Amplification-free targeted enrichment powered by CRISPR-Cas9 and long-read Single-Molecule Real-Time (SMRT) Sequencing can efficiently and accurately sequence challenging repeat expansion disorders

PACBIO®

Abstract #: 916083

Jenny Ekholm, Yu-Chih Tsai, Ian McLaughlin, Ting Hon, Janet Ziegle

PacBio, 1305 O'Brien Drive, Menlo Park, CA 94025

Introduction

Genomic regions with extreme base composition bias and repetitive sequences have long proven challenging for targeted enrichment methods, as they rely upon some form of amplification. Similarly, most DNA sequencing technologies struggle to faithfully sequence regions of low complexity. This has been especially trying for repeat expansion disorders such as Fragile-X disease, Huntington's disease and various Ataxias, where the repetitive elements range from several hundreds of bases to tens of kilobases.

We have developed a robust, amplification-free targeted enrichment technique, called No-Amp Targeted Sequencing, that employs the CRISPR-Cas9 system. In conjunction with SMRT Sequencing, which delivers long reads spanning the entire repeat expansion, high consensus accuracy, and uniform coverage, these previously inaccessible regions are now accessible. This method is completely amplification-free, therefore removing any PCR errors and biases from the experiment. Furthermore, this technique also preserves native DNA molecules, allowing for direct detection and characterization of epigenetic signatures. The No-Amp targeted sequencing method is a two-day protocol that is compatible with multiplexing of up to 15 targets and 48 samples in a single reaction, using as little as 0.5 µg of genomic DNA input per sample.

No-Amp Targeted Sequencing Method

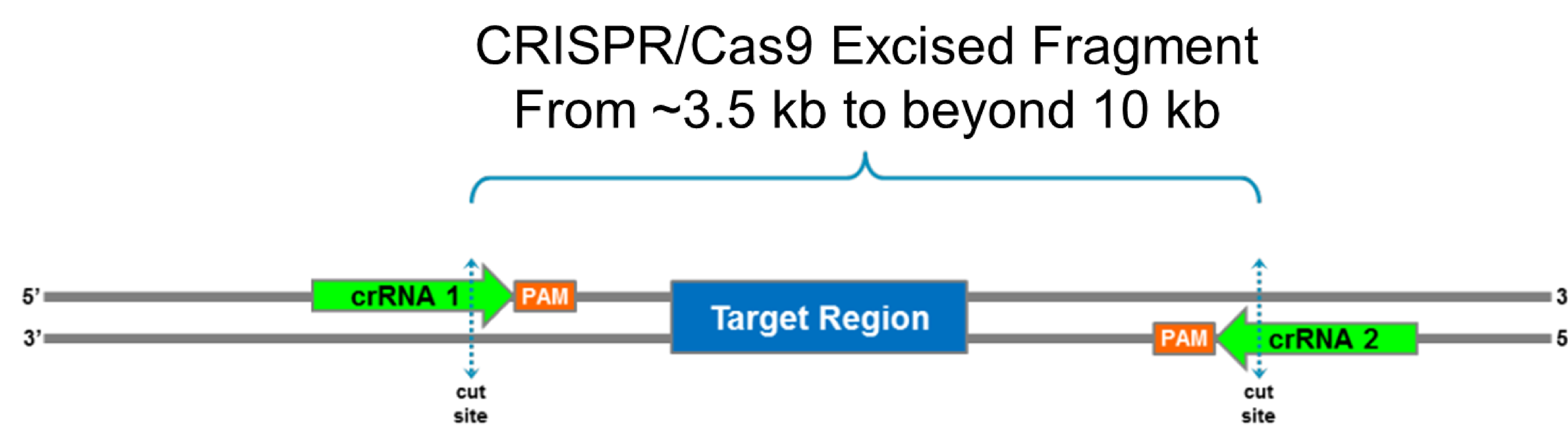


Figure 1. **Simple Experimental Design of Target Region of Interest**. Any region in the genome accessible to the CRISPR-Cas9 nuclease can be targeted for the No-Amp Targeted Sequencing method. The target locus is excised using a pair of guide RNAs – one on each end. With long reads there is flexibility in guide RNA selection. Up to 15 different targets can be multiplexed into one experiment.

