



HAL
open science

Methylation hotspots evidenced by deep sequencing in patients with facioscapulohumeral dystrophy and mosaicism

Stéphane Roche, Camille Dion, Natacha Broucqsault, Camille Laberthonnière, Marie-Cécile Gaillard, Jérôme Robin-Ducellier, Arnaud Lagarde, Francesca Puppo, Catherine Vovan, Charlene Chaix, et al.

► To cite this version:

Stéphane Roche, Camille Dion, Natacha Broucqsault, Camille Laberthonnière, Marie-Cécile Gaillard, et al.. Methylation hotspots evidenced by deep sequencing in patients with facioscapulohumeral dystrophy and mosaicism. *Neurology Genetics*, 2019, 5 (6), pp.e372. 10.1212/NXG.0000000000000372 . hal-02406985

HAL Id: hal-02406985

<https://amu.hal.science/hal-02406985>

Submitted on 12 Dec 2019

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

Methylation hotspots evidenced by deep sequencing in patients with facioscapulohumeral dystrophy and mosaicism

Stéphane Roche, PhD,* Camille Dion, PhD,* Natacha Broucqsault, PhD,* Camille Laberthonnière, MS, Marie-Cécile Gaillard, PhD, Jérôme D. Robin, PhD, Arnaud Lagarde, MS, Francesca Puppo, PhD, Catherine Vovan, BS, Charlene Chaix, BS, Emmanuelle Salort Campana, MD, Shahram Attarian, MD, PhD, Marc Bartoli, PhD, Rafaelle Bernard, MD, Karine Nguyen, MD, PhD,† and Frédérique Magdinier, PhD†

Correspondence

Dr. Magdinier
frederique.magdinier@univ-amu.fr

Neurol Genet 2019;5:e372. doi:10.1212/NXG.0000000000000372

Abstract

Objective

To investigate the distribution of cytosine-guanine dinucleotide (CpG) sites with a variable level of DNA methylation of the D4Z4 macrosatellite element in patients with facioscapulohumeral dystrophy (FSHD).

Methods

By adapting bisulfite modification to deep sequencing, we performed a comprehensive analysis of D4Z4 methylation across D4Z4 repeats and adjacent 4qA sequence in DNA from patients with FSHD1, FSHD2, or mosaicism and controls.

Results

Using hierarchical clustering, we identified clusters with different levels of methylation and separated, thereby the different groups of samples (controls, FSHD1, and FSHD2) based on their respective level of methylation. We further show that deep sequencing–based methylation analysis discriminates mosaic cases for which methylation changes have never been evaluated previously.

Conclusions

Altogether, our approach offers a new high throughput tool for estimation of the D4Z4 methylation level in the different subcategories of patients having FSHD. This methodology allows for a comprehensive and discriminative analysis of different regions along the macrosatellite repeat and identification of focal regions or CpG sites differentially methylated in patients with FSHD1 and FSHD2 but also complex cases such as those presenting mosaicism.

*First coauthors

†Last coauthors

From Aix Marseille University, INSERM, MMG (S.R., C.D., N.B., C.L., M.-C.G., J.D.R., A.L., F.P., E.S.C., S.A., M.B., R.B., K.N., F.M.); Département de Génétique Médicale (A.L., C.V., C.C., R.B., K.N.), AP-HM, Hôpital de la Timone enfants, Marseille; and Centre de référence pour les maladies neuromusculaires et la SLA (E.S.C., S.A.), AP-HM, Hôpital de la Timone, Marseille, France.

Go to [Neurology.org/NG](https://www.neurology.org/NG) for full disclosures. Funding information is provided at the end of the article.

The Article Processing Charge was funded by the authors.

This is an open access article distributed under the terms of the Creative Commons Attribution-NonCommercial-NoDerivatives License 4.0 (CC BY-NC-ND), which permits downloading and sharing the work provided it is properly cited. The work cannot be changed in any way or used commercially without permission from the journal.

Glossary

CG = cytosine-guanine; **CpG** = cytosine-guanine dinucleotide; **FSHD** = facioscapulohumeral dystrophy; **MID** = primers used to investigate the middle region of the D4Z4 repeat; **PBMC** = peripheral blood mononuclear cell.

One of the most frequent genetic diseases associated with epigenetic alterations is facioscapulohumeral dystrophy (FSHD), an autosomal dominant neuromuscular disorder with an estimated prevalence of 1:8,000 to 1:20,000.^{1,2} In 95% of patients (FSHD1, Online Mendelian Inheritance in Man #158900), the dystrophy is linked to deletion of an integral number of repetitive D4Z4 macrosatellites in the subtelomeric 4q35 locus.^{3–5} In the general population, the number of repeats varies between 11 and 150 units, whereas patients with FSHD1 carry 1 to 10 copies. D4Z4 array shortening occurs de novo in 30% of cases, with a rate of 40% of somatic mosaicism among them (i.e., around 4%–12% of all cases).⁶

D4Z4 is extremely GC-rich (70%),⁷ methylated in normal individuals⁸ and decreased by approximately 20% in patients with FSHD1. In 5% of individuals with a typical FSHD phenotype, there is no D4Z4 shortening (type 2 FSHD2; Online Mendelian Inheritance in Man #158901), but these patients display a marked D4Z4 hypomethylation,^{9–12} associated with mutation in the *SMCHD1* gene.^{13,14}

To further investigate D4Z4 methylation at a high-resolution level, we developed a deep sequencing method after sodium bisulfite DNA modification. By applying this technology to the analysis of 65 blood samples from controls, patients with FSHD1, patients with FSHD2 with *SMCHD1* mutation, and mosaic cases, we showed that hypomethylation of the proximal part of the repeat can be evidenced with a strong statistical significance in patients with FSHD1 and FSHD2 in comparison to healthy donors, but also in patients with different levels of mosaicism.

Methods

Data availability

Values for DNA methylation are provided in file e-1, links.lww.com/NXG/A191. Details on samples used in the study are provided in tables e-2 and e-3, links.lww.com/NXG/A190.

Standard protocol approvals, registrations, and patient consents

All individuals have provided written informed consent for the use of the DNA sample for medical research, and the study was conducted in accordance with the Declaration of Helsinki. Controls are randomly selected individuals or patient's relatives. Controls are neither carrier of any genetic mutation nor affected by any muscular pathology. Controls were selected in the same age range and sex representation as patients. All patients were diagnosed at the Department of Medical Genetics, La Timone Hospital, Marseille, either by Southern

blotting or molecular combing.^{15,16} Further details on methods are provided in the supplementary information section.

