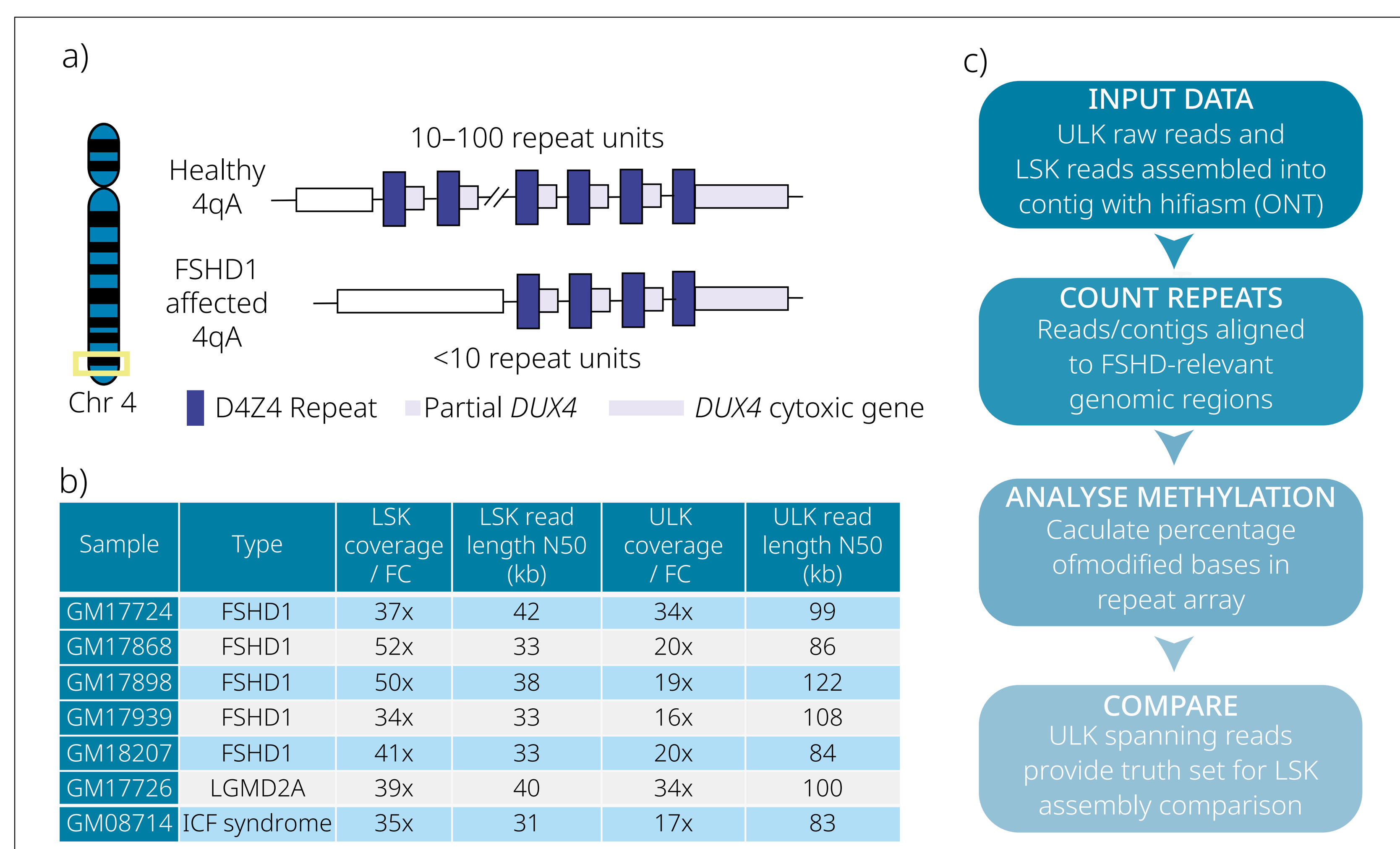
Oxford Nanopore Technologies

Structural and epigenetic characterisation of D4Z4 arrays in FSHD using Oxford Nanopore multiomic sequencing