No-Amp Targeted Sequencing Method

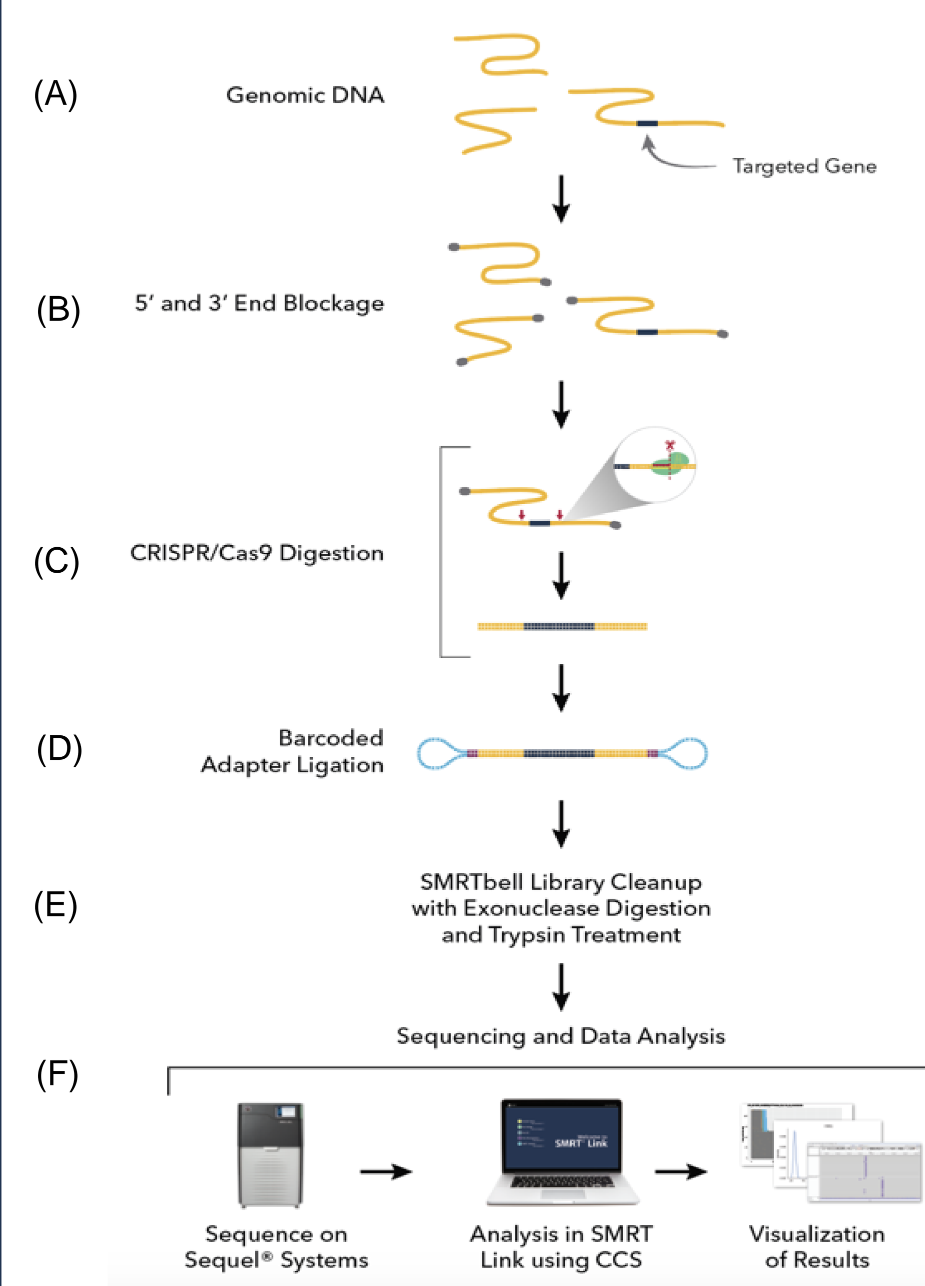


Figure 2. Workflow. In the first step of the No-Amp workflow the 5' and 3' ends the **(A)** gDNA sample will be **(B)** blocked. By **(C)** designing guide RNAs flanking each end of the region/s of interest the CRISPR-Cas9 system will cleave the double stranded gDNA and leave the ends available for sequencing **(D)** adapter ligation. The sequencing library is **(E)** cleaned up before **(F)** sequencing.

