Enrichment of unamplified DNA and long-read SMRT Sequencing in unlocking the underlying biological disease mechanisms of repeat expansion disorders

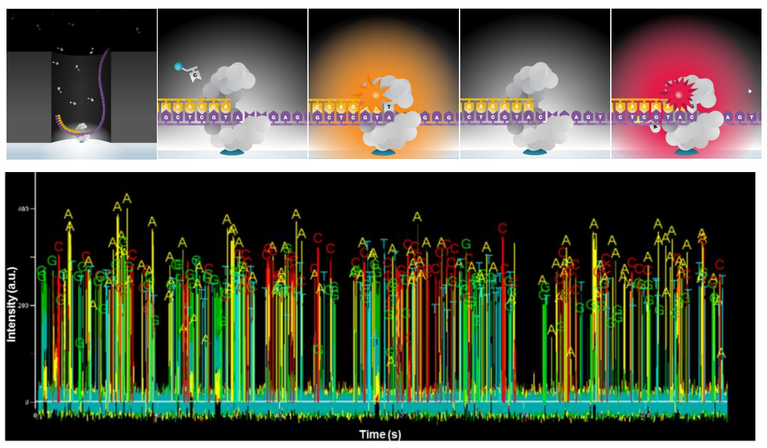
Jenny Ekholm, Yu-Chih Tsai, Tyson A. Clark
PacBio, Menlo Park, CA USA

Introduction

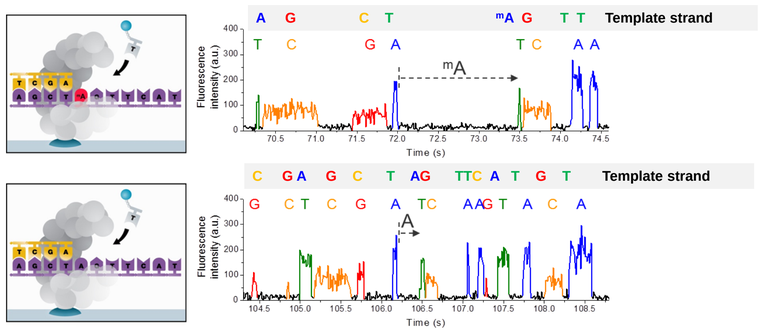
For many of the repeat expansion disorders, the disease gene has been discovered, however the underlying biological mechanisms have not yet been fully understood. This is mainly due to technological limitations that do not allow for the needed base-pair resolution of the long, repetitive genomic regions.

We have developed a novel, amplification-free enrichment technique that uses the CRISPR/Cas9 system to target large repeat expansions. This method, in conjunction with PacBio's long reads and uniform coverage, enables sequencing of these complex genomic regions. By using a PCR-free amplification method, we are able to access not only the repetitive elements and interruption sequences accurately, but also the epigenetic information.

