

Decoding common and complex genotypes of facioscapulohumeral muscular dystrophy using methylation profiling and long-read sequencing

Florentine Scharf,^{1, *} Morghan C Lucas,^{1,2,3, *} Stefanie Gehling,¹ Felix Kleefeld,^{4,5,6} Veronika Schönrock,¹ Anna Benet-Pagès,^{1,7} Hans-Werner Rausch,⁸ Lucas Schirmer,⁸ Jochen Schäfer,⁹ Annika Saak,¹⁰ Joseph Porrmann,¹¹ Sylvia Hütter,¹¹ Christina Emde,¹² Anne Boblest,¹² Svenja Neuhoﬀ,¹³ Tim Hagenacker,¹³ Sebastian Burkart,¹⁴ Sebastian Mingers,¹⁵ Ann-Cathrine Berking,¹⁶ Lisa Olfe,¹⁶ Corinna Hendrich,¹⁶ Kristin Schröck,¹⁷ Elizabeta Sauer,¹⁸ Agata A Mossakowski,⁶ Kai Sendelbach,¹ Udo Koehler,¹ Kerstin Becker,¹ Tobias Bartolomaeus,¹ Ariane Hallermayr,^{1,3} Barbara Klink,¹ Teresa Neuhaus,¹ Maggie C. Walter,² Benedikt Schoser,² Angela Abicht,^{1,2} and Hannes Erdmann^{1,2}

* Florentine Scharf and Morghan C. Lucas contributed equally to this work.

Author affiliations:

1 Medical Genetics Center (MGZ) Munich, 80335 Munich, Germany

2 Friedrich Baur Institute at the Department of Neurology, LMU University Hospital, Ludwig-Maximilians-Universität München, 80336 Munich, Germany

3 Department of Medicine IV, LMU University Hospital, Ludwig-Maximilians-Universität München, 80336 Munich, Germany.

4 Department of Neurology, BG University Hospital Bergmannsheil, Ruhr-University Bochum, 44789 Bochum, Germany

5 BG University Hospital Bergmannsheil, Heimer Institute for Muscle Research, 44789 Bochum, Germany

6 Charité-Universitätsmedizin Berlin, Corporate member of Freie Universität Berlin, Humboldt Universität zu Berlin, and Berlin Institute of Health (BIH), Department of Neurology with Experimental Neurology, 10117 Berlin, Germany

7 Institute of Neurogenomics, Helmholtz Center Munich, 85764 Neuherberg, Germany

8 Department of Neurology, Mannheim Center for Translational Neuroscience, Medical Faculty Mannheim, Heidelberg University, 68167 Mannheim, Germany

- 9 Department of Neurology University Hospital Carl Gustav Carus at TUD Dresden University
of Technology, 01307 Dresden, Germany
- 10 Department of Neurology, University Hospital Linz, 4021 Linz, Austria
- 11 Institute for Clinical Genetics, University Hospital Carl Gustav Carus at TUD Dresden
University of Technology, 01307 Dresden, Germany
- 12 Department of Neuropediatrics, Faculty of Medicine and University Hospital Carl Gustav
Carus, Technische Universität Dresden, 01307 Dresden, Germany
- 13 Department of Neurology and Center for Translational Neuro- and Behavioral Sciences (C-
TNBS), University Hospital Essen, 45147, Essen, Germany
- 14 Institute of Human Genetics, Heidelberg University, 69120 Heidelberg, Germany
- 15 Neurologie Mingers, 52531 Übach-Palenberg, Germany
- 16 Institute for Human Genetics, Hannover Medical School, 30625 Hannover, Germany
- 17 Mitteldeutscher Praxisverbund Humangenetik, 04103 Leipzig, Germany
- 18 Clinical Genetics and Genome Medicine, University Hospital Würzburg, 97080 Würzburg,
Germany

Correspondence to: Dr. med. Hannes Erdmann, M.Sc.
Medical Genetics Center (MGZ) Munich
Bayerstr. 3–5
80335 Munich
Germany
Hannes.erdmann@mgz-muenchen.de

Short title: Common and complex FSHD genotypes

Abstract

Due to the complexity of the genomic region, the relationship between genotype, phenotype, and epigenetics in facioscapulohumeral muscular dystrophy remains incompletely understood, and diagnostics limited in precision.

In this study, we applied Oxford Nanopore Technology long-read sequencing to characterize FSHD1 and FSHD2 patients with common and previously confirmed genotypes ($n = 14$). This approach enabled the simultaneous determination of allele-specific methylation profiles, haplotype, and repeat length. Our results confirm the presence of an asymmetric methylation gradient increasing from the proximal to distal end of 4q35 repeat arrays with a steeper slope in FSHD1 (2.5–8.1%/repeat unit) compared to the two FSHD2 patients (0.3 and 1.2%/repeat unit) studied. The slope is therefore a potential marker to differentiate both types of facioscapulohumeral muscular dystrophy, being less biased by repeat length compared to parameters determined in current bisulfite sequencing-based methylation assays. By determining distal methylation allele-specifically long-read sequencing overcomes a major limitation of such assays. Distal methylation of pathogenic (range 22–45%) and non-pathogenic repeat arrays (range 77–87%) ($P < 0.0001$) is clearly separated and shows a stronger correlation with age-corrected clinical severity ($r = -0.81$, 95%-CI: -0.48 – -0.94) compared to studies based on bisulfite sequencing. Overall, this strengthens distal methylation as a superior diagnostic marker of facioscapulohumeral muscular dystrophy disease status.

Detailed clinical and genetic characterization of a second cohort ($n = 20$) comprising patients with complex genotypes ($n = 11$) and FSHD2 ($n = 9$) was performed using our bisulfite sequencing-based methylation profile analysis, complementary methods, and in three cases, long-read sequencing. Based on these results we highlight underrecognized patient groups at risk of misdiagnosis when methylation analysis is not part of the diagnostic workup: i) patients carrying a 4qA166 repeat array, ii) patients with underrepresented population backgrounds, iii) facioscapulohumeral muscular dystrophy patients carrying 11 repeat units in the absence of global hypomethylation, and iv) FSHD2 patients without a detectable causative variant in known epigenetic suppressor genes (*SMCHD1*, *DNMT3B*, and *LRIF1*). We hypothesize that unknown epigenetic modifiers explain the penetrance of repeat arrays with 8–10 repeat units in the European and North American population, the pathogenicity of some repeat arrays with 11 repeat units, and the hypomethylation of the entire 4q35 repeat array in some FSHD2 patients without a detected variant in known epigenetic suppressor genes.

From our results recommendations for standard diagnostics are derived. Finally, we hypothesize a preliminary model linking DNA methylation, known genetic parameters and disease severity.

Keywords: FSHD, long-read sequencing, methylation, Nanopore sequencing

Introduction

Facioscapulohumeral muscular dystrophy (FSHD, OMIM#158900) is one of the most common muscular dystrophies in Europe.¹ Based on genetic features, FSHD is categorized into two clinically indistinguishable types: FSHD1 and FSHD2. FSHD1 is linked to contractions of the D4Z4 macrosatellite repeat array on chromosome 4q35, while FSHD2 is associated with globally reduced methylation of the whole repeat array, typically caused by pathogenic variants in the epigenetic suppressor genes *SMCHD1*, *DNMT3B*, or *LRIF1*.^{2–5} Both forms require the presence of an intact polyadenylation signal (PAS) encoded by a single nucleotide polymorphism (SNP) within the region flanking the distal repeat unit (RU) that defines the so-called permissive haplotype.⁶

Diagnosis of FSHD typically involves i) testing for a contracted D4Z4 repeat array (≤ 10 RU) on a permissive allele to confirm FSHD1, and ii) testing for pathogenic variants in epigenetic suppressor genes (*SMCHD1*, *DNMT3B* or *LRIF1*) and/or for global hypomethylation in combination with a permissive allele (usually 8–20 RU in length) to confirm FSHD2.⁵ Although diagnostic precision has significantly improved over the last decade, molecular confirmation remains elusive in some patients with distinct FSHD phenotype, currently referred to as “Clinical FSHD, genetic cause unknown”.⁵

The key pathomechanism of both FSHD types is epigenetic derepression of the transcription factor *DUX4*, located within the most distal RU of the 4q35 repeat array and its flanking region. Studies over the last decade confirmed that altered DNA methylation is a molecular correlate of aberrant *DUX4* expression and can serve as a reliable marker of FSHD.^{7–12} Recently, long-read sequencing (LRS) enabled an in-depth characterization of the 4q35 repeat array.^{13–15} By revealing an asymmetric methylation gradient that increases from the proximal to the distal end, these studies provided further insights into the genetic and epigenetic basis of FSHD (Fig. 1A).

Methylation assays are an evolving standalone approach to FSHD testing.^{7,9,12,16,17} Current assays, such as our recently reported FSHD methylation profile analysis (FSHD-MPA), determine two different methylation values of 4q35 repeat arrays: ‘distal’ and ‘global methylation’.^{7,12}

Distal methylation, measured within the most distal RU, is a robust marker of overall disease status (Supplementary Fig. 1A, B, Fig. 1B–D). It does not allow to differentiate FSHD1 from FSHD2. Distal methylation shows an inverse correlation with age-corrected disease severity and is a potential quantitative biomarker.^{7–9,11,12,16–18} The parameter is usually assessed by bisulfite sequencing (BSS)-based assays utilizing primers specific for permissive 4qA (and 4qAL) repeat arrays. However, these approaches lack allele-specificity, particularly limiting sensitivity in patients who are homozygous for the permissive haplotype or who carry somatic mosaicism (Fig. 1C, D). In these cases, distal hypomethylation from the pathogenic repeat array on one allele is diluted by highly methylated, non-pathogenic repeat arrays on the other allele, potentially leading to inconclusive or, in isolated cases, to false-negative results.

