: AG09), BAMS (cyan ; **3**: EB from BAMS 9 ; **4** : EB from BAMS 2 ; 5: EB from BAMS 1) ; FSHD2 (red; **9** : EB from 11491, **10**: EB from 11440, **11**: EB from 14586) and FSHD1 (grey ; **13**: EB from 12759 carrying 7 D4Z4 units). We detected the *DUX4-fl* transcript at variable levels in samples 3,4,5 (BAMS), 9,10,11 (FSHD2). **C**. Expression was confirmed by RT-qPCR in the same samples compared to controls. **D**. We tested expression of three different of DUX4 target genes (*ZSCAN4*, *TRIM43* and *MBD3L2*) in EB from control hiPSCs (dark blue ; **1**: AG08; **2**: AG09), BAMS (cyan ; **3**: EB from BAMS 9 ; **4** : EB from BAMS 2 ; 5: EB from BAMS 1) ; FSHD2 (red; **9** : EB from 11491, **10**: EB from 11440, **11**: EB from 14586) and FSHD1 (grey ; **13**: EB from

12759 carrying 7 D4Z4 units). Expression of the different genes is increased in cells derived from BAMS and FSHD patients (FSHD1or FSHD2). **E**. The TRIM28 (KAP1) transcription factor has been associated with DUX4 repression(16). We tested whether *TRIM28* expression is differentially regulated in hiPSCs (upper panel) or embryoid bodies (lower panel) for controls, BAMS, FSHD2 or FSHD1 by RT-qPCR. Expression was normalized to three housekeeping genes (*GAPDH*, *HPRT* and *PPIA*).

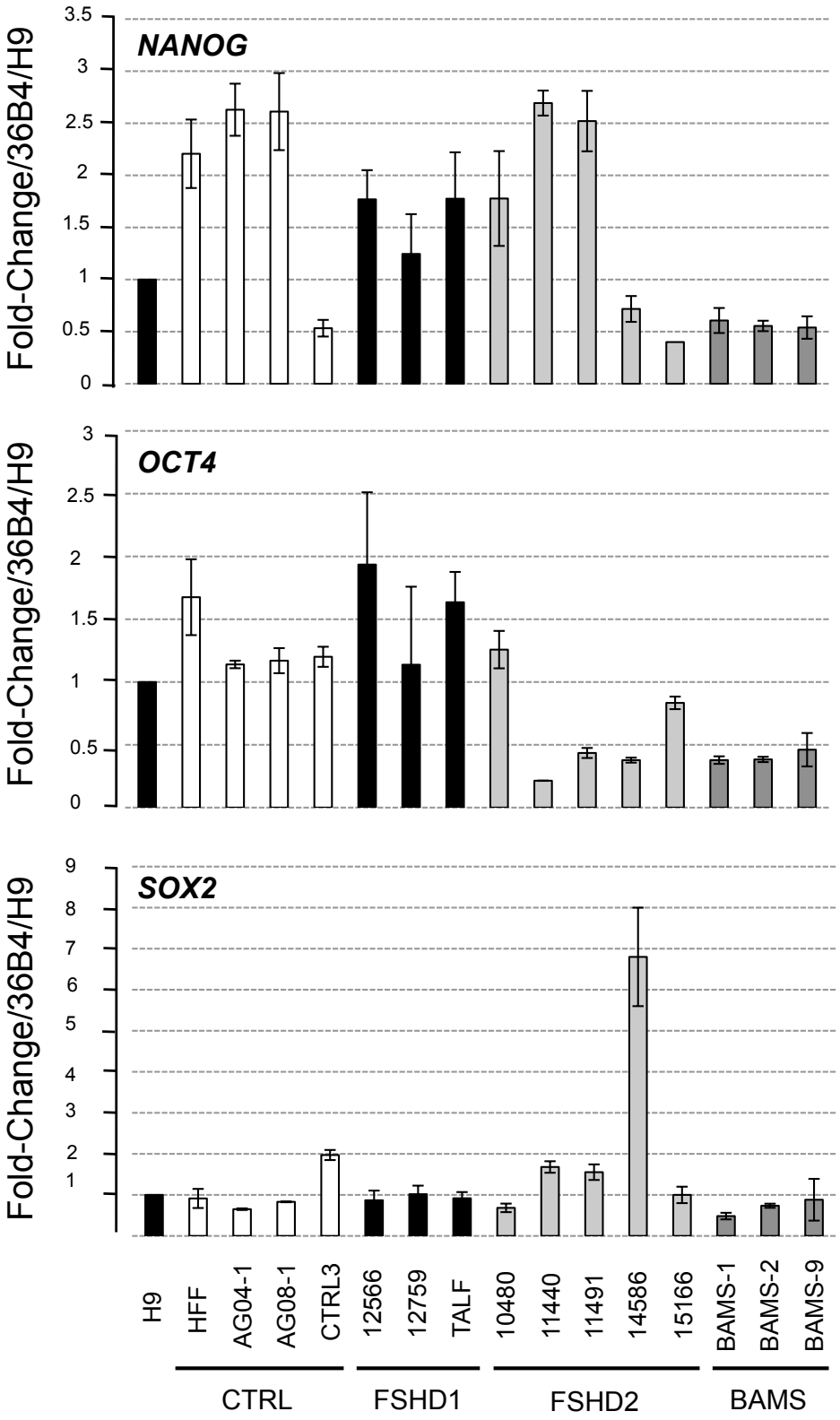
Supplementary figure 12. *DUX4* expression is induced by *SMCHD1* invalidation in the permissive 4qA-type HCT116 cancer cell line.

A. *SMCHD1* expression analyzed by western blotting in the different cell lines (HCT116, HCT116 invalidated for *SMCHD1* (KO), HEK293 cells, HEK293 invalidated for *SMCHD1* (KO). As described in the Materials and Methods section, *SMCHD1* was invalidated after transfection of Zinc finger nucleases targeting SMCHD1 exon 1. Staining of the Lamin B2 nuclear protein was used as loading control. **B.** Histograms display the mean methylation level ± SD (ratio between the number of methylated and unmethylated CpG for each DNA fragment) in the different HCT116 clones for different types of repetitive DNA sequences (AluY, LINE1, RS447, TAR1). **C.** The proximal 4q region is represented by greys rectangles. *ANT1*, *FRG1* and *FRG2* genes are indicated. The p13E11 used for Southern blotting and analysis of the D4Z4 array size is indicated (blue rectangle) upstream of the *D4Z4* array (green triangles). The 4qA and 4qB haplotypes correspond to different genomic elements distal to the D4Z4 and equally common in the population. The 4qA-type haplotype has been associated with FSHD and associated with production of the DUX4 long transcript (*DUX4-fl*) transcribed from the last D4Z4 repeat through the abutting pLAM sequence containing a polyadenylation site. This pLAM sequence is followed by an array of repeated β satellite elements upstream of the telomere (red arrows). The 4qB allele (depicted as a blue rectangle) differs from the 4qA by the absence of β satellite elements upstream of the telomere. **D.** The organization of the 4q35 region in HCT116 cells was analyzed by