You can **multiplex both up to 15 target regions** as well **as up to 48 samples**. The required input gDNA is 5-24 µg/SMRT Cell, which can be subdivided across multiplexed samples. The expected on-target yield is >100 Q20 CCS reads/1 µg sample and the **on-target rate is 40-80+%** per SMRT Cell after demultiplexing which translates to **enrichment factors of 10,000s-100,000s fold.**

(F) The analysis pipeline consists of **CCS analysis** in SMRT Link and command-line based **visual reporting tools** for repeat count, mosaicism characterization, and identification of sequence interruptions.

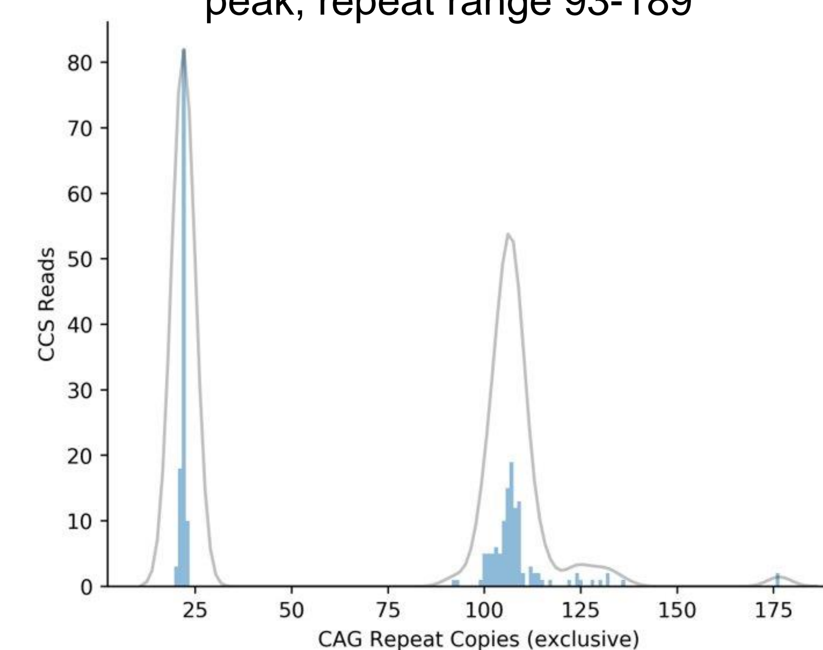
Results

Huntington's Disease (NA20253):

- Normal Allele: 22 CAG repeats
- Expanded Allele: 107 CAG repeats peak, repeat range 93-189

Fragile-X Disease (NA07537):

