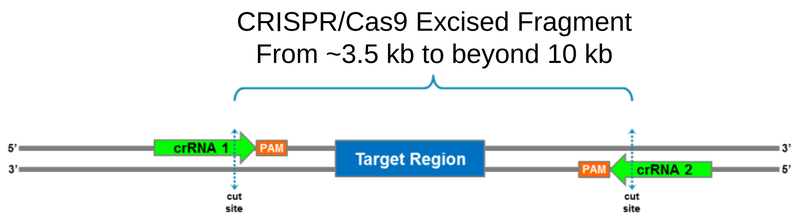
Amplification-free Protocol for Targeted Enrichment of Repeat Expansion Genomic Regions and SMRT Sequencing

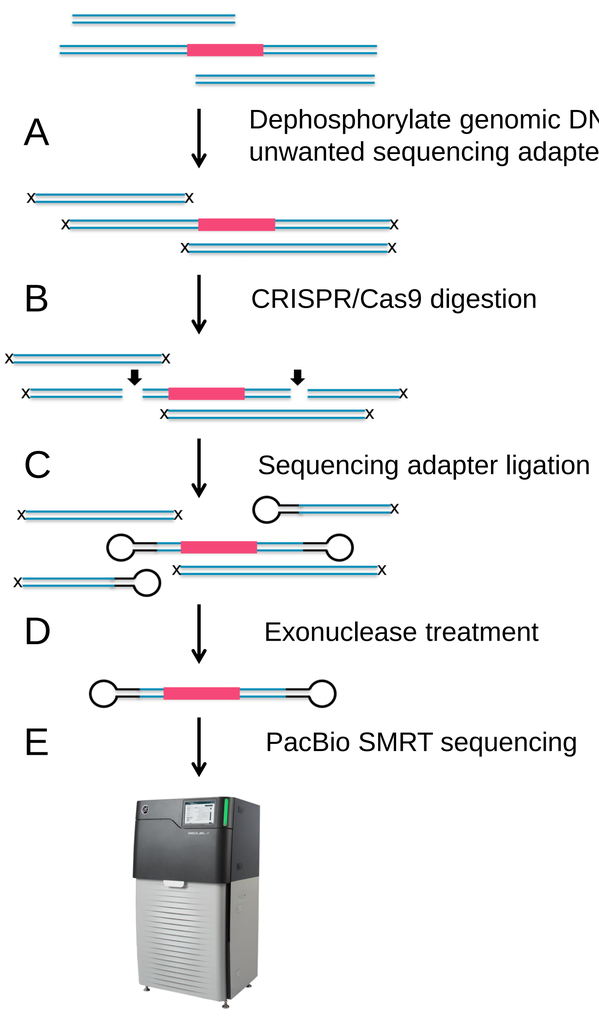
Yu-Chih Tsai, Ian McLaughlin, and Janet Ziegle
PacBio, 1305 O'Brien Drive, Menlo Park, CA 94025

Overview

- The No-Amp sequencing technique harnesses CRISPR/Cas9 system to target hard-to-amplify genomic regions for Single Molecule, Real-Time (SMRT) Sequencing
- Simple experiment design: a pair of guide RNA for each target region

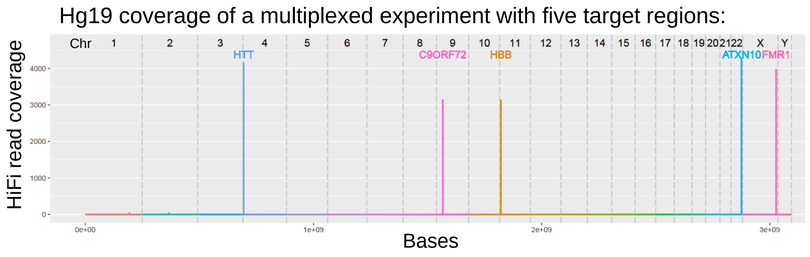


- Experiment workflow:



- HiFi reads are generated with SMRT Link software. Results can be viewed in IGV or analyzed with command-line tools
- Multiple targets and samples can be multiplexed in single sequencing assay

Hg19 coverage of a multiplexed experiment with five target regions:



Multiplexed 4 kb No-Amp Experiment

Target Gene	Associated Disease(s)	Chr	crRNA Coordinates*	Target Size	Repeat Sequence
HTT	Huntington's Disease	4	3074328-3074347 c3078167-3078148	3806 bp	CAG
C9orf72	Familial Frontotemporal Dementia (FTD) and Amyotrophic Lateral Sclerosis (ALS)	9	27571391-27571410 c27575063-27575044	3639 bp	CCCCGG
ATXN10	Spinocerebellar Ataxia Type 10	22	46189510-46189529 c46193116-46193097	3573 bp	Variable ATTCT
FMR1	Fragile X and Fragile X-associated Tremor/Ataxia Syndrome (FXTAS)	X	146993257-146993276 c146997012-146996993	3722 bp	CGG
HBB	Various (Beta thalassemia, Sickle cell disease, and others)	11	5245334-5245353 c5249615-5249596	4248 bp	No repeat expansion

* GRCh37 - Hg19 reference coordinates

Result

Results from one multiplexed Sequel II System experiment with seven Coriell DNA samples and HEK293 DNA as control. Input DNA was 2 µg genomic DNA per cell line:

A. Coriell samples with *HTT* repeat expansion

Sample ID	# On-target HiFi reads		Results comparisons	
	Total	Per normal and expanded allele	Coriell Reported Repeat alleles sizes	PacBio repeat allele sizes
HEK293	928	928, NA	N/A	17/17
NA13505 (M)	619	323, 296	22/50	22/50
NA13509	585	316, 269	15/70	15/72
NA20253 (M)	430	233, 197	22/108	22/108
NA14044 (M)	132	115, 17	19/205	19/141-2237

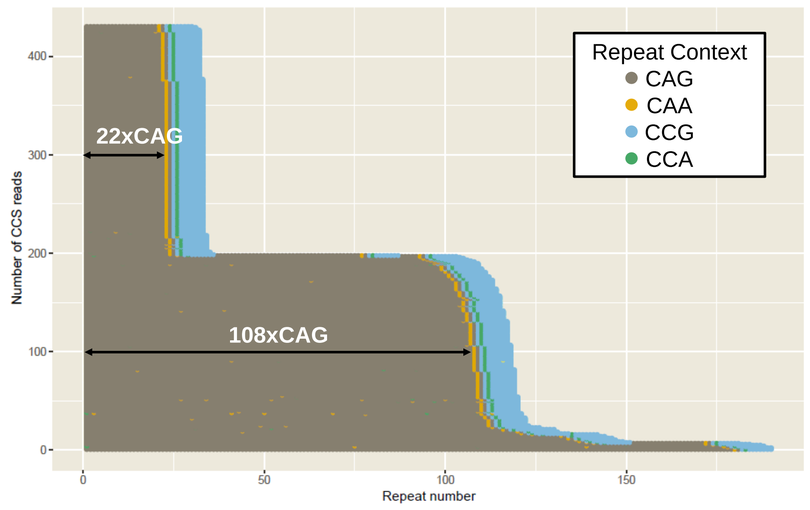
B. Coriell samples with *FMR1* repeat expansion

Sample ID	# On-target HiFi reads		Results comparisons	
	Total	Per normal and expanded allele	Coriell Reported Repeat alleles sizes	PacBio repeat allele sizes
HEK293	1475	1475, N/A	N/A	30
NA13664	365	197, 168	31/53	30/52
NA06896	438	268, 170	23/95-150	23/92-278
NA07537	419	275, 144	28-29/300	29/261-394

Repeat Structure Variation

Repeat sequence structure variation between each DNA molecules can be analyzed by aligning repeat sequences in high accuracy HiFi reads in a graphical presentation after flanking sequences are trimmed. Repeat sequences in each DNA molecule are represented by a series of horizontal dots with different repeat context labeled with different color.