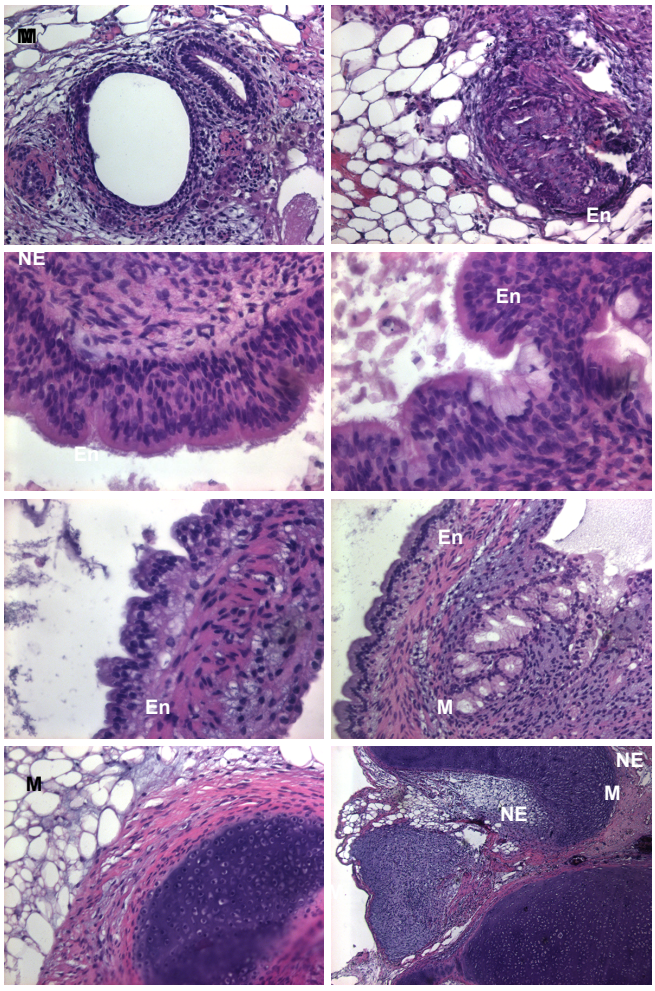
molecular combing. The V3 pink bar-code used to distinguish the different alleles (4qA/B, 10qA/B) is based on a combination of four different colors and different DNA probes encompassing the distal regions up to the telomeric sequence (3). The 4-color barcode comprises 2 probes detected in blue, which hybridize the proximal region common to chromosomes 4 and 10, one 6kb probe detected in red, which hybridizes in the (TTAGGG)n telomeric extremities, and a probe labeled in red that hybridizes the qA-specific β-satellite region, with a variable length (1–5kb). Proximal 4q-specific region is detected by a combination of red and pink probes. The D4Z4 array is detectable by the green probe. The length of the green signal is used to determine the size of the D4Z4 array and number of repeats. Based on this analysis, the HCT116 cell line contains one 4qA-type allele with a D4Z4 array of 103 kb, corresponding to approximately 30 D4Z4 repeated units. The same allele was detected in HCT116 dKO and HCT116 SMCHD1 KO cells. We did not specifically verify the presence of the polyadenylation site but the presence of the polyA signal sufficient to stabilize the DUX4-FI transcript was recently described in these cells(17). **E.** We used nested primers for amplification of the *DUX4-fl* transcript in HCT116, DKO and KO-SMCHD1 cells. Primers are localized in the first exon and downstream region containing DUX4 exon 3 and polyadenylation site. **F.** Representative images of ethidium bromide stained agarose gels are shown. Reaction tubes in which reverse transcriptase was omitted (RT-) are presented as controls. In agreement with the presence of the distal pLAM sequence in which reverse primers were designed, we were able to detect DUX4-fl expression in this non-muscular cells suggesting that *DUX4* expression might be triggered in a non-muscular context. Of note, at the difference of results obtained *in vivo* from patients cells with loss, gain of function mutation or haploinsufficiency, SMCHD1 invalidation in somatic cells causes a significant decrease in the methylation level in the promoter region upstream of the DUX4 sequence which might be compatible with activation of the retrogene.

