Detection of Low-frequency Somatic Variants Using Single Molecule, Real-Time Sequencing

Abstract #: 6033
Primo Baybayan[1], Richard Hall[1], Christine Lambert[1], Shreyasee Chakraborty[1], Yan Guo[1], Catherine Huang[2], Bharathi Anakella[2], Steve Kujawa[1]
[1]PacBio, 1305 O'Brien Drive, Menlo Park, CA 94025, [2]SeraCare Life Sciences, Milford, MA

Introduction

Detection of somatic mutations, especially in heterogeneous tumor samples where variants may be present at a low level, is challenging. Single Molecule, Real-Time (SMRT) Sequencing is ideal for minor variant detection because of its ability to sequence single molecules with very high accuracy (>QV40) using the circular consensus sequencing (CCS) approach.

Here, we characterize the Sequel System for the detection of low-frequency somatic variants using constructs containing mutations in coding regions in *AKT1* and *JAK2* representing substitution variants. Wild type and mutant amplicons, provided by SeraCare, were mixed and serially diluted from 10% down to 0.1% allelic frequency. Independent SMRTbell libraries were constructed for each dilution point, sequenced, and analyzed using SMRT Sequencing to identify the variants and determine the observed frequency. The random error profile and high-accuracy CCS reads make it possible to accurately detect low-frequency somatic variants. We demonstrate sensitivity of the PacBio Systems to detect mutations down to 0.5%.

SMRT Sequencing End-to-End Solution

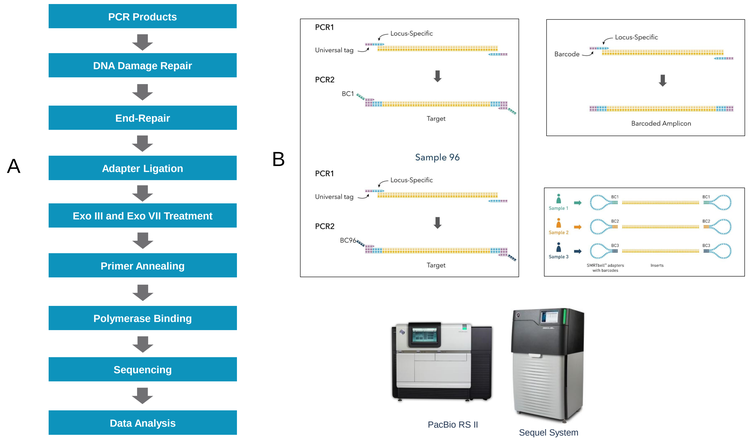

Figure 1. (A) Workflow for minor variant detection on PacBio Systems. **(B)** Barcoding options (Barcoded Universal Primer, Barcoded locus-specific primers, Barcoded Adapter) for multiplexing amplicons.

Highly Accurate Single Molecule Sequencing

The SMRTbell template format enables a consensus sequence to be obtained from multiple passes (both sense and antisense strands) on a single molecule. The number of passes represents how many rounds the same single-molecule is sequenced in a hairpin structure. A Circular Consensus Sequence (CCS) read is a highly accurate error-corrected consensus read derived from multiple subreads belonging to the same single-molecule.

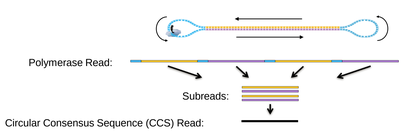

Figure 2. Circular Consensus Sequencing

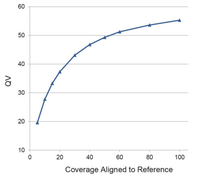

Figure 3. High consensus accuracy with v2.0 Chemistry

The number of CCS reads per SMRT Cell depends on amplicon length, movie collection time and CCS filtering in data analysis. The table below summarizes the number of CCS reads, filtered at varying accuracies from 90 to 99.999, obtained from sequencing three amplicons ranging in size from 450 bp to 4.4 kb, sequenced in the Sequel System using Sequel Sequencing Kit v2.0 with movie collection time of 10 hours.