Global methylation, measured as the average methylation across all 4q35 and 10q26 repeat arrays, is used to differentiate FSHD1 and FSHD2 (Supplementary Fig. 2A, B). The parameter is determined by assaying a region present in each RU of 4q35 and 10q26 repeat arrays (DUX4 in FSHD-MPA and DR1 in other assays, Supplementary Fig. 1A, C).¹⁹ In FSHD2 global methylation is reduced compared to FSHD1 and non-affected individuals. Global methylation levels are biased by repeat length, as shown by a moderate correlation between DR1 methylation level and the combined lengths of the 4q35 and 10q26 repeat arrays in a recent study.²⁰ As such, large repeat arrays can artificially increase global methylation (Supplementary Fig. 2D). This potentially leads to a global methylation in the intermediate range in some FSHD2 patients and therefore inconclusive results. Conversely, additional short repeat arrays may cause artificially low global methylation level within the pathogenic or intermediate range in some FSHD1 cases (Supplementary Fig. 2C), mimicking FSHD2.

In this study, we used Oxford Nanopore Technology (ONT) LRS to characterize the methylation profile of known FSHD genotypes (Cohort 1), confirm previous findings of an asymmetric methylation gradient and highlight methylation as a superior diagnostic marker.^{13–15,21} In a second cohort (Cohort 2), we describe complex genotypes that were primarily characterized using our BSS-based FSHD-MPA method and, in three cases, using ONT LRS. These cases highlight patient groups with a high risk of being misclassified and yield insights into the epigenetic basis of FSHD. We summarize the results of both cohorts by presenting a

model for the relationship between genetics, epigenetics, and phenotype in FSHD and provide recommendations for the diagnostics of FSHD.

Materials and methods

Patients and study approval

A total of 34 individuals previously analyzed in our institute, divided into two cohorts, were analyzed in this study. Cohort 1 (Supplementary Table 1–2) included 14 patients with known genotypes and previously confirmed diagnoses of FSHD1 (12 patients, A1–A12), FSHD2 based on a pathogenic variant in *SMCHD1* (1 patient, B1, referred to as ‘FSHD2_{*SMCHD1*}’), and FSHD1+2 with unknown origin of hypomethylation (1 patient, C1, referred to as ‘FSHD1+2_{no variant}’), all analyzed using ONT LRS to implement the method. Methods used to establish the initial diagnosis varied depending on the time point of analysis and the material available. Cohort 2 comprised 20 patients (Supplementary Table 1–4) with either complex genotypes (11 patients, D1, E1–6, F1, G1–3) or FSHD2 (9 patients, B2–B10). For a comprehensive characterization, all patients were analyzed using established diagnostic methods as summarized in Supplementary Table 1. For FSHD2 patients without a variant in known epigenetic suppressor genes (referred to as ‘FSHD2_{no variant}’) multiplex ligation probe amplification (MLPA) of *SMCHD1* and NGS analysis of FSHD2 candidate genes (*CTCF*, *DNMT1*, *DNMT3A*, *EZH2*, *SUV39H1*) identified in a previous study were performed.²² Additionally, Optical Genome Mapping (OGM) of *SMCHD1*, *DNMT3B*, and *LRIF1* was performed for 5 of 6 patients with appropriate material available. Exome sequencing was performed in all patients of Cohort 2 to identify potential dual diagnoses and rule out alternative ones. ONT LRS was performed for three patients (E2, F1, G1) of this cohort. Additional patients could not be analyzed as ONT R9 flow cells were discontinued during our study and replaced by R10 flow cells, which are largely incompatible with CRISPR/Cas9 target enrichment. Additionally, required amounts of high-molecular-weight DNA were not available for all patients of Cohort 2.

Board-certified neurologists assessed phenotypes. Age-corrected clinical severity (ACSS) was determined as previously described (ACSS = 2000 x CSS /age at examination).^{23,24} Informed consent was obtained from all participants. All genetic analyses and investigations were performed in accordance with the guidelines of the Declaration of Helsinki and approved by

the local ethics committee of the Ludwig-Maximilians-University of Munich (vote no. 24-0789).

Haplotype analysis and methylation profile analysis (FSHD-MPA)

Haplotype analysis following the Tsumagari and KASP assay was performed as previously described.^{12,25} Methylation profile analysis (FSHD-MPA) was performed as previously described with updated cut-off values (Supplementary Table 5).¹² Pathogenic thresholds were defined as the highest value in FSHD1 or FSHD2 patients that did not overlap with values of negative cases and healthy controls. Upper thresholds for the intermediate range were defined as the lowest value in FSHD1- or FSHD2-negative patients that did not overlap with those of positive cases, as previously established.¹⁶

Patients were diagnosed as FSHD-positive when distal methylation fell within the pathogenic range. FSHD2 diagnosis was based on an additional global methylation within the pathogenic range. In cases of intermediate distal or global methylation, confirmation or exclusion of FSHD was determined by additional genetic analyses, including repeat length (FSHD1) and pathogenic variants in FSHD2-associated genes (FSHD2).

OGM

Bionano OGM was performed as previously described to determine D4Z4 repeat lengths, haplotype, and to test for structural variants (SVs) and copy number variants (CNVs) in *SMCHD1*, *DNMT3B*, and *LRIF1* and in FSHD2 candidate genes (*CTCF*, *DNMT1*, *DNMT3A*, *EZH2*, *SUV39H1*).^{22,26–28} Repeat lengths of contracted repeat arrays (<48 kb) determined by OGM show high concordance with Southern blotting, with a variance of ± 1 RU in individual cases, in line with previous studies.²⁹

Southern blotting

Southern blotting was performed as previously described.¹² The accuracy of repeat length was determined as ± 1 RU for repeat arrays <48 kb.

Next-Generation Sequencing (NGS)

Sequencing of *SMCHD1*, *DNMT3B*, *LRIF1*, FSHD2 candidate genes (*CTCF*, *DNMT1*, *DNMT3A*, *EZH2*, *SUV39H1*) as well as exome sequencing, was performed as previously

described.¹² Variants were classified according to the ACMG/AMP (American College of Medical Genetics and Genomics/Association for Molecular Pathology) guidelines.³⁰

SMCHD1-MLPA

FSHD2 patients identified by FSHD-MPA with negative result of NGS sequencing for variants in *SMCHD1*, *DNMT3B*, and *LRIF1* were analyzed by MLPA to test for CNVs in *SMCHD1* or to confirm CNVs detected by NGS. Genomic DNA was analyzed with the Salsa MLPA P493 SMCHD1 kit (MRC-Holland) according to the manufacturer’s instructions. MLPA products were separated by capillary electrophoresis using the ABI-3130XL Genetic Analyzer (Applied Biosystems, Foster City, CA). Data were analyzed with GeneMarker software (Softgenetics, State College, CA), using the MLPA ratio for analysis and peak area for quantification.

ONT LRS

DNA extraction, library preparation and flow cell loading were performed according to the ONT Cas9-targeted sequencing protocol using ONT’s SQK-CS9109 kit and 5 µg of input gDNA. CRISPR RNAs (crRNAs) to enrich the 4q35 repeat array were designed using CHOPCHOP 8 (Supplementary Table 6, Supplementary Fig. 3). Sequencing was performed with ONT FLO-MIN106D R9 flow cells on the GridION X5 sequencer. Base and methylation calling were performed using guppy v.7.4.12+0e5e75c49 using the model dna_r9.4.1_450bps_modbases_5mc_cg_sup.cfg for R9 flow cells and dna_r10.4.1_e8.2_400bps_5khz_modbases_5hmc_5mc_cg_sup.cfg for R10 flow cells.^{27,28} The generated FASTQ files were aligned to the human reference genome (T2T CHM13-v2.0/hs1) using Minimap2 v.2.27-r1193 to pre-filter for reads potentially spanning the region of interest (chr4:193432996–193543640 and chr10:134604599–134733754).²⁹ Of those, reads were further characterized based on specific sequence motifs using blastn v2.11.+. First, reads covering the distal end of the repeat arrays were selected based on the presence of the distal anchor sequence (Supplementary Table 7). These reads were further analyzed for the presence or absence of the restriction sites for *XapI* and *BlnI* to distinguish reads originating from chromosome 4q35 (B-X⁺) from reads originating from chromosome 10q26 (B⁺X⁻). Repeat length was determined based on the number of *XapI* restriction sites in 4q35 reads containing the distal and proximal anchor sequence (full length reads) confirmed by manual inspection of mapped reads. Haplotyping was performed based on the absence of the PAS in 4qB and its presence in the cases of 4qA and 4qAL. 4qAL was distinguished based on an inserted sequence

in intron 2 of *DUX4* as opposed to 4qA. In case of heterozygosity, single reads were assigned to one of the two 4q35 repeat arrays based on the distinct haplotype. In cases of homozygosity, all full-length reads corresponding to the shorter repeat array were assigned to this allele. All reads longer than this repeat array – regardless of whether they are full-length or partial - were assigned to the second allele. Distal methylation was calculated as the average methylation in chr4:193539784–193543078 using modbam2bed v0.9.5 over the mapped reads, separated by allele; accordingly, proximal methylation was calculated in the most proximal RU completely covered by full length reads assigned to one allele, with the repeat unit being flanked by *KpnI* restriction sites. Quality control was done using NanoPlot v1.41.6.³⁰ Repeat arrays were included in the analysis if supported by a minimum of three allele-specific reads (see Statistical Analysis and Supplementary Fig. 4). For distal methylation, all allele-assigned reads were considered; for the methylation gradient's slope, reads were additionally restricted to full-length reads. Sequencing depth and quality parameters are given in Supplementary Tables 8 and 9.

