Comprehensive characterization of *FMR1* mutation status with combination short-read and nanopore long-read confirmation

Jiayi Sun[1], Muh-ching Yee[1], Stephanie Greer[1], Donna Hongo[1], Brynn Levy[1,2], Premal Shah[1], Matthew Rabinowitz[1,3], Akash Kumar[1], Kate Im[1]
[1]MyOme, Inc., Menlo Park, CA, USA, [2]Columbia University Irving Medical Center, New York, NY, USA, [3]Natera, Inc. San Carlos, CA, USA

BACKGROUND

Fragile X Syndrome (FXS) is the most commonly inherited X-linked neurodevelopmental disorder. FXS is caused by gene inactivation of fragile X mental retardation 1 (*FMR1*) leading to loss of the FMR protein (FMRP).

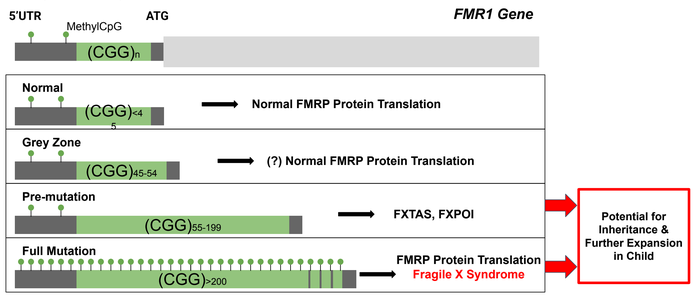

Figure 1. *Transcriptional silencing of FMR1 through short-tandem repeat (STR) CGG expansion*

- Expansion of CGG repeats within the *FMR1* 5'UTR leads to promoter methylation extending through the CGG expansion and transcriptional silencing. Mothers with premutation alleles have a high chance of passing down an expanded allele in the next generation.
- We established a pipeline which assesses *FMR1* status with short-read (SR) sequencing and reflexes potentially expanded samples to Oxford Nanopore (ONT) long-read (LR) sequencing to determine the number of repeats and methylation status.
- This work builds off of MyOme's neurodevelopmental delay copy number variation (CNV) confirmation pipeline[1].

METHODS

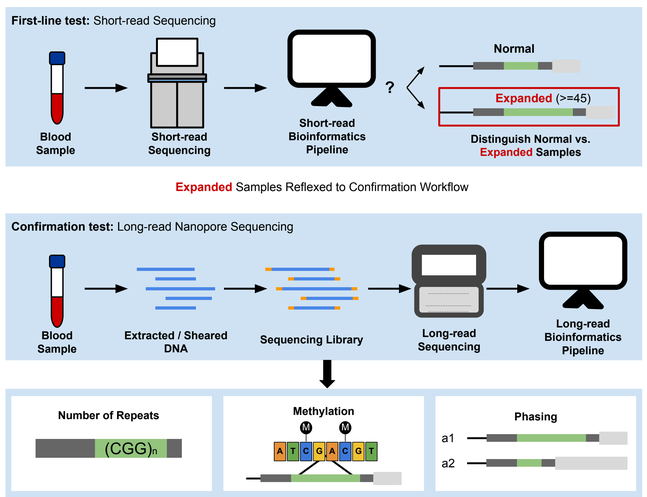

Figure 2. *Short + Long Read Approach Overview.*

Short-Read Development

- External Illumina 30x PCR-free HiSeqX sequencing of 7 control samples and 14 case samples (1 grey, 7 premutation, 6 full-mutation) from Coriell as part of the European Genome Archive (EGA) (study ID EGAS00001002462) were used for testing our short read approach.

Mutation Type	Training	Testing
Normal (<45)	5	2
Grey Zone (45-54)		1
Pre-mutation (55-199)	6	1
Full Mutation (>200)	4	2

Table 1. *Samples Used for SR Development*

- Using a combination of GangSTR, ExpansionHunter and TandemRepeatFinder algorithms, we developed a Nextflow pipeline to detect potentially expanded (grey zone, premutation, full mutation) versus normal samples.

Short-read + ONT targeted long-read accurately identifies *FMR1* repeat status and methylation status

Long-Read Development

- Long read, barcoded libraries were prepared by Diagenode tagmentation of genomic DNA to an average size of 15 kb using the newest LSK114/R10 kit chemistry and flow cell versions.
- LR adaptive sampling was performed on a set of 2 normal, 2 premutation, 2 full mutation samples, with a BED file targeting the *FMR1* gene and flanking regions. As a continuation of our previous ONT CNV work, 5 genomic control regions expected to have normal coverage were also assessed.
- We implemented a Nextflow pipeline which uses the results from Tandem Genotypes, but also outputs results from Tandem Repeat Finder and RepeatHMM. Integrative Genomics Viewer (IGV) was used to visualize the methylation status of the *FMR1* promoter.

RESULTS

Short-Read Results

- Correctly identified normal versus expanded status in **100%** of SR samples with ExpansionHunter (EH), where expanded status was identified as any sample > 45 copies. We found that <50 repeats, EH was accurate within ±1 of the Coriell repeat value.

