Structural and Epigenetic Profiling of D4Z4 Arrays in FSHD using Oxford Nanopore Multiomic Sequencing

Rebecca Stubbs, Bryan Leland, Priyesh Rughani, Carly Tyer, Xiaoguang Dai, Kenney Tuyen, Tatiana O'Connell Tilley, Heather Jeffery, Philipp Rescheneder, Phil James, Scott Hickey, Sissel Juul

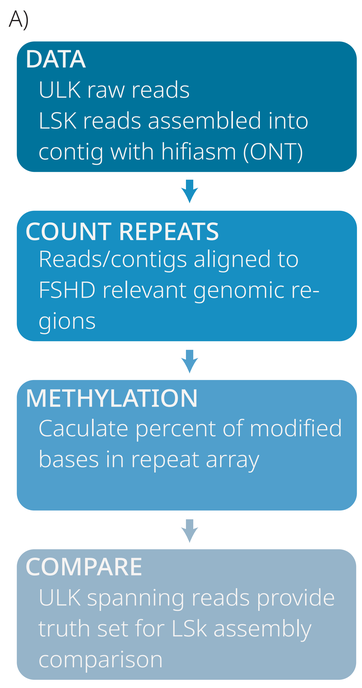
Oxford Nanopore Technologies Inc, New York, NY
Contact: rebecca.stubbs@nanoporetech.com

Abstract

Molecular analysis of facioscapulohumeral muscular dystrophy (FSHD) is challenging due to the highly repetitive D4Z4 macrosatellite array containing the *DUX4* locus, which exists at two nearly identical subtelomeric regions on chromosomes 4q35 and 10q26. FSHD results from aberrant DUX4 expression caused either by contraction of the D4Z4 array on a permissive 4qA haplotype or by pathogenic variants in chromatin-regulating genes, each leading to hypomethylation. Current diagnostic approaches require multiple complementary techniques to assess repeat size, haplotype, sequence variation, and epigenetic state. However, some recent studies have demonstrated methods for leveraging long, spanning reads to better elucidate these repeat arrays[1–3]. In this study, we resolve D4Z4 arrays, quantify number of repeat units, and profile DNA methylation and chromatin accessibility using Oxford Nanopore multiomic sequencing. Seven Coriell cell lines with known FSHD or FSHD-like phenotypes were analyzed using ultra-long (ULK; >80 kb) and standard ligation-based (LSK; 30–40 kb) sequencing, followed by *de novo* assembly. Both these methods accurately resolved all four 4q/10q haplotypes in each sample, ranging from six to 65 D4Z4 repeat units. Epigenetic profiling revealed hypomethylation and increased chromatin accessibility specifically at pathogenic 4qA haplotypes in FSHD samples. Together, these results demonstrate that Nanopore sequencing enables comprehensive, single-platform genetic and epigenetic analysis for FSHD.

Background

We sequenced seven Coriell cell lines with known FSHD or FSHD-like phenotypes (Table 1). Two approaches were used to sequence the repeat arrays (Fig. 1). First, the Ultra-Long DNA Sequencing Kit (ULK) can produce reads averaging >80 kb and was used to unambiguously characterize the array in each haplotype, with individual long reads spanning the entire array. Second, all samples were sequenced with the Ligation Sequencing Kit (LSK) to at least 30x coverage with a read-length N50 of ~30 kb, then assembled into contigs using hifiasm (ONT)[4,5] *de novo* assembly. Both ULK spanning reads and LSK contigs containing D4Z4 repeats were aligned[6], and repeat units were quantified. For both datasets, DNA methylation was detected directly from the sequencing signal and analyzed using Modkit[7]. In addition, chromatin stenciling experiments with exogenous 6mA were performed to investigate chromatin accessibility across the *DUX4* loci[8].