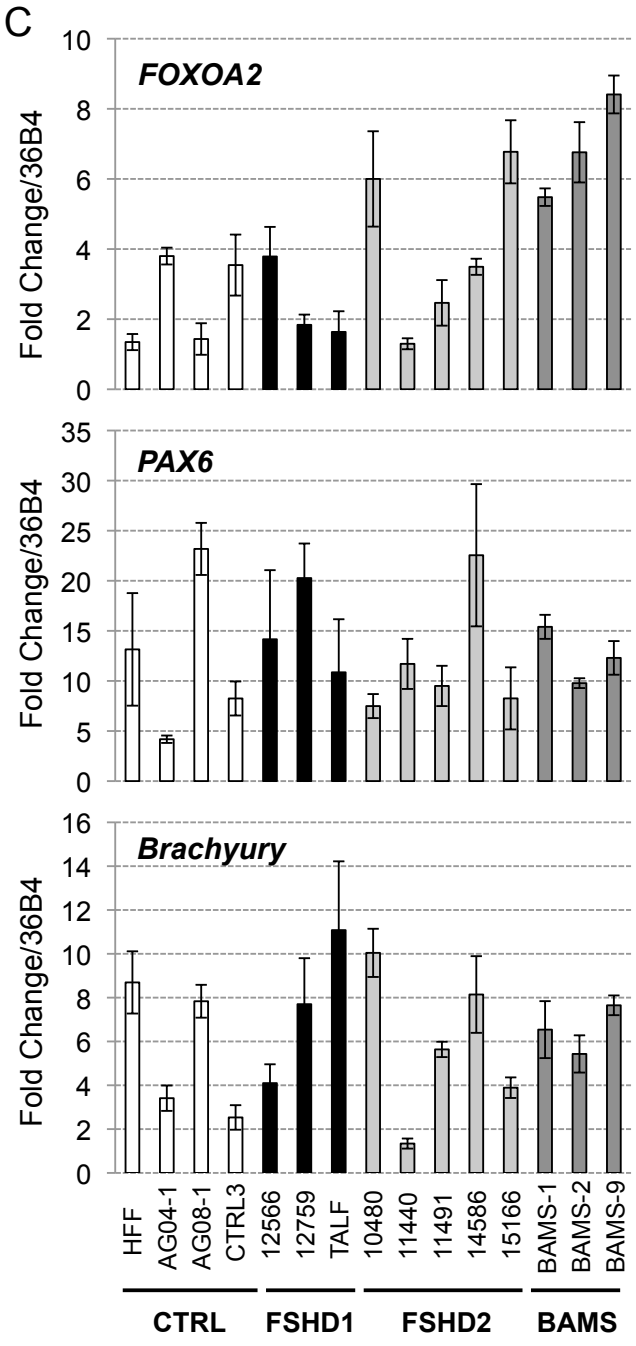
Supplementary references

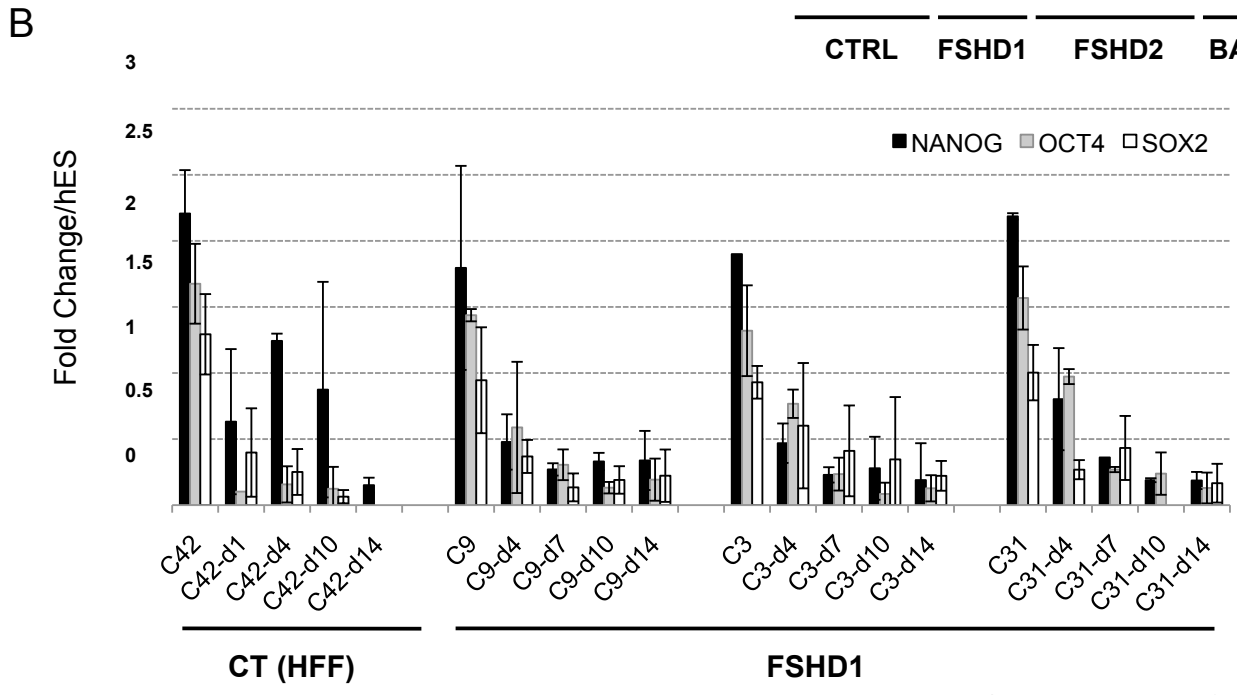
1. Rhee, I., Bachman, K.E., Park, B.H., Jair, K.W., Yen, R.W., Schuebel, K.E., Cui, H., Feinberg, A.P., Lengauer, C., Kinzler, K.W. *et al.* (2002) DNMT1 and DNMT3b cooperate to silence genes in human cancer cells. *Nature*, **416**, 552-556.
2. Marti, M., Mulero, L., Pardo, C., Morera, C., Carrio, M., Laricchia-Robbio, L., Esteban, C.R. and Izpisua Belmonte, J.C. (2013) Characterization of pluripotent stem cells. *Nat Protoc*, **8**, 223-253.
3. Nguyen, K., Walrafen, P., Bernard, R., Attarian, S., Chaix, C., Vovan, C., Renard, E., Dufrane, N., Pouget, J., Vannier, A. *et al.* (2011) Molecular combing reveals allelic combinations in facioscapulohumeral dystrophy. *Ann Neurol*, **70**, 627-633.
4. Magdinier, F., Billard, L.M., Wittmann, G., Frappart, L., Benchaib, M., Lenoir, G.M., Guerin, J.F. and Dante, R. (2000) Regional methylation of the 5' end CpG island of BRCA1 is associated with reduced gene expression in human somatic cells. *FASEB J*, **14**, 1585-1594.
5. Li, L.C. and Dahiya, R. (2002) MethPrimer: designing primers for methylation PCRs. *Bioinformatics*, **18**, 1427-1431.
6. Barnett, D.W., Garrison, E.K., Quinlan, A.R., Stromberg, M.P. and Marth, G.T. (2011) BamTools: a C++ API and toolkit for analyzing and managing BAM files. *Bioinformatics*, **27**, 1691-1692.
7. Bock, C., Reither, S., Mikeska, T., Paulsen, M., Walter, J. and Lengauer, T. (2005) BiQ Analyzer: visualization and quality control for DNA methylation data from bisulfite sequencing. *Bioinformatics*, **21**, 4067-4068.
8. Snider, L., Geng, L.N., Lemmers, R.J., Kyba, M., Ware, C.B., Nelson, A.M., Tawil, R., Filippova, G.N., van der Maarel, S.M., Tapscott, S.J. *et al.* (2010) Facioscapulohumeral dystrophy: incomplete suppression of a retrotransposed gene. *PLoS Genet*, **6**, e1001181.
9. Marsollier, A.C., Ciszewski, L., Mariot, V., Popplewell, L., Voit, T., Dickson, G. and Dumonceaux, J. (2016) Antisense targeting of 3' end elements involved in DUX4 mRNA processing is an efficient therapeutic strategy for facioscapulohumeral dystrophy: a new gene-silencing approach. *Hum Mol Genet*, **25**, 1468-1478.
10. Ludlow, A.T., Robin, J.D., Sayed, M., Litterst, C.M., Shelton, D.N., Shay, J.W. and Wright, W.E. (2014) Quantitative telomerase enzyme activity determination using droplet digital PCR with single cell resolution. *Nucleic Acids Res*, **42**, e104.
11. Ottaviani, A., Schluth-Bolard, C., Rival-Gervier, S., Boussouar, A., Rondier, D., Foerster, A.M., Morere, J., Bauwens, S., Gazzo, S., Callet-Bauchu, E. *et al.* (2009) Identification of a perinuclear positioning element in human subtelomeres that requires A-type lamins and CTCF. *Embo J*, **28**, 2428-2436.
12. Gordon, C.T., Xue, S., Yigit, G., Filali, H., Chen, K., Rosin, N., Yoshiura, K.I., Oufadem, M., Beck, T.J., McGowan, R. *et al.* (2017) De novo mutations in SMCHD1 cause Bosma arhinia microphthalmia syndrome and abrogate nasal development. *Nat Genet*, **49**, 249-255.
13. Badja, C., Maleeva, G., El-Yazidi, C., Barruet, E., Lasserre, M., Tropel, P., Binetruy, B., Bregestovski, P. and Magdinier, F. (2014) Efficient and Cost-Effective Generation of Mature Neurons From Human Induced Pluripotent Stem Cells. *Stem cells translational medicine*.
14. Cody, J.D. and Hale, D.E. (2015) Making chromosome abnormalities treatable conditions. *Am J Med Genet C Semin Med Genet*, **169**, 209-215.
15. Hasi-Zogaj, M., Sebold, C., Heard, P., Carter, E., Soileau, B., Hill, A., Rupert, D., Perry, B., Atkinson, S., O'Donnell, L. *et al.* (2015) A review of 18p deletions. *Am J Med Genet C Semin Med Genet*, **169**, 251-264.
16. De Iaco, A., Planet, E., Coluccio, A., Verp, S., Duc, J. and Trono, D. (2017) DUX-family transcription factors regulate zygotic genome activation in placental mammals. *Nat Genet*, **49**, 941-945.
17. Das, S. and Chadwick, B.P. (2016) Influence of Repressive Histone and DNA Methylation upon D4Z4 Transcription in Non-Myogenic Cells. *PLoS One*, **11**, e0160022.