Statistical Analysis

Statistical analyses were performed using R v.4.0.0. Significance of differences between two groups were assessed using Welch's t-test (two-tailed). A *P*-value < 0.05 was considered statistically significant. Correlation analyses were performed by Pearson's correlation test, 95% confidence intervals of the correlation coefficients were determined and *P*-values were calculated to test the significance of the correlation. Thresholds of distal methylation defining patients as positive or negative for FSHD were determined as average distal methylation $\pm 2 \times$ standard deviation of all disease and non-disease-causing repeat arrays in Cohort 1, respectively. Linear regression analysis was performed to determine the qualitative relationship between average methylation per RU and repeat length. The slope of the resulting function equals the slope of the methylation gradient. To estimate the minimum number of reads required for reliable calculation of distal methylation and slope of the methylation gradient, downsampling experiments were performed on full-length reads of the contracted repeat arrays of Cohort 1 for all samples with more than 50 reads on the contracted repeat array, including only one representative of the familial cases A8–A11 (*n* = 9). For each downsampling level (1–10, 15, and 20 reads per allele), 30 replicates were generated by random sampling using Python's random.sample() function (python3.9), and the interquartile range (IQR) of both distal methylation and methylation gradient's slope was calculated as a measure of dispersion and an

exponential decay function was fitted to the data. The resulting model confirmed a rapid decay in dispersion at low read depths, with the curve approaching saturation from approximately seven reads per allele onwards. Alleles with fewer than three reads showed substantially elevated dispersion and are not recommended for analysis and excluded from this study. Notably, sample A11 showed substantially higher variability in slope estimates across downsampling replicates, as a linear gradient is more susceptible to noise with fewer data points (2 RU). A conservative minimum threshold of 15 reads per allele is recommended for slope analysis for very short repeat arrays, based on visual inspection of dispersion stabilization (Supplementary Fig. 4, Supplementary Code 1).

Results

ONT LRS enables in-depth analysis of the 4q35 repeat array in common FSHD genotypes (Cohort 1)

To enable allele-specific analysis of DNA methylation in FSHD patients, we implemented CRISPR/Cas9-targeted ONT LRS targeting the *p13E11* (proximal) and *pLAM* (distal) regions flanking the D4Z4 repeat array (Supplementary Fig. 3). This approach allows parallel assessment of key diagnostic parameters, including repeat length, haplotype, and methylation across the entire 4q35 repeat array. We applied this method to 14 patients with known FSHD genotypes (Cohort 1, Supplementary Table 1–2).

Repeat length and haplotypes for the shorter repeat array were consistent with those determined by complementary methods, except for patient C1 (4qAL repeat array), where a discrepancy of one RU was noted between OGM and LRS, being within the known error margin of OGM. A proximal-to-distal methylation gradient was consistently observed across all patients. Contracted repeat arrays showed distal hypomethylation (Fig. 2A, B). Average methylation within the most proximal RU of contracted repeat arrays in FSHD1 patients (A1–A12) was $19\% \pm 3.3\%$ (range 10–24%), increasing by an average slope of $4.8\% \pm 1.8\%$ per RU (range 2.5–8.1%/RU), reaching average methylation of $33\% \pm 7.4\%$ (range 22–45%) within the most distal RU, depending on repeat length (Fig. 2B, C, Supplementary Fig. 5, Supplementary Table 2). Due to insufficient full-length coverage, analysis of uncontracted repeat arrays was restricted to the distal end, where average distal methylation ($81\% \pm 3.5\%$, range 77–87%) was

significantly higher compared to contracted repeat arrays ($P < 0.0001$) without any overlap between both groups.

In FSHD2_{SMCHD1} (B1) and FSHD1+2_{no variant} (C1) patients, the shorter permissive repeat array showed distal hypomethylation of 30% and 24%, with corresponding methylation gradients of 0.3%/RU and 1.2%/RU, respectively. For patient C1, the larger 31 RU 4qA repeat array exhibited an approximately linear increase in methylation across the proximal 10 RU, followed by oscillation between 46% and 59% (Supplementary Fig. 5O). The larger repeat array in patient B1 was not sufficiently covered for analysis. Intervals defining patients as positive (18–48%) or negative (74–88%) for FSHD were determined as average distal methylation ± 2 x standard deviation of disease- and non-disease-causing repeat arrays in Cohort 1. As the pathogenicity of the longer repeat array in patient C1 is unknown it was omitted from the calculation.

Genetic findings in complex and high-risk of being misclassified FSHD genotypes (Cohort 2)

To characterize FSHD genotypes at risk of misclassification, we analyzed 20 patients with rare or complex genetic characteristics (Cohort 2) using established diagnostic methods, including FSHD-MPA, and ONT LRS for three cases. We identified four key subgroups prone to diagnostic challenges:

1. Exclusion of FSHD in a patient with a complex 4qA166 repeat array

In patient F1 presenting with a severe neuromuscular phenotype, exclusion of FSHD was only possible through haplotype analysis, including ONT LRS. Patient F1 first presented at age 52 with myalgia and elevated creatine kinase levels (3–5x above normal) without muscular weakness and normal EMG and muscle MRI results of the lower limb (Supplementary Table 3). Due to the patient's Sjögren's syndrome, overlap myositis was suspected as the underlying cause of the symptoms. Despite immunosuppressive therapy, symptoms rapidly progressed to global respiratory insufficiency, proximal weakness in the upper and lower limbs, and axial weakness within three years. Muscle biopsy indicated signs of chronic myositis, EMG revealed myopathic changes, and muscle MRI showed signs of myositis and fatty degeneration in the limb girdle muscles. FSHD was considered as a differential diagnosis. Southern blotting detected a complex repeat contraction (7 RU starting with a non-4q35 RU, Supplementary Table 4). OGM confirmed a contraction of 7 RU and mapped it to chromosome

4q35 *in cis* to the 4qA haplotype without identifying the complex composition of the repeat array. For its further characterization, ONT LRS was performed. The analysis confirmed a contracted repeat array of 7 RU *in cis* to a PAS defining the permissive haplotype (Fig. 2D). Four RU showed the signature of chromosome 4q (absence of the restriction site of *BlnI* and presence of the restriction site of *XapI* (B-X⁺), Supplementary Fig. 6), while three, including the first RU, lacked chromosome 4 and 10 signatures, due to resistance to both *XapI* and *BlnI* (B-X⁻). Based on a proximal B-X⁻ RU, the updated diagnostic guidelines classify the haplotype of the 7 RU repeat array as 4qA166. The haplotype is currently not associated with FSHD despite the presence of a PAS.^{5,31} Therefore, the patient was diagnosed as negative for FSHD. We additionally determined the haplotype based on SNPs in the distal RU and the flanking pLAM region. They were consistent with the haplotype originally denoted as 4qC166H of unknown significance (Supplementary Fig. 7). As the definition of subhaplotypes and hybrid repeat arrays was based on individual samples and different regions, their overlap and variability are currently unknown and require harmonization.

We determined a distal methylation of the 4qA166 repeat array of 60%, which is above the FSHD-positive interval in this study (18–48%). The methylation profile showed a relatively high methylation in the proximal RU (29%) and a methylation gradient’s slope comparable to that in FSHD1 patients (4.0%/RU). A contraction of a 4qA166 repeat array, in combination with high distal methylation, strongly supported the exclusion of FSHD in this patient in line with the atypical clinical phenotype and fast disease progression. Despite comprehensive analysis including exome and genome sequencing, analysis of mitochondrial genes, and additional histopathological examination, the final diagnosis remained elusive.

2. Carrier status detection in a multi-ethnic family with South Asian population background

The complexity of interpreting multiple genetic parameters is exemplified by family G, which has a South Asian population background (referred to as ‘FSHD_{South Asian population}’). The father (G1) presented in his sixties with a clear FSHD phenotype, as the first affected individual in his family, and would have been diagnosed as affected by FSHD1+2 based on distal and global hypomethylation in FSHD-MPA and the detection of a permissive 7 RU repeat array by OGM. However, the average size of all 4q35 and 10q26 repeat arrays was rather short compared to FSHD2_{SMCHD1} patients (12 RU vs. 20 RU in average), making this case suspicious of a potential artificial lowering of global methylation (Supplementary Figure 2). A variant of unknown

significance (VUS) was detected in *DNMT3B* by NGS, with a predicted low impact on *DNMT3B* protein function. Distal methylation determined by LRS was 40%, falling within the FSHD-positive interval. The methylation gradient's slope was 3.3%/RU, within the range of FSHD1 patients in Cohort 1. Final differentiation between FSHD1 and 2 was not possible, also due to the lack of reference values for repeat length and methylation in this population. Exome sequencing to exclude differential diagnoses was negative.

Predictive testing of his daughter (patient G3) was complicated by a) unknown penetrance of a 7 RU repeat contraction in this population, b) unconfirmed validity of the detected global hypomethylation in her father and unknown pathogenicity of his *DNMT3B* variant and c) insufficient validation of methylation assays for predictive testing. High correlation coefficients for distal methylation and clinical severity are observed only after age-adjustment of the latter. This suggests lifetime stability of DNA methylation as prerequisite for predictive testing. For the most precise assessment of the disease risk of patient G3 all parameters were determined collectively. Based on global methylation in the intermediate, distal methylation in the pathogenic range, and the presence of a permissive repeat contraction of 9 RU, the daughter was diagnosed as at risk of developing FSHD during her lifetime. Consistently, careful clinical examination after genetic testing revealed facial weakness as an initial sign of FSHD. Interestingly, the 9 RU repeat array was inherited from the mother (patient G2), who showed normal global and distal methylation and no obvious clinical signs of FSHD despite her higher age.

3. FSHD due to an 11 RU contraction in absence of global hypomethylation (FSHD_{11 RU})

Patient D1, affected by FSHD, had been misdiagnosed with an inflammatory myopathy for more than five decades. Following the revised diagnostic guidelines, she would also have been classified as negative for FSHD.⁵ The patient presented with discrete facial weakness, a foot drop palsy, and a mild proximal weakness, more pronounced in the lower than in the upper limb (Supplementary Table 3). Diagnostic criteria of FSHD1 (presence of a permissive repeat contraction ≤ 10 RU) were not fulfilled as the patient carried a permissive repeat array of 11 RU. Similarly, criteria of FSHD2 were not fulfilled based on the absence of likely pathogenic and pathogenic variants (LP/P) in known epigenetic suppressor genes (*SMCHD1*, *DNMT3B*, and *LRIF1*) and global methylation above the pathogenic range.⁵

FSHD-MPA revealed reduced distal methylation within the pathogenic and global methylation within the intermediate range. Therefore, FSHD-MPA allowed diagnosis of FSHD with inconclusive differentiation between FSHD1 and FSHD2.

Exome sequencing and repeat analysis for myotonic dystrophy type 1 and 2 did not identify an alternative origin of symptoms. Probatory immunosuppressive treatment did not improve the condition.