(A) *HTT* CAG repeat n in NA20253:



Two alleles are identified with varying numbers of CAG repeat. A small fraction of DNA molecules shows greater than 150xCAG repeats.

(B) *FMR1* CGG repeat region in NA07537:

Two alleles are identified with two AGG interruption in the normal allele and one AGG interruption in the expanded allele. Number of repeat units in the expanded allele ranges from 261 to 394 repeat units.

Capture Fragments Longer than 10 kb

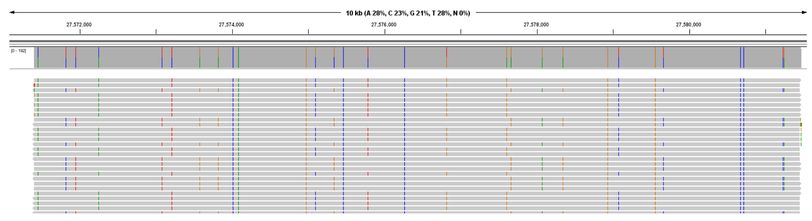
The No-Amp method captured hundreds of high-quality 10 kb HiFi reads at three genomic regions and 16 kb mitochondria from 2 µg of HEK293 DNA.

A. Guide RNA design and HiFi read yield

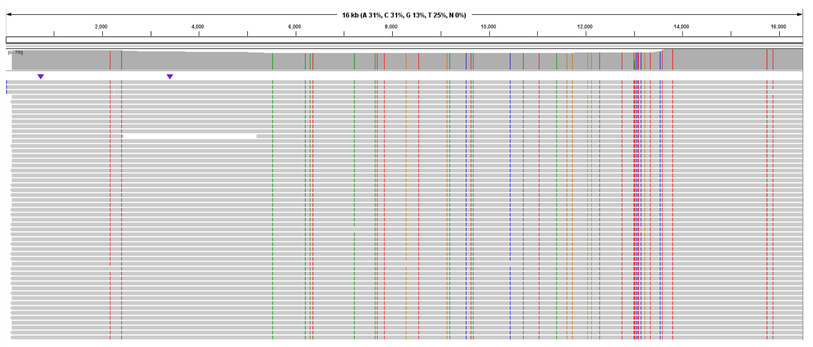
Target Gene	Chr	crRNA Coordinates*	Target Size	HiFi reads
HTT	4	3074328-3074347 c3084475-3084456	10114 bp	305
C9orf72	9	27571391-27571410 c27581486-27581467	10062 bp	192
FMR1	X	146993257-146993276 c147002870-147002851	9575 bp	465
Mitochondria	MT	3294-3313 c776-757	16434 bp	585

* GRCh37 - Hg19 reference coordinates

B. Many single nucleotide differences between two alleles in the C9orf72 target region were resolved and phased by 10 kb HiFi reads.



C. More than five hundred 16 kb mitochondria molecules were sequenced with HiFi reads.



Summary

We have developed a No-Amp method to target and sequence genetic disorder associated genomic regions which is difficult to accurately amplify with PCR.

The method is also applicable to any region of the genome susceptible to CRISPR/Cas9 digestion

The advantage of capturing and sequencing specific genomic regions without amplification is to:

- Avoid PCR amplification bias between normal and mutant alleles
- Target any genomic region regardless of sequence GC contents
- Preserve epigenetic modification

The PacBio® no-amplification (No-Amp) targeted sequencing method is now available on the Sequel and Sequel II Systems as a supported application.[1,2,3]

Application: No-Amp Targeted Sequencing

https://www.pacb.com/applications/targeted-sequencing/no-amp-targeted-sequencing/

References (PacBio Documentation):

1.PacBio Procedure & Checklist - No-Amp targeted sequencing utilizing the CRISPR Cas9 system.

2.PacBio Reference Guide – Designing CRISPR-Cas9 RNA Oligonucleotides for the No-Amp Targeted Sequencing Procedure.

3.PacBio Analysis Procedure – No-Amp Data Preparation and Repeat Analysis.