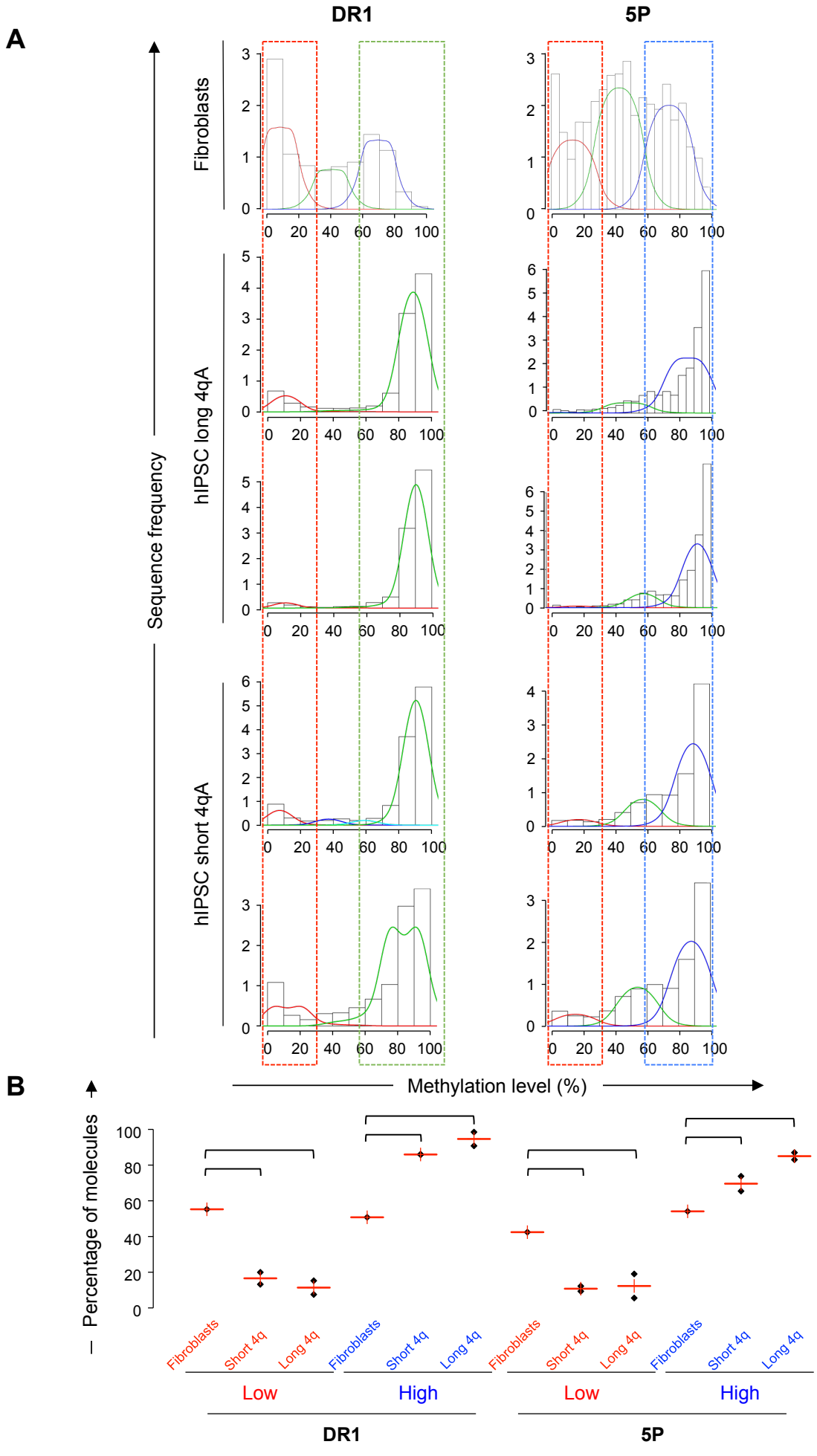


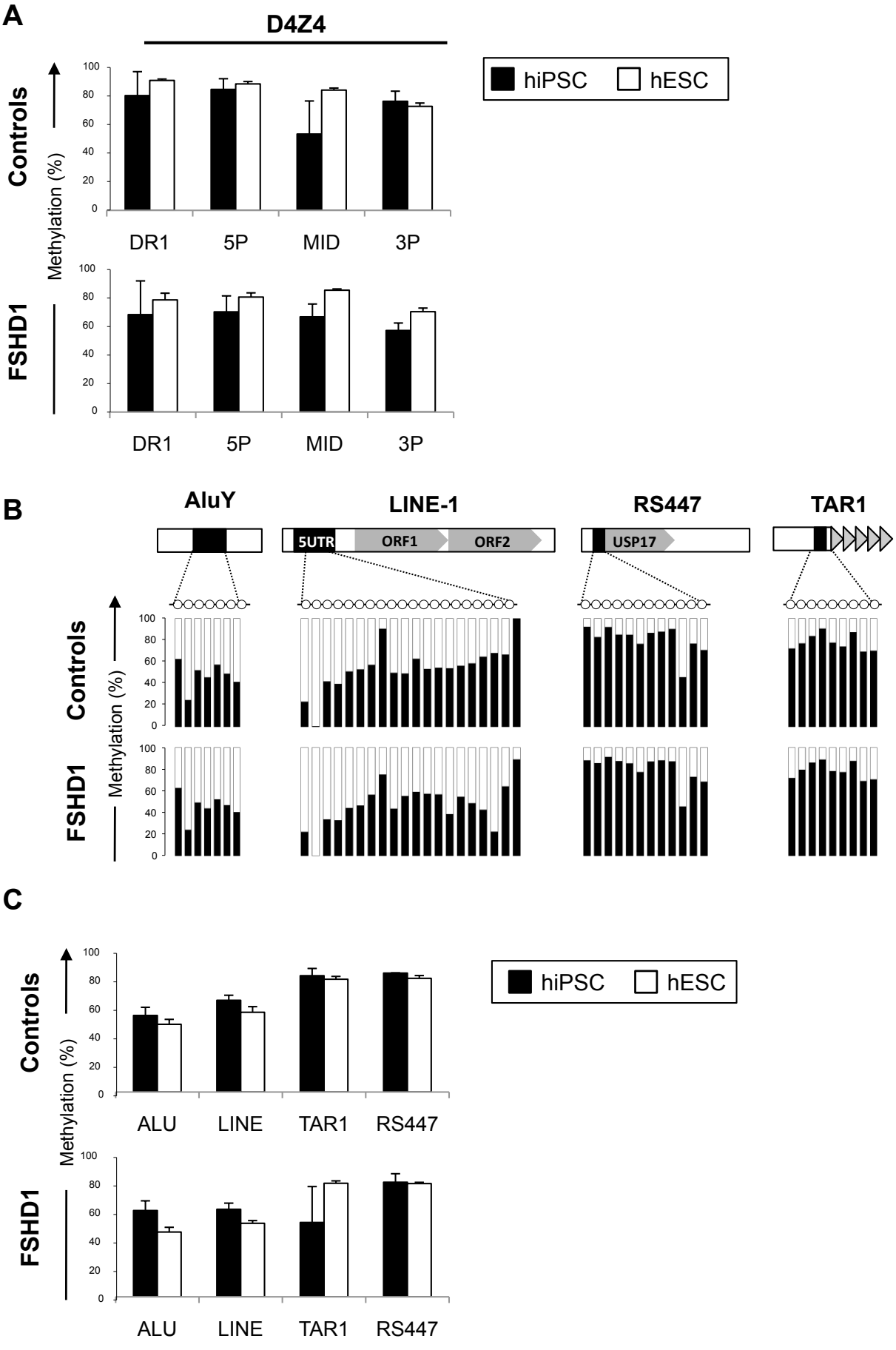
Dion et al. **Supplementary figure 1**

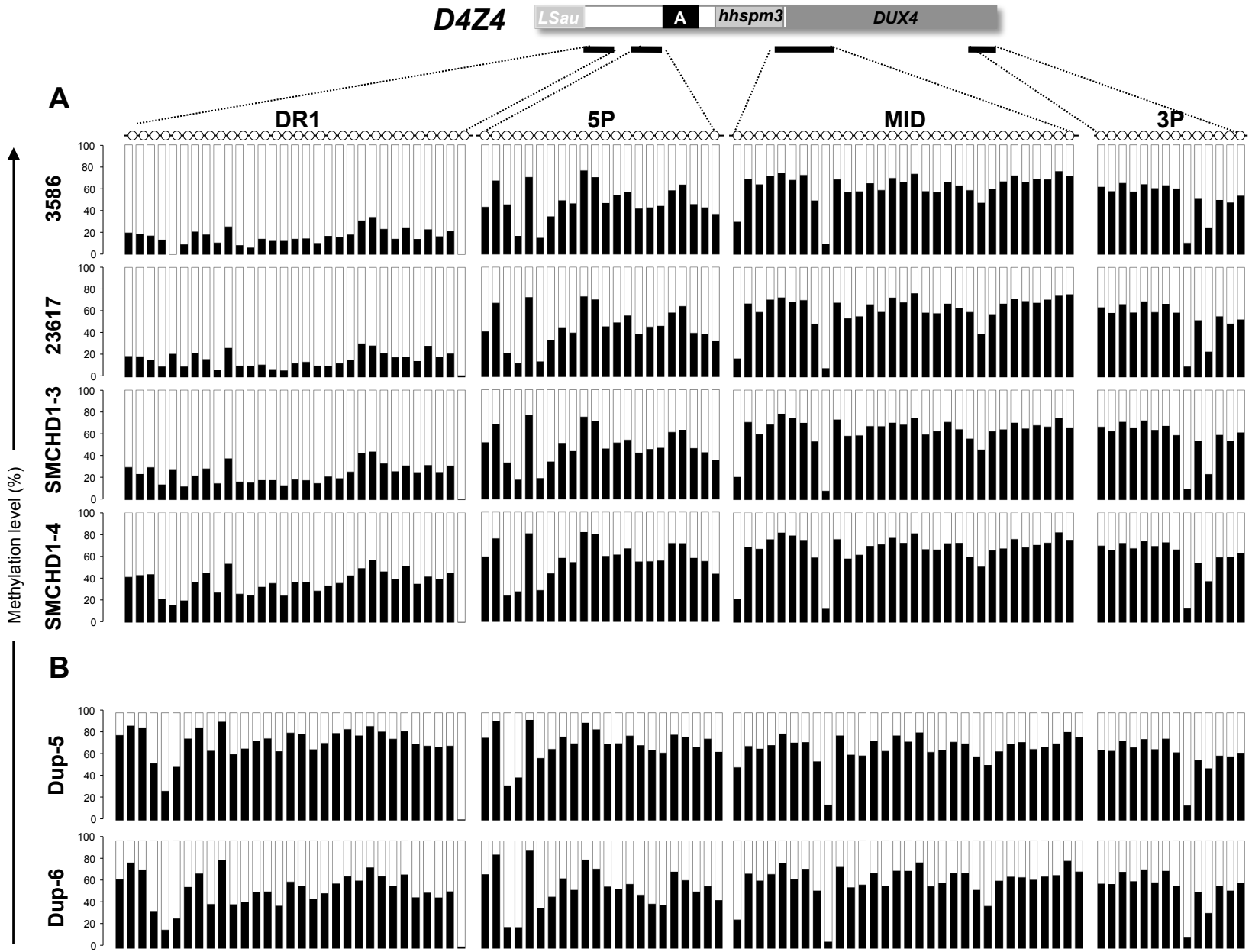

A
M
En
NE
En
En
En
En
M
M
NE
NE
M

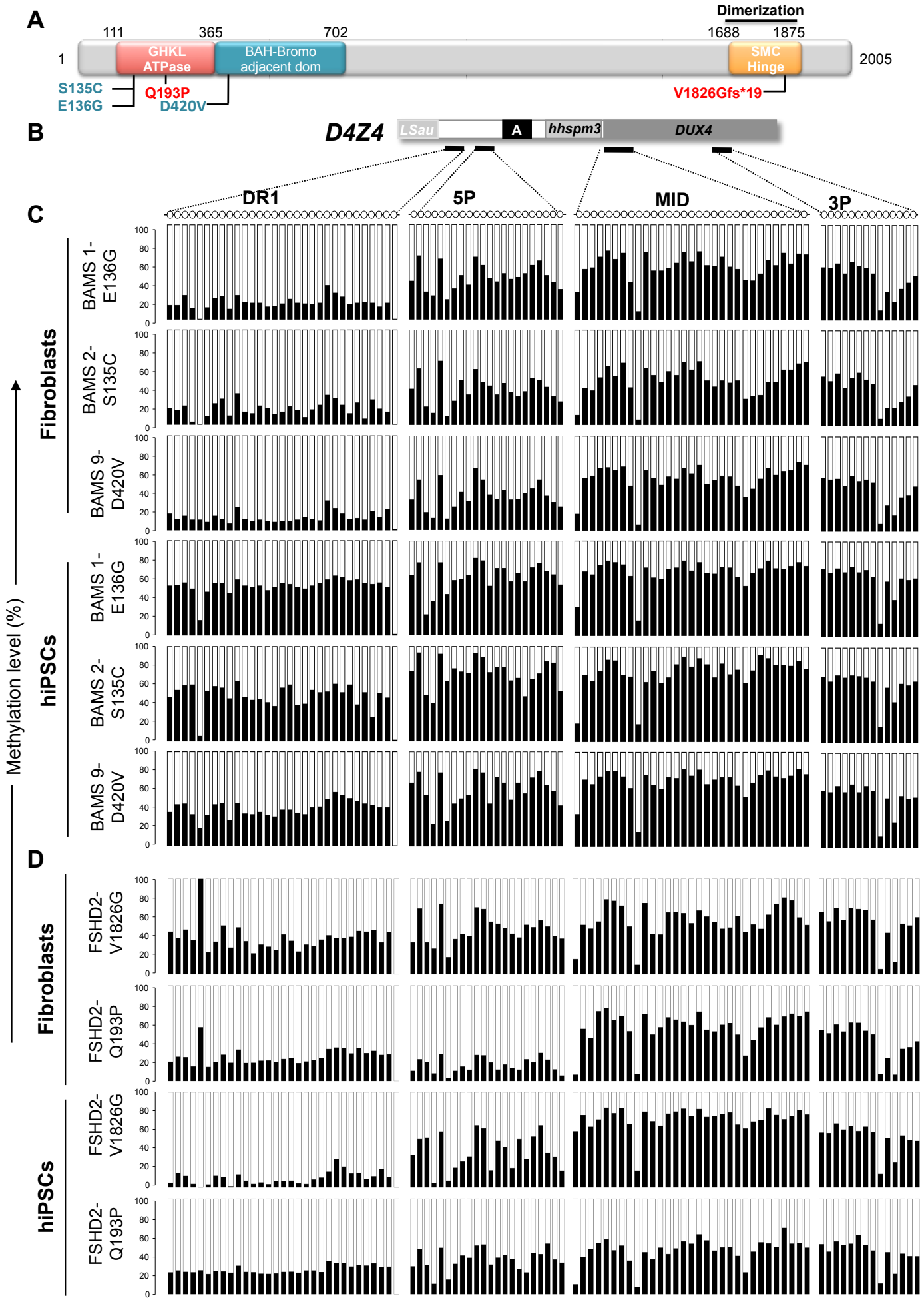

C
FOXOA2
PAX6
Brachyury
Fold Change/36B4
HFF
AG04-1
AG08-1
CTRL3
12566
12759
TALF
10480
11440
11491
14586
15166
BAMS-1
BAMS-2
BAMS-9
CTRL
FSHD1
FSHD2
BAMS


B
Fold Change/hES
NANOG
OCT4
SOX2
C42
C42-d1
C42-d4
C42-d10
C42-d14
C9
C9-d4
C9-d7
C9-d10
C9-d14
C3
C3-d4
C3-d7
C3-d10
C3-d14
C31
C31-d4
C31-d7
C31-d10
C31-d14
CT (HFF)
FSHD1


A
DR1
5P
Fibroblasts
hIPSC long 4qA
hIPSC short 4qA
Sequence frequency
Methylation level (%)
B
Percentage of molecules
Fibroblasts
Short 4q
Long 4q
Low
High
DR1
5P