4. FSHD2 without a causative variant in known epigenetic suppressor genes (FSHD2_{no variant})

As no variants (SNVs and CNVs) in known epigenetic suppressor genes were identified by NGS, six patients (E1–E6) could only be identified as affected by FSHD2 based on FSHD-MPA results (Supplementary Table 3,4). Four patients (E1, E2, E5, E6) had typical FSHD phenotypes (ACSS: range 48–157), while two patients (E3 and E4) showed additional symptoms uncommon for FSHD. Patient E3 presented with a right-sided scapula alata since birth. In line with the diagnosis of a perinatal Erb’s palsy, EMG of the right shoulder girdle revealed neurogenic changes, whereas the left side was normal. However, progressive bilateral weakness of the shoulder girdle developed during adolescence, which was inconsistent with an isolated perinatal trauma and raised the suspicion of FSHD. Patient E4 displayed FSHD features in addition to early hypotonia, mild motor delay, and exercise intolerance. All six patients (E1–E6) had negative family histories for FSHD and other neuromuscular disorders. All patients fulfilled the diagnostic criteria of FSHD2, including global hypomethylation and the presence of at least one permissive repeat array ≥ 8 RU. We compared the average repeat size of the four 4q35 and 10q26 repeat arrays in FSHD2_{no variant} and FSHD2_{SMCHD1} patients, as repeat length can bias the global methylation level. Generally, FSHD2_{no variant} patients showed shorter mean average lengths (16 ± 2.6 RU, range 13–21 RU) than FSHD2_{SMCHD1} patients (20 ± 5.6 RU, range 14–32 RU, $P = 0.10$). However, differences were not significant and both groups overlapped. FSHD disease status is confirmed by distal hypomethylation in the pathogenic (four patients) or intermediate range (two patients, both homozygous for the permissive haplotype 4qA).

We characterized both 4q35 repeat arrays in FSHD2_{no variant} patient E2 using ONT LRS. The shorter 16 RU 4qA repeat array had a methylation gradient’s slope of 1.2%/RU. Distal methylation was 41% and as such within the FSHD-positive interval determined in Cohort 1. Interestingly, the distal methylation of the second repeat array (19 RU on 4qAL) had a similar

slope (1.4%/RU) but a relatively high distal methylation of 60%, despite the small length difference of 3 RU.

To investigate the origin of global hypomethylation, we first performed MLPA analysis of *SMCHD1* to detect CNVs that might have been missed by NGS, which was negative in all patients (E1–E6). SVs and CNVs in genes including *SMCHD1*, *DNMT3B*, and *LRIF1* were further assessed via *de novo* assembly of OGM data in five patients (E1–E5) with suitable material available. No such variants were identified in any of these patients. Additionally, screening of previously identified FSHD2 candidate genes (*CTCF*, *DNMT1*, *DNMT3A*, *EZH2*, and *SUV39H1*) from NGS data (patients E1–E6) revealed no potentially pathogenic variants.²² In patient E4, additional short-read whole genome sequencing did not reveal a cause of global hypomethylation, but two compound heterozygous LP/P variants in *DOK7*. This establishes a dual diagnosis with congenital myasthenic syndrome (CMS) explaining atypical symptoms overlapping with FSHD.

Upper threshold of repeat length in FSHD2_{SMCHD1} patients

Beyond the complexity of rare genotypes, the genetic basis of FSHD2 remains incompletely understood. FSHD2 follows a digenic mode of inheritance, requiring an LP/P variant in *SMCHD1*, *DNMT3B*, or *LRIF1* or global hypomethylation of other origin, in combination with a permissive repeat array with a size in the lower FSHD1-negative range. Current diagnostic guidelines assume 20 RU as the upper length threshold in most FSHD2 cases.⁵ To examine whether this threshold also applies to FSHD2_{SMCHD1} in this study (patients B2–10), we determined the repeat lengths using OGM (Supplementary Table 4). The shorter 4qA repeat array ranged from 11 to 27 RU (average 19 ± 6 RU), with five of nine patients exceeding 20 RU in line with previous reports.^{20,32,33} The high number of repeat arrays larger than 20 RU was partly influenced by the familial case of patient B5 and B6 (mother and son). Notably, these patients illustrate how a large second repeat array can bias global methylation. Patient B5 (second repeat array: 16 RU on 4qB) had a pathogenic global methylation, whereas patient B6 (second repeat array: 37 RU on 4qB) showed global methylation in the intermediate range. Both patients had a dual diagnosis of FSHD2 and *EYA4*-associated progressive bilateral sensorineural hearing loss since childhood. Of the five patients with more than 20 RU on their shorter 4qA-repeat array, four had a common FSHD phenotype indistinguishable from that of other FSHD2_{SMCHD1} patients. The fifth patient (patient B6) had a relatively mild phenotype without facial weakness as previously described.³⁴ Average ACSS for all other FSHD2_{SMCHD1}

patients was similar for those with less or more than 20 RU on their shorter repeat array (140 ± 49 vs. 141 ± 83 , $P > 0.05$).

Correlation of allele-specific distal methylation and age-corrected clinical severity

We further analyzed the correlation between allele-specific distal methylation determined across the complete distal RU by ONT LRS and age-corrected clinical severity for patients of Cohort 1 and patients E2 and G1 of Cohort 2. Patients A7 and B1 were excluded due to incomplete clinical data, patient A8 was excluded due to somatic mosaicism of the contracted repeat array. For these 13 patients a Pearson’s correlation coefficient of -0.81 (95-CI: $-0.48 - -0.94$) was determined (Fig. 2E). The threshold for pathogenic distal methylation was estimated at $\leq 48\%$ based on the x-intercept of the function ($ACSS = 0$).

Discussion

Diagnosing FSHD remains challenging due to its biological complexity and technical limitations. However, correct identification of patients affected by FSHD is not only important in view of upcoming therapeutic options, but also to avoid injurious therapies based on an incorrect differential diagnosis. Interestingly, in at least four FSHD patients in this study inflammatory myositis was considered as the disease origin after an initial negative genetic result for FSHD. In all patients, immunosuppressive therapy was initiated, likely causing liver fibrosis in one patient and facilitating active tuberculosis and a non-Hodgkin lymphoma in another. In addition, the diagnosis cannot be based solely on the presence of a classic clinical phenotype, as dual diagnoses, as observed in our study (patients E3, E4 and B5, B6), are common and associated with atypical symptoms. Furthermore, atypical FSHD phenotypes, as part of the disease spectrum, require reliable molecular markers to confirm the diagnosis.

By enabling allele-specific analysis of the methylation profile of individual repeat arrays, LRS overcomes a key limitation of current methylation assays, in line with previous reports. Our study further supports distal methylation as a superior diagnostic parameter for determining overall disease status as indicated by the following findings: i) A strong correlation of allele-specific distal methylation with age-corrected disease severity was observed. ii) Methylation levels of disease-causing and non-pathogenic repeat arrays in FSHD1 patients were completely

separated. This allowed independent confirmation of FSHD1 in patients A1, A5 and A6, which had an inconclusive result in FSHD-MPA due to homozygosity for the permissive haplotype. iii) Disease status was correctly determined based on distal methylation in all patients of Cohort 2, in contrast to a diagnostic only based on repeat length and variants in epigenetic suppressor genes. We propose the slope of the methylation gradient (Fig. 1A, 2C) as an alternative parameter to differentiate FSHD1 from FSHD2, potentially replacing global methylation in BSS-based assays. In this study, FSHD1 patients (slopes of 2.5–8.1%/RU) were separated from the two FSHD2 patients in Cohort 1 (slopes of 0.3 and 1.2%/RU) based on the slope, consistent with a previous study reporting slopes of 1.9–3.9%/RU in FSHD1 and 0.7–1.6%/RU in FSHD2.¹⁴ As the slope is largely independent of the average repeat length of the 4q35 and 10q26 repeat arrays as opposed to global methylation (illustrated in Supplementary Fig. 2), it may serve as a quantitative marker for the strength of epigenetic modifiers. Given the small sample size of Cohort 1, confirmation of our findings in larger and genetically diverse cohorts is required to validate the slope of the methylation gradient as a diagnostic parameter.

The methods currently used to diagnose FSHD each have distinct advantages and limitations (Supplementary Table 10). As illustrated by patient F1, OGM does not detect intra-array SVs, as labeling occurs only proximal and distal to the 4q35 and 10q26 repeat arrays. This leads to misclassification of 4qA166 repeat arrays or SVs such as *in cis* duplications and 4q-10q rearrangements. Patient F1 showed rapid disease progression and severe respiratory insufficiency, which are atypical for FSHD. However, initial symptoms were compatible with the disease. Together with the detection of a contracted repeat array on an apparently permissive and pathogenic repeat array by OGM, this could have led to a false FSHD diagnosis. Based on high distal methylation (60%) exceeding the FSHD-positive interval, FSHD was correctly excluded by ONT LRS. Additionally, the patient carried the 4qA166 haplotype *in cis* to the repeat contraction, which is currently not associated with FSHD despite carrying a PAS. Non-pathogenicity of 4qA166 repeat arrays was derived in previous studies from two older, unrelated and unaffected carriers of 4qA166 repeat arrays with 5 and 8 RU.³⁵ Patient F1 in our study showed a methylation gradient's slope similar to FSHD patients, but an unusually high proximal methylation. The first six RU showed average methylation within the FSHD1-positive interval, implying that a similar, but shorter repeat array, might show distal methylation in the pathogenic range. As such, general non-pathogenicity of 4qA166 repeat arrays should be assumed with caution. It may, in principle, be possible that lower cut-offs exist for the pathogenicity of 4qA166 repeat arrays.