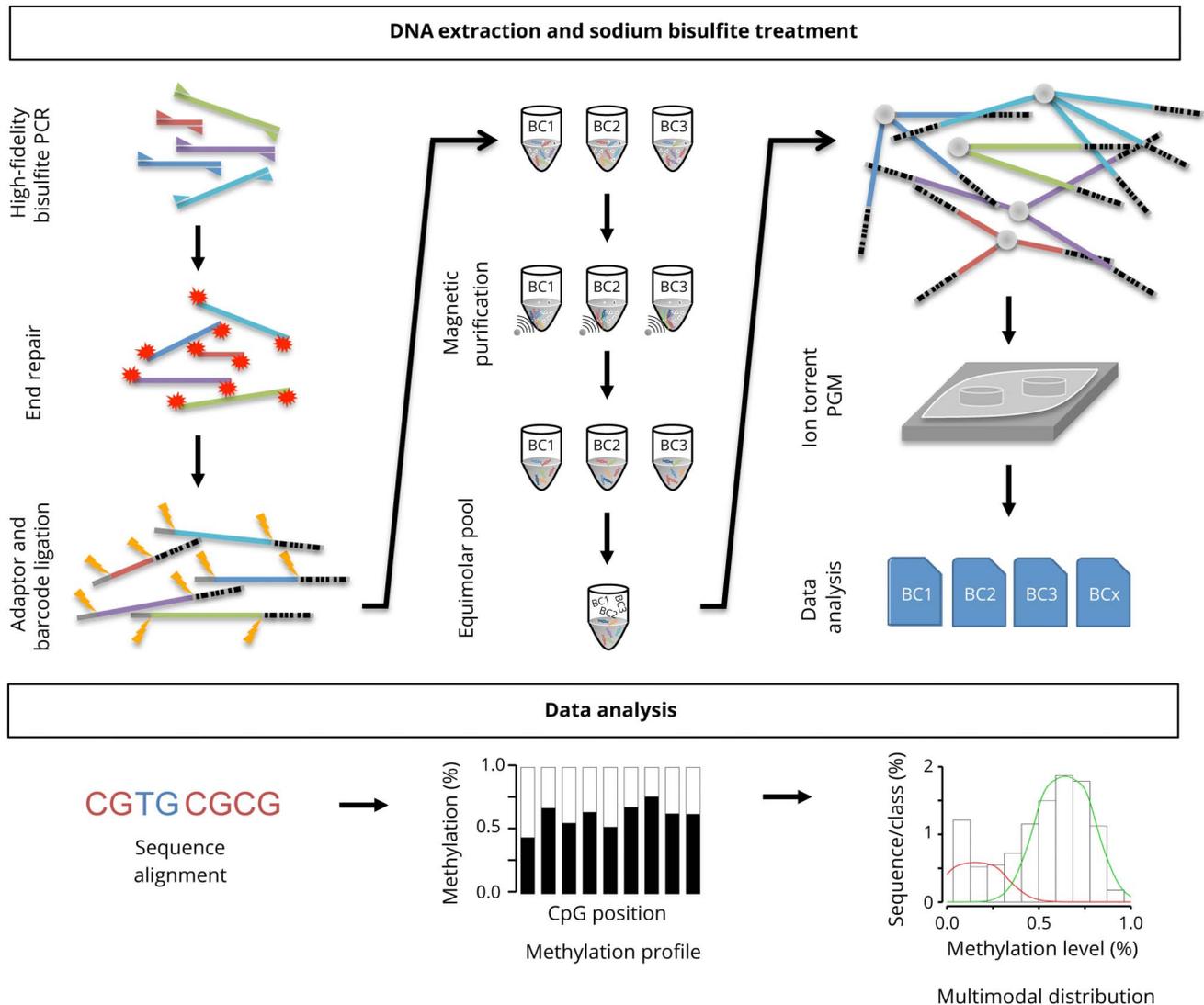
Results

Deep sequencing for analysis of DNA methylation after sodium bisulfite modification

Given the importance of epigenetic changes in deciphering the molecular mechanisms of FSHD, we developed a novel strategy for in-depth D4Z4 methylation analysis based on sodium bisulfite modification followed by deep sequencing (figure 1). The sodium bisulfite methodology based on the chemical modification of cytosines allowing DNA methylation profiling at a single base and single molecule level is considered as a gold standard for quantitative DNA methylation analysis.¹⁷ After deep sequencing, data are analyzed using a specific workflow (figure e-1A, links.lww.com/NXG/A190). Alignment of bisulfite-treated sequences is performed using BiQ Analyzer HiMod that converts aligned methylated cytosine-guanine dinucleotide (CpG) to 1 (cytosine-guanine [CG]) or 0 for unmethylated CpG (TG).¹⁸ Unaligned or unsequenced CGs, noted “x”, are not further considered except for determination of the sequencing coverage (figure e-1B, links.lww.com/NXG/A190). Then, using a custom algorithm in R, we generated the methylation profile of each region CpG by CpG and calculated the methylation level for each sequence (figure 1; figure e-1A, links.lww.com/NXG/A190). Only sequences with >95% of conversion are kept for analysis. We used stringent filters allowing only analysis of highly matching sequences with a minimum of 500 DNA molecules per barcode. Depth of coverage does not affect the accuracy of DNA methylation level determination (figure e-1B, links.lww.com/NXG/A190).

Methylation is determined at 3 different levels. The first gives the methylation level for each CpG. The second corresponds to the level of methylation of each DNA fragment (figure e-1C, links.lww.com/NXG/A190), and the third to methylation for each biological sample calculated as the ratio of methylated CG divided by the number of CGs analyzed. Homogeneity in the distribution of methylation or presence of sequences with a low or high methylation level is estimated using the mixtools package in R using a random sampling method and confidence intervals calculation to visualize the multimodal distribution of sequenced molecules based on their methylation level (figure 1, figure e-1D, links.lww.com/NXG/A190). When performing the entire assay in triplicate (DNA extraction, bisulfite conversion, PCR, sequencing, and quantification), we found a reduced variability with an SD of 3.6% for DR1, 6.5% for 5P, 1.6% for primers used to investigate the Middle region of the D4Z4 repeat (MID), and

Figure 1 Workflow for DNA methylation analysis



Genomic DNA was extracted from blood cells from different donors. After extraction, DNA was modified by sodium bisulfite treatment, PCR amplified, and analyzed by deep sequencing. Specific repetitive elements are amplified using a high-fidelity PCR in bisulfite-treated samples. After preparation of a first equimolar pool of PCR fragments and end repair, adaptors and barcodes are ligated. Samples are collected and cleaned up by magnetic purification. An equimolar pool of all barcoded samples is prepared before Ion Torrent PGM sequencing. Barcoded samples are submitted to an emulsion PCR and loaded onto a 316v2 Ion Torrent chip. After 650 flows of sequencing, barcodes and regions are split, and CpG alignment is performed using the BiQ Analyzer HiMod software.¹⁸ For each region, the global methylation level and methylation profile are determined. Random sampling method and confidence interval calculation identified 10 classes of molecules, depending on their methylation level from strongly hypomethylated (0%–10% of methylated CG per molecule) to highly methylated sequences (90%–100% of methylated CpG) allowing identification of subpopulations of DNA molecules with a low (red line) or high methylation (green line) level calculated as the area under each curve. CG = cytosine-guanine; CpG = cytosine-guanine dinucleotide.

5.1% for 3P in the calculation of the mean methylation level, validating the performance of the method.

Overall, this methodology requires less DNA than a Southern blot, centered on the analysis of a few CpGs in a single repeat. Compared with bisulfite sequencing requiring cloning and 10–15 colony picking or pyrosequencing, deep sequencing limits potential bias in sample selection, dilution, or masking of the methylation signal of underrepresented sequences because CGs are sequenced several times. Furthermore, direct sequencing allows samples multiplexing and analysis of heterogeneous biological samples.

