High Resolution Analysis Of D4Z4 Repeat Regions For Studying Facioscapulohumeral Muscular Dystrophy (FSHD) Using Whole Genome Optical Mapping

Suresh Shenoy[1], Yi Hao Liu[1], Mark Oldakowski[2], Alex Hastie[2], Ernest Lam[2], Henry Sadowski[2], Satish Khadilkar[3], Rashna Dastur[4], Pradnya Gaitonde[4], Rajiv Rose[5], Madhuri Hegde[6], Silvere Van der Maarel[7], R.J.L.F. Lemmers[7], Alka Chaubey[6].

[1]Revvity Omics, Pittsburgh, PA, USA, [2]Bionano Genomics, San Diego, CA, USA, [3]Department of Neurology, Bombay Hospital, Mumbai, India, [4]Centre for Advanced Molecular Diagnostics in Neuromuscular Disorders, Mumbai, India, [5]Revvity Omics, Chennai, India, [6]Revvity Omics, Duluth, GA, [7]Leiden University Medical Center, Leiden, The Netherlands.

BACKGROUND

Facioscapulohumeral muscular dystrophy (FSHD), is one of the most common forms of autosomal dominant, progressive muscular dystrophies affecting approximately 1/10,000-1/25,000 individuals worldwide. The clinical features of this condition vary from a severe infantile form to a less severe adult onset form, include muscular dystrophy involving progressive wasting of the muscles of the face, shoulder blades, upper arms as well as the lower legs. Detection and diagnosis of FSHD is very challenging by Next Generation Sequencing (NGS) methodologies due to the presence of polymorphic microsatellite repeat cluster (D4Z4) at two chromosomal locations in the human genome (4q and 10q). The gold standard for FSHD testing is the accurate detection of the D4Z4 repeat contractions by Southern blot analysis. Due to the time consuming and labor-intensive costs required for detecting the D4Z4 repeat contractions by conventional methodologies, we have evaluated a high-throughput, optical mapping methodology to validate and accurately map the repeats on chromosomes 4q and 10q using the Bionano Saphyr® genome imaging platform

METHODS

Briefly, whole genome optical mapping involves isolating large molecules of the DNA (150kb to ≥1Mb), uniformly labeling them at a specific 6-base sequence motifs, loading the labeled DNAs into a cartridge, where the molecules are electrophoretically linearized and imaged multiple times using the Bionano Genomics Saphyr® platform. Using the captured images, a *de novo* genome map indicating the positions of the labels is constructed and compared to a reference genome to detect structural differences in the 2 maps. Molecules aligning to regions of interest in chromosome 4 and chromosome 10 were extracted and assembled. The resulting consensus genomic regions of interest maps were used for the Bionano EnFocus™ FSHD Analysis. The repeat arrays were sized, and the permissive and non-permissive alleles (4qA and 4qB) assigned. A total of 44 DNA samples were utilized for LDT evaluation of this assay including 6 Coriell FSHD positive cell lines (performed in triplicate), 6 normal controls (3 males and 3 females), and 14 clinically diagnosed FSHD patients Additional structural variants and copy number gains and losses were interrogated in the proximity of the D4Z4 repeat array on chromosome 4. Copy number gains and losses in the proximity of the *SMCHD1* gene on chromosome 18 were also interrogated.

Figure 1. BioNano Genomics Workflow for Optical Mapping using the Saphyr® System

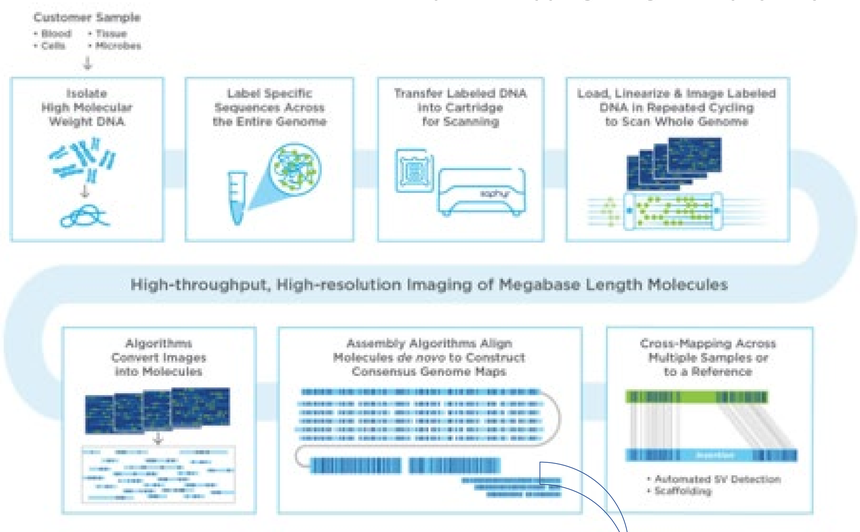
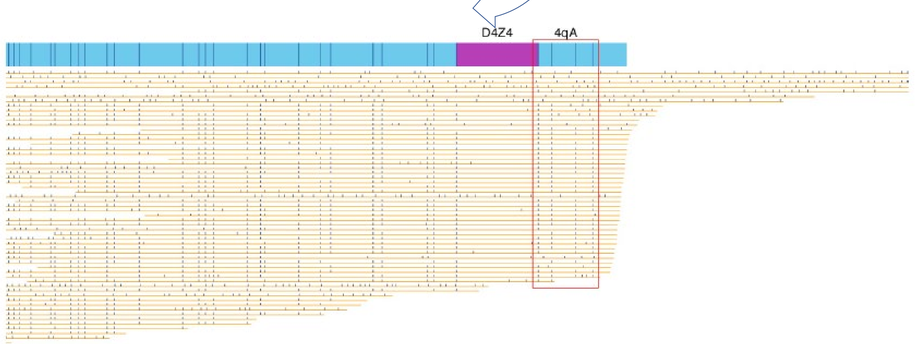


