A simple segue from Sanger to high-throughput SMRT Sequencing with a M13 barcoding system

Andres Larrea[1], George Yuan[1], Kimberly Wagner[2], Bryan Butt[2], Brian Norvell[2], **Lori Aro[1]**, Jenny Gu[1], Roberto Lleras[1], Swati Ranade[1], Jonas Korlach[1], Brian Krueger[2], Cheryl Heiner[1]

[1]Pacific Biosciences, 1305 O'Brien Drive, Menlo Park, CA 94025

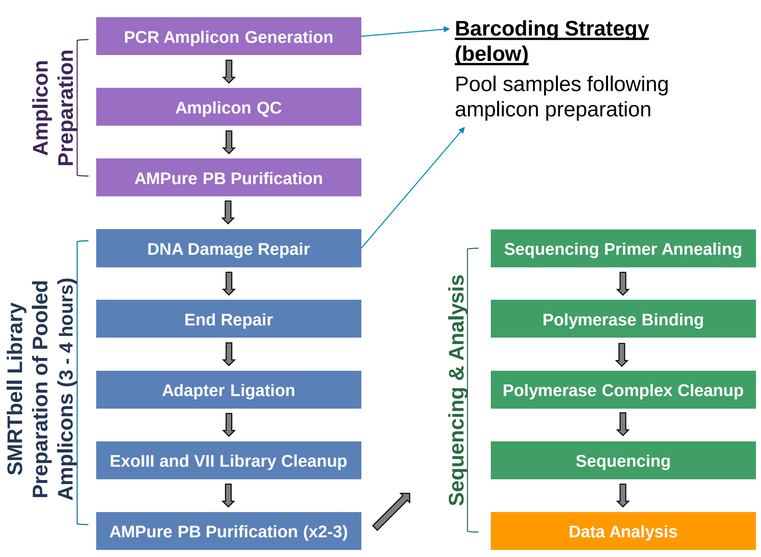
[2]LabCorp, 1440 York Ct, Burlington, NC 27215

Introduction

High-throughput NGS methods are increasingly utilized in the clinical genomics market. However, short-read sequencing data continues to remain challenged by mapping inaccuracies in low complexity regions or regions of high homology and may not provide adequate coverage within GC-rich regions of the genome. Thus, the use of Sanger sequencing remains popular in many clinical sequencing labs as the gold standard approach for orthogonal validation of variants and to interrogate missed regions poorly covered by second-generation sequencing. The use of Sanger sequencing can be less than ideal, as it can be costly for high volume assays and projects. Additionally, Sanger sequencing generates read lengths shorter than the region of interest, which limits its ability to accurately phase allelic variants. High-throughput SMRT Sequencing overcomes the challenges of both the first- and second-generation sequencing methods. PacBio's long read capability allows sequencing of full-length amplicons with multiple passes per molecule for accurate genotyping with phasing of allelic variants.

In the present work, we have developed a universal M13 barcoding system for SMRT Sequencing that utilizes the M13 sequences commonly used to tail PCR amplicons for Sanger sequencing. We demonstrate both workflow efficiency and cost-efficiency of SMRT Sequencing for a broad range of clinically relevant targets up to 20 kb. This provides a simple, cost effective and highly accurate segue from Sanger to SMRT Sequencing for existing or new clinical sequencing assays to interrogate a spectrum of polymorphisms across a wider span of amplicon size, regardless of the genomic sequence context. This project was completed using 2.1 sequencing chemistry and SMRT Link v5.1 analysis tools.