Comprehensive analysis of D4Z4 methylation for determination of the D4Z4 methylation pattern at a single-base resolution

We first validated the different primers used for sodium bisulfite sequencing (DR1, 5P, MID, and 3P,^{10,11} table e-1, links.lww.com/NXG/A190) by assessing the power of discrimination between methylated and unmethylated molecules of our next generation sequencing assay and bioinformatics pipeline using an *in vitro* methylation assay. To this aim, we mixed different proportions of an *in vitro* methylated plasmid carrying 1 copy of the D4Z4 repeat and unmethylated plasmid DNA (0%–100% of *in vitro* methylated molecules). After bisulfite

conversion, PCR amplification, next generation sequencing, and sequence alignment, we analyzed the methylation profile in the different conditions and observed only low CpG to CpG variations for the different proportions of samples tested (figure e-2, links.lww.com/NXG/A190). Of note, the level of methylation did not reach 100% but 96.2% of global methylation indicating a capability of discriminating sequences with very high level of methylation (>95%) or molecules with a low methylation level (<10%). Furthermore, we demonstrated a linear relationship between theoretical and measured methylation, with a correlation of 0.979 for DR1, 0.987 for 5P, 0.998 for MID, and 0.995 for 3P (figure e-2, links.lww.com/NXG/A190).

We then applied our methodology to the analysis of D4Z4 methylation of peripheral blood mononuclear cell (PBMC) DNA from healthy donors (controls, $n = 10$), patients with FSHD1 ($n = 29$), and patients with FSHD2 mutated for *SMCHD1* ($n = 10$) (figures 2A and B, table e-2, links.lww.com/NXG/A190; file e-1, links.lww.com/NXG/A191) for the 4 different regions across the macrosatellite (DR1, 5P, MID, and 3P).^{10,11} The DNA methylation level of individual CpG sites is not uniform across D4Z4 (figure 2A). Mean DR1 and 5P methylation is high ($66.5\% \pm 19$ and $72.8\% \pm 10.2$) in controls but significantly decreased in FSHD1 (DR1: $46.1\% \pm 12.5$, $p = 3.1 \times 10^{-6}$ and 5P: $59.8\% \pm 7.8$, $p = 9.5 \times 10^{-7}$, respectively) and FSHD2 (DR1: $12.8\% \pm 5$, $p = 1.7 \times 10^{-5}$ and 5P: $42.7\% \pm 4.5$, $p = 9.2 \times 10^{-5}$) (figure 2B, file e-1, links.lww.com/NXG/A191).

Euclidean correlation and hierarchical clustering analysis reveal 3 major dendrogram branches that separate samples in 3 groups, from high to low methylation corresponding, respectively, to controls, FSHD1, and FSHD2 (figure 2C, figure e-3A, links.lww.com/NXG/A190). The heatmap of the relative methylation level for each individual CpG across the different D4Z4 subregions (figure e-3A, links.lww.com/NXG/A190) enabled the identification of critical CG hypomethylated in the disease, mostly clustered at DR1 and 5P (figure e-3B, links.lww.com/NXG/A190). As for the global methylation level (figure 2A, B), methylation of individual CGs is not different for the MID and 3P regions in the different groups of samples with exception for the MID-CpG.1, MID-CpG.23, and 3P-CpG.11, less methylated in patients with FSHD1 and FSHD2 compared with controls (figure 2A, figures e-3B, links.lww.com/NXG/A190).

Our primers amplify D4Z4 elements present both at the 4q35 and 10q26 regions (figure e-4, links.lww.com/NXG/A190) but do not amplify D4Z4 homologous elements present elsewhere in the genome and containing numerous polymorphisms.¹⁹ Given the absence of significant methylation change at the MID and 3P regions (figure 2D), we used methylation of these regions for normalization of methylation values to the total number of D4Z4 repeats on the 4q and 10q (table e-2, links.lww.com/NXG/A190). We confirmed the significant differences in the methylation ratio for the DR1 and 5P after normalization to the MID (figure 3A) or 3P methylation levels

(figure 3B) showing that analyses of different regions across D4Z4 can be used as internal standard for normalization of methylation changes and calculation of a mean methylation ratio for individual samples. When methylation was plotted as a function of the total number of repeats at the 4q and 10q in controls, FSHD1, and FSHD2, (table e-2, links.lww.com/NXG/A190), methylation for the DR1 and 5P significantly separated samples in the 3 different categories (controls, FSHD1, and FSHD2) but not MID and 3P (figure 2D), indicating that deep sequencing of bisulfite-converted DNA can be reliably used for analysis of D4Z4 methylation changes in patients with FSHD.

By analyzing the distribution of hypo- or hypermethylated molecules using random sampling and confidence interval calculation, the number of molecules in each subclass confirms the variability in D4Z4 methylation. Controls display a low proportion of hypomethylated molecules (mean methylation level <25%, red curves) and a higher proportion of highly methylated molecules (green curves, figure e-5, links.lww.com/NXG/A190). Patients with FSHD1 display an average of 50% of molecules with a low methylation level, whereas the number of hypomethylated molecules is strongly increased in patients with FSHD2.

Hypomethylation is restricted to D4Z4 in patients with FSHD

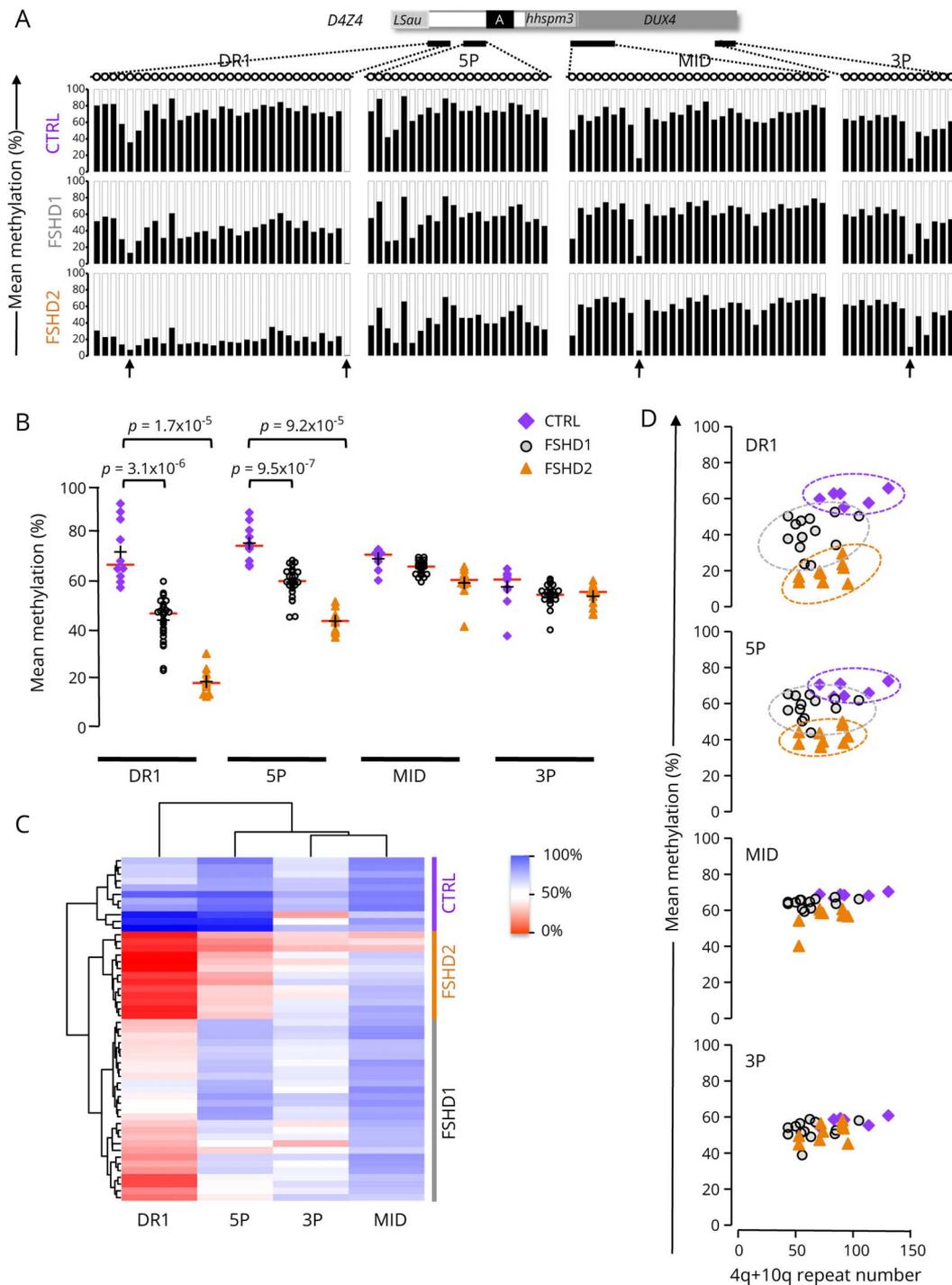
Repetitive DNA sequences represent the largest proportion of methylated DNA sequences in the human genome.²⁰⁻²³ Hence, to assess more globally methylation changes in the different conditions, we included in our analysis primers for interspersed short repetitive sequences such as AluY or the 5' UTR of LINE-1 retrotransposon dispersed throughout the genome, often used as a surrogate marker for global methylation analysis^{24,25} or additional repetitive elements such as the RS447 macrosatellite or TARI subtelomeric element (figure e-6A, links.lww.com/NXG/A190; table e-1, links.lww.com/NXG/A190).

