**Supplementary information**

**Materials and Methods.**

**Sodium Bisulfite sequencing.**

DNA was extracted from the different types of samples using the Qiagen DNA prep kit. For bisulfite modification, 1 μg of genomic DNA was denatured for 30 minutes at 37°C in NaOH 0.4N and incubated overnight in a solution of 3M Sodium bisulfite pH5 and 10mM Hydroquinone using a previously described protocol [1](#_ENREF_1). Converted DNA was then purified using the Wizard DNA CleanUp kit (Promega) following manufacturer’s recommendation and recovered after ethanol precipitation for 5 hours at -20°C. After centrifugation, the DNA pellet was resuspended in 40μL of water and stored at -20°C until use.Converted DNA was amplified using primer sets designed with the MethPrimer software [2](#_ENREF_2) avoiding the presence of CpGs in the primer sequence in order to amplify both methylated and unmethylated DNA with the same efficiency (Table e1). For Ion Torrent sequencing, amplicons were modified by addition at the 5’ end of barcode/adaptors sequences and at the 3’ end of a Tag for sequence amplification. Amplification was carried out using the High Fidelity Taq polymerase (Roche) according to manufacturer’s instructions. After initial denaturation at 94°C for two minutes, amplification was done after 10 cycles at 94°C for 20 seconds, 56°C for 30 seconds, 72°C for one minute and 25 cycles at 94°C for 20 seconds, 54°C for 30 seconds, and an amplification step of 4 minutes and 30 seconds for the first cycle and an implement of 10 seconds at each subsequent cycle. At the end of the program, a final extension step at 72°C for 7 minutes was done. For primers described in [3](#_ENREF_3), nested PCR was performed as described, with 10% of the first amplification used in the second amplification step.

PCR products were cleaned up and purified using the AMPure XP kit following manufacturer’s instruction and recommendation for NGS applications (Beckman Coulters). After purification, DNA was quantified using the Qubit® dsDNA High Sensitivity Assay Kit (Thermofisher). An equimolar mix of all regions and samples/barcodes is done. DNA high-throughput sequencing was performed using Ion PGM Sequencing 400 kit (Life Technologies) reagents. Ion Sphere Particles (ISP) were loaded onto an Ion 316R sequencing chip (Life Technologies) and DNA sequencing was performed with the Ion PGM instrument at 650 run flows. Raw data were processed using the Ion Torrent platform-specific pipeline software Torrent Suite v5.0.4. Results are split in different files to isolate data assigned to each barcode using the Ion Torrent server. Sequences were identified by the presence of a 17-base pair index at the beginning or at the end of the sequence (table e-1). Trimming of barcodes is performed using the fastx\_barcode\_splitter.pl software and fastq\_to\_fasta software (Barcode Splitter & FASTX\_ToolKit, 0.0.13, <https://github.com/agordon/fastx_toolkit>) to prepare data for the BiQ Analyser HiMod software (http://biq-analyzer.bioinf.mpi-inf.mpg.de)[4](#_ENREF_4). Sense and antisense sequences were assembled in a single sequence and bam file were converted to fastq file. After trimming, data are analyzed using the Bamtools software (BamTools 2.4.0, <https://github.com/pezmaster31/bamtools> [5](#_ENREF_5)). After processing, data were exported in “tab” separated value or “tsv file” into the R software (version 3.4.2). The first column contains the Ion Torrent ID. The last two columns correspond to the name of the region amplified and sample number. Others columns correspond to the methylation status of each CpG. Aligned sequences with a conversion rate >90% and a recovery rate >50% were further analyzed. BiQ Analyzer HiMod converts sequencing data by using « 1 » for a methylated CG, « 0 » for unmethylated and « x » in case of misalignment. The percentage of coverage is calculated as the ratio between the numbers of « x » divided by the total number of letters per sequence. Three methylation score are calculated, (*i*) **the CpG methylation score** corresponds to the percentage of methylation for each CpG in the reference sequence is calculated as the ratio of number of methylated CpGs on the number of CpGs aligned to the reference sequence, (*ii*) **the sequence methylation score**, corresponds to the average methylation level of each sequence calculated as the ratio of methylated CpG of all aligned CpG for a given sequence and (*iii*) **the global methylation score** corresponds to the global level of methylation for each biological sample in a given region calculated as the ratio of methylated CpG with the number of aligned CpG for all sequences and CpG positions for a given biological sample. Graphics representations were performed using barplot, boxplot and hist functions.