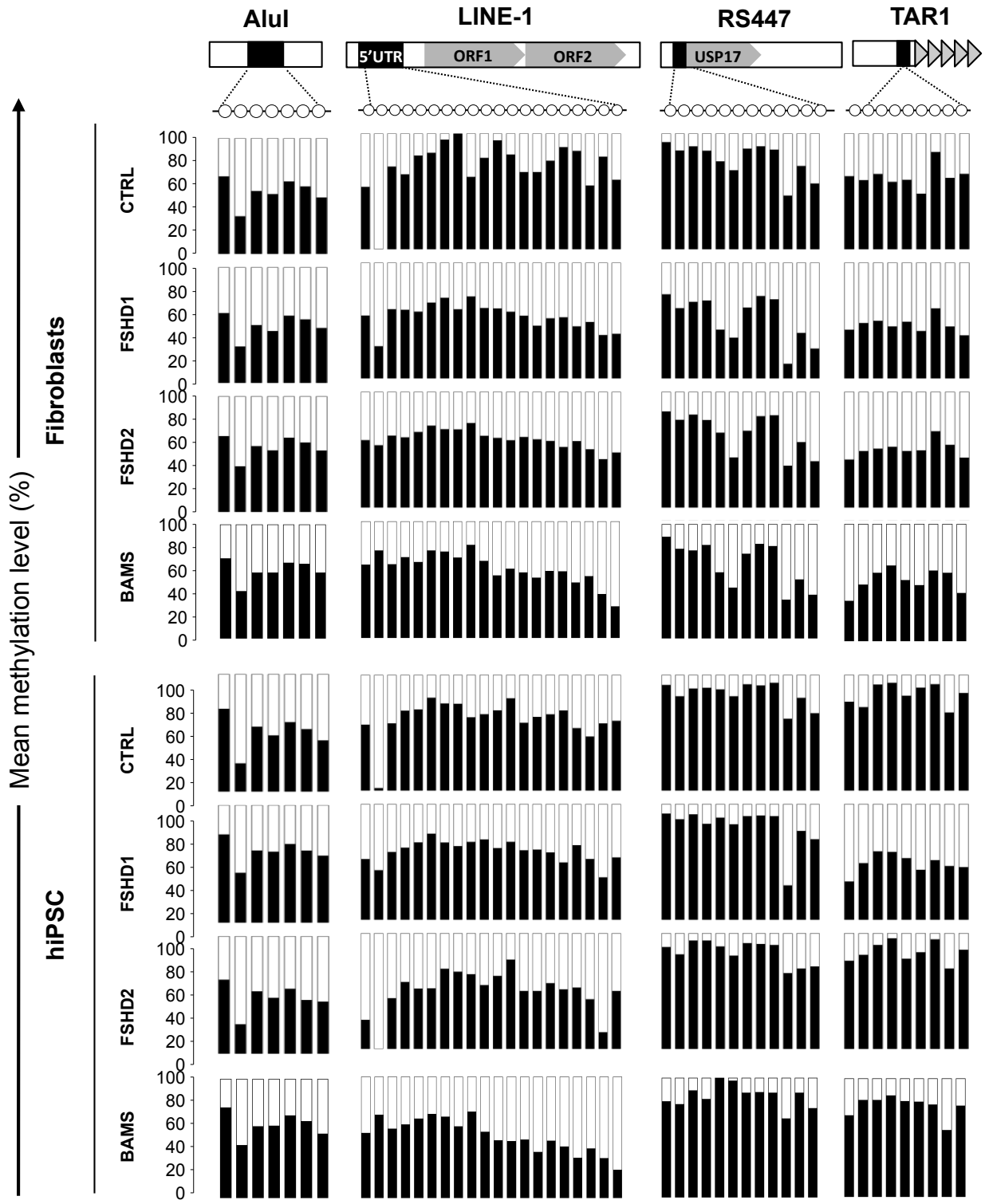

A
D4Z4
hiPSC
hESC
Controls
FSHD1
Methylation (%)
DR1
5P
MID
3P
B
AluY
LINE-1
RS447
TAR1
5UTR
ORF1
ORF2
USP17
C
ALU
LINE
TAR1
RS447

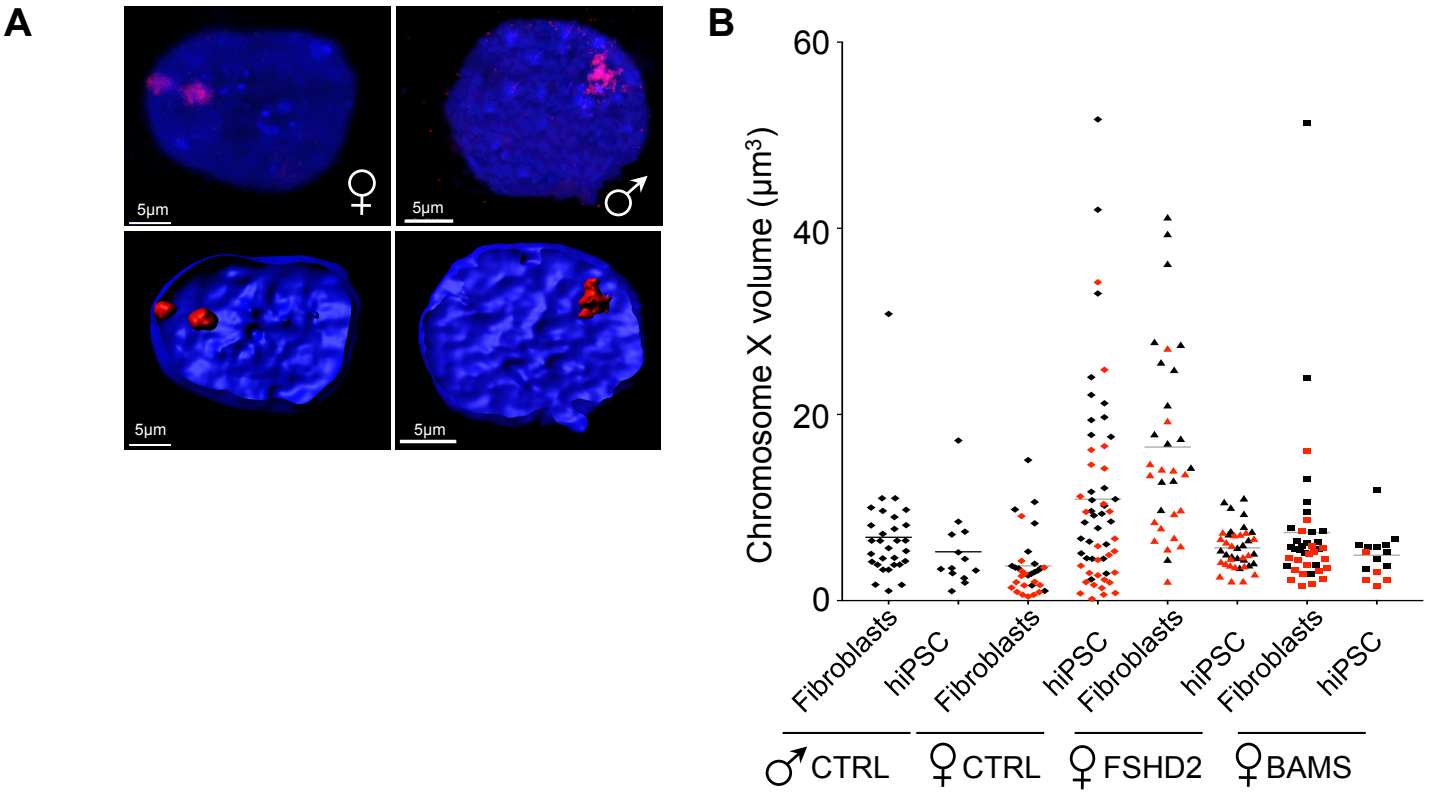

D4Z4
LSau
A
hhspm3
DUX4
A
DR1
5P
MID
3P
3586
23617
SMCHD1-3
SMCHD1-4
Methylation level (%)
B
Dup-5
Dup-6