Coriell ID	Sex	Coriell Repeats	GangSTR	EH	TRF
NA03200	male		[200]	158	1
NA05185	male		[153]	78	1
NA04025	male	645	[185]	112	2
NA06852	male	>200	[20]	74	1
NA06889	female	30/23	[30, 30]	23/30	0
NA06890	male	30	[30]	30	0
NA07539	male	23	[23]	23	0
NA07543	female	29/20	[20, 20]	20/29	0
NA20244	male	41	[39]	40	0
NA06891	male	118	[124]	110	0
NA06894	female	78/30	[27, 57]	30/79	0
NA06896	female	95-140/23	[23, 64]	23/81	1
NA06907	female	85/29	[27, 59]	29/90	1
NA06968	female	107/32	[20, 61]	33/99	1
NA06905	female	70/23	[17, 57]	23/66	0
NA07537	female	28-29/>200	[29, 59]	29/72	0
NA09145	male		[162]	125	4
NA20230	male	53	[80]	65	1
NA20234	female	31/46	[28, 28]	31/46	1
NA20238	female	29/30	[29, 29]	29/30	0
NA20231	male	76	[128]	84	2

Table 2. *SR results using GangSTR, ExpansionHunter (EH) and TandemRepeatFinder (TRF). TRF results indicate number of high quality reads with CGG repeats passing expanded threshold. Coriell reported normal samples are indicated in green.*

Long Read Results

Coriell ID	Coriell Repeats	Repeats	Methyl. %	Public[2]	CTRL Cov.	*FMR1* Cov.
CD00014	56*	57	1.3	57	19.7	11.4
NA06905	70/23	23/78	47.1	78/23	12.4	10.2
NA05131	>200* (606)	738	64.5	689	12.7	11.1
NA06890	30			31		
NA06893	23/30			23/30		
NA04026	>200* (654)	720	72.8	444	15.0	6.6

Table 3*. LR adaptive sampling depth and repeat results. Two normal samples from external reference indicated in green[2]. Methylation % was from 304 CpG sites between X:146992568-146994628. CTRL refers to the 5 control regions assessed for estimating average coverage.*

- In comparison to the R9 flow cells, we observed a 50% higher average QC PoreCount, 0.66 correlation between QC PoreCount before and after loading (versus 0.47 with R9) and a 50% higher average on-target coverage.
- NA06905 is a pre-mutation female which was reported as having a methylated allele, despite being pre-mutation$_2$. We were able to confirm pre-mutation and methylation status.

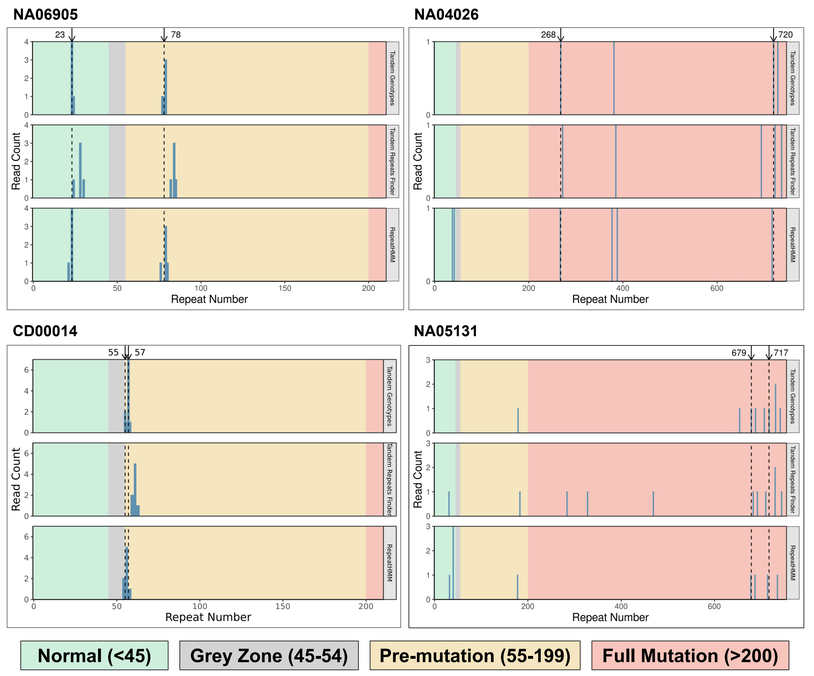

Figure 2. *LR FMR1 repeat results using Tandem Genotypes, Tandem Repeat Finder and RepeatHMM.*

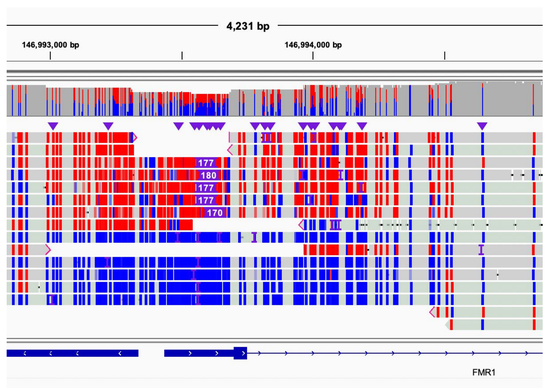

Figure 3. *IGV view of NA06905 pre-mutation methylation values. Methylated sites (red) are only observed in the reads with expanded CGG repeats (purple insertion).*

CONCLUSIONS AND FUTURE DIRECTIONS

- Developed short+long read workflow that accurately assessed *FMR1* repeat expansions and methylation status
- Combining short-read and long-read sequencing for the detection and confirmation of STRs and other difficult variants can streamline diagnosis and confirmation of genetic disease, potentially shortening the time to a medically actionable diagnosis
- Planning full clinical validation and incorporation into existing clinically-validated ONT CNV confirmation workflow

1. Greer, S. et al. Implementation of Nanopore sequencing as a pragmatic workflow for copy number variant confirmation in the clinic. J Transl Med (2023).
2. Stevanovski, I. et al. Comprehensive genetic diagnosis of tandem repeat expansion disorders with programmable targeted nanopore sequencing. Sci Adv (2022).

Correspondence: Monika Sun, jmsun@myome.com