**Statistical analysis.**

Statistical analyses were done with R (version 3.4.2) and RStudio softwares, dplyr library and ComplexHeatmap library from Bioconductor (version 3.6). DNA methylation levels were compared using a Wilcoxon non-parametric test (wilcox.test function). The significance threshold (α=0.05) was corrected for multiple comparisons using the Bonferroni method for False Discovery Rate (FDR) determination. Only values showing a significant correlation with a p-value < 0.05 are presented. Heatmaps and Manhattan distance calculation were used to simultaneously visualize clusters of samples and global methylation of the different sequences. Clusters of CpG by CpG differentially methylated were determined using an Euclidean distance calculation. For each sample, we used the Mixtools version 1.1.0 package in R that estimates the sampling distribution using a random sampling method and confidence intervals calculation. Bootstrap was used to determine a confidence interval for the ratio of methylated CG divided by the number of CG analyzed for each sample by constructing an interval centered at a point estimate with a margin of error equal to twice the standard error.

***In vitro* DNA methylation.**

For validation of the different primers to amplify methylated and unmethylated DNA with the same efficiency, 2 µg of a plasmid containing one repetition of the D4Z4 macrosatellite was methylated *in vitro* using the M.SssI CpG methylase (NEB) in a final of 50µL containing 5µL of enzyme buffer, 0.5µL of S-Adenosyl Methionine (SAM) (NEB). To ensure a good efficiency of the M.SssI enzyme, the plasmid was purified using the Ampure XP kit (Beckman Coulters) and submitted to a second step of methylation using the same protocol.

**Table e-1:** Sequence of the primers used for Sodium bisulfite PCR.

D4Z4 5P, MID and 3P were described in [6](#_ENREF_6); DR1 primers were described in [7](#_ENREF_7); 4qA-Ext and 4qA-Int were described in [3](#_ENREF_3).

| **Sequence** | **Name** | **Primers** | **Temperature** | **Size** |
| --- | --- | --- | --- | --- |
|
| D4Z4 | 5P | AAATATGTAGGGAAGGGTGTAAGTTGGAGAGAGGGTTTGGTATATTTAAG | 56°C | 275 |
| D4Z4 | MID | ATTCATGAAGGGGTGGAGCCTCAGAGAACGGCTGGCCCAGGCCAT | 56°C | 353 |
| D4Z4 | 3P | GTTTTGTTGGAGGAGTTTTAGGATGGGATTTTTGTTTTTTAGGTTTAG | 56°C | 171 |
| D4Z4 | DR1 | GAAGGTAGGGAGGAAAAGACTCAACCTAAAAATATACAATCT | 56°C | 254 |
| D4Z4 | 4qA-Ext | GTTTTGTTGGAGGAGTTTTAGGAAACATTCAACCAAAATTTCACRAAA | 54°C | 770 |
| D4Z4 | 4qA-Int | GTTTTGTTGGAGGAGTTTTAGGAAACAAAAATATACTTTTAACCRCCAAAAA | 54°C | 596 |
| RS447 | 5P | TGTGATTTTTTTTTGAATTGAGCTCACCTCCCAAATAAAATAAA | 56°C | 200 |
| LINE | 5UTR | GTAAGGGGTTAGGGAGTTTTTTTTATCTATACCCTACCCCCAAA | 56°C | 366 |
| ALU | - | GGATTATTTGAGGTTAGGAGATTCCCRAATAACTAAAACTACAA | 56°C | 146 |
| Telomere | TAR1 | GGAGTAGAGTTTTTTTTAGGTTAGATTAAACAAACAATACCCCCAAC | 56°C | 150 |