The challenges of assessing the pathogenicity of permissive and contracted repeat arrays based solely on repeat length, particularly in underrepresented population backgrounds, are exemplified by family G. In Asian populations different thresholds for FSHD1 have been reported, e.g., 1–6 RU in Korea and Japan or 1–9 RU in China.^{36–38} The 7 RU repeat array of the index patient (G1) was classified as pathogenic only through the additional finding of distal hypomethylation. Likewise, disease risk of his daughter (G3) could only be established by detecting distal hypomethylation of the 9 RU repeat array inherited from the mother (G2). In the mother, the 9 RU repeat array showed a distal methylation within the normal range, in line with the absence of clinical signs of FSHD despite her higher age. Results from this family indicate that methylation might represent a more general parameter of disease status than repeat length and may be less dependent on population background. However, the applicability of identical methylation thresholds across different populations requires validation in larger and more diverse cohorts. The findings in family G further indicate the presence of epigenetic modifiers causing pathogenicity of the maternally inherited 9 RU repeat array. This is also reflected by the pronounced distal hypomethylation in patient G3, in contrast to her mother. Although the daughter is heterozygous for the permissive haplotype while her mother is homozygous, it is unlikely that the second permissive repeat array of the mother masks pathogenic hypomethylation of her 9 RU repeat array. If that was the case, we would expect a result in the intermediate rather than in the normal range. As the quantitative relationship between global methylation and repeat size is currently unknown, the ability to conclude the presence of epigenetic modifiers based on reduced global methylation in the index patient and his daughter compared to the mother is limited, also given the larger average 4q35 and 10q26 repeat length (16 RU) in the mother compared to the index patient and his daughter (12 RU). Whether the variant detected in *DNMT3B* serves as the respective epigenetic modifier remains unknown, with bioinformatic prediction rather indicating non-pathogenicity. The slope of the methylation gradient of 3.3% in the index patient does not support the presence of strong but rather moderate epigenetic modifiers, although reference values for different populations remain to be established.

The pathogenicity of permissive repeat arrays of 11 RU in absence of global hypomethylation remains controversial. While in some cases they were regarded as borderline with reduced penetrance, current diagnostic guidelines define the FSHD1 cut-off at 10 RU for the European and North American population.^{5,39} In this study, we report one patient with an FSHD phenotype (patient D1) carrying a permissive repeat array of 11 RU. Based on global

hypomethylation within the intermediate range, we hypothesize the presence of additional weak to moderate epigenetic modifiers contributing to pathogenicity.

Cases of FSHD2 marked by global hypomethylation but without detectable variants in known epigenetic suppressor genes are well described and account for up to 15% of FSHD2 patients.^{2,4,10,40} Similarly, in our study, the cause of global hypomethylation remained elusive in six FSHD2_{no variant} patients (patients E1–E6). These cases highlight the need to routinely assess global methylation, as recommended in the revised diagnostic guidelines.⁵ Restricting FSHD2 testing to sequencing of known epigenetic suppressor genes, as still practiced in many diagnostic laboratories, may miss such cases. Given the substantial number of FSHD2 patients with an unknown cause of hypomethylation, deriving pathogenicity of VUS in *SMCHD1* based solely on co-occurrence with global hypomethylation should be performed with caution. This is particularly critical in prenatal and preimplantation diagnostics, where accurate variant classification is essential, since methylation patterns likely change during early embryonic development and are not valid as diagnostic parameters.

In principle, missed SVs, deep intronic or regulatory variants in *SMCHD1*, *DNMT3B*, and *LRIF1*, or other candidate genes could explain global hypomethylation in FSHD2_{no variant} patients. In these cases, long-read whole genome and RNA sequencing of blood and muscle could further be applied to search for causative variants. However, given the broad variety of methods applied, this appears unlikely in all FSHD2_{no variant} cases. Alternatively, currently unknown epigenetic suppressor genes could exist. Interestingly, patients with FSHD2_{no variant} had, on average, shorter 4q35 and 10q26 repeat arrays than patients with FSHD2_{SMCHD1}. However, it is unlikely that the reduced global methylation observed in FSHD2_{no variant} patients is only a technical artifact of the shorter repeat arrays, as the average repeat lengths of both groups largely overlap. Rather, we hypothesize that weaker epigenetic modifiers in FSHD2_{no variant} patients are present, that can only be detected in the absence of large 4q and 10q repeat arrays, which would artificially increase global methylation. That epigenetic modifiers in FSHD2_{no variant} patients are weaker than in FSHD2_{SMCHD1} patients is supported by i) a larger slope of the methylation gradient in FSHD2_{no variant} patient E2 (1.2%) compared to FSHD2_{SMCHD1} patient B1 (0.3%) and ii) two FSHD2_{no variant} patients having a distal methylation in the intermediate range which is atypically high for FSHD2, even in patients being homozygous for the permissive haplotype. However, additional FSHD2_{SMCHD1} and FSHD2_{no variant} patients need to be analyzed by LRS to confirm these differences. It additionally needs to be studied whether FSHD2_{no variant} patients might on average show milder phenotypes.

1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60

Both FSHD_{2no variant} and FSHD_{11 RU} patients are equally source of “clinical FSHD, genetic cause unknown” diagnoses as defined in the updated diagnostic guidelines.⁵ Especially when large 4q35 and 10q26 repeat arrays are present, moderate to weak epigenetic modifiers of global hypomethylation might be masked. In such cases, global and distal methylation might be in the intermediate range and diagnosis of FSHD remains elusive. This is observed in a small number of patients with distinct FSHD phenotype in our institution (data not shown). We hypothesize that SNPs not fulfilling the classification of a LP/P variant within genes responsible for methylating the 4q35 repeat array could act as weak and moderate epigenetic modifiers by potentially reducing the efficiency of the associated gene. In FSHD_{2no variant} patients, the accumulation of multiple such variants may exert a combined effect on the overall methylation similar to a LP/P variant in known epigenetic suppressor genes. Such oligogenic mode of inheritance would be supported by the negative family history of all FSHD_{2no variant} patients. Weak to moderate epigenetic modifiers could not only explain the pathogenicity of some repeat arrays with a size of 11 RU, but also why repeat contractions of 8–10 RU are penetrant in only a subset of individuals of the European and North American population.^{41,42} SNPs in epigenetic suppressor genes with opposite effect might also explain lower pathogenic thresholds for FSHD1 in other populations, as certain variants in *SMCHD1* are associated with higher methylation levels of the 4q35 repeat array.⁴³ In addition, variants in epigenetic regulators influencing chromatin structure are candidates that impact disease status in ways not reflected by DNA methylation.

While the relationship between DNA methylation and empirically identified genetic parameters (4q35 repeat contractions and LP/P variants in epigenetic suppressor genes) has remained incompletely understood, it can now be elucidated by an asymmetric methylation gradient at 4q35 and 10q26 repeat arrays (Fig. 1A, 3).^{13–15} We propose that repeat length, LP/P variants in known epigenetic suppressor genes, and unknown epigenetic modifiers collectively shape the methylation profile of a D4Z4 repeat array. A hypothesis-driven model in line with our observations is presented in Figure 3. In FSHD1 patients A1–A12 in our study, the average methylation within the most proximal RU is 19% and increases by on average 4.8% per RU. As such, distal methylation would on average exceed the FSHD-positive interval of 18–48% after the 7th RU, which equals the known repeat length threshold for fully penetrant repeat arrays in FSHD1 in the studied population. Variability in the methylation gradient’s slope (range: 2.5–8.1%/RU; average: 4.8 ± 1.8%/RU) likely reflects the influence of weak and moderate epigenetic modifiers. Consequently, distal methylation thresholds are reached after a varying number of RU, potentially explaining inter-individual and population-specific repeat

length thresholds in FSHD1. Additionally, this could explain variations in disease severity among patients with similar repeat lengths.

The presence of pathogenic variants in known epigenetic suppressor genes as strong epigenetic modifiers likely causes a flatter gradient's slope in FSHD2. Consequently, methylation levels in the non-pathogenic range are likely reached after a higher number of RU (Fig. 1A, 3B). This suggests the existence of an upper threshold for repeat length in FSHD2, which may differ depending on the origin of global hypomethylation with respect to the specific variant and the epigenetic suppressor gene. A recent study confirmed that *SMCHD1* variants differ in their impact on global methylation levels without identifying differences depending on the type of variant.²⁰ We hypothesize, that variants causing strong global hypomethylation, potentially reflected in low slopes of the methylation gradient, lead to FSHD manifestation even with longer permissive repeat arrays (e.g. 27 RU in this study), while milder variants and modifiers associated with steeper gradients would require shorter permissive repeat arrays to cause FSHD, possibly the case in FSHD2_{no variant} patients. This proposed model is in line with the concept of FSHD as a molecular continuum (Fig. 3B), where repeat length predominates in FSHD1 (1–7 RU in the European and North American population), epigenetic modifiers are primary drivers in FSHD2. Both factors likely interact in the incomplete penetrance range (8–10 RU in the European population), in FSHD patients with 11 RU in the absence of global hypomethylation and potentially some FSHD2_{no variant} patients.

While our study presents LRS as superior approach to study the FSHD locus, the presented workflow cannot yet be used as diagnostic test for several reasons: i) The number of samples analyzed with LRS in our study is limited. ii) Target enrichment via CRISPR/Cas9 is largely incompatible with current ONT R10 flow cells. This, together with limited availability of sufficient amounts of high molecular weight DNA is the reason why only three patients of Cohort 2 underwent LRS. Future studies should focus on alternative strategies for target enrichment, e.g. adaptive sampling as presented in other studies.^{44–46} iii) The method and the methylation gradient's slope as a diagnostic parameter require further validation regarding reproducibility, stability, accuracy and required minimum sequencing depth. iv) Accuracy and limits of detection need to be determined for repeat contractions present as somatic mosaicism and complex structural variants. v) The determined cut-off values require refinement in larger cohorts with diverse population background and controls not affected by a neuromuscular disease.

1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60

We summarized the advantages and disadvantages of current standard diagnostic methods in Supplementary Table 10. Based on this analysis and the results of our study, we recommend performing FSHD-MPA or similar assays according to our previously described workflow, with confirmatory repeat length analysis for FSHD1 and/or sequencing of epigenetic suppressor genes for FSHD2.¹² In case of a negative result (global and distal methylation in the normal range), despite a clear FSHD phenotype, Southern Blot with pulsed field gel electrophoresis and different probes should be performed to test for somatic mosaicism and indicators of rearrangements and other SVs.¹² In case of negative results following a traditional diagnostic workflow with repeat length analysis and sequencing of epigenetic suppressor genes as first-tier analyses, global methylation should routinely be analyzed when a permissive haplotype has been detected. However, as a long-term goal, implementation of long-read sequencing as stand-alone, first-tier approach for FSHD diagnostics should be pursued, when technically and economically available.

