No-Amp Targeted SMRT Sequencing Using a CRISPR-Cas9 Enrichment Method

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

Introduction

While PCR amplification techniques can be used successfully with genomic regions of moderate to high complexity, low-complexity regions and regions containing repetitive elements can lead to poor or failed PCR-based enrichment. Many genetic diseases caused by repeat element expansions have proven to be challenging enrichment targets with PCR amplification, used either alone or in combination with a hybridization capture method.

We have developed an enrichment method for targeted SMRT Sequencing on the PacBio Sequel System using the CRISPR-Cas9 system that requires no PCR amplification[1]. This method, paired with SMRT Sequencing's long reads, high consensus accuracy, and uniform coverage, allows sequencing of genomic regions regardless of challenging sequence context that cannot be investigated with other technologies. The method is amenable to analyzing multiple samples and/or targets in a single reaction. In addition, this method also preserves epigenetic modifications allowing for the detection and characterization of DNA methylation which has been shown to be a key factor in the disease mechanism for some repeat expansion diseases.

Here we present results of our latest No-Amp Targeted Sequencing procedure applied to the characterization of CAG triplet repeat expansions in the *HTT* gene responsible for Huntington's Disease.

No-Amp Targeted SMRT Sequencing

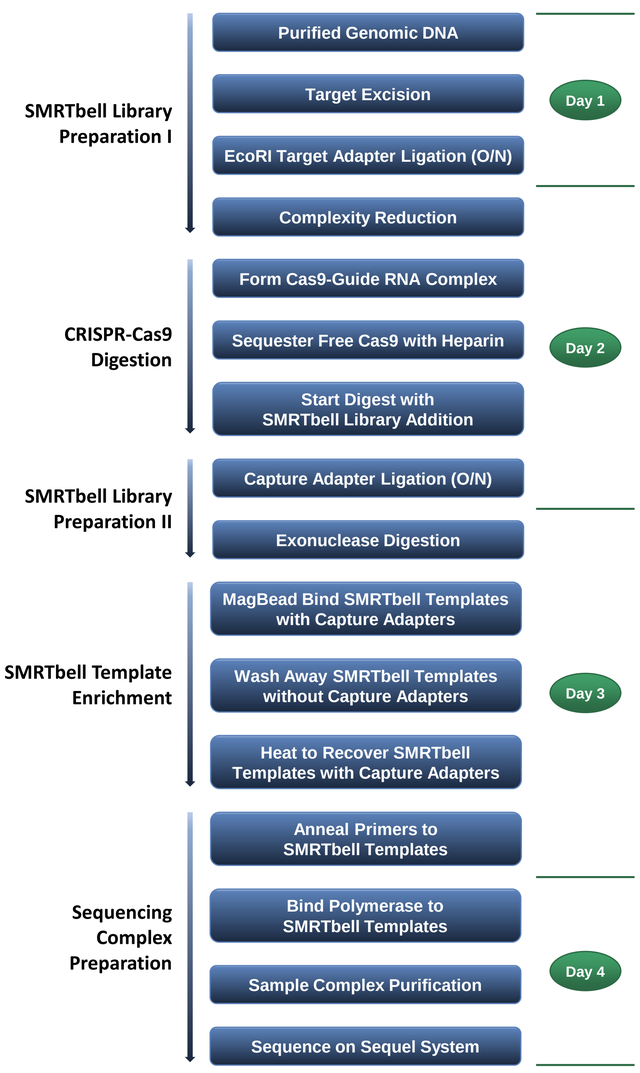

Figure 1. Workflow for No-Amp Targeted SMRT Sequencing. If multiplexing samples, barcoded EcoRI target adapters are used to form a uniquely barcoded library for each sample. Barcoded libraries are pooled prior to the Complexity Reduction step where background gDNA not associated with SMRTbell templates containing the targeted region is digested with a mixture of restriction enzymes. Target multiplexing is facilitated by the use of multiple guide RNAs in the CRISPR-Cas9 Digestion step.

