Optical mapping enables high-throughput analysis of pathogenic repeats

Ernest T Lam, Jian Wang, Dong Zhang, Tom Wang, Andy WC Pang, Henry B Sadowski, Alex R Hastie, Mark Oldakowski
Bionano Genomics, San Diego, California, USA

Introduction

Repetitive elements are abundant and diverse in the human genome. They are also genetically unstable. Repeat expansions and contractions could impact protein structure and gene regulation. They underlie disorders such as facioscapulohumeral muscular dystrophy (FSHD) and amyotrophic lateral sclerosis (ALS). However, sequence analysis of these regions can be challenging. Clinical laboratories often rely on Southern blotting, which is labor intensive and typically requires radioactive staining.

Optical mapping with Bionano Genome Imaging provides an alternative high-throughput workflow that overcomes these limitations. The use of high molecular weight DNA molecules up to mega-basepairs in size allows large repeat structures to be spanned and elucidated. The repeats can thus be accurately sized. The single-molecule, amplification-free method also allows mosaic repeat alleles to be analyzed.

Methods

We obtained 12 FSHD-positive cell lines from the Coriell Cell Repositories with known pathogenic repeats. We collected optical mapping data on the cell lines and developed a pipeline to automatically analyze the FSHD-relevant D4Z4 repeat on chromosome 4q35 and the haplotype background (Figure 1). We also developed quality-control metrics that can point to sample or data quality issues. As proof of concept, we analyzed other repeats in the genome using a similar method.

Table 1: Analysis of Coriell samples

Sample	Repeat length (U)	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 differed by 2 units
GM17868	5	A	Yes
GM17898	6	A	Repeat differed by 2 units
GM17939	4	A	Yes
GM18027	4	A	Yes

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

Sensitivity (Table 1): We evaluated sensitivity with 12 Coriell FSHD samples. We detected D4Z4 repeat contractions on all samples. The repeat size and haplotype assignment were consistent with annotation.

Specificity (Figure 2): We analyzed 58 control samples with no FSHD phenotypes from San Diego Blood Bank (45 samples) and 1000 Genomes Project. None had clear pathogenic repeat contractions.

Consistency and reproducibility: We selected a subset of 6 Coriell samples and performed reproducibility analysis by running the samples in triplicates. In all cases, we produced equivalent results. We generated data with a second enzyme for the 12 Coriell samples and 20 control samples. In all cases, the repeat size and haplotype assignments were consistent between the two enzymes. We tested that the assay was robust to different shipping conditions for blood samples.

Figure 1: Example FSHD results

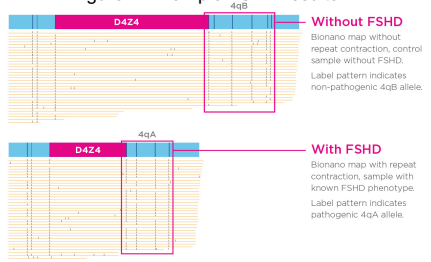
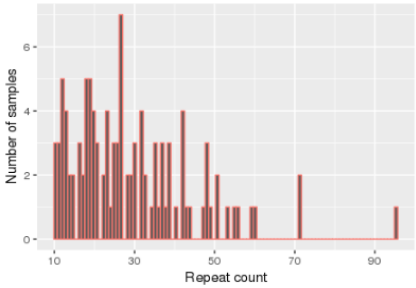
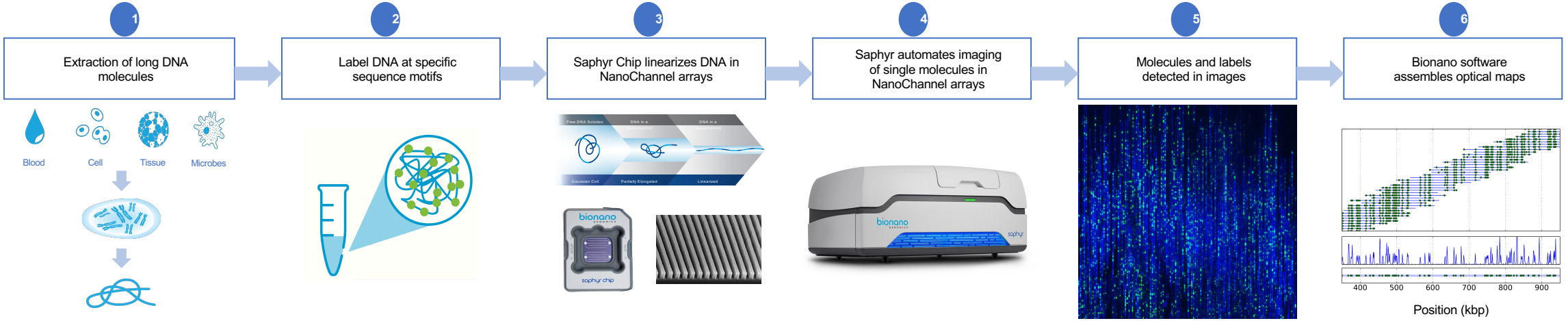


Figure 2: Repeat count distribution in controls



Overview



(1) Long DNA molecules 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 into *de novo* genome maps.