Data availability

Anonymized data from this study are available from the corresponding author on reasonable request. Code used for the downsampling analysis and statistical analyses is provided as Supplementary Code S1. Code for filtering reads is based on the published method of Butterfield *et al.*¹⁴

Funding

No funding was received towards this work.

Competing interests

The authors report no competing interests.

Supplementary material

Supplementary material is available at *Brain Communications* online.

References

1. Deenen JCW, Arnts H, van der Maarel SM, et al. Population-based incidence and prevalence of facioscapulohumeral dystrophy. *Neurology*. 2014;83(12):1056.
2. Lemmers RJLF, 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(12):1370-1374.
3. van den Boogaard ML, Lemmers RJLF, Balog J, et al. Mutations in DNMT3B Modify Epigenetic Repression of the D4Z4 Repeat and the Penetrance of Facioscapulohumeral Dystrophy. *Am J Hum Genet*. 2016;98(5):1020-1029.
4. Hamanaka K, Šikrová D, Mitsuhashi S, et al. Homozygous nonsense variant in LRIF1 associated with facioscapulohumeral muscular dystrophy. *Neurology*. 2020;94(23):e2441.
5. Giardina E, Camaño P, Burton-Jones S, et al. Best practice guidelines on genetic diagnostics of facioscapulohumeral muscular dystrophy: Update of the 2012 guidelines. *Clin Genet*. 2024;106(1):13-26.
6. Lemmers RJLF, van der Vliet PJ, Klooster R, et al. A Unifying Genetic Model for Facioscapulohumeral Muscular Dystrophy. *Science*. 2010;329(5999):1650-1653.
7. 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(1):23.
8. 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(1):37.
9. Calandra P, Cascino I, Lemmers RJLF, et al. Allele-specific DNA hypomethylation characterises FSHD1 and FSHD2. *J Med Genet*. 2016;53(5):348.

10. Lemmers RJLF, 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.* 2014;24(3):659-669.

11. Zheng F, Qiu L, Chen L, et al. Association of 4qA-Specific Distal D4Z4 Hypomethylation With Disease Severity and Progression in Facioscapulohumeral Muscular Dystrophy. *Neurology.* 2023;101(3):e225-e237.

12. Erdmann H, Scharf F, Gehling S, et al. Methylation of the 4q35 D4Z4 repeat defines disease status in facioscapulohumeral muscular dystrophy. *Brain.* 2023;146(4):1388-1402.

13. Hiramuki Y, Kure Y, Saito Y, et al. Simultaneous measurement of the size and methylation of chromosome 4qA-D4Z4 repeats in facioscapulohumeral muscular dystrophy by long-read sequencing. *J Transl Med.* 2022;20(1):517.

14. Butterfield RJ, Dunn DM, Duval B, Moldt S, Weiss RB. Deciphering D4Z4 CpG methylation gradients in facioscapulohumeral muscular dystrophy using nanopore sequencing. *Genome Res.* 2023;33(9):1439-1454.

15. Erdmann H, Scharf F, Hallermayr A, et al. Reply: An epigenetic basis for genetic anticipation in facioscapulohumeral muscular dystrophy type 1. *Brain.* 2023;146(12):e111-e114.

16. Strafella C, Megalizzi D, Trastulli G, et al. Integrating D4Z4 methylation analysis into clinical practice: improvement of FSHD molecular diagnosis through distinct thresholds for 4qA/4qA and 4qA/4qB patients. *Clin Epigenetics.* 2024;16(1):148.

17. Caputo V, Megalizzi D, Fabrizio C, et al. D4Z4 Methylation Levels Combined with a Machine Learning Pipeline Highlight Single CpG Sites as Discriminating Biomarkers for FSHD Patients. *Cells.* 2022;11(24):4114.

18. Zheng F, Qiu L, Chen L, et al. An epigenetic basis for genetic anticipation in facioscapulohumeral muscular dystrophy type 1. *Brain.* 2023;146(12):e107-e110.

19. Hartweck LM, Anderson LJ, Lemmers RJ, et al. A focal domain of extreme demethylation within D4Z4 in FSHD2. *Neurology.* 2013;80(4):392.

20. Gérard L, Delourme M, Tardy C, et al. SMCHD1 genetic variants in type 2 facioscapulohumeral dystrophy and challenges in predicting pathogenicity and disease penetrance. *Eur J Hum Genet.* 2025;33:784-792.

21. Huang M, Zhang Q, Jiao J, et al. Comprehensive genetic analysis of facioscapulohumeral muscular dystrophy by Nanopore long-read whole-genome sequencing. *J Transl Med.* 2024;22(1):451.
22. Strafella C, Caputo V, Bortolani S, et al. Whole exome sequencing highlights rare variants in *CTCF*, *DNMT1*, *DNMT3A*, *EZH2* and *SUV39H1* as associated with FSHD. *Front Genet.* 2023;14:1235589.
23. Ricci E, Galluzzi G, Deidda G, et al. Progress in the molecular diagnosis of facioscapulohumeral muscular dystrophy and correlation between the number of KpnI repeats at the 4q35 locus and clinical phenotype. *Ann Neurol.* 1999;45(6):751-757.
24. Van Overveld PGM, Enthoven L, Ricci E, et al. Variable hypomethylation of D4Z4 in facioscapulohumeral muscular dystrophy. *Ann Neurol.* 2005;58(4):569-576.
25. Tsumagari K, Chen D, Hackman JR, Bossler AD, Ehrlich M. FSH dystrophy and a subtelomeric 4q haplotype: a new assay and associations with disease. *J Med Genet.* 2010;47(11):745.
26. Barseghyan H, Eisenreich D, Lindt E, et al. Optical Genome Mapping as a Potential Routine Clinical Diagnostic Method. *Genes (Basel).* 2024;15(3):342.
27. van der Sanden B, Neveling K, Pang AWC, et al. Optical Genome Mapping for Applications in Repeat Expansion Disorders. *Curr Protoc.* 2024;4(7):e1094.
28. Stence AA, Thomason JG, Pruessner JA, et al. Validation of Optical Genome Mapping for the Molecular Diagnosis of Facioscapulohumeral Muscular Dystrophy. *J Mol Diagn.* 2021;23(11):1506-1514.
29. Guraju NM, Jump V, Lemmers R, et al. Molecular Diagnosis of Facioscapulohumeral Muscular Dystrophy in Patients Clinically Suspected of FSHD Using Optical Genome Mapping. *Neurol Genet.* 2023;9(6):e200107.
30. Richards S, Aziz N, Bale S, et al. Standards and guidelines for the interpretation of sequence variants: a joint consensus recommendation of the American College of Medical Genetics and Genomics and the Association for Molecular Pathology. *Genet Med.* 2015;17(5):405-423.
31. Lemmers RJLF, van der Vliet PJ, van der Gaag KJ, et al. Worldwide Population Analysis of the 4q and 10q Subtelomeres Identifies Only Four Discrete Interchromosomal Sequence Transfers in Human Evolution. *Am J Hum Genet.* 2010;86(3):364-377.

32. Rieken A, Bossler AD, Mathews KD, Moore SA. CLIA Laboratory Testing for Facioscapulohumeral Dystrophy. *Neurology*. 2021;96(7):e1054–e1062.
33. Lemmers RJLF, van der Vliet PJ, Vreijling JP, et al. Cis D4Z4 repeat duplications associated with facioscapulohumeral muscular dystrophy type 2. *Hum Mol Genet*. 2018;27(20):3488-3497.
34. He J -J., Lin X -D., Lin F, et al. Clinical and genetic features of patients with facial-sparing facioscapulohumeral muscular dystrophy. *Eur J Neurol*. 2018;25(2):356-364.
35. Lemmers RJLF, Wohlgemuth M, van der Gaag KJ, et al. Specific Sequence Variations within the 4q35 Region Are Associated with Facioscapulohumeral Muscular Dystrophy. *Am J Hum Genet*. 2007;81(5):884-894.
36. Park HJ, Hong JM, Lee JH, et al. Low D4Z4 copy number and gender difference in Korean patients with facioscapulohumeral muscular dystrophy type 1. *Neuromuscul Disord*. 2015;25(11):859-864.
37. Goto K, Nishino I, Hayashi YK. Rapid and accurate diagnosis of facioscapulohumeral muscular dystrophy. *Neuromuscul Disord*. 2006;16(4):256-261.
38. Wang Z, Qiu L, Lin M, et al. Prevalence and disease progression of genetically-confirmed facioscapulohumeral muscular dystrophy type 1 (FSHD1) in China between 2001 and 2020: a nationwide population-based study. *Lancet Reg Health West Pac*. 2022;18:100323.
39. Preston MK, Tawil R, Wang LH. Facioscapulohumeral Muscular Dystrophy. In: Adam MP, Ardinger HH, Pagon RA, Wallace, Stephanie E, eds. GeneReviews®. University of Washington;1993–2025. Accessed June 7, 2025. <https://www.ncbi.nlm.nih.gov/books/NBK1443/>
40. Lemmers RJLF, van den Boogaard ML, van der Vliet PJ, et al. Hemizygosity for SMCHD1 in Facioscapulohumeral Muscular Dystrophy Type 2: Consequences for 18p Deletion Syndrome. *Hum Mutat*. 2015;36(7):679-683.
41. Scionti I, Greco F, Ricci G, et al. Large-Scale Population Analysis Challenges the Current Criteria for the Molecular Diagnosis of Facioscapulohumeral Muscular Dystrophy. *Am J Hum Genet*. 2012;90(4):628-635.
42. de Greef JC, Lemmers RJLF, van Engelen BGM, et al. Common epigenetic changes of D4Z4 in contraction-dependent and contraction-independent FSHD. *Hum Mutat*. 2009;30(10):1449-1459.