CRISPR-Cas9 Enrichment Methodology

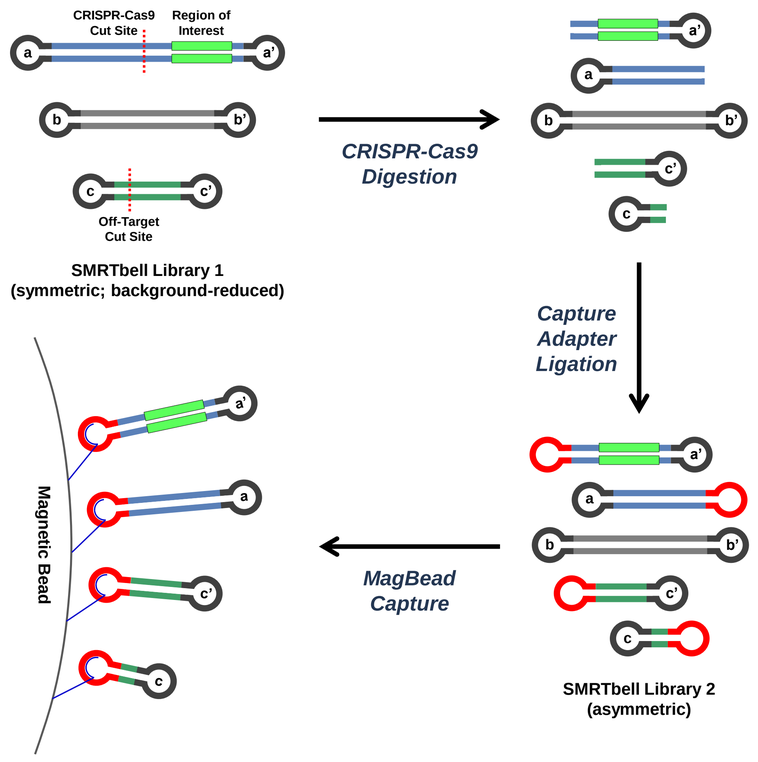

Figure 2. CRISPR-Cas9 enrichment methodology. The CRISPR-Cas9 system is used to target a region of interest in a background-reduced collection of SMRTbell templates followed by capture adapter ligation and magnetic bead enrichment of the resulting asymmetric SMRTbell library. Enriched SMRTbell templates are subsequently prepared for SMRT Sequencing on a Sequel System.

Results – Non-Expanded *HTT* Alleles

Two Cas9 Digest input amounts were tested with a single multiplex set consisting of 5 samples. The same control sample was used for each Target Excision reaction in the multiplex set to study protocol reproducibility. Sample pooling occurred after ligation of barcoded EcoRI target adapters.

Input into Target Excision was 10 µg each sample. Yield of SMRTbell Library 1 was 1,459 ng or 2.9% of total multiplexed gDNA input (50 µg).

Template input amount for Cas9 Digest Reaction Set 1 and Set 2 was 1,000 ng and 368 ng, respectively, which equates to 6.6 µg and 2.4 µg initial protocol input per sample.

Multiplex Reaction Set	Cas9 Digest Reaction Set	Cas9 Input	Sample	Barcode	Number Mapped Reads	*HTT* Target Count	%On-Target	Enrichment Factor	CAG Repeat Count
1	1	1,000 ng	M94-1	bc1001	46,987	3,203	6.8%	86,000	17
			M94-2	bc1002	40,746	2,844	7.0%	89,000	17
			M94-3	bc1004	43,501	2,906	6.7%	85,000	17
			M94-4	bc1005	44,434	3,192	7.2%	91,000	17
			M94-5	bc1006	31,967	2,079	6.5%	82,000	17
	2	368 ng	M94-1	bc1001	15,211	1,070	7.0%	87,000	17
			M94-2	bc1002	13,314	1,036	7.8%	96,000	17
			M94-3	bc1004	14,143	1,013	7.2%	90,000	17
			M94-4	bc1005	14,784	1,114	7.5%	94,000	17
			M94-5	bc1006	10,241	727	7.1%	88,000	17

Table 1. *HTT* results for a multiplexed sample set testing two different Cas9 digestion input amounts. *HTT* target counts are consistent within Cas9 Digest sets and scale proportionally to Cas9 input amount. The %On-Target values are also consistent between Cas9 Digest sets revealing that with only 368 ng input, approximately 1,000 target reads per sample can be achieved. This suggests that a greater multiplexing level is possible with the protocol as currently only 50 reads per allele are required to provide an accurate repeat call estimate.