- Normal Allele: 27 CGG repeats
- Expanded Allele: 254-386 CGG repeats



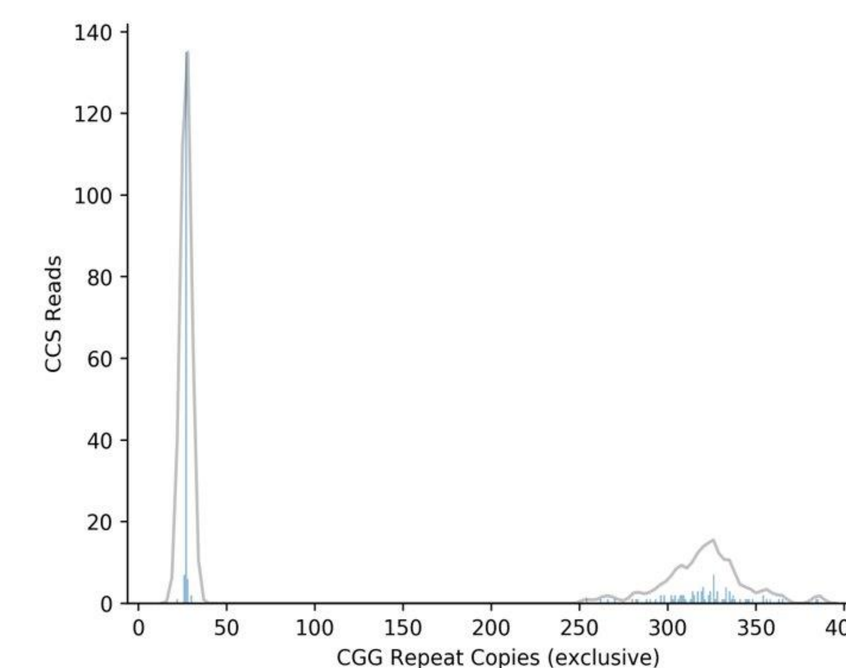


Figure 3. Repeat Count and Characterization of Mosaicism. The motif plots allow for easy visualization of repeat counts per allele and characterization of mosaicism which is known to be present in the mutated expanded disease allele.

Results

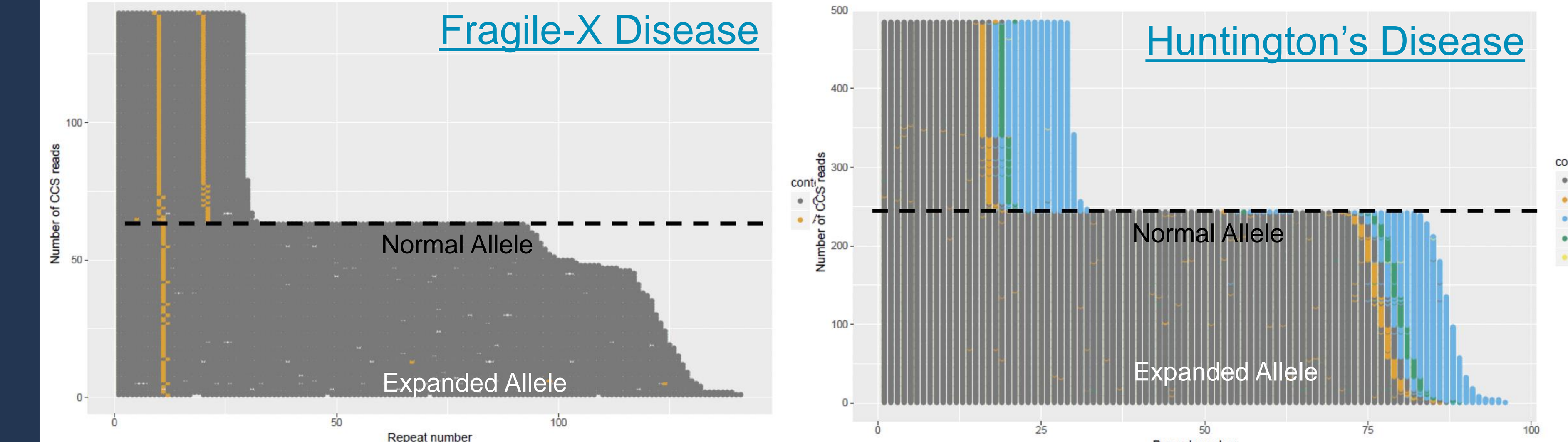


Figure 4. Visualize Repeat Structure and Identify Interruption Sequences. The waterfall plots allow for easy visualization of the repeat structure of the sample and identification of interruption sequences present in the sample.