SMRT Sequencing overview

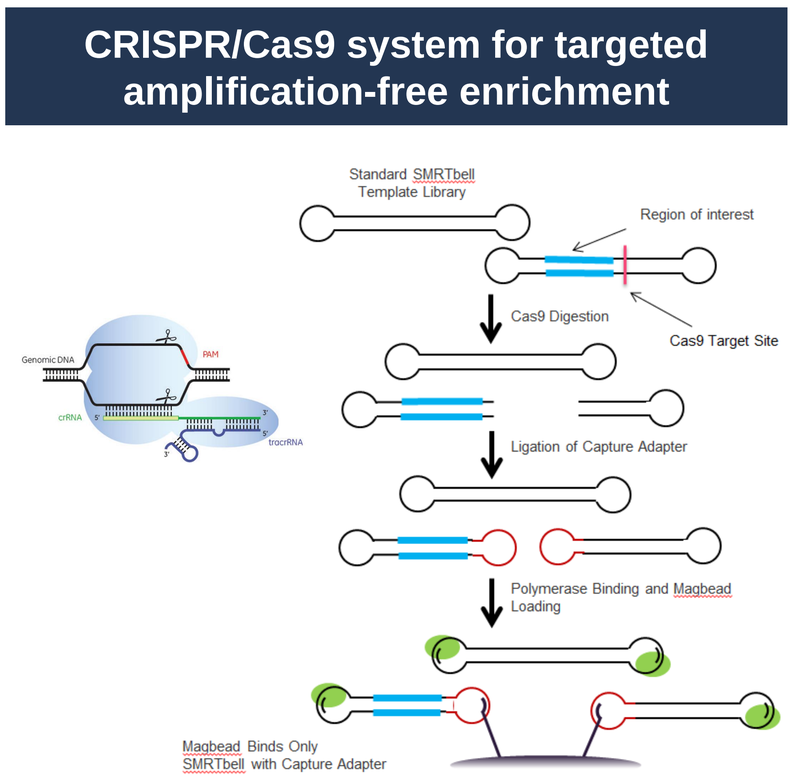


In addition to calling the bases, SMRT Sequencing uses the kinetic information from each nucleotide to distinguish between modified and native bases.



Example: N^6-methyladenine

CRISPR/Cas9 system for targeted amplification-free enrichment



1. Prepare a standard SMRTbell library
2. Use guide RNA to target the region of interest
3. Digest the SMRTbell template using Cas9
4. Ligate a capture adapter to the SMRTbell template that can be used to select for the SMRTbell templates that contain the region of interest
5. Sequence the SMRTbell templates

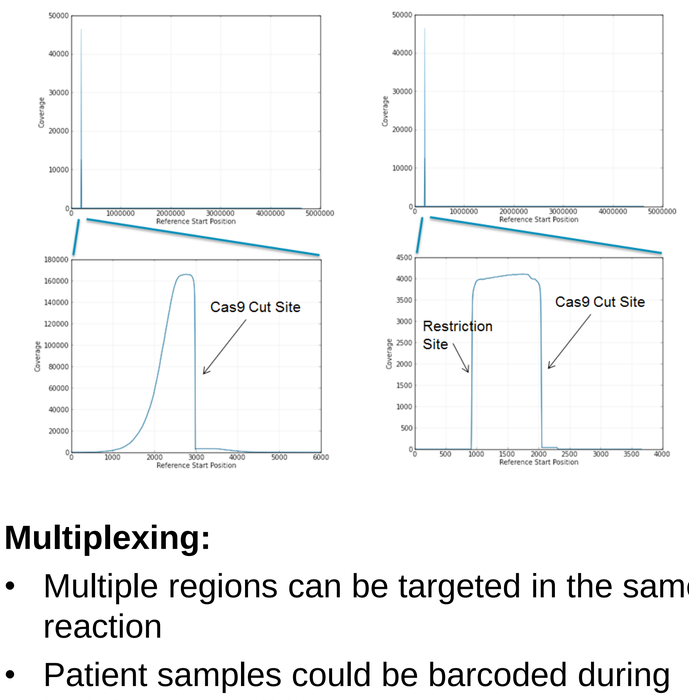
CRISPR/Cas9 system for targeted amplification-free enrichment

Complexity Reduction	DNA Input	Total Reads 1 SMRT Cell	HTT Chr4 CAG	ALS Chr9 GGGGCC	FMR1 ChrX CGG	SCA10 Chr22 ATTCT (various)	% of Reads "On Target"
None	1-5 µg	44,031	162	41	224	518	2.1%
2 RE	5-10 µg	51,806	574	174	607	1254	5.0%
4 RE	10-20 µg	45,676	1145	203	1016	1971	9.5%

Number of individual molecules sequenced

Improved on-target rate with complexity reduction:

- Restriction enzymes are used to degrade unwanted SMRTbell templates
- Additional starting DNA is required to maintain input into Cas9 digestion step