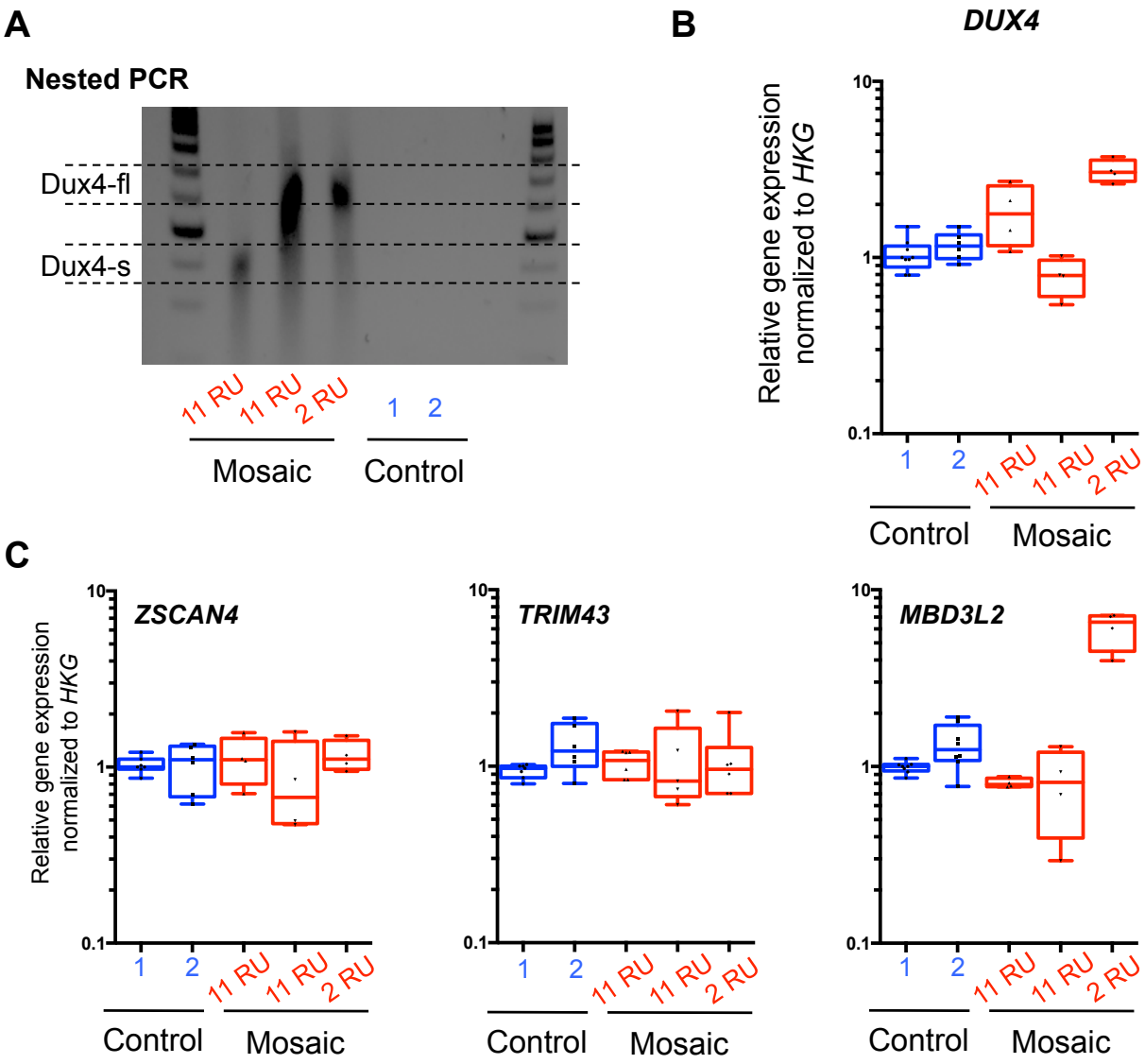


Dion et al.

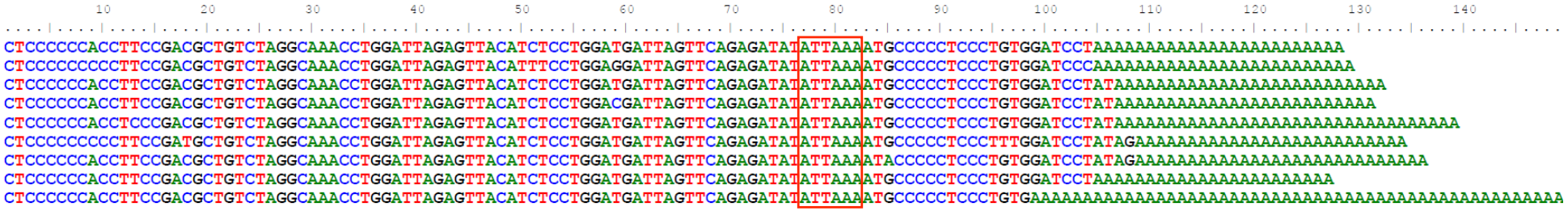
Supplementary figure 6


AluI
LINE-1
RS447
TAR1
5'UTR
ORF1
ORF2
USP17
Mean methylation level (%)
Fibroblasts
hiPSC
CTRL
FSHD1
FSHD2
BAMS
0
20
40
60
80
100

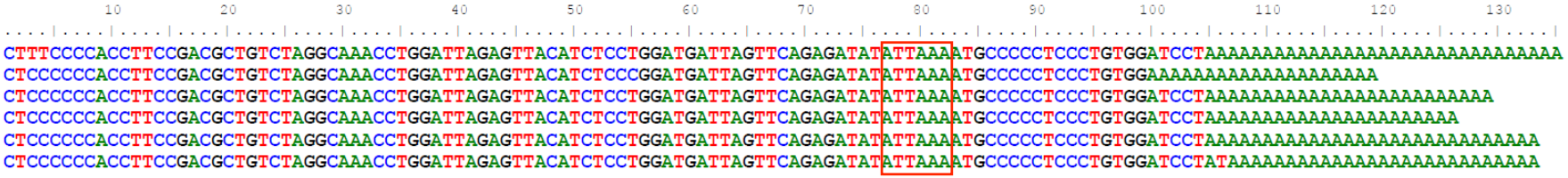

A
5µm
5µm
5µm
5µm
B
Chromosome X volume (µm3)
60
40
20
0
Fibroblasts
hiPSC
Fibroblasts
hiPSC
Fibroblasts
hiPSC
Fibroblasts
hiPSC
CTRL
CTRL
FSHD2
BAMS


A
Nested PCR
Dux4-fl
Dux4-s
11 RU
11 RU
2 RU
1
2
Mosaic
Control
B
DUX4
Relative gene expression normalized to HKG
10
1
0.1
1
2
11 RU
11 RU
2 RU
Control
Mosaic
C
ZSCAN4
TRIM43
MBD3L2
Relative gene expression normalized to HKG
10
1
0.1
1
2
11 RU
11 RU
2 RU
Control
Mosaic