**Table e-2:** Blood samples (PBL) from healthy donors, FSHD1 and FSHD2 patients.

Samples for which distribution is presented in supplementary figure 6 are highlighted in gray. Most cases have been characterized by Molecular combing. Samples characterized by Southern blotting or for whom only the size of the shortened allele is available are indicated by \*.

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
|  | **Diagnosis** | **Genotype** | ***SMCHD1* status** | **Age** | **Gender** |
| 18248 | Control | >1095 RU | No mutation | 53 | Male |
| 18304 | Control | >10120RU | No mutation | 59 | Female |
| 18488 | Control | >1080RU | No mutation | 52 | Male |
| 18521 | Control | >1070RU | No mutation | 52 | Female |
| 18553 | Control | >10120RU | No mutation | 46 | Female |
| 21383 | Control | >10 | No mutation | 65 | Female |
| N1\* | Control | >10 | No mutation | 35 | - |
| N2\* | Control | >10 | No mutation | 38 | - |
| N4\* | Control | >10 | No mutation | 59 | - |
| 15165 | Control | >1092RU | No mutation | 37 | Female |
| 11016 | FSHD1 | 4qA 2-3RU4qA 19RU10qA 12RU10qA 14RU | No mutation | 37 | Female |
| FJR | FSHD1 | 4qA 5-6RU4qA 10RU10qA 24 RU10qA 19RU | No mutation | 55 | Male |
| 14557 | FSHD1 | 5-6RU | No mutation | 29 | Male |
| 12759 | FSHD1 | 6RU4qA 6RU4qA 34RU10qA 13RU10qB 18RU | No mutation | 51 | Female |
| BOR.JU | FSHD1 | 4qA 5RU4qB 13RU10qA 19RU10qA >140kb | No mutation | 33 | Female |
| 11104 | FSHD1 | 9RU | No mutation | 32 | Male |
| 17743 | FSHD1 | 4qA 6RU4qA 34 RU10qA 13RU10qB 18RU | No mutation | 80 | Male |
| 14383 | FSHD1 | 4qA 9RU4qA 9RU10qA 9RU10qA 32RU | No mutation | 26 | Male |
| L0400\* | FSHD1 | 7RU | No mutation | 44 | Female  |
| L 0700\* | FSHD1 | 9RU | No mutation | 71 | - |
| LA0100\* | FSHD1 | 7RU | No mutation | 55 | Female  |
| LA0200\* | FSHD1 | 8RU | No mutation | 56 | Female  |
| 16214\* | FSHD1 | 6RU | No mutation | 25 | Female |
| 15432 | FSHD1 | 4qA 7RU4qA 43RU10qB 17RU10qA 44RU | No mutation | 33 | Male |