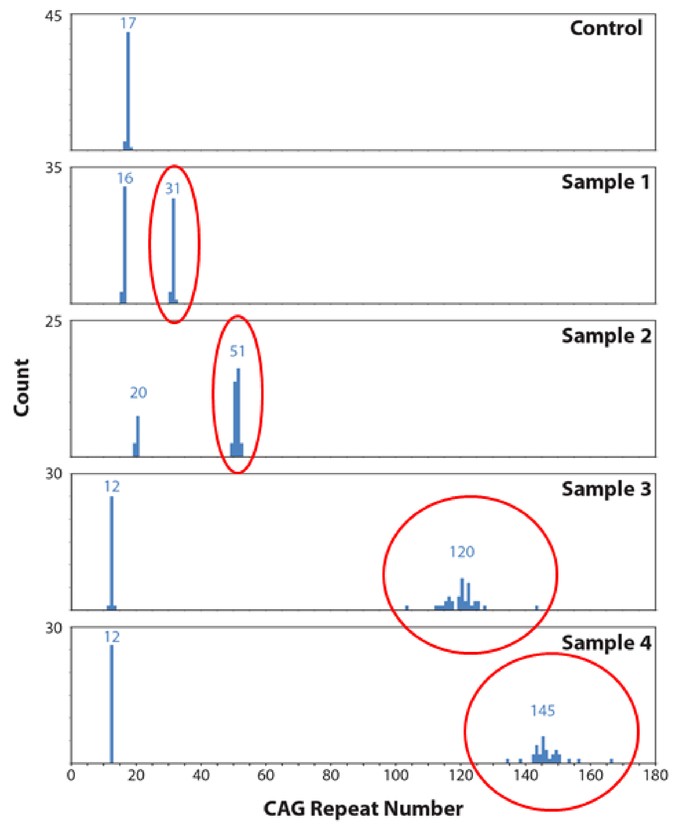


Multiplexing:

- Multiple regions can be targeted in the same reaction
- Patient samples could be barcoded during initial SMRTbell library preparation

Repeat counting

The *Huntingtin* gene CAG repeat counts in HD patients

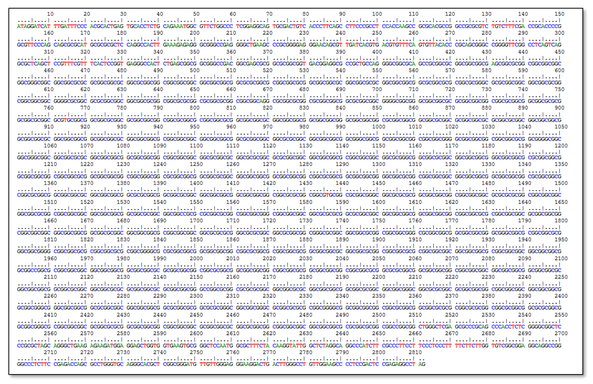


- CRISPR/Cas9 captures both the normal and the mutated allele successfully
- Widening repeat number distribution at the mutated allele is a known biological mechanism
- Obtained roughly equal number of sequenced molecules for normal and mutated alleles

Characterization of *FMR1* pre-mutation sample

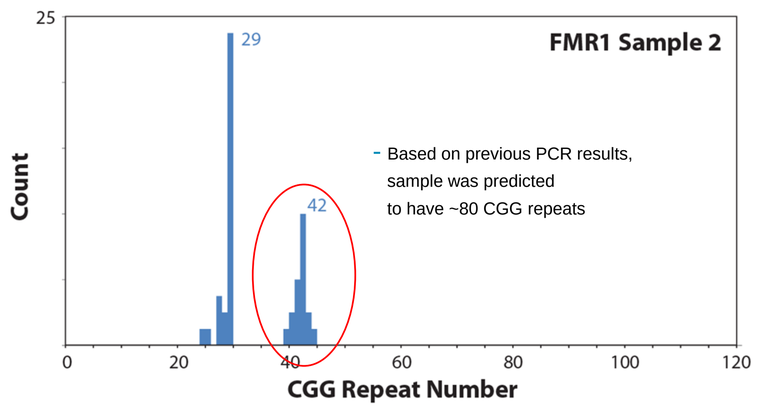
Repeat expansion disorders are challenging to interrogate due to the long repetitive regions. Using CRISRP/Cas9 we were able to access the repeat counts, interruption sequences, as well as epigenetic information, without introducing PCR bias. In addition, long-read sequencing allows for assembly-free, full-length isoform characterization.