Nanopore sequencing enables comprehensive resolution of D4Z4 repeat structure, haplotype, methylation, and chromatin accessibility, providing a consolidated approach to genetic and epigenetic characterisation of FSHD



Sample	Type	LSK coverage / FC	LSK read length N50 (kb)	ULK coverage / FC	ULK read length N50 (kb)
GM17724	FSHD1	37x	42	34x	99
GM17868	FSHD1	52x	33	20x	86
GM17898	FSHD1	50x	38	19x	122
GM17939	FSHD1	34x	33	16x	108
GM18207	FSHD1	41x	33	20x	84
GM17726	LGMD2A	39x	40	34x	100
GM08714	ICF syndrome	35x	31	17x	83

Fig. 1 a) Schematic of FSHD1 pathogenesis, b) sequencing statistics, and c) analysis workflow.

Single-platform, multiomic sequencing resolves facioscapulohumeral muscular dystrophy (FSHD)

Genetic analysis of FSHD is challenging due to the presence of highly repetitive D4Z4 macrosatellite arrays, which exist at two nearly identical subtelomeric regions on chromosomes 4q35 and 10q26. FSHD Type 1 is caused by aberrant *DUX4* expression from contraction of the D4Z4 array on a permissive 4qA haplotype (**Fig. 1a**). Current diagnostic approaches require multiple complementary techniques to assess repeat size, haplotype, sequence variation, and epigenetic status. Here, we resolved the full-length D4Z4 arrays, quantified the number of repeats, and profiled DNA methylation and chromatin accessibility in FSHD cell lines (**Fig. 1b**). Repeat counting and methylation quantification were performed on raw ultra-long reads (ULK; read length N50 >80 kb) and hifiasm-assembled contigs from standard ligation sequencing reads (LSK; read length N50 ~30 kb) (**Fig. 1c**).