| 15431 | FSHD1 | 4qA 7RU4qA 43RU10qA 15RU10qA 44RU | No mutation | 28 | Female |
| M0400 | FSHD1 |  4qA 7RU4qA 43RU10qB 15RU10qA 44RU | No mutation | 74 | Male |
| M1200 | FSHD1 | 9RU | No mutation | 64 | Female |
| M1303\* | FSHD1 | 7RU | No mutation | 49 | Female |
| M2000 | FSHD1 | 4qA 7RU4qB 28RU10qA 12RU10qA 24RU | No mutation | 66 | Female |
| M2100 | FSHD1 | 4qA 7RU4qA 21RU10qA 12RU10qA17RU | No mutation | 45 | Female |
| M2300 | FSHD1 | 7RU | No mutation | 73 | Female |
| 16881 | FSHD1 | 8-9RU | No mutation | 58 | Male |
| 16802 | FSHD1 | 4qA 9RU4qA 30RU10qA 7RU10qA 44RU | No mutation | 23 | Female |
| S0100\* | FSHD1 | 7RU | No mutation | 63 | Male |
| T1100\* | FSHD1 | 6RU | No mutation | 65 | Female |
| 14173 | FSHD1 | 4qA 2-3RU4qA 19RU10qA 12RU10qA 14RU | No mutation | 32 | Female |
| 15165 | FSHD1 | 4qA 9RU4qB 12RU10qA 21 RU10qA 21RU | No mutation | 55 | Male |
| 18161 | FSHD1 | 7RU | No mutation | 30 | Male |
| 14589 | FSHD2 | >104qA 11RU; 4qB 21RU10qA 11RU; 10qA 35RU  | c.4151G>T; p.Gly1384Val | 51 | Male |
| 14586 | FSHD2 | >104qA 13RU4qA 13RU10qA 14RU10qA 37RU | c.578A>C; p.Glu193Pro | 67 | Male |
| 11440 | FSHD2 | >104qA 11RU4qA 33RU10qA 14RU10qA 17RU | c.2338+4A>G; p.S754\* | 37 | Male |
| 18267 | FSHD2 | >104qA 15RU4qA 33RU10qA 14RU10qA 37RU | c.1436G>A ; pR479Q | 68 | Male |
| 17073 | FSHD2 | >104qA 11RU4qA 23RU10qA 18RU10qA 25RU | c.1580 C>T ; p.T527M | 24 | Female |
| 10480 | FSHD2 | >104qA 15RU4qA 48RU10qA 9RU10qA 30RU | c.2260+6T>C | 53 | Female |
| 11491 | FSHD2 | >104qA 22RU4qA 22RU10qA 6RU10qA 25RU | c.5476+3 A>G; r.5477\_5547del; p.V1826Gfs\*19 | 66 | Female |
| 15166 | FSHD2 | >104qA 11RU4qB 20RU10qA 25RU10qA 21RU | r.4614\_4615insTATAATA; p.A1539Yfs\*4 | 67 | Female |
| 17143 | FSHD2 | >104qA 14RU4qA 14RU10qA 7RU10qA 24 RU | c.2328delT; P776L\* | 53 | Female |
| 14581 | FSHD2 | >104qA 22RU4qA 35RU10qA 11RU10qA 4RU | c.4892A>T; D1631V | 51 | Male |

