Improved liquid biopsy assay performance using sequencing by binding (SBB) on the PacBio Onso system

Abstract # 1875435
Dan Nasko[1], Phillip Pham[1], Stuti Joshi[1], Kristi Kim[1], Nairi Pezeshkian[1], Young Kim[1], **Alexandra Sockell[1]**, and Jonas Korlach[1]
[1]Pacific Biosciences, Menlo Park, CA

Introduction

Liquid biopsy is revolutionizing the field of early cancer detection research through non-invasive detection of tumor DNA in the blood. However, existing liquid biopsy assays are limited in their sensitivity for ctDNA detection at low variant allele frequencies (VAFs). Here we describe the application of the PacBio Onso short-read sequencing system to help enable detection of ctDNA at low VAFs using the SeraCare Complete ctDNA Mutation Mix reference sample.

Improved sequencing accuracy with SBB

The PacBio Onso system uses SBB chemistry, which decouples the interrogation and incorporation steps of the sequencing cycle. This allows for more optimized chemistry for each step and reduces molecular scarring caused by the incorporation of modified bases into the growing strand of DNA, as is done in traditional sequencing by synthesis (SBS). In contrast, in SBB the fluorescently-labelled nucleotides bind transiently for imaging, and then are washed away and replaced with unlabeled nucleotides, which get incorporated into the DNA. This results in quality scores (Q scores, phred scale) that are 10-100x greater than with existing technologies. As a result, 100% of bases in the SBB data are at Q40+ (probability of error = 0.001), with ~85% of bases at Q50+ (probability of error = 0.0001).

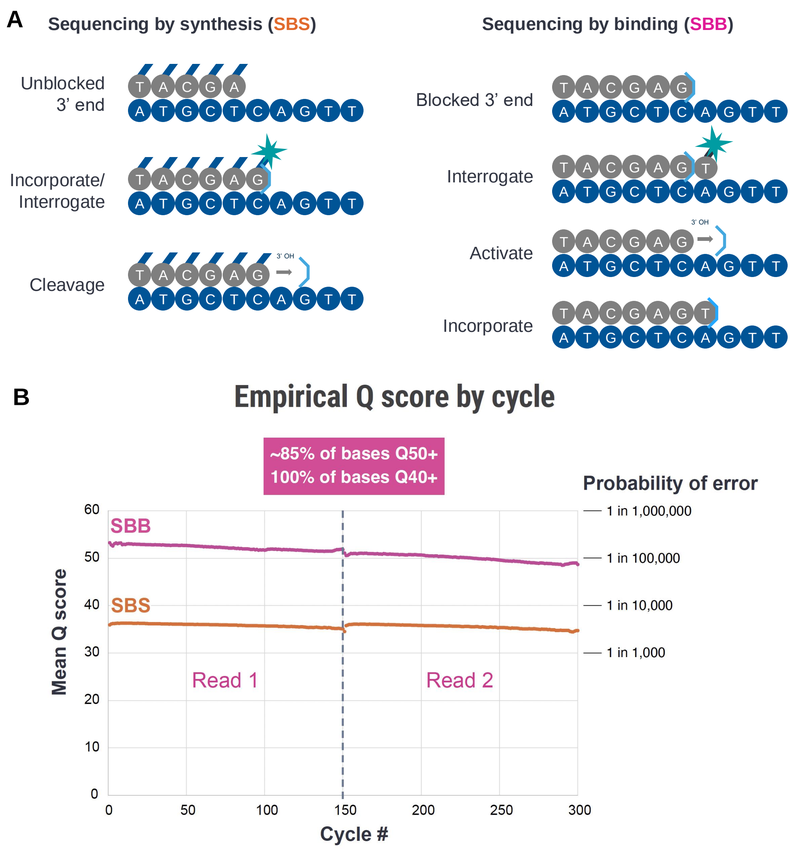

Figure 1. A) Schematic showing each step of the sequencing cycle for SBS (left), and SBB (right). **B)** Empirical Q scores for top 99% of bases with SBB (magenta) vs. SBS (orange), using PCR-free HG002 data for comparison.

Experimental overview

We performed a head-to-head comparison of variant detection performance using the Seraseq ctDNA Complete Mutation Mix reference sample. To minimize variability due to random sampling of molecules during library prep and capture, we first prepared the libraries and performed targeted capture. Post-capture libraries were split into 2 aliquots for the comparison, one of which was converted using the Onso library conversion kit to add Onso-compatible adapter sequences. The converted product is then sequenced on Onso, while the other was sequenced on a NovaSeq 6000 followed by subsampling, variant calling, and comparison of variant detection performance.

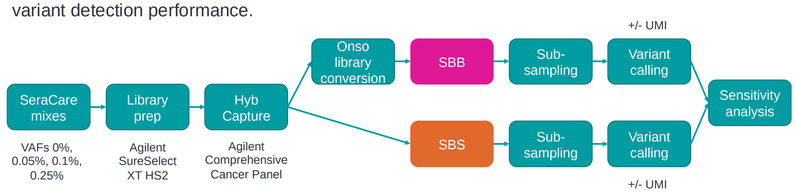

Figure 2. Overview of experimental workflow. Libraries were prepared for reference samples at multiple VAFs: 0% (WT), 0.05%, 0.1%, and 0.25% using the Agilent SureSelect XT HS2 kit. Targeted capture was performed using the Agilent ClearSeq Comprehensive Cancer Panel. Library conversion was performed on one of the post-capture aliquots using 5 cycles of PCR to add Onso-compatible adapter sequences, followed by sequencing on Onso. The other post-capture aliquot was sequenced on a NovaSeq 6000. 10-fold subsampling to 24,000X, 12,000X, and 6,000X coverage was performed, followed by variant calling with and without duplex UMI correction. Sensitivity was measured as the percentage of known variants in the reference sample detected by each technology.