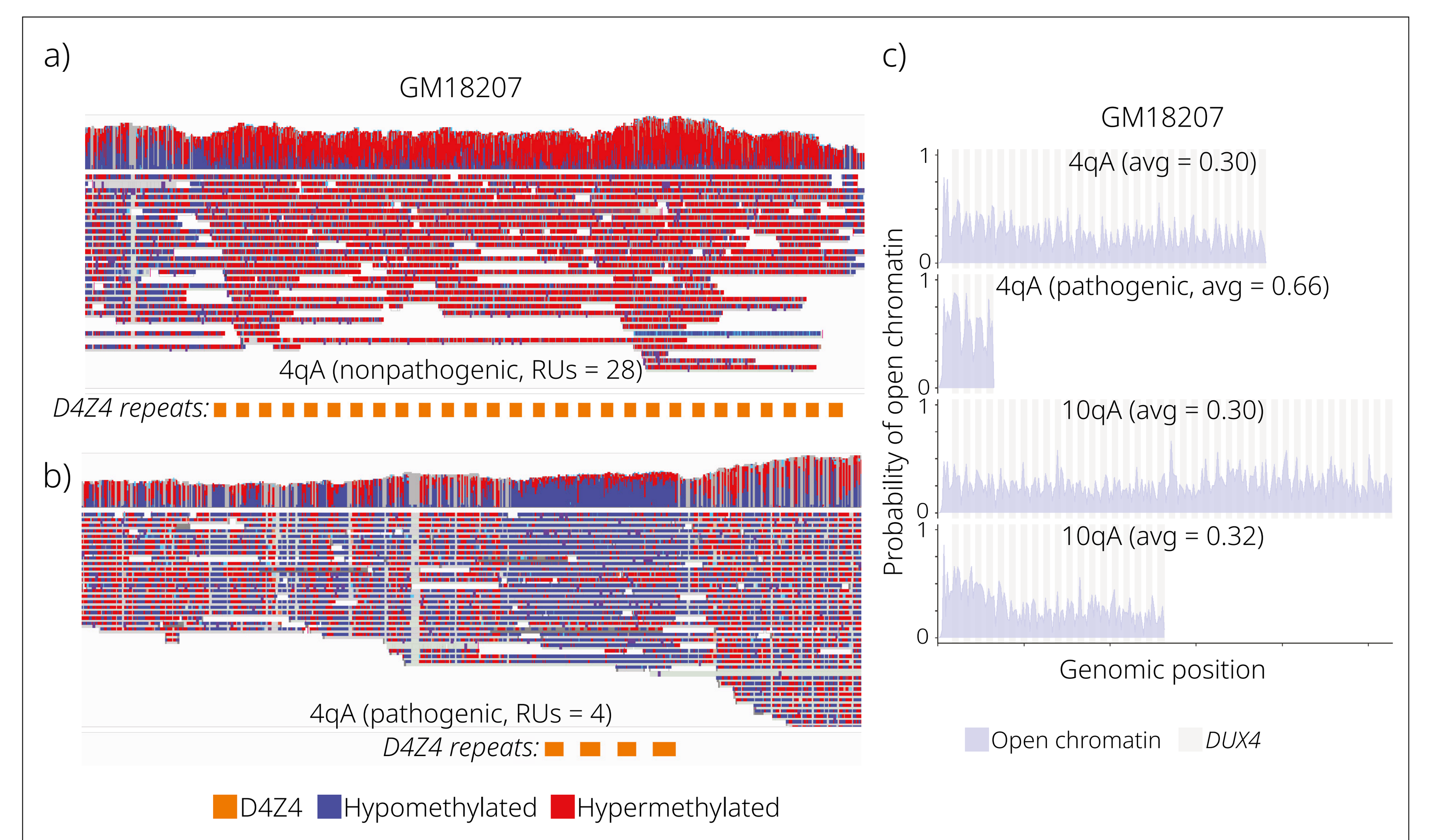


Fig. 2 a) Increased methylation, b) reduced methylation, and c) chromatin accessibility in GM18207.

Contracted D4Z4 haplotypes have reduced methylation and more open chromatin than wild-type alleles

Epigenetic dysregulation of the D4Z4 repeat array is central to the disease mechanism in FSHD. To assess methylation patterns, reads were mapped to the *de novo* LSK genome assemblies and DNA methylation levels were detected directly from the sequencing signal using Modkit. Our analyses showed that contracted, pathogenic 4qA haplotypes were hypomethylated, whereas nonpathogenic haplotypes were hypermethylated. An example from GM18207 shows a nonpathogenic allele with 28 D4Z4 repeat units with hypermethylation (**Fig. 2a**), while the pathogenic allele, with only four D4Z4 repeats, has hypomethylation (**Fig. 2b**). We also labelled open chromatin with 6mA to highlight regulatory regions using the Oxford Nanopore chromatin stenciling assay. As predicted, pathogenic haplotypes had increased chromatin accessibility, while nonpathogenic haplotypes had compacted chromatin (**Fig. 2c**).