**Table e-3:** Blood samples (PBL) from mosaic FSHD1 patients. The genotype and percentage of mosaicism is indicated for each patient. The short allele is indicated in bold.

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **Patient ID** | **Frequency of contracted allele** | ***Genotype*** | **Year of birth** | **Gender** |
| 16111 | 31% | **4qA 6RU**-4qA 20RU-4qB 23RU10qA 9 RU-10qA 26 RU | 1953 | Female |
| 17706 | 10% | **4qA 2RU**-4qA 11RU-4qB 22RU10qA 38 RU-10qA 14 RU | 1962 | Female |
| 19179 | 13% | **4qA 1RU**-4qA 39RU-4qB 46RU10qA 17 RU-10qA 35 RU | 1977 | Male |
| 19180 | 28% | **4qA 3RU**-4qA 21RU-4qB 14RU10qA 6 RU-10qA 35 RU | 1965 | Male |
| 19181 | 29% | **4qA 3RU**-4qA 48RU-4qB 19RU10qA 8 RU-10qA 12 RU | 1958 | Male |
| 19182 | 34% | **4qA 3RU**-4qA 32RU-4qB 19RU10qA 18 RU-10qA 18 RU | 1993 | Male |
| 19183 | 14% | **4qA 4RU**-4qA 28RU-4qA 33RU10qA 18 RU-10qA 21 RU | 1974 | Male |
| 19184 | 23% | **4qA 2RU**-4qA 27RU-4qA 41RU10qA 28 RU-10qA 28 RU | 1974 | Male |
| 19186 | 21% | **4qA 3RU**-4qA 32RU-4qB 14RU10qA 19 RU-10qA 23 RU | 1966 | Male |
| 19187 | 45% | **4qA 2RU**-4qA 22RU-4qA 32RU10qA 15 RU-10qA 25 RU | 1958 | Female |
| 19271 | 35% | **4qA 3RU**-4qA 20RU-4qA 29RU10qA 22 RU-10qA 23 RU | 1965 | Female |
| 19272 | 45% | **4qA 4RU**-4qA 31RU-4qA 34RU10qA 22 RU-10qA 16 RU | 1964 | Female |
| 19274 | 52% | **4qA 3RU**-4qA 20RU-4qB 20RU10qA 5 RU-10qA 24 RU | 1986 | Male |
| 19275 | 42% | **4qA 11 RU**-4qA28 RU-4qB 27 RU10qA 27 RU-10qA 16 RU | N/A | Male |
| 19276 | 35% | **4qA 6RU**-4qA 47RU-4qB 16RU10qA 14 RU-10qA 27 RU | N/A | Male |
| 19279 | 42% | **4qA 2RU**-4qA 12RU-4qA 44RU10qA 12 RU-10qA 32 RU | 1978 | Female |

**Legends to the supplementary figures.**

**Figure e-1. Workflow for analysis of DNA methylation level after deep-sequencing.**

**A.** Raw data were processed using the Ion Torrent platform-specific pipeline software Torrent Suite v5.0.4. After trimming of each barcode using the fastx\_barcode\_splitter.pl software and fastq\_to\_fasta software, data are analyzed using a the BamTools 2.4.0 software [5](#_ENREF_5) to convert bam file to fastq file . Data are then processed using the BiQ Analyser HiMod software and R for statistical analysis. For each sequence, the percentage of recovery and percentage of conversion are calculated. **B.** Only sequences with a percentage of recovery > 50% and percentage of conversion >90% are kept for further analysis. Three methylation scores are calculated. **The CpG methylation score** corresponds to the percentage of methylation for each CpG in the reference sequence calculated as the ratio of number of methylated CpGs on the number of CpGs aligned to the reference sequence. **The sequence methylation score** corresponds to the average methylation level of each sequence, calculated as the ratio of methylated CpG of all aligned CpG for a given sequence. **The global methylation score** is the global level of methylation for each biological sample in a given region calculated as the number of methylated CpG with the number of aligned CpG for all sequences and CpG positions for a given biological sample. Graphical representations were performed using barplot, boxplot and hist functions. **C.** Histogram showing the percentage of methylation for a given sequence. Each bar corresponds to a CpG position in a given sequence. Black corresponds to the percentage of methylated CpG, white to the percentage of unmethylated CpG. **D.** Distribution of the number of sequence per methylation level. For each sample, we used the Mixtools package in R that estimates the distribution of methylated fragments using a random sampling method and confidence intervals calculation. Ten classes were defined depending on the score of sequence methylation (0-10 to 90-100 %). The frequency of molecule within each class is calculated as the area under the curve. The intersection between the two curves gives the average percent of methylation. The frequency of molecules with a low methylation score (considered as hypomethylated) is calculated as the area under the red curve and frequency of molecules with a high methylation score (considered as methylated), as the area under the green curve.

**Figure e-2. Validation of primers efficiency for determination of mean methylation level and percentage of hypomethylated sequences**. We performed *in vitro* methylation of a plasmid carrying one repeat of the D4Z4 macrosatellite using the M.SssI methylase and mixed different proportions of methylated and unmethylated plasmids. After sodium bisulfite modification, PCR amplification and deep sequencing, we determined the global level of methylation and proportion of hypomethylated sequences using our deep sequencing and bioinformatic workflow. **A.** Representation of D4Z4 with the 4 regions (DR1, 5P, MID and 3P). **B.** Relationship between the global methylation level and the proportion of hypomethylated D4Z4 plasmid for the different D4Z4 primers. We observed a linear correlation between the quantity of unmethylated plasmids and number of hypomethylated reads. **C**. We plotted the percentage of hypomethylated molecules detected after deep sequencing to the percentage of hypomethylated plasmid used in the assay.