	Number of Primary Reads	Minimum Predicted Accuracy	Number of CCS Reads	CCS Read Score (mean)
450 bp Amplicon	6,744,805	90	247,436	99.5387
		99	213,442	99.8577
		99.9	136,847	99.9811
		99.99	86,465	99.9988
		99.999	63,809	99.9999
2 kb Amplicon	2,787,070	90	172,976	99.5305
		99	145,959	99.8364
		99.9	90,315	99.9734
		99.99	33,539	99.9970
		99.999	12,523	99.9997
4.4 kb Amplicon	2,080,217	90	153,148	99.3597
		99	120,037	99.7118
		99.9	34,356	99.9421
		99.99	832	99.9938
		99.999	13	99.9993

Table 1. Number of CCS Reads for Different Size Amplicons filtered at varying accuracies

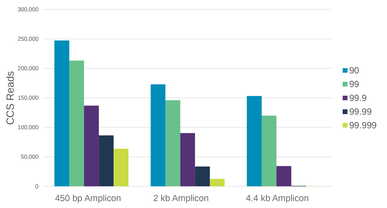

Figure 4. Histogram of the number of CCS Reads for 450 bp, 2 kb and 4.4 kb amplicons

Data Analysis Workflow

The output files of the Sequel System, BAM files, are submitted to SMRT Analysis which generates CCS reads and aligns them for variant calling. This integrated workflow will be available in SMRT Analysis in mid-2017.

Results

To study limit of detection for a single SNP, two genes with single mutations were sequenced in the Sequel System. Two single mutations were introduced into plasmids containing the AKT1 (49G>A) and JAK2 (1849G>T) coding regions (SeraCare). PCR amplified WT and Mutant constructs (~200 bp) were mixed and serial-dilutions made down to 0.1 % allelic frequency. Independent SMRTbell libraries were constructed for each dilution point and sequenced in the Sequel System.

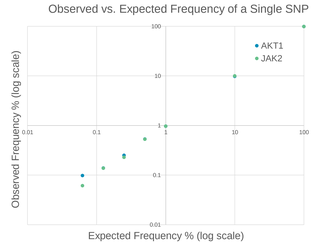

Figure 5. Serial dilution experiment results showing linearity and sensitivity of mutations detected down to 0.5%.

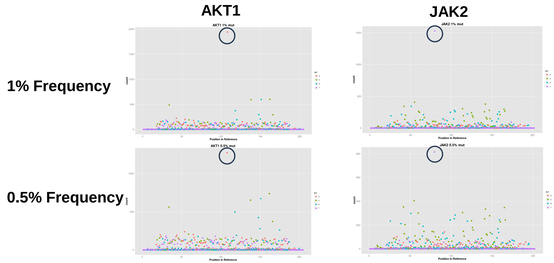

Figure 6. Limit of detection for a single SNP at 1% and 0.5% frequency. At 1% frequency, the SNP is above background. At 0.5%, the SNP is still discernable above background.

Conclusions

- The Sequel System is ideal for applications that require detection of minor variants including somatic mutations and tumor heterogeneity
- PacBio provides end-to-end solution from sample preparation to data analysis
- The number of CCS reads obtained from a Sequel SMRT Cell 1M is a function of amplicon length
- With CCS, single molecule reads >100,000 with >99% accuracy can be obtained from amplicons up to 4.5 kb
- It is possible to detect minor variants as low as 0.5% allelic frequency using high accurate CCS reads

Resources

Data Analysis
https://github.com/PacificBiosciences/pbccs
https://github.com/PacificBiosciences/ccscheck
https://github.com/PacificBiosciences/PacBioFileFormats/wiki/BAM-recipes#splitting

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

For Research Use Only. Not for use in diagnostics procedures. © Copyright 2017 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. FEMTO Pulse and Fragment Analyzer are trademarks of Advanced Analytical Technologies. All other trademarks are the sole property of their respective owners.