Resolution of the D4Z4 repeat responsible for facioscapulohumeral muscular dystrophy with HiFi sequencing

Poster 4104F

Xiao Chen[1], Zev Kronenberg[1], Joseph M. Devaney[2], Jessica Noya[2], April S. Berlyoung[2], Shamila Yusuff[2], Solomon Lynch[2], Keith Nykamp[2], Amanda S. Lindy[2], Egor Dolzhenko[1], Michael A. Eberle[1]

1. PacBio, Menlo Park, CA, USA 2. GeneDx, Gaithersburg, MD, USA

Introduction

The D4Z4 repeat is a variable number tandem repeat (VNTR) that contains some of the most difficult-to-resolve medically relevant alleles in the human genome. D4Z4 has a repeat unit of 3.3 kb (encoding the *DUX4* gene) that is repeated between 1-100 times per allele. Facioscapulohumeral muscular dystrophy (FSHD) is caused by ectopic expression of *DUX4*, mediated by contraction of D4Z4 to less than 11 copies (FSHD1, 95% of FSHD cases) or hypomethylation of D4Z4 induced by mutations in the epigenetic machinery such as *SMCHD1* (FSHD2, 5% of FSHD cases). D4Z4 is present on both chromosome 4 and chromosome 10, while *DUX4* can only be expressed from the permissive A haplotype that usually occurs on chromosome 4. The D4Z4 repeat size is conventionally measured by Southern blot analysis, which is low-throughput and can be confounded by genetic polymorphism. Due to the large repeat size, high polymorphism and homology issues, D4Z4 remains difficult to resolve by sequencing technologies and population analysis is lacking. Here we present a computational tool, Kivvi, to resolve D4Z4 using PacBio HiFi whole-genome sequence data.

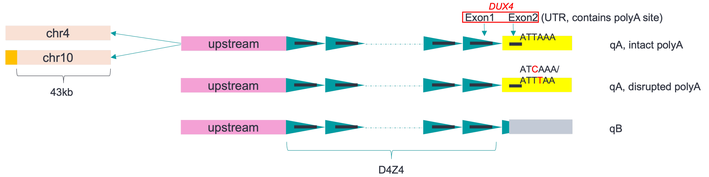

Figure 1. Overview of the D4Z4 locus. D4Z4 is present on both chromosome 4 and chromosome 10, with a homology region that extends upstream of D4Z4 for 43 kb. A D4Z4 repeat unit encodes one exon of *DUX4*. The distal haplotype downstream of D4Z4 controls *DUX4* expression via the polyadenylation site.

Kivvi workflow

Kivvi identifies unique copies of the repeat unit and assembles them into alleles. For each D4Z4 allele, Kivvi identifies the chromosome (4 vs. 10), distal haplotype (A vs. non-permissive haplotypes) and the methylation level.

Kivvi on GitHub

Figure 2. Unique copies of the D4Z4 repeat units in a sample. HiFi reads are realigned to the D4Z4 reference sequence. Reads from the qA haplotype extend into the last exon overlapping the polyA site, while reads from the qB haplotype do not overlap the last exon.

Application of Kivvi in FSHD patient samples

- 11 FSHD1 samples, allele size known from Bionano optical genome mapping and/or Southern Blot.
- 6 samples with *SMCHD1* variants - candidates for FSHD2.
- 473 population samples from five ancestral populations.

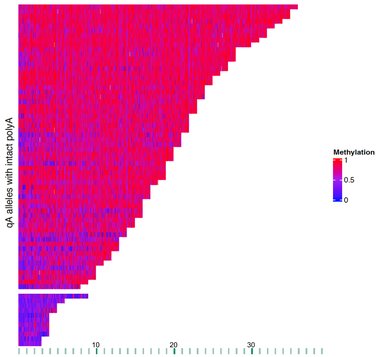

Figure 3. Repeat size and methylation level of assembled D4Z4 alleles with intact polyA. Shown at the bottom are 11 alleles with copy numbers 3-9 in 11 FSHD1 patient samples orthogonally confirmed.

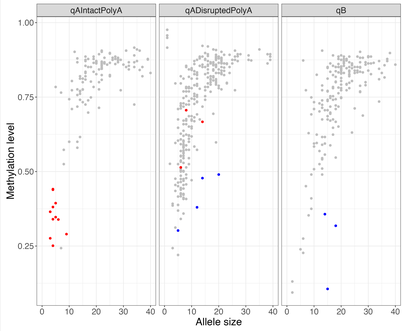

Figure 4. Fully assembled D4Z4 alleles and their methylation levels.

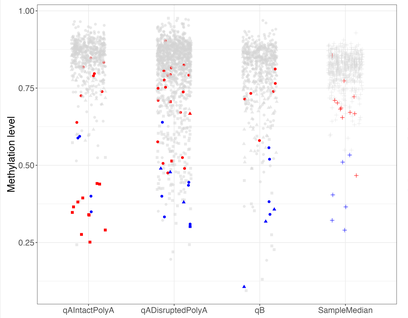

Figure 5. Methylation levels of all D4Z4 alleles, including partially assembled ones, and summary methylation values for each sample (right column).

Sequence analysis across populations

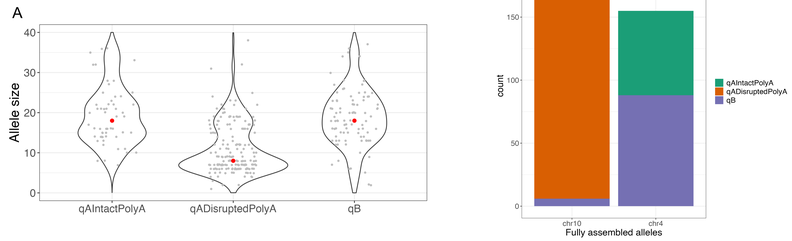

Figure 6. Distribution of D4Z4 alleles. (A) Copy number. **(B)** Proportions of distal haplotypes on chr4 and chr10.

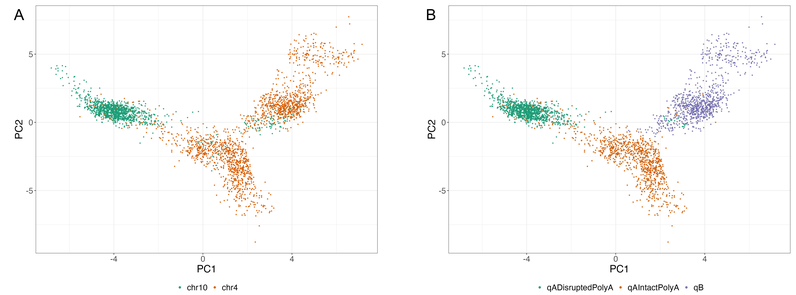

Figure 7. Principal component analysis of D4Z4 repeats based on sequence variants. Each dot is a D4Z4 repeat unit. **(A)** Separation of chr4 and chr10 alleles. **(B)** Separation of distal haplotypes.

In cis duplications of D4Z4 arrays

Figure 8. Sample with four assembled D4Z4 alleles, two of which are in cis with one or two duplicated D4Z4 arrays downstream.

References

1. *Lemmers* et al. A unifying genetic model for facioscapulohumeral muscular dystrophy. *Science*, 2010.
2. *Lemmers* et al. Digenic inheritance of an *SMCHD1* mutation and an FSHD-permissive D4Z4 allele causes facioscapulohumeral muscular dystrophy type 2. *Nat Genet*. 2012.
3. *Lemmers* et al. Autosomal dominant in cis D4Z4 repeat array duplication alleles in facioscapulohumeral dystrophy. *BRAIN* 2024.