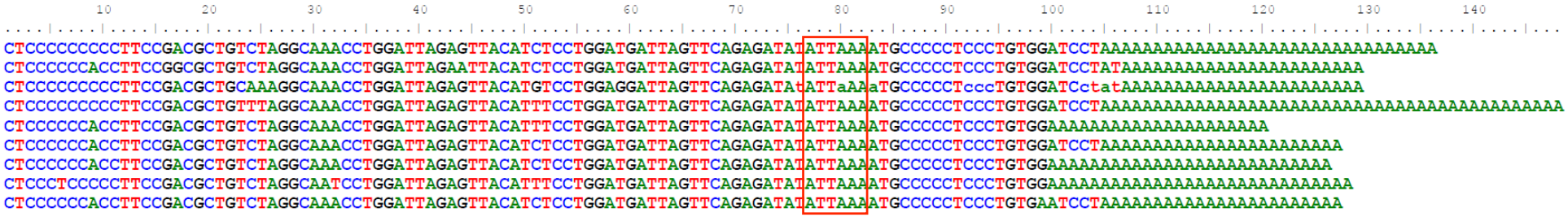
FSHD2-11440



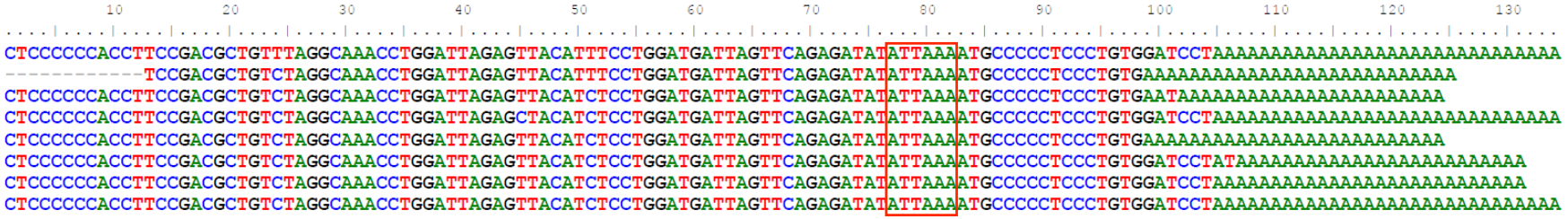
FSHD2-11491

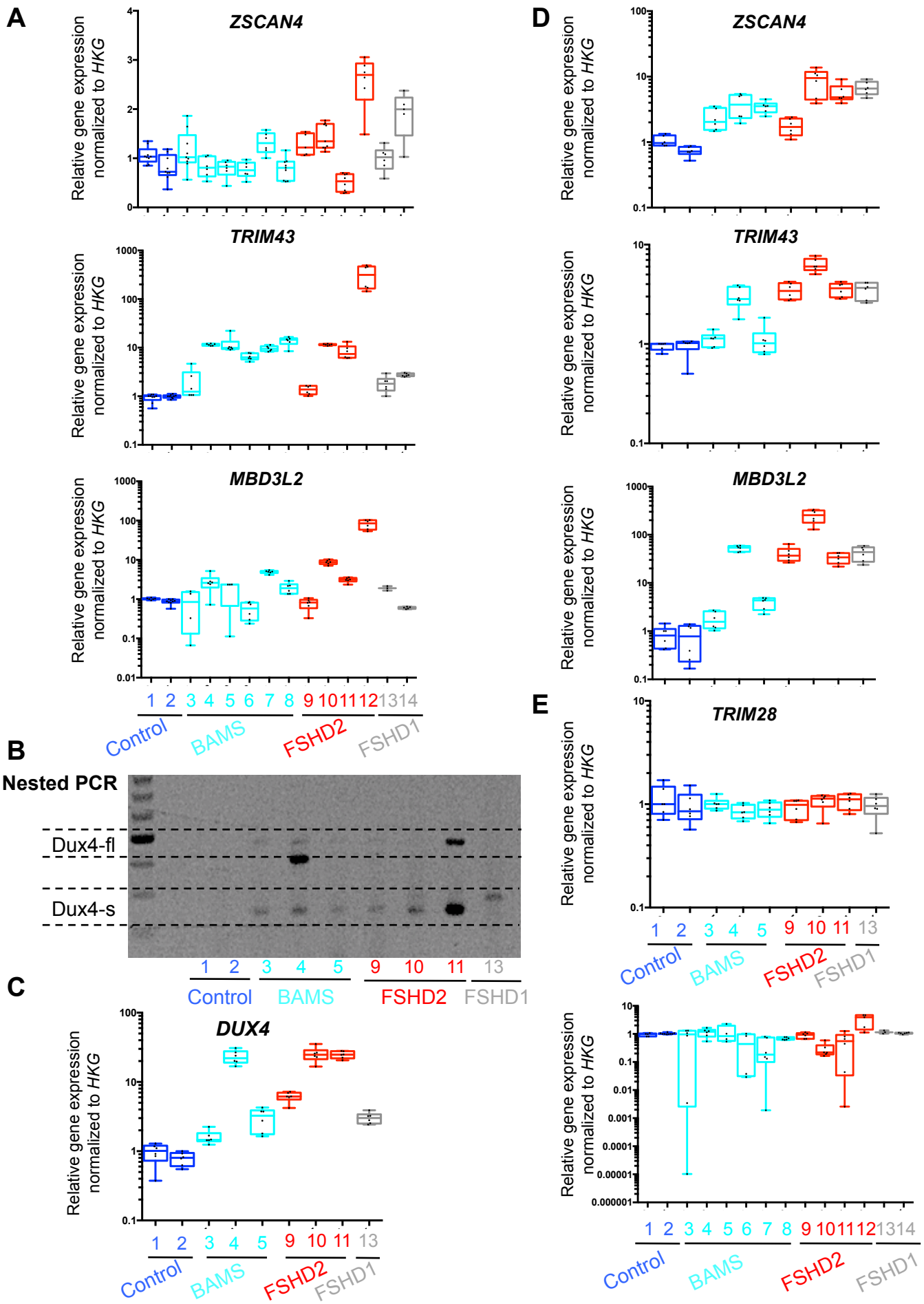


FSHD2-14586



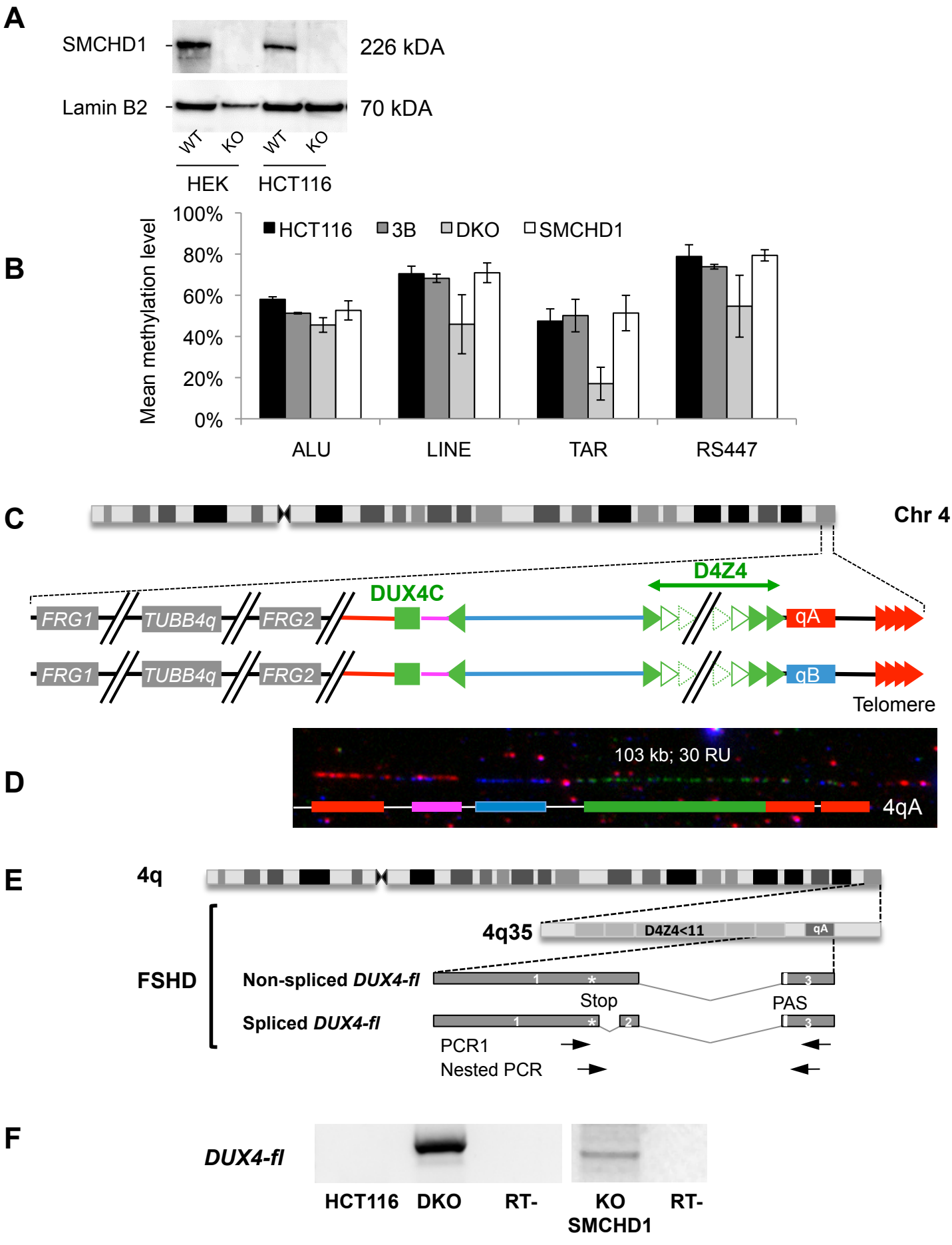
BAMS-2





Dion et al.

Supplementary figure 11



Dion et al.

Supplementary Figure 12