High throughput analysis of tandem repeat contraction associated with Facioscapulohumeral Muscular Dystrophy (FSHD) by optical mapping

Jian Wang, Ernest T Lam, Andy WC Pang, Tom Wang, Dong Zhang, Henry B Sadowski, Alex R. Hastie, Mark Oldakowski

Bionano Genomics, San Diego, California, United States of America

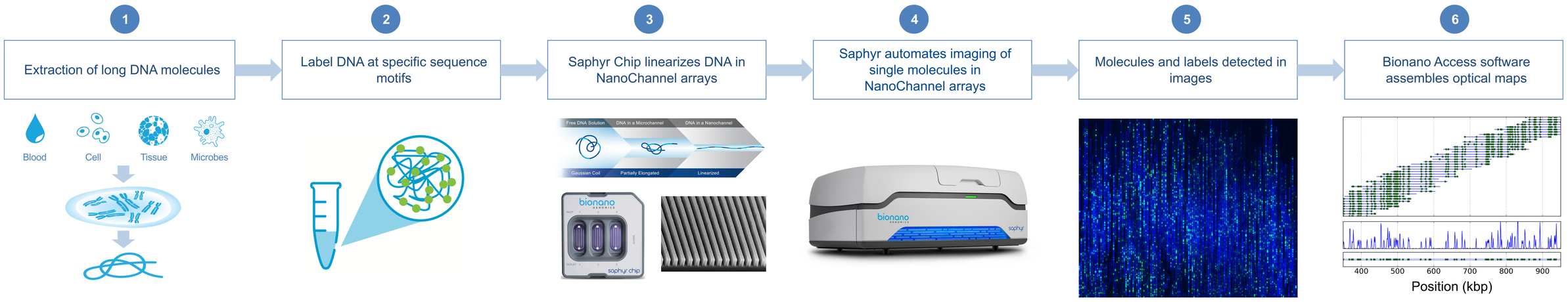
Abstract

Tandem repeats play important roles in gene regulation and chromosome structures, and are associated with various diseases. PCR and sequencing can characterize short repeats, but are ineffective when repeat size exceeds the PCR amplicon or sequence read. For large repeats, gel electrophoresis plus Southern blot analysis and fluorescent in situ hybridization (FISH) are used. While relatively effective, these procedures are laborious and highly specialized to each disease.
We introduced a new workflow based on optical mapping on the Bionano Genomics Saphyr platform to assay for repeat-instability disorders. As an example, we examined samples with Facioscapulohumeral muscular dystrophy (FSHD). It is the third most common genetic diseases of skeletal muscle. FSHD can be diagnosed by looking for a contraction of the D4Z4 repeats at the chromosome region of 4q35 with the permissive haplotype, commonly referred to as the 4qA haplotype.
We showed that in 12 samples with known FSHD phenotypes we can correctly detect the D4Z4 repeat contractions and assign the correct haplotype in the disease allele. We further show that in 58 control samples without known FSHD, we size the repeat and determine the haplotypes correctly and detect no FSHD-type contraction in these samples. Reproducibility experiments on a subset of these samples show that we can consistently obtain equivalent results in all tested cases.
Bionano offers sample preparation, DNA imaging and genomic data analysis technologies combined into one streamlined workflow that enables high-throughput analysis of tandem repeat regions of interest.

Background

Generating high-quality finished genomes replete with accurate identification of structural variation and high completion (minimal gaps) remains challenging using short read sequencing technologies alone. The Saphyr™ system provides direct visualization of long DNA molecules in their native state, bypassing the statistical inference needed to align paired-end reads with an uncertain insert size distribution. These long labeled molecules are *de novo* assembled into physical maps spanning the entire diploid genome. The resulting provides the ability to correctly position and orient sequence contigs into chromosome-scale scaffolds and detect a large range of homozygous and heterozygous structural variation with very high efficiency.