**Figure e-3. Heatmap representation of D4Z4 methylation level in blood samples. A.** Methylation level was calculated for each individual CpG within the four D4Z4 subregions in blood DNA from controls, FSHD1 and FSHD2 patients. **B.** Schematic representation of the D4Z4 repeat and of the different regions analyzed by sodium bisulfite sequencing (DR1, 5P, MID, 3P). CpG sites significantly hypomethylated in FSHD1 and FSHD2 patients are colored in red. These CpGs are mainly clustered in the DR1 and 5P regions.

**Figure e-4. Alignment of D4Z4 with the different primers used for bisulfite sequencing.** The DR1, 5P, MID and 3P primers were blasted to the D4Z4 sequence present on chromosomes 4 and 10. Primers do not amplify D4Z4-like sequences present on other chromosomes.

**Figure e-5. Distribution of the number of sequence with low or high methylation in blood samples from controls, FSHD1 or FSHD2 patients.** For each sample, we used the Mixtools package in R to estimate the distribution of molecules with a low (red curve) or high (green curve) methylation level using a random sampling method and confidence intervals calculation. At least ten classes were defined depending on the methylation score (0-10 to 90-100 %). The global methylation level is calculated as the number of methylated CpG divided by the number of aligned CpG for all sequences for a given biological sample. The frequency of molecules with a low methylation score (considered as hypomethylated) is calculated as the area under the red curve and the frequency of molecules with a high methylation score (considered as methylated), as the area under the green curve.Representative data for one control, one FSHD1 and one FSHD2 patient are presented (highlighted in grey in table e-2). The number of hypomethylated molecules is statistically significant for the DR1 sequence in FSHD1 (p< 0.04) and FSHD2 (p< 0.001) and for the 5P region (FSHD1, p< 0.0001; FSHD2, p<0.004) compared to controls.

**Figure e-6. Methylation profiles of different repetitive DNA sequences in control, FSHD1 and FSHD2 samples. A.** Schematic representation of the genomic structure of the different types of repetitive DNA sequences.The genomic structure of the RS447 macrosatellite (Genbank accession number D38378.1) or TAR1 subtelomeric element (Genbank accession number M57753.1; 150bp fragment) is represented. For AluY (AluY, Repbase) short and LINE1 (L1HS, Repbase) long interspersed elements, consensus sequences are schemed. White arrows represent the open reading frame present within the repetitive elements. D4Z4 encodes the DUX4 double homeobox transcription factor. RS447 encodes the USP17 deubiquitinating enzyme [8](#_ENREF_8). Primers amplify a 200bp encompassing the *USP17* promoter region contained within RS447. Grey boxes correspond to CG-rich regions. The different repetitive sequences contain short repetitive elements such as L*Sau* repeats, (CCA)n or (CA)n microsatellites. **B.** We analyzed the methylation profile of different types of repetitive elements after sodium bisulfite modification and deep-sequencing. In particular, we focused our attention on AluY and LINE1, the RS447 macrosatellite element and the TAR1 subtelomeric repeat. A schematic representation of the different sequences is provided. The region analyzed is indicated together with the number of CpG within each target sequence. For AluY, we analyzed a 123bp region, a 350bp region within the 5’ UTR of LINE-1 elements, a 200bp region in the promoter region of the RS447 macrosatellite and a 150bp of the TAR1 subtelomeric repeat. Histograms display the percentage of methylated (black) and unmethylated (white) CpG for each position. Analysis was done in blood DNA samples from healthy donors (CTRL), FSHD1 and FSHD2 patients.