For AluY, LINE-1, or RS447, we observed a stable level of methylation (figure e-6B; links.lww.com/NXG/A190; file e-1, links.lww.com/NXG/A191) in the different groups of samples. The global methylation level is not modified in patients with FSHD, including patients with FSHD2 mutated for *SMCHD1*, underlining a specific role for this factor in D4Z4 regulation.²⁶ Of interest, we observed a slight methylation decrease for the TARI subtelomeric element in FSHD1 blood samples ($41.4\% \pm 8.3$) compared with controls ($52.9\% \pm 11.9$, $p < 0.0034$) and patients with FSHD2 ($56.7\% \pm 7.6\%$, $p < 0.0005$), suggesting that besides D4Z4 FSHD patients might be also prone to hypomethylation of other subtelomeres (figure e-6, links.lww.com/NXG/A190; file e-1, links.lww.com/NXG/A191).

Methylation of the DR1 sequence for identification of hypomethylated alleles in patients with mosaicism

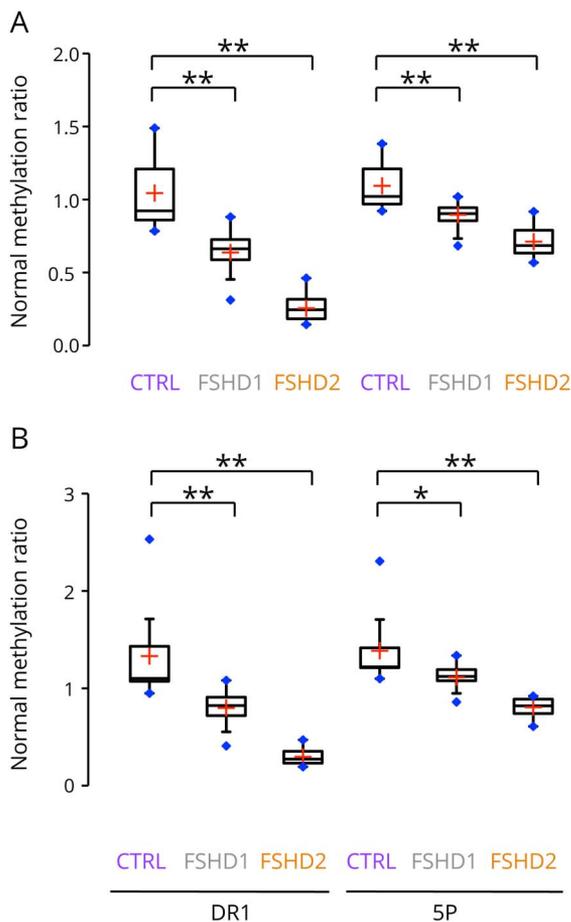
FSHD is characterized by a high proportion of somatic mosaicism. To test whether our methodology enabled estimation

Figure 2 D4Z4 methylation profile in blood cells from healthy donors, patients with FSHD1, or patients with FSHD2



(A) For each sample, DNA methylation was determined after sodium bisulfite for 4 regions within D4Z4 (DR1, [254bp; 31 CGs], 5P [275bp, 22 CGs], MID [353bp, 31 CGs], and 3P [171bp, 14 CGs]) after PCR amplification and deep sequencing. Primers were designed by avoiding the presence of CGs in the primers sequence to amplify methylated and unmethylated DNA with the same efficiency. For each sequence, dots above the histograms represent individual CpGs. Histogram bars represent the percentage of methylated (black) or unmethylated (white) CpG for each position. Histograms correspond to the average methylation level of blood cells DNA from 10 healthy donors, 29 patients with FSHD1, and 10 patients with FSHD2. CpGs with a low coverage are indicated by a black arrow and were discarded from the calculation of the mean methylation level. (B) Scatterplot distribution of the D4Z4 methylation level at DR1, 5P, MID, and 3P in control (n = 10), FSHD1 (n = 29), and FSHD2 (n = 10) PBMC DNA samples. Horizontal red lines correspond to the median for each group. Data sets were compared using the Wilcoxon nonparametric test, and brackets identify significantly different groups based on post hoc Dunn comparison and Bonferroni correction. *p* Values are indicated for each group of samples. (C) Heatmap analysis of the global methylation level in blood from healthy donors (n = 10, purple stars), patients with FSHD1 (n = 29, gray circles), and patients with FSHD2 (n = 10, orange triangles). The dendrogram representation discriminates the different subgroups of patients based on their respective level of methylation. (D) Scatterplot showing the percentage of methylation for the different D4Z4 regions as a function of the cumulative number of repeats from chromosomes 4 and 10 for the different subregions in blood samples. For the DR1 and 5P region, ellipses were used to discriminate the 3 groups (controls, FSHD1, and FSHD2). Methylation levels are not significantly different for the MID and 3P regions with overlap between the different groups of samples. FSHD = facioscapulohumeral dystrophy; PBMC = peripheral blood mononuclear cell.