Improved sequencing accuracy reduces noise at variant sites

Increased sequencing accuracy achieved by SBB reduced noise caused by sequencing errors at variant sites. We compared performance at a single variant site in the 0.1% VAF sample by looking at the number of base calls for each alternate allele. In the SBS data, all three alternate alleles received a similar number of base calls, obscuring the true variant. In SBB data, we only observed base calls for the correct allele, and the lack of errors revealed the variant is a T→C.

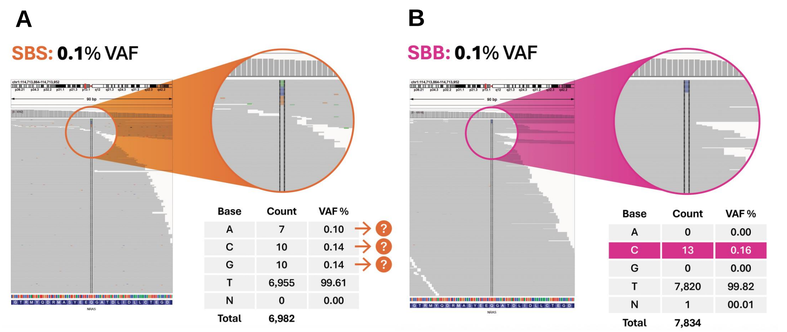

Figure 3. Representative IGV screenshots and allele counts at one of the known variant sites (NRAS G12D) for **A)** SBS and **B)** SBB for one subsampling replicate at 6,000X coverage. A similar allele frequency is observed for all three alternate alleles in the SBS data, whereas only the known allele is called in the SBB data.

Conclusions

Increased sequencing accuracy with PacBio SBB results in:

- Fewer false positives
- Improved VAF estimation
- Increased sensitivity for rare variants
- 4-fold reduction in required coverage

PacBio SBB reduces false positives and improves VAF estimate

We compared the observed VAF across the 12,000X subsampling replicates in the 0% (WT) vs 0.1% VAF sample. Significantly fewer false positives (FPs) were observed in the Onso data compared to the SBS data. This reduction in false positive variant calls for the WT sample for SBB results in a lower limit of detection and increased recall at high specificity. Additionally, the observed VAF of the true positive variant calls in SBB data is closer to the known VAF of the sample compared to SBS, where sequencing errors may result in inflation of the estimated VAF.

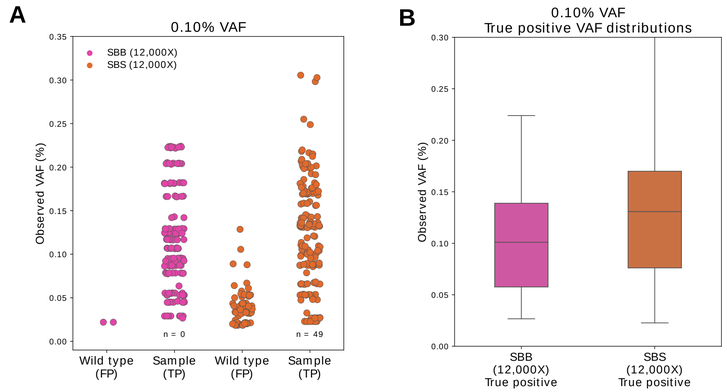

Figure 4. A) Observed VAF for variants called in the WT (FPs) vs. 0.1% VAF sample (TPs) across all 10 subsampling replicates at 12,000X coverage. Dashed boxes indicate variant calls in the sample that would have been missed with a specificity threshold of 100% (n=0 missed calls for SBB, n=49 for SBS). **B)** Same 0.1% VAF sample data plotted as a boxplot, black line = median observed VAF.

SBB increases specificity and reduces coverage requirements

We compared variant detection sensitivity for SBB compared to SBS. SBB achieves significantly higher sensitivity for variants at 0.05% and 0.1% VAF compared to SBS at matched coverages. Additionally, Onso achieved similar rare variant detection performance with a quarter of the sequencing depth, even without error correction.

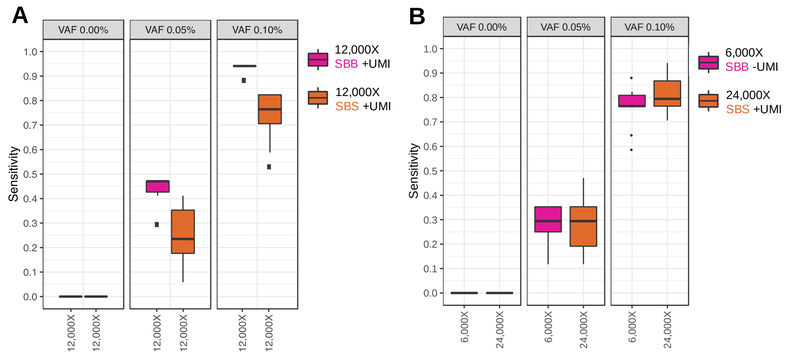

Figure 5. Comparison of sensitivity measured as the fraction of known variant sites detected by each technology for **A)** 12,000X SBB +UMI correction vs 12,000X SBS +UMI correction, and **B)** 6,000X SBB -UMI correction vs 24,000X SBS +UMI correction.

Acknowledgements

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