**Figure e-7. Analysis of the distribution of methylation sequences for the DR1 region to discriminate mosaic patients.** Representative distributions are presented for controls, FSHD1 and mosaic patients. For each sample, the number of sequenced molecules was plotted in different classes defined by their mean methylation level, from low (0-10% of methylated CpGs) to high (90-100% of methylated CpG). When the distribution does not follow a normal distribution, it is possible to divide sequenced molecules in two or more well-separated peaks using the "Mixtools" bootstrap algorithm in R. Curves display molecules showing distinct levels of methylation with differences between the overall mean of methylation and the mean methylation of each peak. Representative examples are presented for controls, FSHD1 and mosaic patients. Fourteen out of 16 mosaic patients display and intermediate level of methylation for the DR1 region (green curve, mosaic patients).

**Figure e-8. Alignment of the last D4Z4 repeat and adjacent qA sequence with primers used for bisulfite sequencing.** The 3P and 4qBSS primers were blasted. The 4q BSS sequence overlaps with the 3P region in its proximal part with 14 CGs common to both amplicons..

**Figure e-9**. **Analysis of CpG present at DUX4 splicing sites in the distal 4qA D4Z4 repeat.** **A**. Schematic representation of the D4Z4 repeat and position of the different primers used for bisulfite sequencing, including the distal 4q BSS and 3P primers. Since changes in methylation of exonic region are often associated with alternative splicing sites usage, we indicated CG sites present at splicing junctions. **B**. We plotted the mean methylation levels of the 4q BSS sequence in control blood samples (n=6), FSHD1 (n=10) and FSHD2 (n=5). Histogram bars correspond to individual CpG with black bars corresponding to methylated CGs and white bars to unmethylated sites. CG1-3; 14 and 56 overlap with exon boundaries (indicated by a red box). CG1-3 are also present in the region amplified with the 3P primers. **C.** Mean methylation level of the different CG encompassing exon boundaries in the distal part of the D4Z4 repeat and adjacent 4qA region. In the different types of samples determined using the 4qA BSS primers. Methylation is highly variable for CG1-3 present within D4Z4 in FSHD1 and 2 compared to control but the difference is not significantly different (p value = 0.059 between controls and FSHD1).

**References**

1. Magdinier F, Billard LM, Wittmann G, et al. Regional methylation of the 5' end CpG island of BRCA1 is associated with reduced gene expression in human somatic cells. FASEB J 2000;14:1585-1594.

2. Li LC, Dahiya R. MethPrimer: designing primers for methylation PCRs. Bioinformatics 2002;18:1427-1431.

3. Jones TI, Yan C, Sapp PC, et al. Identifying diagnostic DNA methylation profiles for facioscapulohumeral muscular dystrophy in blood and saliva using bisulfite sequencing. Clin Epigenetics 2014;6:23.

4. Bock C, Reither S, Mikeska T, Paulsen M, Walter J, Lengauer T. BiQ Analyzer: visualization and quality control for DNA methylation data from bisulfite sequencing. Bioinformatics 2005;21:4067-4068.

5. Barnett DW, Garrison EK, Quinlan AR, Stromberg MP, Marth GT. BamTools: a C++ API and toolkit for analyzing and managing BAM files. Bioinformatics 2011;27:1691-1692.

6. Gaillard MC, Roche S, Dion C, et al. Differential DNA methylation of the D4Z4 repeat in patients with FSHD and asymptomatic carriers. Neurology 2014.

7. Hartweck LM, Anderson LJ, Lemmers RJ, et al. A focal domain of extreme demethylation within D4Z4 in FSHD2. Neurology 2013.

8. Saitoh Y, Miyamoto N, Okada T, et al. The RS447 human megasatellite tandem repetitive sequence encodes a novel deubiquitinating enzyme with a functional promoter. Genomics 2000;67:291-300.