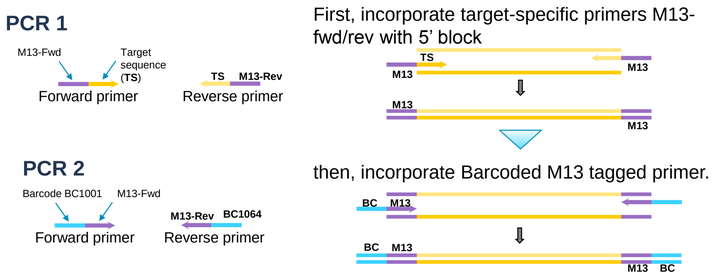
Multiplexed Amplicon Library Preparation



Non-specific amplicons can be removed by AMPure PB purification (with standard protocol), gel purification, Sage Blue Pippen or ELF Systems size selection. Purification can occur after amplicon generation or after library preparation.

Barcoding – M13 Tails

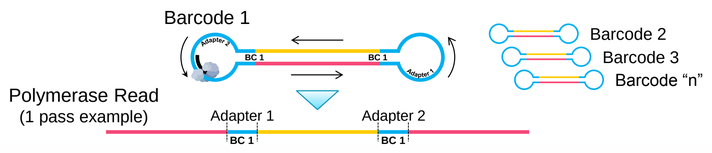
Barcoded Universal Primers



384 16-bp barcodes are available. Different barcodes may be used for forward and reverse primers, enabling >1000-fold multiplexing.

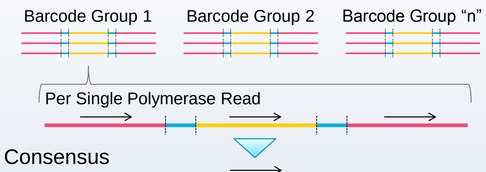
Sequencing Analysis Workflows

Circular Consensus Sequencing (CCS) Analysis

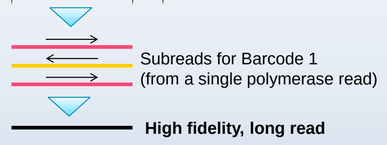


In SMRT Analysis:

1. Pre-Process Filtering (Analysis Parameters)
2. Demultiplex



3. Generate Circular Consensus
   - The CCS analysis method combines multiple passes from a single molecule resulting in high individual read accuracy (>99%)
   - High fidelity, long reads ready for further analysis (alignment, variant calling, etc.) with standard informatic tools



Results

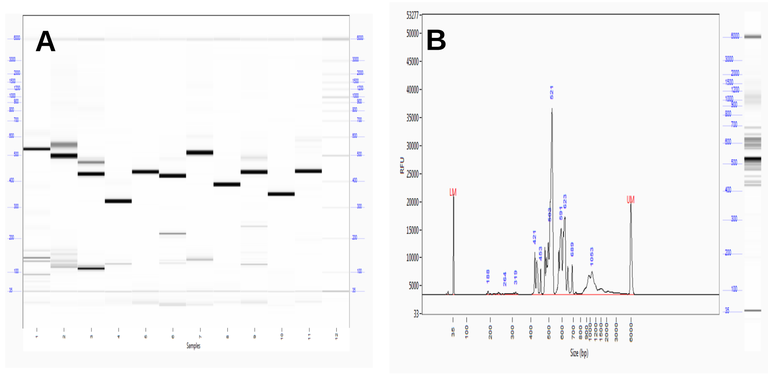


Figure 1 – PCR and Library Construction and QC. (A) Results from Fragment analyzer of 12 samples chosen at random from across all 384 samples. In all lanes, major bands can be seen of approximately the expected size. In almost all cases additional minor bands are also seen. **(B)** Results from Fragment Analyzer following equimolar pooling of all 384 samples and final SMRTbell library construction.

Movie Length (hr)	Total Bases (Gb)	Polymerase read length	# of productive ZMWs
10	13.32	24,523	543,160 (53%)

Table 1. Primary sequencing metrics from one Sequel SMRT Cell 1M loaded at 2.75 pM using sequencing chemistry 2.1 and run for 10 hours.