Methods

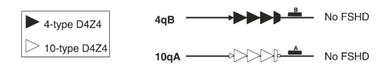


(1) Long molecules of DNA are labeled with Bionano reagents by (2) incorporation of fluorophores at a specific sequence motif throughout the genome. (3) The labeled genomic DNA is then linearized in the Saphyr Chip using NanoChannel arrays (4) Single molecules are imaged by Saphyr and then digitized. (5) Molecules are uniquely identifiable by distinct distribution of sequence motif labels (6) and then assembled by pairwise alignment into *de novo* genome maps.

Introduction to Facioscapulohumeral muscular dystrophy

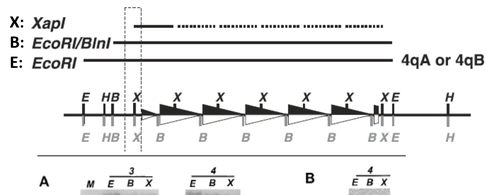
FSHD is one of the most common inherited muscular dystrophy that affects face, scapula and upper arms. Life expectancy can be threatened by respiratory insufficiency, and up to 20% of affected individuals become severely disabled. About 95% of case of FHSD are associated with the contraction of the D4Z4 regions at subtelomeric region of chromosome 4q35. In addition, the disease only occurred when the contraction is presented on an allele with permissive haplotype (4qA). There is a homologous D4Z4 repeat in chromosome 10 which is not related to FSHD, which can confound the repeat analysis.

Basics for the genetic diagnosis of FSHD

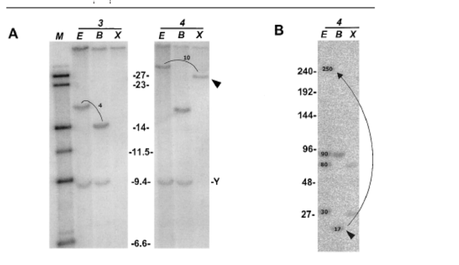


Current standard of care to confirm FSHD diagnosis is mainly through a southern blot assay which measures the size of D4Z4 repeat region.

Digestion site of different enzyme in D4Z4 region [Lemmel et.al. 2012]



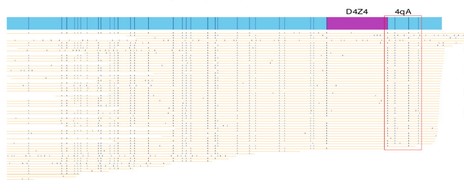
Example of a **A)** Linear gel analysis and **B)** Pulse Filed Gel analysis from Southern Blot assay [Lemmel et.al. 2001]



Tandem repeat characterization with Bionano optical maps

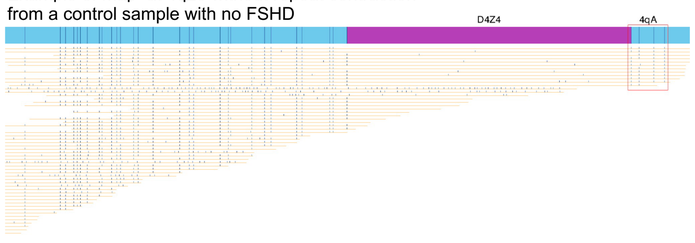
We performed reference-guided assembly of molecules near the D4Z4 regions, generating consensus maps around the D4Z4 repeat regions for both chromosome 4 and 10.

- Chromosome-specific label pattern allow us assign each map to the correct chromosome location.
- The ultra-long molecules spanning the whole repeat region allow us to accurately size the D4Z4 repeat.
- The haplotype-specific label pattern at the end of the map allow us to distinguish the permissive haplotype (4qa) from the non-permissive one (4qb).

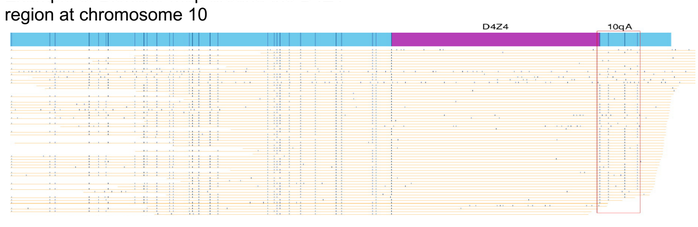


Example of Bionano map with a contraction in the D4Z4 repeat region from sample with known FSHD phenotype.