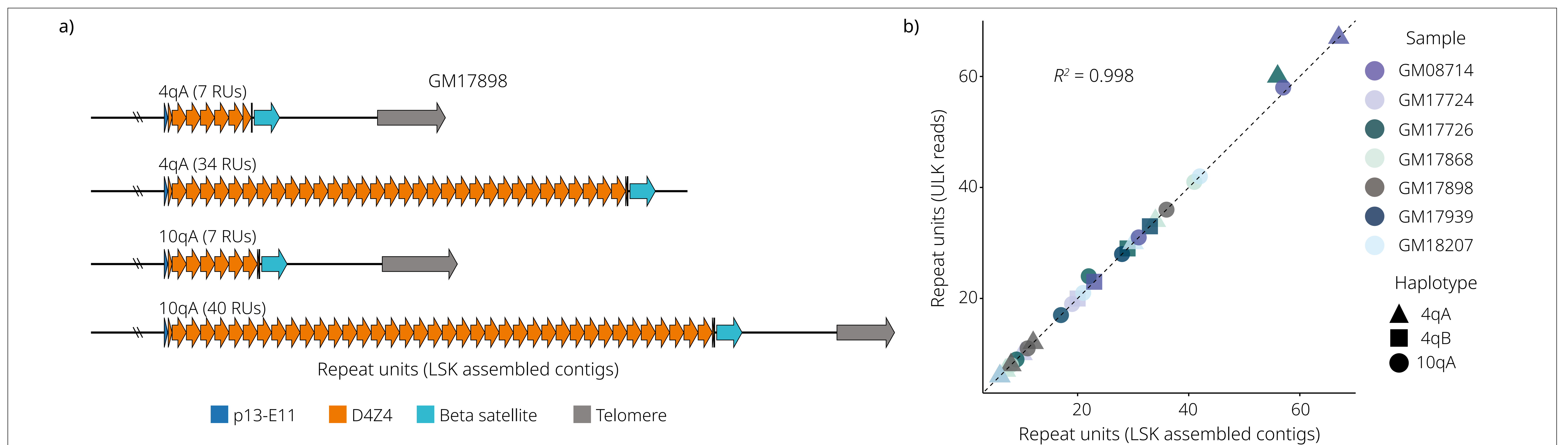


Fig. 3 a) Schematic of assembled contigs for the four haplotypes with D4Z4 repeats in GM17898 and b) comparison of the number of D4Z4 repeat units in the LSK-assembled contigs and the ULK raw reads.

Ultra-long raw reads and ligation sequencing-assembled contigs fully spanned the 10q/4q D4Z4 repeat arrays in eight FSHD samples, successfully distinguishing repeat lengths in all four haplotypes

Oxford Nanopore sequencing enables single-platform FSHD analysis by resolving D4Z4 repeat units across all four haplotypes. Two approaches were used to sequence the repeat arrays in eight cell lines (**Fig. 1c**). First, reads generated from Ultra-Long DNA Sequencing Kit-prepared libraries were used to unambiguously characterise the array in each haplotype, with individual raw reads spanning the entire array. Secondly, all samples were prepared using the Ligation Sequencing Kit (LSK) and sequenced to at least 30x coverage (~30 kb N50), then assembled using hifiasm (ONT). Using both raw ULK reads and LSK-assembled contigs, we successfully identified and counted the repeat units across the four target haplotypes in all cell lines, accurately resolving all four 4q/10q haplotypes in each sample and showing a range of 6–67 D4Z4 repeat units (**Fig. 3a,b**). For all haplotypes in all samples, we obtained at least one read in the ULK dataset that spanned the entire repeat array. The longest array spanned by a raw ULK read was 219 kb (67 D4Z4 repeat units), and the longest spanning read was 565 kb. To evaluate assembly accuracy, we compared D4Z4 repeat unit counts from the raw ULK reads with those from the LSK-assembled contigs and found that the counts were highly concordant (R^2 = 0.998; **Fig. 3b**). Taken together, both ULK and LSK workflows spanned entire arrays, allowing for accurate downstream haplotyping and variant analysis. Our findings demonstrate a single-platform, comprehensive approach for FSHD analysis using Oxford Nanopore multiomic sequencing.

Information correct at time of publication. May be subject to change.
Oxford Nanopore Technologies, the Wheel icon, AmPORE-TB, EPI2ME, GridION, MinION, MinKNOW, PromethION, P2 Solo, and P2 are registered trademarks or the subject of trademark applications of Oxford Nanopore Technologies plc in various countries. Information contained herein may be protected by copyright, patents or patents pending of Oxford Nanopore Technologies plc. All other brands and names contained are the property of their respective owners. © 2026 Oxford Nanopore Technologies plc. All rights reserved. Oxford Nanopore Technologies products are RUO. Products labelled/branded as Oxford Nanopore Diagnostics may be RUO or may be regulated as in-vitro diagnostic devices in some jurisdictions, please check individual product labelling.

PO_1341(EN)_V1_19May2026