43. Tapia del Fierro A, den Hamer B, Benetti N, et al. SMCHD1 has separable roles in chromatin architecture and gene silencing that could be targeted in disease. *Nat Commun.* 2023;14(1):5466.
44. Huang M, Zhang Q, Wu S, et al. Accurate detection of D4Z4 repeats, methylation and allele haplotype in facioscapulohumeral muscular dystrophy 1 using nanopore long-read adaptive sampling sequencing: a pilot study. *J Med Genet.* 2025;62(11):669-676.
45. Gamaarachchi H, Stevanovski I, Hammond JM, et al. Targeted sequencing and iterative assembly of near-complete genomes. *Nat Commun.* 2025;16(1):10406.
46. Löwenstern T, Madritsch S, Horner D, et al. DUCKS4: a comprehensive workflow for Nanopore sequencing analysis of facioscapulohumeral muscular dystrophy (FSHD). *Hum Genomics.* 2026;20(1):48.

Figure legends

Figure 1. Illustration of the recently discovered asymmetric methylation gradient of 4q35 repeat arrays and potential biases in determining distal methylation in bisulfite sequencing-based methylation assays. (A) Schematic representation of the increase in methylation from the proximal to the distal end of 4q35 repeat arrays in FSHD1 and FSHD2 patients. In FSHD1 patients a larger increase in methylation per RU represented by a steeper slope of the methylation gradient ($\Delta\text{methylation}/\Delta\text{RU}$) is observed. Methylation within the most distal RU of a contracted repeat array (blue line) is below the pathogenicity threshold and therefore causative for FSHD1. Distal methylation of non-pathogenic, uncontracted repeat arrays (cyan line) is above the pathogenicity threshold. In FSHD2 patients the methylation gradient's slope is flatter (purple and pink line). Repeat arrays causing FSHD2 (pink line) are usually shorter than 20 RU and therefore have a distal methylation below the pathogenicity threshold. Large repeat arrays (purple line) can be non-disease causing likely associated with a distal methylation above the threshold of pathogenicity. The exact threshold for repeat length in FSHD2 remains to be determined and likely differs inter-individually. (B) Average distal methylation (black bar, right diagram) solely reflects the contracted permissive repeat array (blue bar) as the second non-permissive repeat array is not amplified leading to correct pathogenic results. (C) Reduced distal methylation of the disease-causing contracted repeat

arrays (blue bar) is diluted by a highly methylated, non-contracted permissive repeat array on the second allele (cyan bar), also carrying the permissive haplotype. Consequently, average distal methylation (black bar) is often in the intermediate, inconclusive range. (D) Reduced distal methylation of a contracted repeat array present as somatic mosaicism (blue bar) is strongly diluted by permissive repeat arrays present on the other alleles and in different cell populations (cyan bars), respectively. In case of a low allele frequency, average distal methylation (black bar) is within the normal range, potentially leading to a false-negative result.

Figure 2. Results of ONT LRS of patients of Cohort 1 and 2. (A) Methylation profile of both 4q35 repeat arrays of patient A5 carrying one contracted permissive repeat array with 6 RU (in dark blue) and an uncontracted non-pathogenic repeat array (in light blue). One dot represents average CpG methylation within a sliding window of 250 base pairs over all reads at respective chromosomal positions relative to the polyadenylation signal within the most distal RU. (B) Average methylation of all CpGs within the most distal RU determined over all reads from ONT LRS of all patients of Cohort 1. Left: methylation within the shorter and pathogenic repeat array (n = 14, patients A1–12, B1, C1), right: non-pathogenic and longer repeat array, if covered in the analysis (n = 7). As the pathogenicity of the longer repeat array of patient C1 is unknown, it was not included. Dashed lines represent average methylation over all samples $\pm 2 \times$ standard deviation defined as interval for FSHD-positive (left, 18–48%) and FSHD-negative (right, 74–88%) repeat arrays. (C) Course of CpG methylation over the whole repeat array for all patients of Cohort 1 and patients F1, G1 and E2 from Cohort 2 (n = 17) determined by regression analysis from average CpG methylation per RU over all reads of the shorter allele and repeat length. While the slope of the methylation gradient is steep in FSHD1 patients (range 2.5–8.1%) and for the complex 4qA166 repeat array (4.0%), it is flatter in the FSHD2_{SMCHD1} (0.3%), FSHD2_{no variant} (1.2%) and FSHD1+2_{no variant} (1.2%) patients. In the patient with South Asian population background (patient G1) it is within the range of FSHD1 patients (3.3%). (D) Composition of the complex 4qA166 repeat array in patient F1 characterized by ONT LRS and in agreement with Southern blotting. Structure was determined based on presence or absence of restriction sites for the enzymes *BlnI* and *XapI*. The 4qA166 haplotype is currently not associated with FSHD. (E) Correlation analysis of allele-specific distal methylation of the shorter, FSHD-causing repeat array with the ACSS for patients of Cohort 1 and 2 (n = 13), if detailed clinical data were available. Respective Pearson's

correlation coefficients, their 95% confidence intervals and *P*-values are given. Grey areas indicate 95% confidence interval of the regression.

Figure 3. Linkage of genetic parameters, epigenetics (methylation) and phenotype in FSHD. (A) Hypothesis-driven model of repeat length, LP/P variants in epigenetic suppressor genes and unknown epigenetic modifiers impacting the methylation profile of the FSHD locus linked to disease status. (B) Hypothesized model of the proposed impact of repeat length (x-axis) and strength of epigenetic modifiers (y-axis) on distal methylation as quantitative biomarker for FSHD disease status. In previous studies, low distal methylation levels were, on average, associated with a severe FSHD phenotype (shown in deep blue), while medium distal methylation levels were, on average, associated with a mild FSHD phenotype (shown in white).^{11,12} High distal methylation level were associated with the absence of FSHD (shown in grey). In our model, short repeat arrays and strong epigenetic modifiers are proposed to cause low distal methylation levels (shown in deep blue). Following this hypothesis, the longer the repeat array and the weaker the epigenetic modifiers, the higher the distal methylation. At a certain point (marked by the black line within the diagram) distal methylation levels cross the border of the pathogenic and normal, non-pathogenic range. Strong epigenetic modifiers can be LP/P variants in *SMCHD1*, while the nature of weak modifiers is currently unknown and could hypothetically be resembled by non-LP/P variants in epigenetic suppressor genes. The dashed line indicates hypothetical inverse epigenetic modifiers that potentially increase distal methylation and might explain lower thresholds of pathogenicity in other populations. Bottom scale indicates FSHD as molecular continuum with FSHD1 and FSHD2 as the two extremes. FSHD1 is primarily caused by short repeat arrays (represented in cyan, 1–7 RU in the European population), while FSHD2 is primarily caused by epigenetic modifiers such as LP/P variants in *SMCHD1*, *DNMT3B* and *LRIF1* (represented in pink) in FSHD2. Both are likely relevant in case of individuals affected by FSHD carrying about 8–10 RU. 20 RU is assumed as upper threshold for pathogenicity in most FSHD2 patients, however larger permissive repeat arrays can hypothetically lead to FSHD manifestation in case of strong epigenetic modifiers.

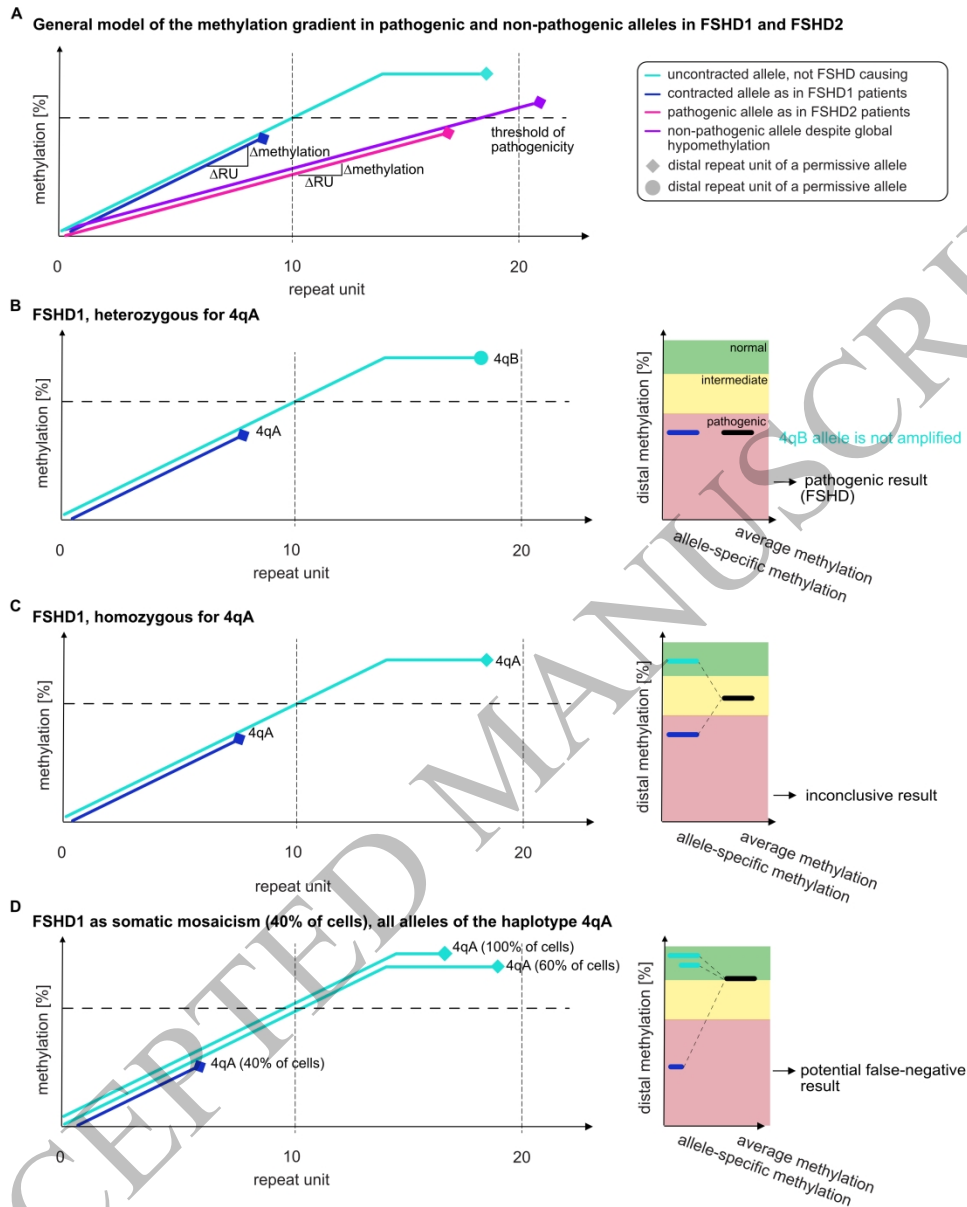


Figure 1 Illustration of the recently discovered asymmetric methylation gradient of 4q35 repeat arrays and potential biases in determining distal methylation in bisulfite sequencing-based methylation assays.