Figure 3 D4Z4 methylation level in the proximal end of the repeat discriminates healthy donors from patients with FSHD1 and FSHD2



(A and B). Our primers amplify D4Z4 elements present on both chromosomes 4 and 10. Therefore, we used the methylation level determined for the MID and 3P regions to normalize the results obtained in the DR1 and 5P regions in the different groups of samples (controls, n = 10, FSHD1, n = 29, and FSHD2, n = 10). (A) Ratio of methylation in the DR1 and 5P regions normalized to the methylation level of the MID region. (B) Ratio of methylation in the DR1 and 5P regions normalized to the methylation level of the 3P region. Statistical significance was determined using a Mann-Whitney non-parametric test, ** $p < 0.0001$; * $p < 0.005$. CTRL = control; FSHD = facioscapulohumeral dystrophy.

of the degree of hypomethylation in this subset of complex cases, we analyzed methylation of the DR1, 5P, MID, and 3P regions in 16 patients for which the percentage of mosaicism and size of the different 4q and 10q alleles have been precisely determined by molecular combing¹⁶ (table e-3, links.lww.com/NXG/A190). By plotting the mean methylation level in these patients compared with controls, patients with FSHD1, and patients with FSHD2 for the different regions across D4Z4, we show that patients with mosaicism are significantly hypomethylated compared with controls (figure 4A, p value < 0.01) with a level of methylation for the DR1 sequence intermediate between the levels in controls and patients with FSHD1. We then compared the distribution of the number of sequence as a function of the percentage of methylation. In 14 of the 16 mosaic cases, we observed a 3 classes distribution

with a proportion of molecules with an intermediate level of methylation for DR1 (figure e-7, links.lww.com/NXG/A190). The remaining 2 samples correspond to patients with the higher proportion of short D4Z4 allele (42% and 51% of mosaicism, respectively). By plotting the percentage of molecules in the different classes of methylated molecules to the percentage of short allele for the different patients (figure 4B), we obtained a positive correlation between the level of methylation for DR1 and the percentage of mosaicism. We concluded that deep sequencing of bisulfite-converted DNA discriminates efficiently alleles with a decreased level of methylation from highly methylated ones in patients with mosaicism.

Methylation changes at the distal D4Z4 repeat and adjacent A-type sequence

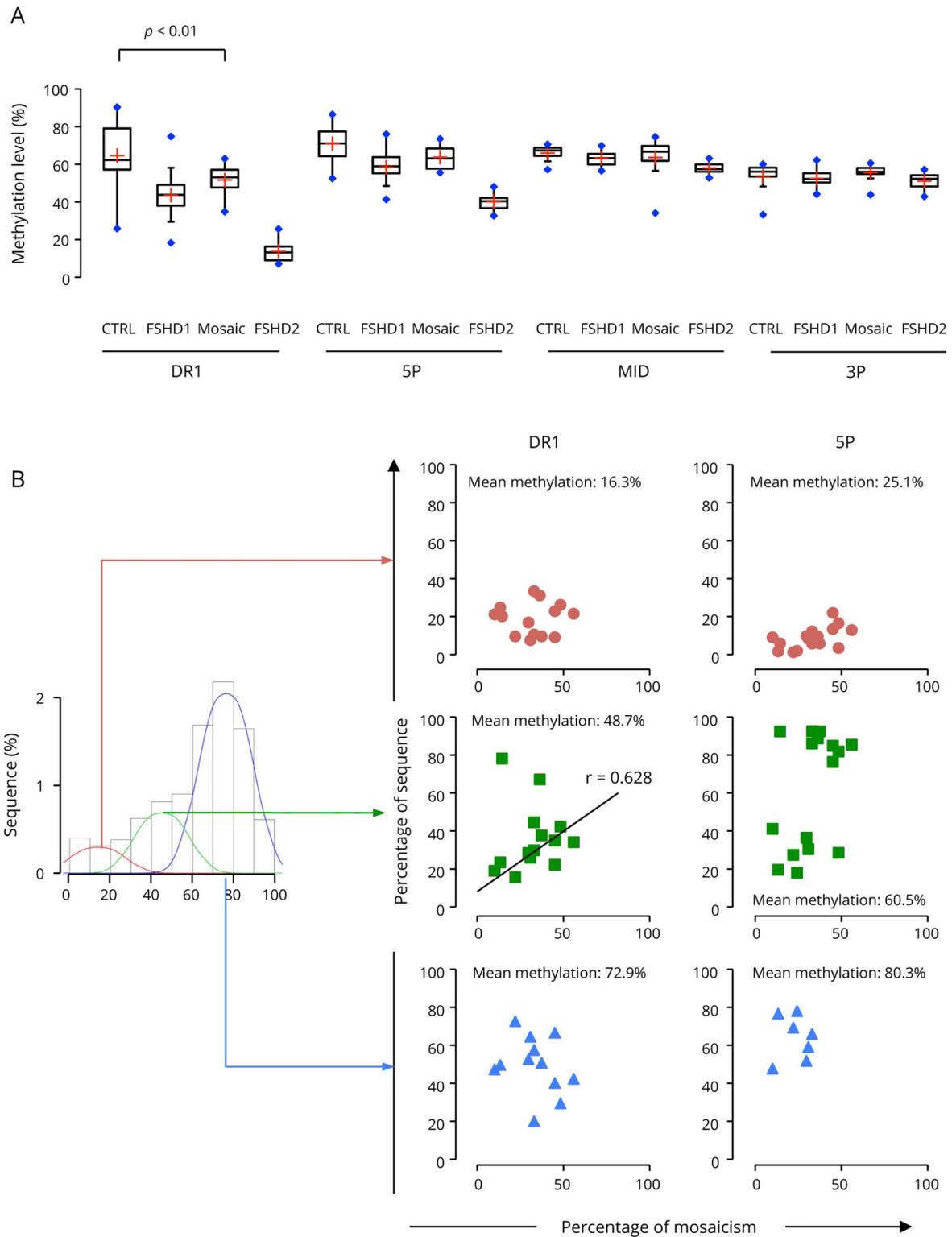
Taking advantage of published results^{27,28} and to deepen our approach, we applied the same methodology to investigate the methylation of the most distal D4Z4 unit and adjacent 4qA sequence containing the *DUX4* polyadenylation site. We used previously published primers and adapted the initial procedure designed for nested PCR amplification of a 594-bp sequence encompassing 57 CpGs (figure 5A),^{27,28} subcloning, and Sanger sequencing of clones to deep sequencing. In the different groups of samples, we observed a slight significant difference ($p = 0.048$) between controls (with a range of 20.5%–40.8% of methylation) and patients with FSHD1 (14.1%–31% range) (figure 5B; file e-1, links.lww.com/NXG/A191).

Of interest, the 5' forward primer located within D4Z4 is almost identical to our 3P-forward primers. Thus, the 14 proximal CGs of this 4qA bisulfite sequence are identical to those analyzed using our 3P primers (figure e-8, links.lww.com/NXG/A190). To compare the last D4Z4 repeat carried by 4qA alleles to the most distal part of all sequenced D4Z4, we first analyzed separately the most proximal part of the 4qA-specific sequence, which matches the 3P sequence (CpGs 1–14, figure e-8, links.lww.com/NXG/A190; figure 5A). This distal part of the last D4Z4 repeat is not differentially methylated in FSHD1 or FSHD2 compared with controls, as observed when all D4Z4 are considered (3P primers, figure 2).

Hypomethylation seems more pronounced in the region encompassing CGs 15–56 and corresponding to *DUX4* exons 2 and 3 and intronic sequences (pLAM and *DUX4* 3' UTR, p value = 0.027 between controls and FSHD1 blood samples (figure 5B, figure e-9, links.lww.com/NXG/A190) indicating epigenetic changes in the 3' UTR of *DUX4*.