Example of map at 4q23 without repeat contraction from a control sample with no FSHD



Example of Bionano map around the D4Z4 region at chromosome 10



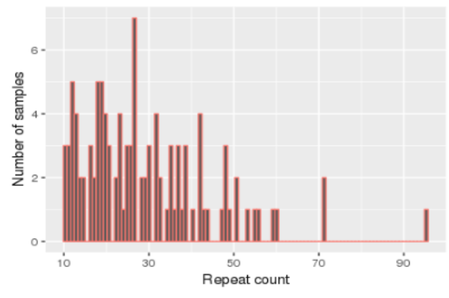
Evaluation of using Bionano optical map to characterize D4Z4 tandem repeats

Sensitivity: We evaluated the sensitivity of our pipeline in 12 Coriell samples with known FSHD phenotype. We are able to detect the D4Z4 repeat contraction on all the samples. Furthermore, the repeat size and haplotype assignment are consistent with Southern blot assays.

Sample Name	Repeat length	Haplotype	Consistent with annotation*
GM16250	5	A	Yes**
GM16283	5	A	Yes**
GM16334	4	A	Yes**
GM16337	4	A	Yes**
GM16348	3	A	Yes**
GM16354	8	A	Yes**
GM16420	5	A	Yes**
GM17724	8	A	Repeat differs by 2 units
GM17868	5	A	Yes
GM17898	6	A	Repeat differs by 2 units
GM17939	4	A	Yes
GM18027	4	A	Yes

*Annotation are extracted from Coriell sample website: https://www.coriell.org/
** Samples with annotation from southern blot assay

Specificity: We evaluated our pipeline on 58 control samples without known FSHD phenotypes and did not detect any FSHD-type contraction in 56 samples. The remaining two cases have a repeat count of 10 on 4qa alleles which are considered borderline cases that may not show FSHD phenotypes in the diagnostic literature [Butz et. al. 2003].



Consistency and Reproducibility:
To further validate our results, we generate data with a second enzyme for the 12 Coriel samples and 20 control samples and assemble the same D4Z4 locus. In all cases the repeat size and haplotype assignments are consistent between the two enzymes.

We also selected a subset of six Coriell samples and perform reproducibility analysis by running the samples in triplicates. In all cases we produced equivalent results.

Conclusions

Bionano offers sample preparation, DNA imaging and genomic data analysis technologies combined into one streamlined workflow that enables high-throughput analysis of tandem repeat regions of interest. Together, these components allows for efficient analysis of diseases associated with repeat expansion and contraction.

Reference

1) Cao, H., et al., Rapid detection of structural variation in a human genome using NanoChannel-based genome mapping technology. Gigascience (2014); 3(1):34.
2) Hastie, A.R., et al. Rapid Genome Mapping in NanoChannel Arrays for Highly Complete and Accurate De Novo Sequence Assembly of the Complex Aegilops tauschii Genome. PLoS ONE (2013); 8(2): e55864.
3) Das, S. K., et al. Single molecule linear analysis of DNA in nano-channel labeled with sequence specific fluorescent probes. Nucleic Acids Research (2010); 38: 8.
4) Xiao, M et. al. Rapid DNA mapping by fluorescent single molecule detection. Nucleic Acids Research (2007); 35:e16.
5) Butz, Miriam, et al. "Facioscapulohumeral muscular dystrophy." Journal of neurology 250.8 (2003): 932-937.
6) Lemmers, Richard JLF, et al. "Best practice guidelines on genetic diagnostics of Facioscapulohumeral muscular dystrophy: workshop 9th June 2010, LUMC, Leiden, The Netherlands." *Neuromuscular disorders* 22.5 (2012): 463-470.