432x533mm (236 x 236 DPI)

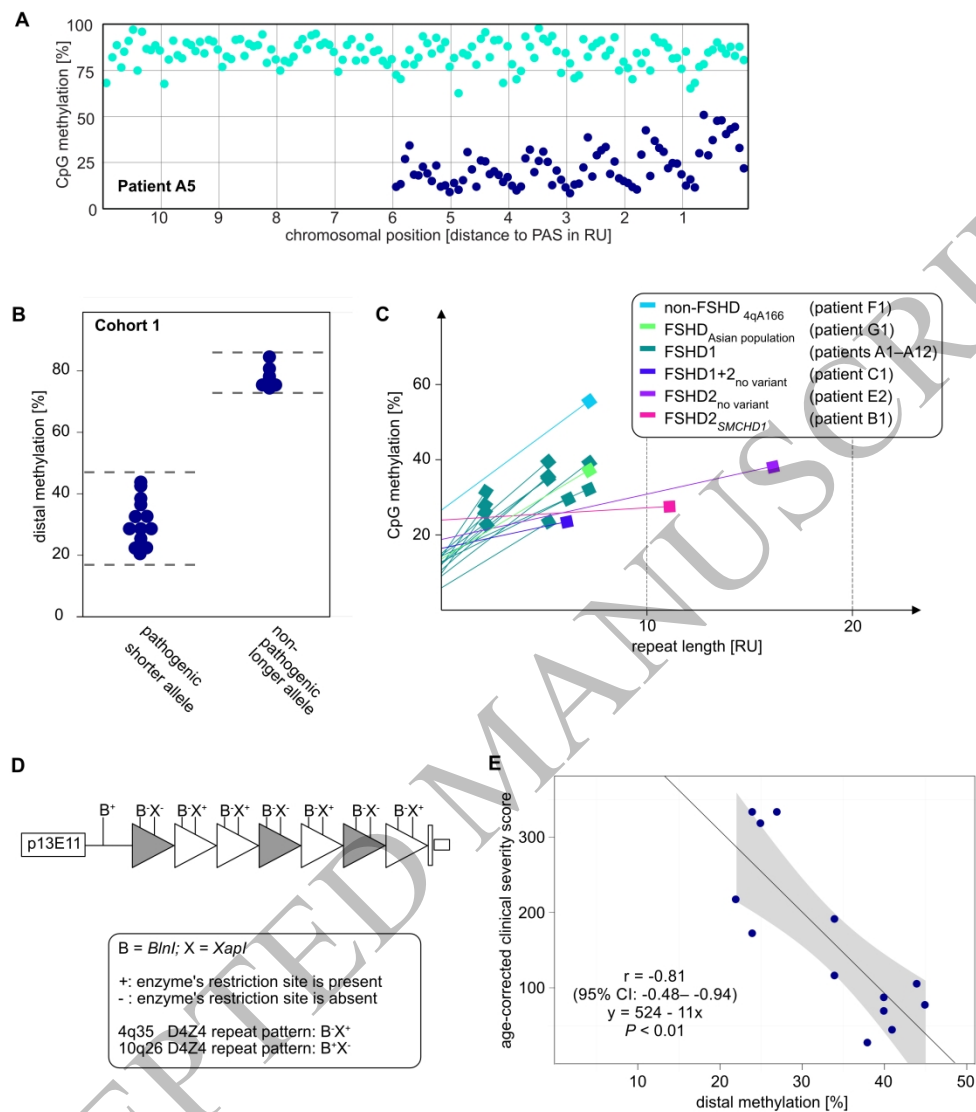


Figure 2 Results of ONT LRS of patients of Cohort 1 and 2.

432x533mm (236 x 236 DPI)

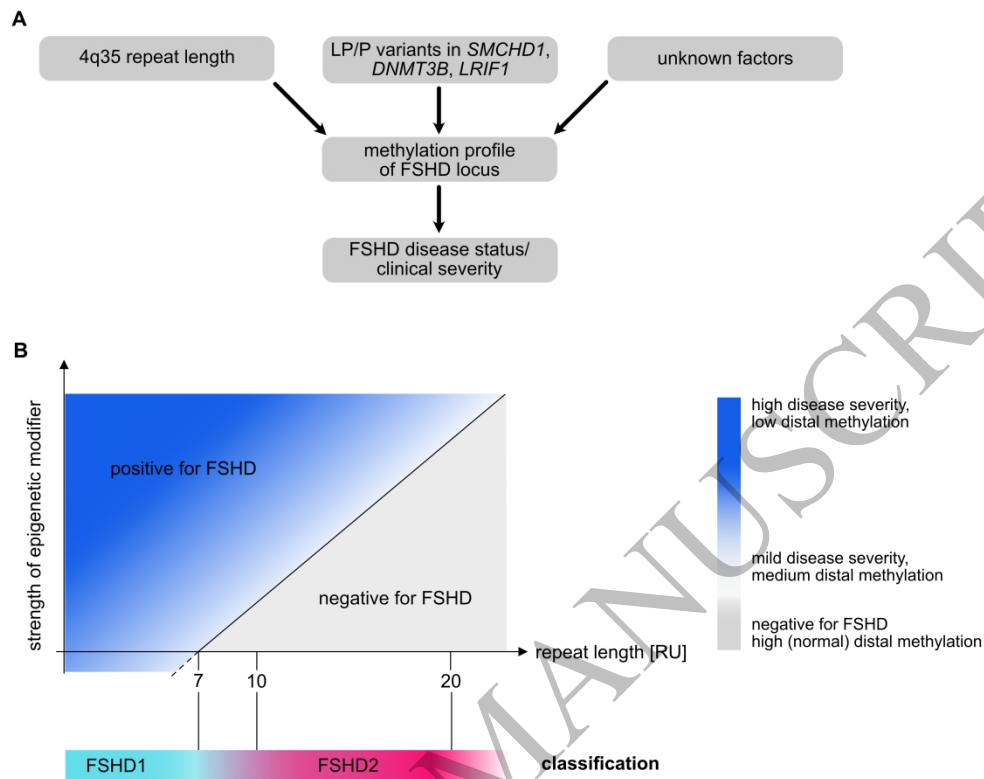
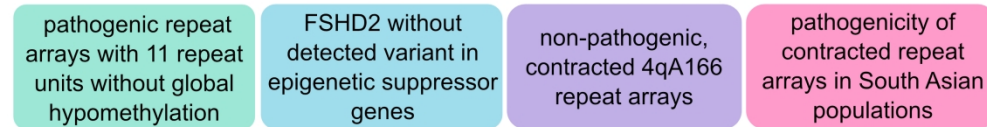


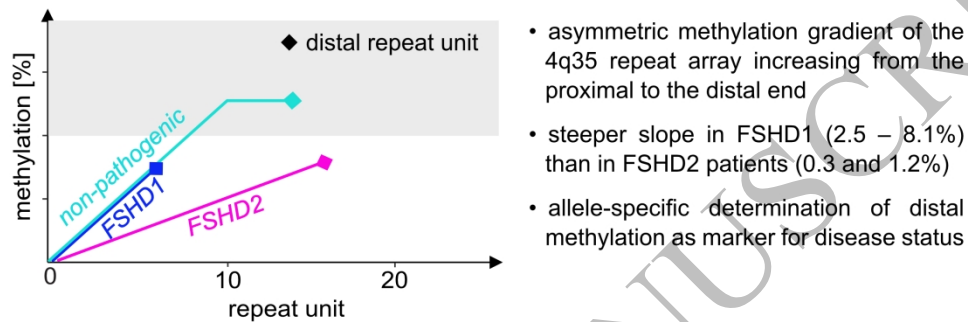
Figure 3 Linkage of genetic parameters, epigenetics (methylation) and phenotype in FSHD.

432x356mm (236 x 236 DPI)

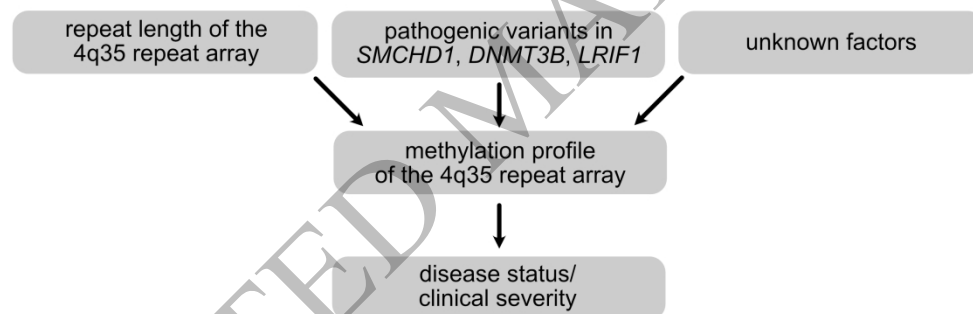
Identification of complex facioscapulohumeral muscular dystrophy (FSHD) genotypes by methylation analysis with high risk of misclassification



Long-read sequencing of common genotypes



Hypothesized linkage of genetics, epigenetics and phenotype



Graphical Abstract

305x305mm (236 x 236 DPI)