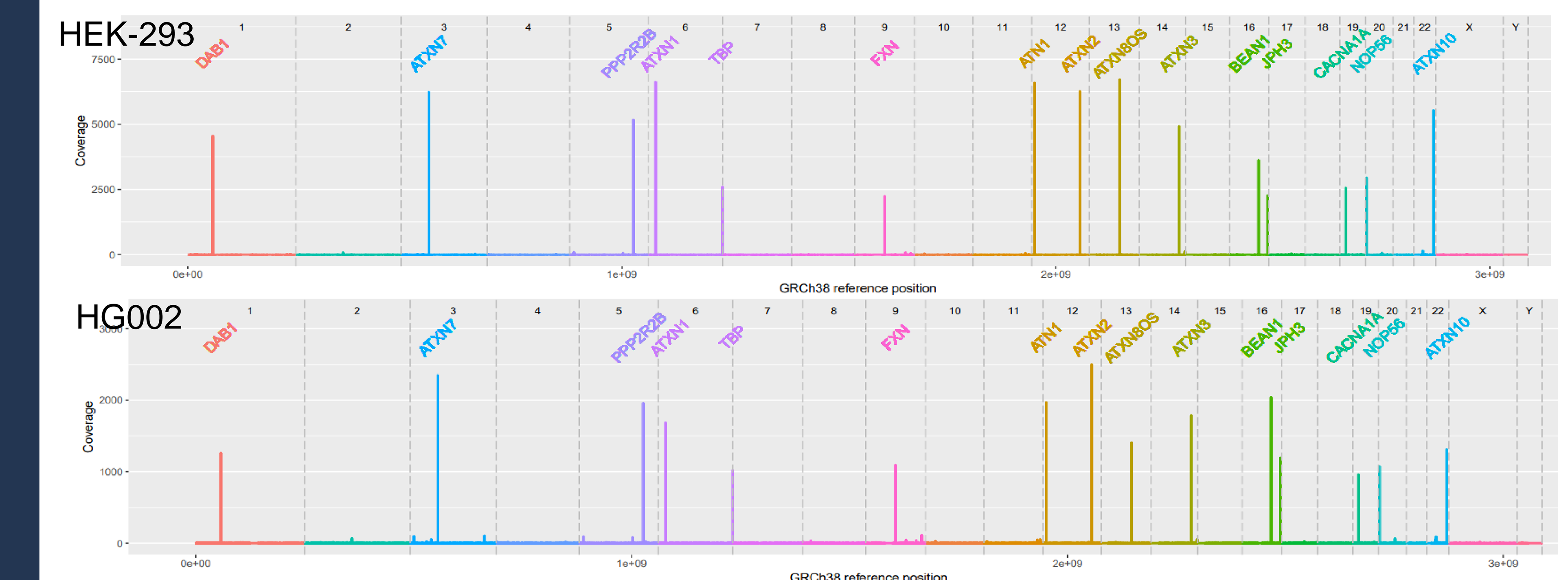


Figure 5. Ataxia No-Amp Panel. Fifteen ataxia loci were targeted in one experiment and sequenced/SMRT Cell 8M. This panel is compatible with multiplexing of 48 samples, bringing the cost per sample and target to under $6.00 USD.

Conclusions

No-Amp Targeted Sequencing: Amplification-free targeted enrichment of hard-to-amplify regions using CRISPR-Cas9

- No PCR-bias or errors

SMRT Sequencing provides*

- Long reads to span the region of interest
- High accuracy to accurately call every base
- Uniform coverage to access previously unsequenceable regions
- Single-molecule resolution that enables characterization of mosaicism

Repeat Expansion Disorders - holistically capture all needed information in one experiment

- Accurately repeat count for both alleles
- Identification of interruption sequences
- Characterization of mosaicism

*Study design, sample type, and level of multiplexing may affect the number of SMRT Cells 8M required. All prices are listed in USD and cost may vary by region. Pricing includes library and sequencing reagents run on your Sequel II System and does not include instrument amortization or other reagents

For Research Use Only. Not for use in diagnostic procedures. © Copyright 2020 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. Pacific Biosciences does not sell a kit for carrying out the overall No-Amp Targeted Sequencing method. Use of the No-Amp method may require rights to third-party owned intellectual property. FEMTO Pulse and Fragment Analyzer are trademarks of Agilent Technologies Inc. All other trademarks are the sole property of their respective owners.