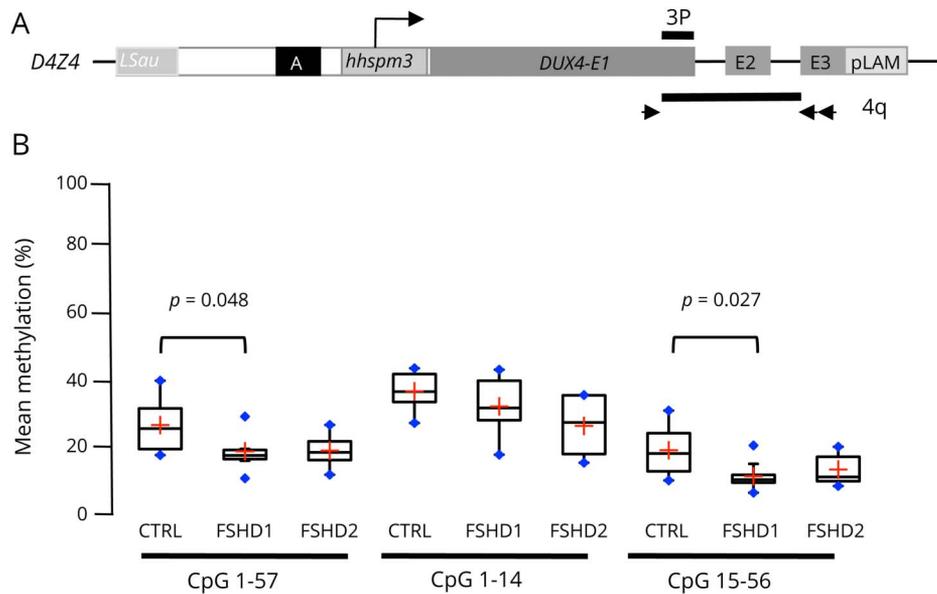
By analyzing CG sites overlapping with splice junction, no significant difference in methylation was observed, suggesting that alternative splicing of *DUX4* 3' UTR is not dependent on methylation changes. Hence, by comparing the methylation in the last D4Z4 unit, our approach reveals that most of the hypomethylation occurs in the most proximal part of the repeat in the 3 groups of patients investigated (FSHD1, FSHD2, and mosaic cases).

Figure 4 Methylation level of the proximal part of D4Z4 discriminates mosaic individuals from patients with FSHD1



(A) Boxplot representation of the mean methylation level for controls ($n = 10$), patients with FSHD1 ($n = 28$), patients with FSHD2 ($n = 10$), and patients with mosaicism ($n = 16$) for the DR1, 5P, MID, and 3P sequences. Horizontal black lines correspond to the median for each group, and the red cross to the mean for each group. Data sets were compared using the Wilcoxon nonparametric test, and brackets identify significantly different groups based on post hoc Dunn comparison and Bonferroni correction. The p values are indicated. (B) Scatterplot showing for the different patients with mosaicism ($n = 16$), the percentage of molecule with a low mean methylation, DR1: 16.1%, 5P: 25.1% (upper panel, red circles), intermediate mean methylation, DR1: 48.7%, 5P: 60.5% (middle, green squares), and high mean methylation, DR1: 72.9%, 5P: 80.3% (lower panel, blue triangles) for DR1 and 5P regions as a function of the percentage of mosaicism (or percentage of short D4Z4 4q allele) in blood samples. The correlation was calculated using the “cor” function of the R software and Spearman method. A positive correlation was found between the percentage of DR1 sequence with a medium level of methylation (48.7%) and the percentage of mosaicism ($r = 0.628$). FSHD = facioscapulohumeral dystrophy.

Figure 5 Analysis of the methylation pattern of the distal end of the D4Z4 array after sodium bisulfite modification and deep sequencing



We adapted the protocol of bisulfite modification, nested PCR amplification of the distal part of the last D4Z4 repeat and adjacent pLAM sequence, cloning and sequencing of selected clones to deep sequencing described in Ref. 27. (A) The position of the region amplified is indicated by black arrows. The 4q BSS sequence overlaps with the 3P region, with 14 CpG in common between the 2 regions. (B) Quantification of the mean methylation level of the 4q BSS sequence in DNA extracted from blood cells from controls, patients with FSHD1, and patients with FSHD2. The red line corresponds to the mean value. Data sets were compared using the Wilcoxon nonparametric test. Brackets identify significantly different groups based on post hoc Dunn comparison and Bonferroni correction ($p = 0.048$ between controls and patients with FSHD1). Quantification of the mean methylation level in the most proximal part of the 4q BSS sequence, encompassing the 3P region is displayed separately (CpG.1–14, as well as the most distal part of the sequence (CpG 15–56). Histograms display the mean methylation level in DNA from PBMCs from controls ($n = 6$), patients with FSHD1 ($n = 10$), and patients with FSHD2 ($n = 6$). BSS = bisulfite sequencing; CG = cytosine-guanine; CpG = cytosine-guanine dinucleotide; FSHD = facioscapulohumeral dystrophy; PBMC = peripheral blood mononuclear cell.

Discussion

Hypomethylation has been either linked to D4Z4 array shortening or mutation in *SMCHD1*. We analyzed DNA methylation of this macrosatellite and other repetitive elements in patients with FSHD using a custom deep sequencing method based on sodium bisulfite treatment of genomic DNA. For D4Z4, we combined analysis of regions with variable methylation (DR1 and 5P)^{10,11} and regions not differentially methylated.¹⁰ By comparing different groups of individuals, we demonstrate that changes in the level of methylation in the proximal part of the repeat can be quantified by sodium bisulfite and deep sequencing. The decreased methylation consistently observed in patients with FSHD at the DR1 and 5P regions indicates that hypomethylation occurs for the vast majority of sequenced fragments and not only at the distal *DUX4*-encoding unit^{27–29} or proximal repeat.³⁰ We inferred from our results and analysis of other regions that D4Z4 hypomethylation is not uniform but focalized in the proximal part of the repeat. We further showed that analysis of regions with a constant methylation level (MID and 3P) can be used as internal standard and for normalization to D4Z4 copy number.

Since the discovery of the “FSHD locus”, a decrease in DNA methylation of the shortened D4Z4 array has been

systematically associated with the disease. Several methodologies have been developed for analysis of this epigenetic modification going from Southern blot after digestion with methylation sensitive enzymes to sodium bisulfite sequencing. When probing D4Z4 methylation, the difficulty resides in the comprehensive analysis of the 4q-causing allele. By probing methylation of a single *FseI* site in the most proximal unit without distinction of the 4q and 10q chromosomes, Lemmers et al³⁰ defined the Delta Score for evaluation of hypomethylation in patients. Besides, 2 groups recently proposed analyses of the most distal D4Z4 unit and adjacent A-type sequences by sodium bisulfite sequencing using primers that distinguish 4q from 10q alleles^{27,29} and revealed that the distal *DUX4*-coding sequence (exon 2–3 and intronic sequences) is also hypomethylated.^{27–29} Nevertheless, in these studies, amplification of both the long and shortened arrays yields a potential bias in the estimation of the methylation level for individuals carrying 2 A-type alleles compared with 4qA/4qB patients carrying only a short 4qA allele.^{27–29} Overall, these observations on the methylation of the most distal repeat raise question of the role of methylation in *DUX4* splicing because methylation changes in coding sequences correlate with splicing sites usage than activation of transcription.³¹

Despite the systematic hypomethylation in patients with FSHD,^{8–10,32–34} the impact of this epigenetic modification

in FSHD pathomechanism remains poorly understood and its use as a disease biomarker debated. In some cases, methylation might be more reduced than expected,³⁰ and there is no correlation between the extend of hypomethylation and *DUX4* expression with a >50-fold range of increase in the *DUX4*-fl level but only small differences in the methylation level of the last D4Z4 unit at 4qA regions as reported.^{27,28}