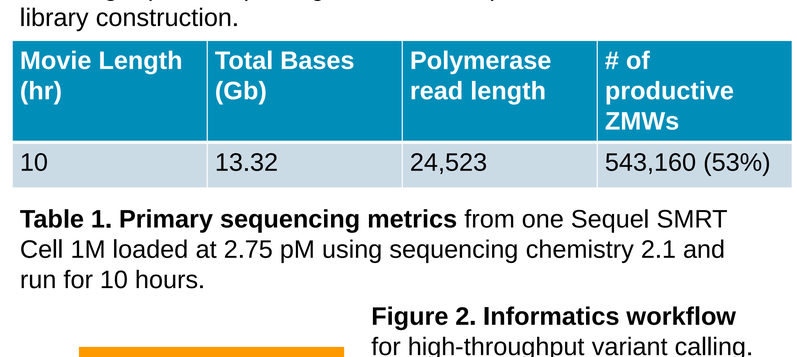


Figure 2. Informatics workflow for high-throughput variant calling. Raw data from PacBio Sequel System was converted into high fidelity reads using the CCS method in SMRT Link v5.1. High fidelity reads were then aligned using both BWA and Minimap2 (for comparison) and the resultant aligned BAM files were analyzed with HaplotypeCaller, a part of the GATK package. All tools are freely available and can be readily deployed and automated for large scale, high-throughput analysis.

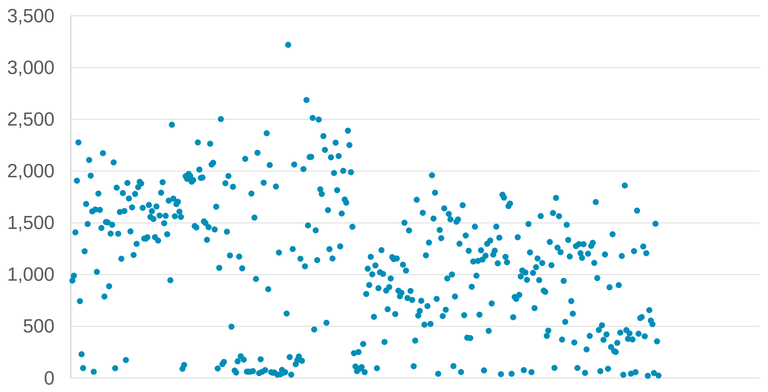


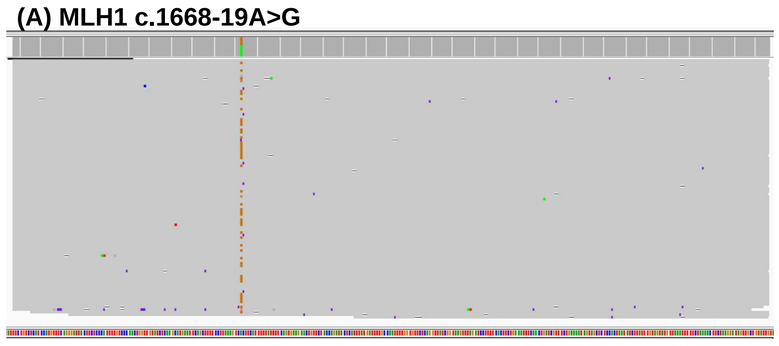
Figure 3. De-multiplexed high fidelity reads results for all 384 barcodes filtered for Q30 reads. On average each barcode had >1000 CCS reads with only 12 of 384 (3%) having less than 50.