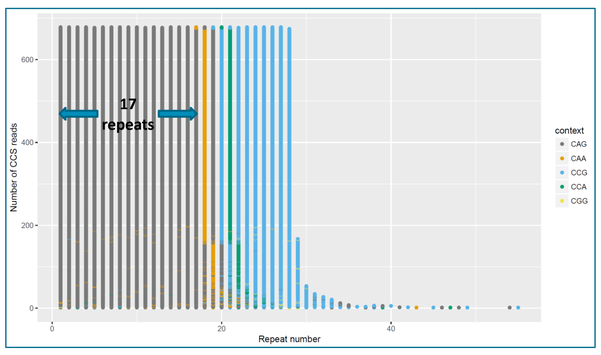

Figure 3. Representative repeat region size histogram for M94-5 of Cas9 Digest Reaction Set 2. There are 17 contiguous CAG repeats for the vast majority of CCS reads. Other repeat contexts are included in the plot to study additional sequence elements associated with the CAG repeat.

Results – Non-Expanded *HTT* Alleles

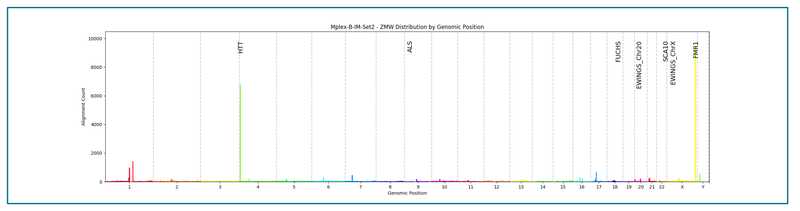

Figure 4. Representative genomic coverage by ZMW plot for all 5 samples in Cas9 Digest Reaction Set 2. While *HTT* is the focus of the current protocol, other targets are occasionally tested such as *FMR1* in the plot above. Other relatively small peaks such as in chromosomes 1, 7, 17, and Y may represent SMRTbell templates containing sequence homologous to the guide RNA sequence. These off-target molecules are a common feature of any CRISPR-Cas9 system and do not impact the ability to obtain on-target reads in our optimized protocol.

Results – Expanded *HTT* Alleles

Sample	Expected CAG Repeat Alleles
bc1001	15,24
bc1002	22,108
bc1004	22,50
bc1006	15,70
bc1007	18,60
bc1008	45,47

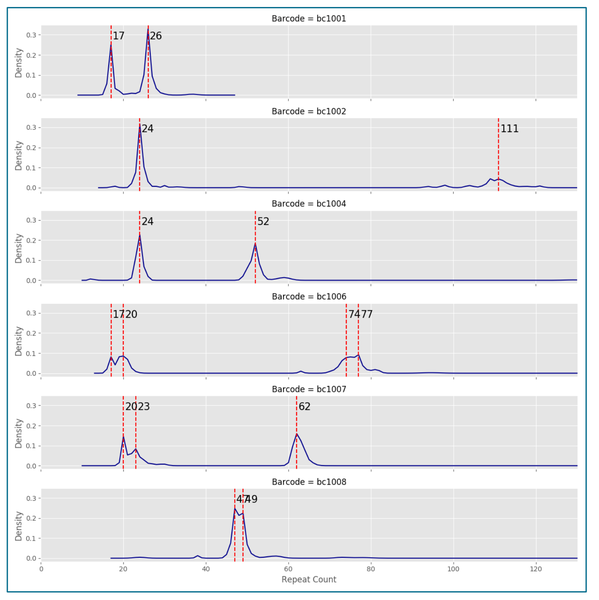

Figure 5. CAG repeat count estimate plots for expanded *HTT* target alleles. These results come from a separate experiment testing the same No-Amp protocol with a multiplex of samples containing expansion alleles. Protocol input was 10 µg per sample. Expanded alleles were detected and match the expected values within a few repeat units. The count variation in each size class observed with the No-Amp protocol could reflect biological variation present in different cell populations (mosaicism).

Conclusions

- No-Amp Targeted SMRT Sequencing is a powerful tool for capturing and assessing sequence variation in repeat expansion elements.
- Thousands of target reads can be obtained without the need for amplification enabling an accurate assessment of the natural variation present in a population of repeat elements which is often impossible to do with a PCR-based approach.
- With less than 3 µg of input DNA, greater than 1,000 *HTT* target molecules can be sequenced. We estimate that an accurate repeat count can be achieved with as little as 50 reads per allele. These results indicate that higher sample multiplexing than demonstrated here is possible.

References

1. Tsai Y, et al. (2017). Amplification-free, CRISPR-Cas9 Targeted Enrichment and SMRT Sequencing of Repeat-Expansion Disease Causative Genomic Regions. *bioRxiv* 203919.

No-Amp Targeted Sequencing

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

Acknowledgements

The authors would like to thank everyone who helped generate data for the poster.