Optical Genome Mapping for High Throughput Analysis of Repeat Expansion Disorders

Martin D Muggli[1], Bahar Ramandi, Neil Miller, Dong Zhang, Ernest T Lam, Jian Wang, Tom Wang, Joyce Lee, Andy WC Pang, Henry B Sadowski, Alex R Hastie, Mark Oldakowski

Bionano Genomics, San Diego, USA

Background:

Genomic tandem repeat expansions and contractions are associated with degenerative disorders such as facioscapulohumeral muscular dystrophy (FSHD) and Fragile X syndrome (FXS). Early detection provides psychosocial benefits and enables research into treatments. The repetitive and polymorphic nature of these regions presents difficulties for both polymerase chain reaction (PCR) and next generation sequencing (NGS). Optical genome mapping (OGM) in nanochannel arrays offers several advantages. Large repeat arrays are spanned with single ultra-long molecules, and the distance between the flanking labels is precisely measured to size the repeats independent of the sequence content.

Methods and results:

Two regions are analyzed and described here:

1. the D4Z4 repeat on chromosome 4q35, contraction of which is diagnostic for FSHD when present on a permissive haplotype. We analyzed 30 FSHD-positive cases, 28 of which had repeat contractions on the 4qA (permissive) haplotype. We also analyzed 58 FSHD-negative samples; none had repeat counts in the contracted/permissive structure.
2. the microsatellite repeats in X-linked dominant FMR1 gene, expansion of which is associated with FXS. We analyzed 45 cases with orthogonal results from Southern blot, 10 simulated heterozygotes and 20 negative controls to make a total validation set of 75. Our analysis shows diagnostic performance of 100% PPV and 97% sensitivity.

Conclusions:

Optical genome mapping offers sample preparation, DNA imaging and genomic data analysis in a single streamlined workflow that enables high-throughput analysis of tandem repeat regions of interest. Together, these components allow for efficient analysis of diseases associated with repeat expansion and contraction.

References:

1. Zheng Y et al., 2019, Rapid prenatal diagnosis of Facioscapulohumeral Muscular Dystrophy 1 by combined Bionano optical mapping and karyomapping
2. Dai Y et al., 2019, Single-molecule optical mapping enables quantitative measurement of D4Z4 repeats in facioscapulohumeral muscular dystrophy (FSHD)
3. Zhang Q et al., 2019, Clinical application of single-molecule optical mapping to a multigeneration FSHD1 pedigree
4. Sahajpal N et al., 2021, Optical Genome Mapping as a Next-Generation Cytogenomic Tool for Detection of Structural and Copy Number Variations for Prenatal Genomic Analyses

FSHD EnFocus™ Analysis

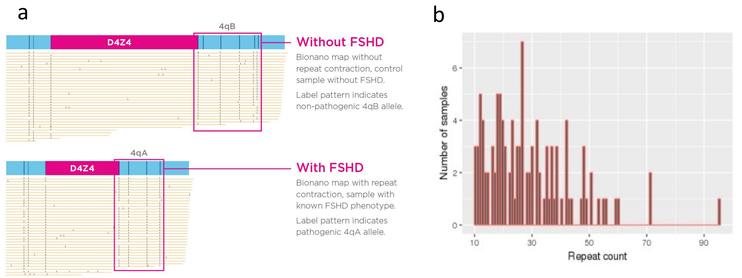


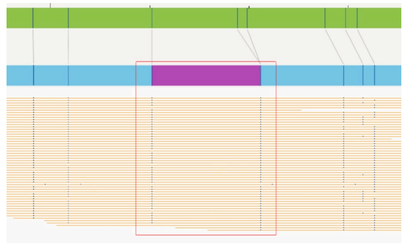
Fig. 1 (a) FSHD haplotype and (b) Specificity. We analyzed 58 control samples with no reported FSHD phenotypes from donors (45 samples) and 1000 Genomes Project. None had pathogenic repeat contractions.

Sample	Repeat Units	Haplotype	Consistent with annotation*?
GM16250	5	4QA	Yes**
GM16283	5	4QA	Yes**
GM16334	4	4QA	Yes**
GM16337	4	4QA	Yes**
GM16348	3	4QA	Yes**
GM16354	8	4QA	Yes**
GM16420	5	4QA	Yes**
GM17724	8	4QA	Repeat differed by 2 units
GM17868	5	4QA	Yes
GM17898	6	4QA	Repeat differed by 2 units
GM17939	4	4QA	Yes
GM18027	4	4QA	Yes

Table 1 Evaluation of 12 Coriell samples with known FSHD genotypes. We detected D4Z4 repeat contractions consistent with annotation in all samples.

*Annotation from Coriell website: https://www.coriell.org/
** Samples with Southern Blot data

Fragile X EnFocus™ Analysis



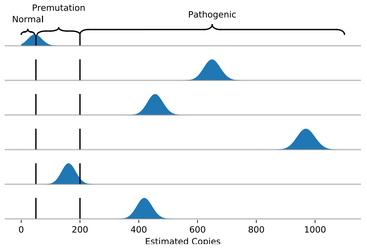


Fig. 2
(a) Bionano map with FMR1 expansion (645 units).

(b) Visualization of calculated likelihoods for six Coriell samples. Analysis includes the most likely repeat count, 99% credible interval, and probability the sample is pathogenic.

(c) Confusion matrix describing overall diagnostic performance in 75 samples. Analytical sensitivity was 97% with 100% PPV. Analysis produced one false negative sample annotated as 200 repeats, at the limit of full *FMR1* mutation

c

Annotation \ Prediction	+	-
+	36	1
-	0	38

	CI < 195			CI [195, 215]			CI > 215				Coverage level summary		
	TN	FN	Accuracy	TN	FN	Accuracy	TP	FN	FP	Accuracy	Sens.	PPV	Accuracy
400 Gb	33	0	100.0%	5	1	83.3%	36	0	0	100.0%	97.3%	100.0%	**98.7%**
800 Gb	13	0	100.0%	3	1	75.0%	22	0	0	100.0%	95.7%	100.0%	**97.4%**
1.2 Tb	12	0	100.0%	3	1	75.0%	21	0	0	100.0%	95.5%	100.0%	**97.3%**
CI Summary			100.0%			78.6%				100.0%			

Table 2 Validation results for all samples and replicates at three coverage tiers and stratified by repeat size credible interval (CI). The assay produced 100% PPV with no false positives at all coverage levels. Results show 100% accuracy when CI is less than 195 or greater than 215 with one false negative (FN) call in the interval between 195 and 215.

Corresponding Author: Martin Muggli, mmuggli@bionanogenomics.com