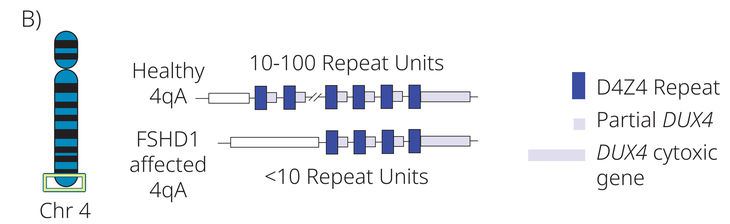


Figure 1. Overview of FSHD and analysis workflow. A) Analysis framework to capture FSHD-related genomic regions, variants, and epigenetics. B) Schematic of the pathogenesis of FSHD1. The D4Z4 macrosatellite array at the chromosome 4q subtelomere, which comprises 1-100 3.3 kb repeat units, is the causative locus of FSHD.

Table 1. Samples, sequencing coverage, and read-length N50. LSK = Ligation Sequencing Kit, ULK = Ulta-Long DNA Sequencing Kit.

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

Accurate Assemblies of D4Z4 Repeats

Using both raw ULK array-spanning reads and LSK-assembled contigs, we successfully identified and counted the repeat units across the four target haplotypes in all cell lines (Fig. 2A,B). For all samples, we obtained at least one read in the ULK dataset that spanned the entire repeat array, with the longest individual spanning read being 565 kb. To evaluate assembly accuracy, we compared D4Z4 repeat unit counts from the ULK reads with those from the LSK-assembled contigs, and the counts were highly concordant (R^2 = 0.998; Fig. 2B).

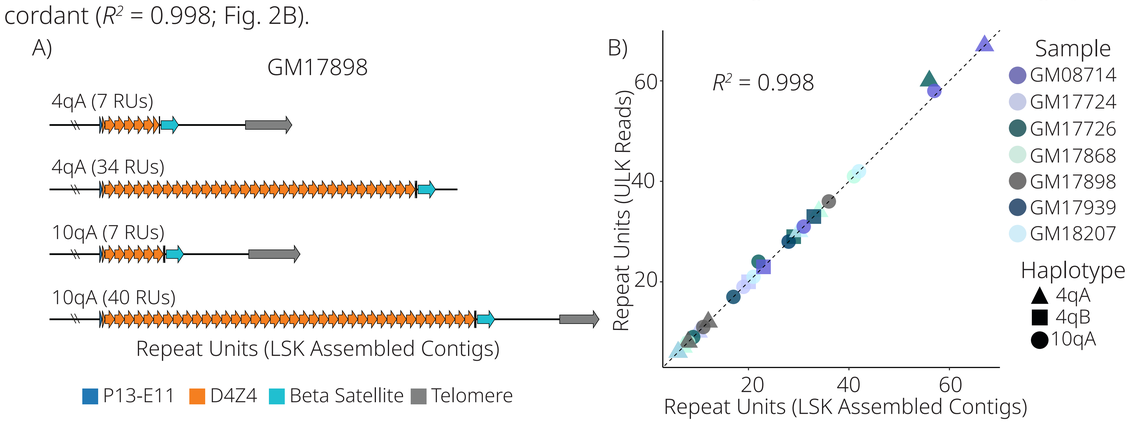


Figure 2. Assembly of LSK reads accurately represents the number of D4Z4 repeats. A) Schematic of assembled contigs for the four haplotypes with D4Z4 repeats. Assemblies created with hifiasm (ONT) and LSK reads (>30kb read-length N50). Each schematic includes key features of the repeat and number of repeat units (RUs). B) Comparison of number of D4Z4 RUs in the LSK-assembled contigs and the ULK reads for every sample in the study.

Epigenetic States Distinguish Pathogenic Haplotypes

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. Our analyses showed that pathogenic 4qA haplotypes were hypomethylated, whereas nonpathogenic haplotypes were hypermethylated (Fig. 3A,B). Further, we labeled open chromatin with 6mA to highlight regulatory regions using the Oxford Nanopore chromatin stenciling assay[8]. Pathogenic haplotypes had increased chromatin accessibility while nonpathogenic haplotypes had compacted chromatin (Fig. 3C).