Results

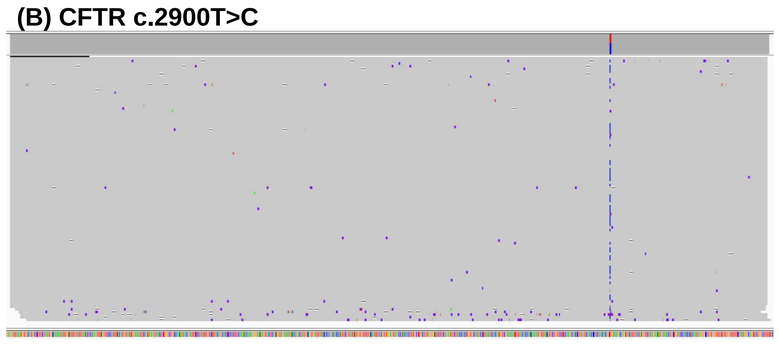
Sample	Gene	Variant	Exon	Type	Zygosity	PacBio Result
1	MLH1	c.1668-19A>G	15	Intronic	Het	Concordant
2	MLH1	c.1668-19A>G	15	Intronic	Het	Concordant
3	CFTR	c.2900T>C	17	Missesne	Het	Concordant
4	CFTR	c.3870A>G	23	Syn	Het	Concordant
5	CFTR	c.869+11C>T	7	Intronic	Het	Concordant
6	CFTR	c.869+11C>T	7	Intronic	Het	Concordant
7	CFTR	c.869+11C>T	7	Intronic	Het	Concordant
8	CFTR	c.869+11C>T	7	Intronic	Het	Concordant
9	CFTR	c.-8G>C	1	5' UTR	Het	Concordant
10	CFTR	c.-8G>C	1	5' UTR	Het	Concordant
11	MLH1	c.1668-19A>G	15	Intronic	Het	No amplicon
12	CFTR	c.1408G>A	11	Missense	Hom	No amplicon

Table 2. Comparison of Sanger results with PacBio results. Variants were 100% concordant except for the amplicons that failed to amplify.

(A) MLH1 c.1668-19A>G



(B) CFTR c.2900T>C



(C) CFTR c.869+11C>T

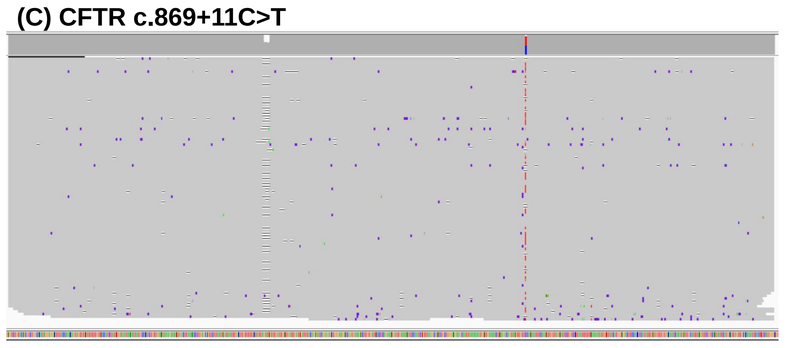


Figure 4. Representative amplicon alignments visualized in IGV. **(A)** Benign MLH1 c.1668-19A>G **(B)** VUS CFTR c.2900T>C **(C)** Benign CFTR c.869+11C>T

Conclusions

- Targeted amplicon sequencing with barcoded M13-tailed primers is compatible with SMRT Sequencing on the Sequel System.
- SMRT Sequencing targeted amplicon sequencing results are highly concordant with Sanger variant calls
- Hundreds of thousand highly accurate (>99%) single molecule reads generated on one SMRT Cell for Sanger sequencing designed amplicons using 2.1 sequencing chemistry and SMRT Link v5.1.
- Future areas of development:
  - Higher purity amplification primers to improve PCR yield and specificity
  - Equimolar pooling
  - With latest Chemistry 3.0 and SMRT Link v6.0, generate highly accurate single molecule reads (>99%) for amplicons up to 20 kb.

Resources / Acknowledgements

Targeted Sequencing
- Procedure & Checklist - Amplicon Template Preparation and Sequencing
- Application Brief – Targeted Sequencing for Amplicons Best Practices

Barcoding
- Product Note, Barcoding Solutions: Multiplexing Amplicons Up To 10 kb
- Document: SMRT Analysis Barcoding Overview

Circular Consensus Sequencing
- Tutorial: Circular Consensus Sequence analysis application