Sequencing of *FMR1* gene, >700x CGG



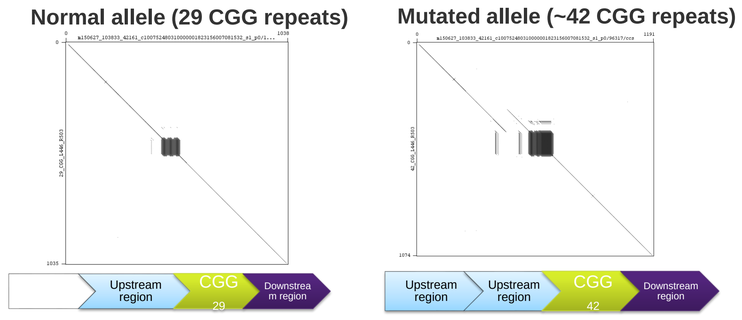
2 …CGG CGG CGG CGG *AGG* CGG CGG CGG CGG CGG CGG CGG CGG CGG CGG *AGG* CGG …
1 …CGG CGG CGG CGG *AGG* CGG CGG CGG CGG CGG CGG CGG CGG CGG CGG CGG …
0 …CGG CGG CGG CGG CGG CGG CGG CGG CGG CGG CGG CGG CGG CGG CGG CGG …

Repeat counting of *FMR1* gene

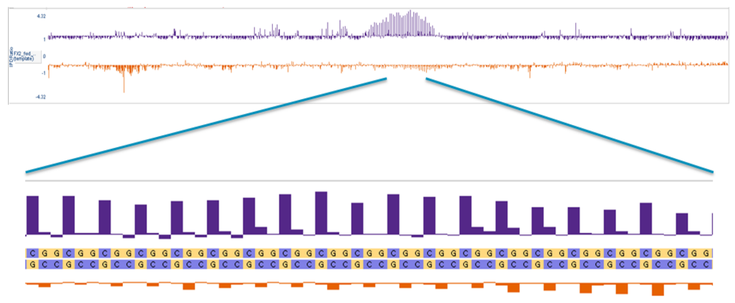


The mutant allele in this sample has a ~110 bp duplicated region preceding the CGG repeats.

Normal allele (29 CGG repeats) | Mutated allele (~42 CGG repeats)



Direct Methylation Detection of *FMR1* gene

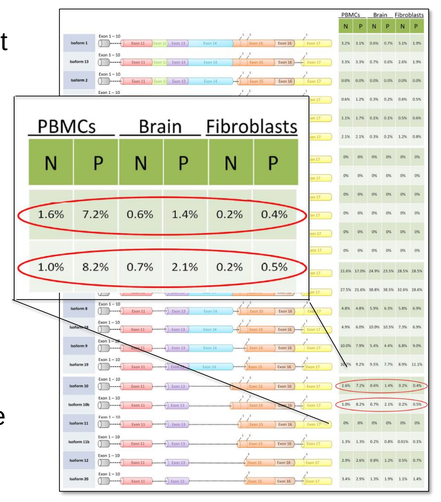


CGG repeat region appears to be heavily methylated (5mC)

Full Landscape of *FMR1* mRNA Isoforms

Pretto et al. [1] provided the first comprehensive map of FMR1 gene isoforms

- Identified novel alternative splicing event
- RNA toxicity may arise from a relative increase of all FMR1 mRNA isoforms. However 2 isoforms were identified with differential expression in premutation carriers suggesting a functional relevance of these in the pathology of FMR1-associated disorders



References and Acknowledgements

[1] Pretto DI et al . (2015) Differential increases of specific FMR1 mRNA isoforms in premutation carriers.. J Med Genet.52(1):42-52.

Samples provided by Vanessa Wheeler (Harvard Medical School), Paul Hagerman (UC Davis)

For Research Use Only. Not for use in diagnostics procedures. © Copyright 2016 by Pacific Biosciences of California, Inc. All rights reserved. Pacific Biosciences, the Pacific Biosciences logo, PacBio, SMRT, SMRTbell, Iso-Seq, and Sequel are trademarks of Pacific Biosciences. BluePippin and SageELF are trademarks of Sage Science. NGS-go and NGSengine are trademarks of GenDx. All other trademarks are the sole property of their respective owners.