Simplified and robust library construction for high-throughput HiFi sequencing for human variant detection

Harsharan Dhillon, Christine Lambert, Cairbre Fanslow, Lei Zhu, Aaron Wenger, Primo Baybayan, Jonas Korlach

PacBio, 1305 O'Brien Drive, Menlo Park, CA 94025

Introduction

- Laboratories worldwide are adopting HiFi reads due to high accuracy (>Q30) and long read lengths making it ideal for the accurate detection of single nucleotide variants and large structural variants.
- The quantity of DNA available for a cohort can place constraints on a given study especially when the source of the sample is already deceased or access to additional DNA is not possible.
- The current published workflow requires 5 µg of high-quality gDNA and size-selection using traditional gel-based system is recommended to achieve highly accurate long reads with subread lengths between 15 kb -18 kb. Using this approach, post-size selection recovery yield is often low (10 %) making it impossible to generate sufficient library from <4 µg of gDNA.
- At PacBio, we have evaluated a new workflow using samples from Rady's Children Institute of Genomic Medicine with a wide range of gDNA input ranging from 0.9 - 6 µg.

Objectives

- **Reduce Input DNA Requirements**: Optimization of the HiFi library prep workflow using <4 µg of human genomic DNA.
- **Custom DNA Shearing Conditions**: Implementation of double shearing and different shearing speeds using the Megaruptor 3 System (Diagenode) that achieves a narrow and tight library size distribution appropriate for generating high-quality HiFi reads.
- **Custom nuclease treatment**: Adjustment of nuclease treatment based on sample yield post ligation to maximize recovery.
- **Custom Size Selection**: Utilization of different lower-end cutoffs on PippinHT System (Sage Science) and AMPure PB bead purification based on the sample mass.

HiFi SMRTbell library prep workflow

Figure 1. Library prep workflow. Fragment DNA to smaller target size distribution mode (15-18 kb) for consistent loading and lower on-plate loading concentration.

Consistent size distribution with custom shearing conditions

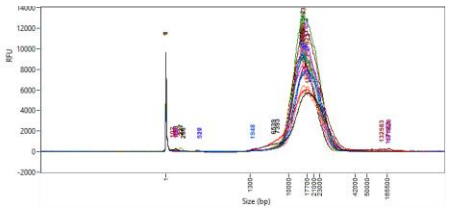

Figure 2. Shearing distribution of 23 genomic DNA samples sheared with Megaruptor 3 (Diagenode) using speed settings of 31 + 32 or 30 + 31. Input DNA of 0.9 – 6 µg in 60 µl volume.

Custom nuclease treatment based on post ligation yield

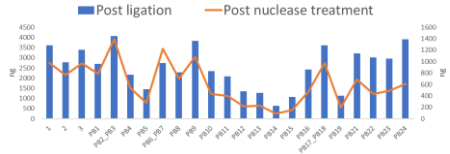

Figure 3. Library yields post-ligation and nuclease treatment. Post- ligation yields range 630 ng – 4056 ng which necessitated adjustment of the enzyme cleanup mix 2.0 to maximize recovery. Recommendation is 2 µl of SMRTbell Enzyme Clean Up Mix per 1 µg of DNA for low mass samples (<3 µg).

Custom size selection maximizes library recovery

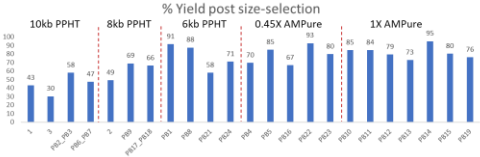

Figure 4. % recovery post size-selection. For SMRTbell library <500 ng, use AMPure PB bead purification. For >500 ng SMRTbell library, adjust lower cutoffs to 6 kb, 8 kb or 10kb on PippinHT using 6-10 kb High Pass Marker 75E.

gDNA with minimal presence of <5 kb fragments required for AMPure PB bead purification

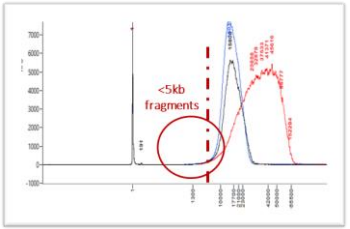

Figure 5. As a general rule, AMPure PB bead purification may replace traditional gel-based size-selection methods provided presence of fragments <5 kb is minimal. Use gel-based size selection methods for libraries with excess <10 kb inserts.

AMPure PB bead purification generates high-quality HiFi reads

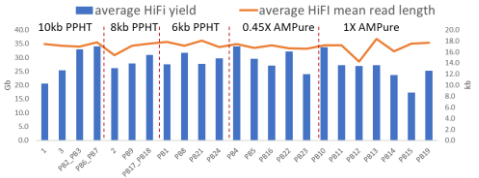

Figure 6. Average HiFi yield and HiFi mean read length for samples sequenced on Sequel IIe system. Average HiFi yields for samples with AMPure PB-bead based purification were 27.3 Gb versus 28.6 Gb for samples size-selected using PippinHT.

Conclusions

- **We have developed a simple workflow using AMPure PB bead size selection for generating high-quality HiFi reads for samples with <4 µg gDNA input. This workflow requires high-quality, high-molecular weight gDNA with minimal presence of <5 kb fragments.**
- **The resulting HiFi SMRTbell libraries yield an average of 27 Gb of HiFi reads.**
- **The mean HiFi read lengths were 16.9 kb with median read quality of Q33.**
- **This workflow can be automated from HiFi library construction to AMPure PB bead purification without the need of traditional gel-based systems providing an end-to-end solution for library construction**.

Acknowledgements

The authors would like to thank Rady's Children Institute of Genomic Medicine for providing the samples used to generate data for the poster.