Furthermore, the marked hypomethylation in *SMCHD1*-mutated patients with FSHD2 is not accompanied with either a more severe phenotype and earlier onset^{9,11,34} or with a muscle phenotype in patients with bosma arhinia microphthalmia syndrome.^{26,35,36} Furthermore, we recently showed that D4Z4 methylation is dynamically regulated in an *SMCHD1*-dependent manner opening new questions on the regulation of this epigenetic mechanism and *DUX4* activation.²⁶

There is no ideal method for evaluation of D4Z4 array methylation in patients with FSHD, given the repetitive nature of the region. Compared with other methods, our approach requiring a limited amount of DNA offers the possibility of analyzing simultaneously different regions across D4Z4 including the distal A-type region.^{10,11,27} It is noteworthy that all statistical assays or normalization methods applied to our data set lead to the same conclusion in the discrimination between controls, FSHD1, and FSHD2 but also mosaic cases. Importantly, we report here for the first time the analysis of DNA methylation in patients with mosaicism, accounting for approximately 40% of FSHD1 de novo cases (4%–12% of all FSHD1) with a correlation between the level of methylation with the percentage of mosaicism or short allele.

These observations would not be possible if one assume that a single D4Z4 unit is markedly hypomethylated (i.e., either the first unit as probed by Southern blotting or the *DUX4* encoding more distal unit), indicating that hypomethylation is not limited to the proximal or distal repeats but occurs within all repeats. The consistency and significance of results obtained using different methods of quantification argues against any potential bias in data interpretation and evaluation D4Z4 methylation despite the amplification of 4q- and 10q-derived repeats.

Thus, besides the clinical evaluation and molecular diagnosis, evaluation of D4Z4 methylation appears as an interesting tool for the evaluation of the patients, given the wide diversity in the FSHD-associated genotypes.^{16,30,37}

Acknowledgment

The authors are indebted and thank all patients for participating in this study. They thank Mr. Armand Tasmadjian for technical helps. They acknowledge the Marseille Maladies Rares (MarMaRa) institute for funding.

Disclosure

S. Roche, C. Dion, N. Brouqsault, M.-C. Gaillard, J.D. Robin, F. Puppo, E.S. Campana, S. Attarian, M. Bartoli, R. Bernard, K. Nguyen, F. Magdinier, C. Laberthonnière, A. Lagarde, C. Vovan,

and C. Chaix report no disclosures. Go to Neurology.org/NG for full disclosures.

Study funding

This study was funded by “Association Française contre les Myopathies” (AFM; MNHDecrypt and TRIM-RD), Agence Nationale de la Recherche (ANR, FSHDecipher, ANR-13-BSV1-0001). CD was the recipient of a fellowship from the French Ministry of Higher Education and Research and FSH Society. CL is the recipient of a fellowship from the French Ministry of Higher Education and Research. We acknowledge the Marseille maladie rare institute (MarMaRa) for funding.

Publication history

Received by *Neurology: Genetics* February 7, 2019. Accepted in final form October 4, 2019.

Appendix Authors

Name	Location	Role	Contribution
Stéphane Roche, PhD	Aix-Marseille University, INSERM, MMG U1251, Marseille	Author	Designed and conceptualized the study; analyzed the data; and drafted the manuscript for intellectual content
Camille Dion, PhD	Aix-Marseille University, INSERM, MMG U1251, Marseille	Author	Major role in the acquisition of data
Natacha Brouqsault, PhD	Aix-Marseille University, INSERM, MMG U1251, Marseille	Author	Major role in the acquisition of data
Camille Laberthonnière, MS	Aix-Marseille University, INSERM, MMG U1251, Marseille	Author	Major role in the acquisition of data
Marie Cécile Gaillard, PhD	Aix-Marseille University, INSERM, MMG U1251, Marseille	Author	Major role in the acquisition of data
Jérôme D. Robin, PhD	Aix-Marseille University, INSERM, MMG U1251, Marseille	Author	Revised the manuscript for intellectual content
Arnaud Lagarde, MS	Aix-Marseille University, INSERM, MMG U1251, Marseille	Author	Designed and conceptualized the study
Francesca Puppo, PhD	Aix-Marseille University, INSERM, MMG U1251, Marseille	Author	Role in the acquisition of data
Catherine Vovan, BS	Department of Medical Genetics, La Timone Hospital, Marseille	Author	Major role in the acquisition of data

Continued

Appendix (continued)

Name	Location	Role	Contribution
Charlène Chaix, BS	Department of Medical Genetics, La Timone Hospital, Marseille	Author	Major role in the acquisition of data
Emmanuelle Salort Campana, MD	Aix-Marseille University, INSERM, MMG U1251, Marseille Centre de reference pour les maladies neuromusculaires et la SLA, Hôpital de la Timone, Marseille	Author	Major role in the acquisition of samples
Shahram Attarian, MD, PhD	Aix-Marseille University, INSERM, MMG U1251, Marseille Centre de reference pour les maladies neuromusculaires et la SLA, Hôpital de la Timone, Marseille	Author	Major role in the acquisition of samples
Marc Bartoli, PhD	Aix-Marseille University, INSERM, MMG U1251, Marseille	Author	Role in the acquisition of data
Rafaelle Bernard, MD	Aix-Marseille University, INSERM, MMG U1251, Marseille Department of Medical Genetics, La Timone Hospital, Marseille	Author	Major role in the acquisition of data
Frédérique Magdinier, PhD	Aix-Marseille University, INSERM, MMG U1251, Marseille	Author	Designed and conceptualized the study; analyzed the data; and drafted, wrote, and revised the manuscript for intellectual content
Karine Nguyen, MD, PhD	Aix-Marseille University, INSERM, MMG U1251, Marseille Department of Medical Genetics, La Timone Hospital, Marseille	Author	Designed and conceptualized the study; major role in the acquisition of data; and revised the manuscript for intellectual content

References

- Wang LH, Tawil R. Facioscapulohumeral dystrophy. *Curr Neurol Neurosci Rep* 2016;16:66.
- Deenen JC, Arnts H, van der Maarel SM, et al. Population-based incidence and prevalence of facioscapulohumeral dystrophy. *Neurology* 2014;83:1056–1059.
- Sarfarazi M, Wijmenga C, Upadhyaya M, et al. Regional mapping of facioscapulohumeral muscular dystrophy gene on 4q35: combined analysis of an international consortium. *Am J Hum Genet* 1992;51:396–403.
- Upadhyaya M, Lunt P, Sarfarazi M, Broadhead W, Farnham J, Harper PS. The mapping of chromosome 4q markers in relation to Facioscapulohumeral Muscular Dystrophy (FSHD). *Am J Hum Genet* 1992;51:404–410.
- Wijmenga C, Hewitt JE, Sandkuijl LA, et al. Chromosome 4q DNA rearrangements associated with facioscapulohumeral muscular dystrophy. *Nat Genet* 1992;2:26–30.
- van der Maarel SM, Deidda G, Lemmers RJ, et al. De novo facioscapulohumeral muscular dystrophy: frequent somatic mosaicism, sex-dependent phenotype, and the