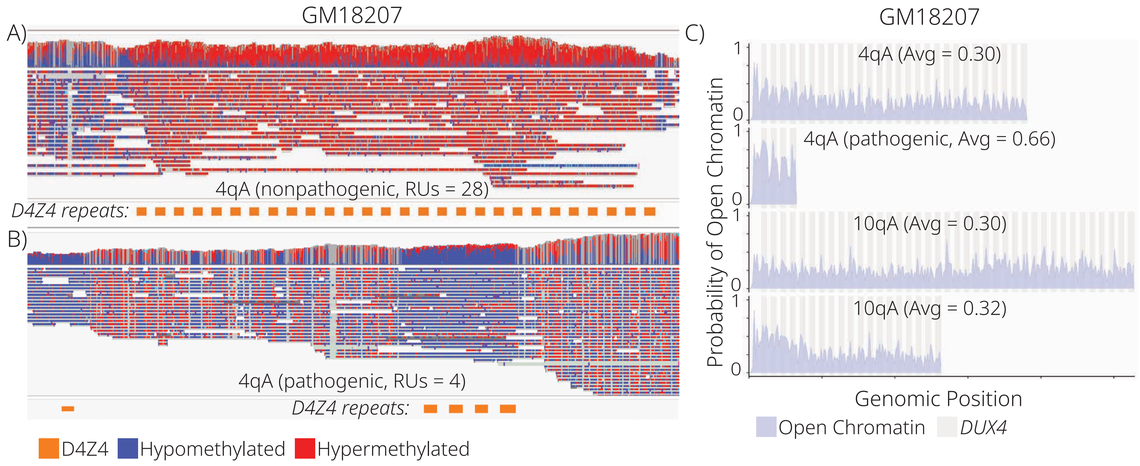


Figure 3. Pathogenic and nonpathogenic haplotypes show epigenetic differences in both methylation profiles and open chromatin. A) and B) LSK read pileup colored by methylation status. A) Nonpathogenic 4qA haplotype with 28 D4Z4 repeats has increased methylation at the repeat array. B) Pathogenic 4qA haplotype with four D4Z4 repeats has reduced methylation at the repeat array. C) Plot of chromatin accessibility across *DUX4* repeats per haplotype. The probability of open chromatin is higher in pathogenic haplotypes compared to nonpathogenic haplotypes.

Conclusion

Oxford Nanopore sequencing enables single-platform FSHD testing by resolving and counting D4Z4 repeat units across all four *DUX4* alleles. Both ULK (array-spanning reads) and LSK (*de novo* assembly with hifiasm (ONT)) workflows spanned entire arrays, allowing for accurate downstream haplotyping and variant analysis. From ULK spanning reads, we recover complete repeat array size and D4Z4 unit counts, while the LSK assembly returns concordant results using ~30 kb reads. Direct detection of native DNA methylation revealed the expected hypomethylation of the pathogenic array, and chromatin accessibility analysis using chromatin stenciling identified open chromatin at the pathogenic contracted allele. Taken together, our findings demonstrate a single-platform, comprehensive approach for FSHD analysis using Oxford Nanopore multiomic sequencing.

References

1. Xiao, L. C. et al. D4Z4End2End: complete genetic and epigenetic architecture of D4Z4 macrosatellites in FSHD, BAMS and reference cohorts.
2. Yeow, D. et al. Targeted long-read sequencing enables comprehensive analysis of the genetic and epigenetic landscape of inherited myopathies.
3. Löwenstern, T. et al. DUCKS4: a comprehensive workflow for Nanopore sequencing analysis of facioscapulohumeral muscular dystrophy (FSHD). Hum. Genomics (2026).
4. Cheng, H. et al. Efficient near-telomere-to-telomere assembly of nanopore simplex reads. Nature (2026).
5. Cheng, H., Concepcion, G. T., Feng, X., Zhang, H. & Li, H. Haplotype-resolved de novo assembly using phased assembly graphs with hifiasm. Nat. Methods 18, 170–175 (2021).
6. Li, H. Minimap2: pairwise alignment for nucleotide sequences. Bioinformatics 34, 3094–3100 (2018).
7. Modkit. Oxford Nanopore Technologies (2026). <https://nanoporetech.github.io/modkit/>
8. Chromatin accessibility know-how document. Oxford Nanopore Technologies (2026). < https://nanoporetech.com/document/requirements/chromatin-accessibility-know-how-document>