Figure 2. Example of a consensus optical map of a true positive showing the repeat count of 3 at 4qA



RESULTS

Table 1. Performance of True Positives

Cell Line	Calculated Repeat Counts Short Allele					Calculated Repeat Counts Long Allele				
	Coriell Annotation	Observed Run 1	Observed Run 2	Observed Run 3	Observed Run 4	Coriell Annotation	Observed Run 1	Observed Run 2	Observed Run 3	Observed Run 4
GM16520	6	5	5	5	5	NA	17	17	17	17
GM16337	5	4	4	4	4	NA	42	42	42	42
GM16348	4	3	3	3	3	NA	28	28	27	28
GM16354	9	8	8	8	8	NA	38	38	38	38
GM17868	5	5	5	5	5	31	32	33	32	32
GM18027	3	4	4	4	4	27	28	28	28	28

[1]Repeat count units have ±1 margin

True positives show contraction in 4qA repeat number and a normal range of 4qB repeat counts

Table 2. Performance of True Negatives

Normal Human Blood	Calculated Repeat Counts 4qA		Calculated Repeat Counts 4qB	
Identifiers	Run 1	Run 2	Run 1	Run 2
420633			21, 38	21, 38
426036	19	19		
426039	22	22	47	47
431307	>20	69	10	10
447281			15, 18	15, 18
447269	68	68	19	19

[1]Repeat count units have ±1 margin

True negatives show normal range of 4qA 4qB repeat counts

Table 3. 4qA and 4qB Repeat Count in 14 Patients with Clinical Features of FSHD

	Allele 1		Allele 2		Additional maps/alleles
Sample Identifier	Haplotype	Repeat count (U)	Haplotype	Repeat count (U)	Haplotype, Repeat count
772	4qA	3	4qA	19	4qB, 29 U, mosaic
773	4qA	36	4qA	53	None
774	4qA	3	4qB	23	None
817	4qA	4	4qA	14	None
818	4qA	8	4qB	22	None
829	4qA	6	4qA	43	None
830	4qA	2	4qB	34	4qA, 19 U, mosaic
867	4qA	7	4qA	33	None
910	4qA	5	4qB	14	None
918	4qA	17	4qA	21	None
1021	4qA	5	4qA	37	None
1025	4qA	5	4qA	37	None
1026	4qA	5	4qA	34	None
1028	4qA	7	4qB	26	None

[1]Repeat count units have ±1 margin

14 cases with clinical diagnosis of FSHD; 12 show contraction in 4qA repeat count, and 2 exhibit mosaicism

Table 4. Confirmation of Optical Mapping Data using Southern Blot Data

	Allele1 (4qA)				Allele 2 (4qB)				Allle3 (4qA or 4qB)						
Sample Identifier	Bionano Genomics	Perkin Elmer Genomics	Lieden Univ. Med Center, Netherlands Southern Blot Analysis	Diff	Bionano Genomics	Perkin Elmer Genomics	Lieden Univ. Med Center, Netherlands. Southern Blot Analysis	Diff	Bionano Genomics		Perkin Elmer Genomics		Lieden Univ. Med Center, Netherlands. Southern Blot Analysis		Diff
772	3	3	3	0	19	19	22	3	29	4qB	29	4qB	30	4qB	1
773	36	37	36	-1	53	53	53	0							
774	3	3	3	0	23	23	23	0							
817	4	4	5	1	14	14	14	0							
818	8	8	8	0	22	22	22	0							
829	6	6	7	1	43	43	42	-1							
830	2	2	3	1	34	34	33	-1	19	4qA	19	4qA	18	4qA	-1
867	7	7	At analysis		33	33	At analysis								
910	5	5	6	1	14	14	15	1							
918	17	17	16	-1	21	21	20	-1							
1021	5	5	5	0	37	37	37	0							
1025	5	5	At analysis		37	37	At analysis								
1026	5	5	6	1	34	34	34	0							
1028	7	7	7	0	26	26	26	0							

Optical mapping data is confirmed using Southern Blot analysis

[1]Repeat count units have ±1 margin

DISCUSSION

We demonstrate 100% analytical accuracy and precision of this assay using the FHSD positive Coriell cell lines with an accuracy of ±1 repeat. Normal male and female control samples revealed the D4Z4 repeats to be within normal range (15-48 D4Z4 repeats of either the 4qA or 4qB haplotype, data not shown). In 14 clinically diagnosed FSHD cases, 12 cases were positive for the repeat contractions (ranging from 2-8 repeats) and were reproducible across intra-site, inter-site and inter-instrument and inter method comparison. Two of these positive cases also showed a mosaic pattern of the contracted allele, Chr 10q (not shown) and the normal 4qB alleles were also deemed highly reproducible across different runs performed at 2 sites. The results of our study demonstrate 100% reproducibility and precision of the samples utilized for the LDT evaluation of Bionano's Saphyr® genome imaging platform as a high resolution, high-throughput and cost-effective next generation cytogenomic tool.