- role of mitotic transchromosomal repeat interaction between chromosomes 4 and 10. *Am J Hum Genet* 2000;66:26–35.
- Lyle R, Wright TJ, Clark LN, Hewitt JE. The FSHD-associated repeat, D4Z4, is a member of a dispersed family of homeobox-containing repeats, subsets of which are clustered on the short arms of the acrocentric chromosomes. *Genomics* 1995;28:389–397.
- van Overveld PG, Lemmers RJ, Sandkuijl LA, et al. Hypomethylation of D4Z4 in 4q-linked and non-4q-linked facioscapulohumeral muscular dystrophy. *Nat Genet* 2003;35:315–317.
- de Greef JC, Lemmers RJ, Camaño P, et al. Clinical features of facioscapulohumeral muscular dystrophy 2. *Neurology* 2010;75:1548–1554.
- Gaillard MC, Roche S, Dion C, et al. Differential DNA methylation of the D4Z4 repeat in patients with FSHD and asymptomatic carriers. *Neurology* 2014, 83:733–742.
- Hartweck LM, Anderson LJ, Lemmers RJ, et al. A focal domain of extreme demethylation within D4Z4 in FSHD2. *Neurology* 2013, 80:392–399.
- Gaillard MC, Puppo F, Roche S, et al. Segregation between SMCHD1 mutation, D4Z4 hypomethylation and Facio-Scapulo-Humeral Dystrophy: a case report. *BMC Med Genet* 2016;17:66.
- Lemmers RJ, Tawil R, Petek LM, et al. Digenic inheritance of an SMCHD1 mutation and an FSHD-permissive D4Z4 allele causes facioscapulohumeral muscular dystrophy type 2. *Nat Genet* 2012;44:1370–1374.
- Blewitt ME, Vickaryous NK, Hemley SJ, et al. An N-ethyl-N-nitrosourea screen for genes involved in variegation in the mouse. *Proc Natl Acad Sci USA* 2005;102:7629–7634.
- Nguyen K, Walrafen P, Bernard R, et al. Molecular combing reveals allelic combinations in facioscapulohumeral dystrophy. *Ann Neurol* 2011;70:627–633.
- Nguyen K, Brocqsault N, Chaix C, et al. Deciphering the complexity of the 4q and 10q subtelomeres by molecular combing in healthy individuals and patients with facioscapulohumeral dystrophy. *J Med Genet* 2019;56:590–601.
- Frommer M, McDonald LE, Millar DS, et al. A genomic sequencing protocol that yields a positive display of 5-methylcytosine residues in individual DNA strands. *Proc Natl Acad Sci USA* 1992;89:1827–1831.
- 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.
- Zeng W, de Greef JC, Chen YY, et al. Specific loss of histone H3 lysine 9 trimethylation and HP1gamma/cohesin binding at D4Z4 repeats is associated with facioscapulohumeral dystrophy (FSHD). *PLoS Genet* 2009;5:e1000559.
- Charlesworth B, Sniegowski P, Stephan W. The evolutionary dynamics of repetitive DNA in eukaryotes. *Nature* 1994;371:215–220.
- Jurka J, Kapitonov VV, Kohany O, Jurka MV. Repetitive sequences in complex genomes: structure and evolution. *Annu Rev Genomics Hum Genet* 2007;8:241–259.
- Rollins RA, Haghhighi F, Edwards JR, et al. Large-scale structure of genomic methylation patterns. *Genome Res* 2006;16:157–163.
- Eden A, Gaudet F, Waghmare A, Jaenisch R. Chromosomal instability and tumors promoted by DNA hypomethylation. *Science* 2003;300:455.
- Lisanti S, Omar WA, Tomaszewski B, et al. Comparison of methods for quantification of global DNA methylation in human cells and tissues. *PLoS One* 2013;8:e79044.
- Horard B, Eymery A, Fourel G, et al. Global analysis of DNA methylation and transcription of human repetitive sequences. *Epigenetics* 2009;4:339–350.
- Dion C, Roche S, Laberthonnière C, et al. SMCHD1 is involved in de novo methylation of the DUX4-encoding D4Z4 macrosatellite. *Nucleic Acids Res* 2019, 47:2822–2839.
- 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.
- Jones TI, King OD, Himeda CL, et al. Individual epigenetic status of the pathogenic D4Z4 macrosatellite correlates with disease in facioscapulohumeral muscular dystrophy. *Clin Epigenetics* 2015;7:37.
- Calandra P, Cascino I, Lemmers RJ, et al. Allele-specific DNA hypomethylation characterises FSHD1 and FSHD2. *J Med Genet* 2016;53:348–355.
- Lemmers RJ, Goeman JJ, van der Vliet PJ, et al. Inter-individual differences in CpG methylation at D4Z4 correlate with clinical variability in FSHD1 and FSHD2. *Hum Mol Genet* 2015;24:659–669.
- Lev Maor G, Yearim A, Ast G. The alternative role of DNA methylation in splicing regulation. *Trends Genet* 2015;31:274–280.
- Sacconi S, Camaño P, de Greef JC, et al. Patients with a phenotype consistent with facioscapulohumeral muscular dystrophy display genetic and epigenetic heterogeneity. *J Med Genet* 2012;49:41–46.
- van Overveld PG, Enthoven L, Ricci E, et al. Variable hypomethylation of D4Z4 in facioscapulohumeral muscular dystrophy. *Ann Neurol* 2005;58:569–576.
- de Greef JC, Lemmers RJ, van Engelen BG, et al. Common epigenetic changes of D4Z4 in contraction-dependent and contraction-independent FSHD. *Hum Mutat* 2009;30:1449–1459.
- Gordon CT, Xue S, Yigit G, et al. De novo mutations in SMCHD1 cause Bosma arhinia microphthalmia syndrome and abrogate nasal development. *Nat Genet* 2017;49:249–255.
- Shaw ND, Brand H, Kupchinsky ZA, et al. SMCHD1 mutations associated with a rare muscular dystrophy can also cause isolated arhinia and Bosma arhinia microphthalmia syndrome. *Nat Genet* 2017;49:238–248.
- Nguyen K, Puppo F, Roche S, et al. Molecular combing reveals complex 4q35 rearrangements in Facioscapulohumeral dystrophy. *Hum Mutat* 2017;38:1432–1441.

Neurology[®] Genetics

Methylation hotspots evidenced by deep sequencing in patients with facioscapulohumeral dystrophy and mosaicism

Stéphane Roche, Camille Dion, Natacha Broucqsault, et al.

Neurol Genet 2019;5;

DOI 10.1212/NXG.0000000000000372

This information is current as of November 14, 2019

Updated Information & Services	including high resolution figures, can be found at: http://ng.neurology.org/content/5/6/e372.full.html
References	This article cites 37 articles, 7 of which you can access for free at: http://ng.neurology.org/content/5/6/e372.full.html##ref-list-1
Subspecialty Collections	This article, along with others on similar topics, appears in the following collection(s): All Genetics http://ng.neurology.org/cgi/collection/all_genetics Gene expression studies http://ng.neurology.org/cgi/collection/gene_expression_studies Muscle disease http://ng.neurology.org/cgi/collection/muscle_disease
Permissions & Licensing	Information about reproducing this article in parts (figures, tables) or in its entirety can be found online at: http://ng.neurology.org/misc/about.xhtml#permissions
Reprints	Information about ordering reprints can be found online: http://ng.neurology.org/misc/addir.xhtml#reprintsus

Neurol Genet is an official journal of the American Academy of Neurology. Published since April 2015, it is an open-access, online-only, continuous publication journal. Copyright © 2019 The Author(s). Published by Wolters Kluwer Health, Inc. on behalf of the American Academy of Neurology. All rights reserved. Online ISSN: 2376-7839.

