Poster Reprint

ASHG 2025
Poster number 9025T
PR7003-306

A Streamlined library prep workflow with Comprehensive Genome Profiling (CGP) panel, and PacBio long-read sequencing, provide a holistic view of tumor biology

Bahram Arezi[1], Brandyn Clark[1], Adam Janssen[1], Jeff Fox[1], Nedda F. Saremi[2], Harsharan Dhillon[3], Kristi Stephenson[2], Kelle Hammock[2], Camille Conner[3], Deborah Moine[3] , Cindy Hon[2]

[1]Agilent Technologies, La Jolla, CA [2]Agilent Technologies, Santa Clara, CA [3]PacBio, Menlo Park, CA

Introduction

The effectiveness of next-generation sequencing in cancer research is influenced by the availability and quality of DNA and RNA, varying tumor fraction in samples as well as practical considerations such as workflow turnaround time, ease of use, and overall assay performance. Moreover, due to certain short-read sequencer limitations (e.g., mapping rate in repetitive regions, etc.), enabling long-read sequencing allows for better detection of structural variants, complex rearrangements, and variants in highly repetitive or polymorphic regions. However, challenges remain in terms of cost effectiveness and analysis of whole genome long-read sequencing. Applying target enrichment up front of long-read sequencing is a useful approach to circumvent these challenges and further enable the implementation of long-read sequencing. We have developed a flexible and automation-friendly workflow solution that accommodates DNA input as low as 200 ng and is compatible with mechanical shearing and fast hybridization (90 minutes). Coupled with the Agilent SureSelect Cancer CGP panel (comprising 679 genes globally curated from leading cancer databases and in partnership with key cancer researchers) and long-read HiFi sequencing from PacBio, we can detect multiple types of genetic alterations (SNV, CNV, Indels, and translocations) in tumor samples using a single assay. We demonstrate high enrichment efficiency (on-target rate of >80% for libraries containing inserts as large as 4-5 kb) using a novel fast hybridization chemistry that maintains complexity even when multiple libraries are pooled before capture. Furthermore, we demonstrate examples of challenging regions where enriched long-read sequencing generates greater coverage compared to short-read sequencing.

Experimental

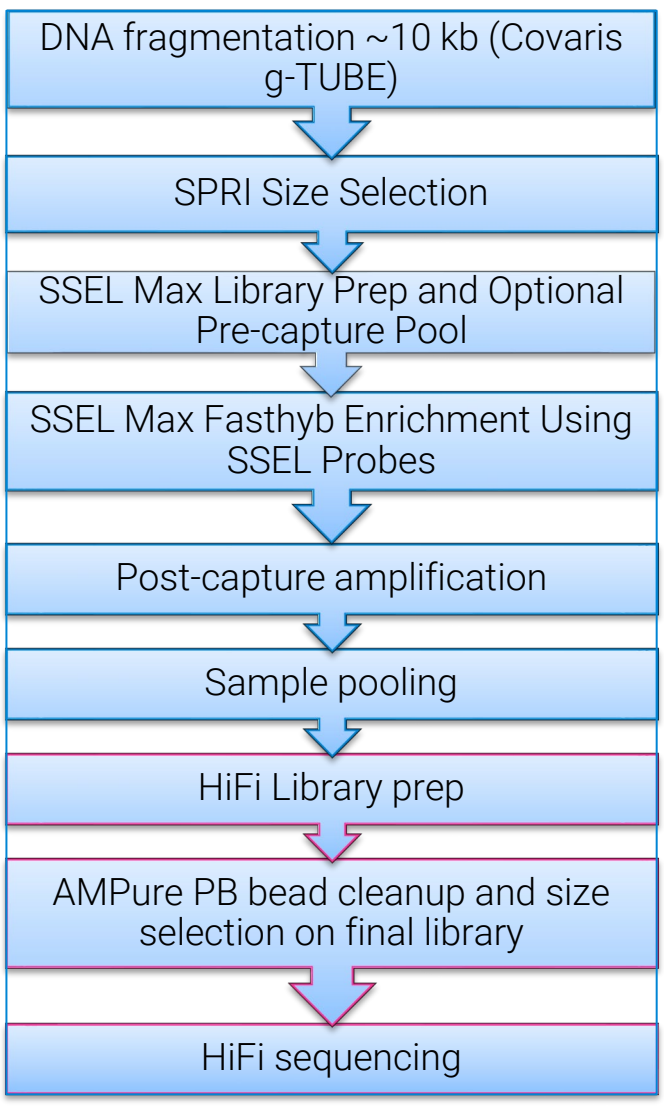


Figure 1. Workflow diagram for library prep and capture with SureSelect Max chemistry, followed by PacBio HiFi sequencing.

Experimental

Sample Details

Coriell samples NA12878, NA24385, NA12143, and NA12149 were used. 24 samples were prepped in total; 8 samples were captured individually; 16 samples were pooled together into a single capture pool using Agilent SureSelect Max library prep and target enrichment chemistry. All captures were performed using the Agilent SureSelect Cancer CGP DNA panel. After capture, samples were prepared for PacBio sequencing using HiFi library prep and sequenced on one SMRT® cell on the Revio® system.

Picard[1] CollectHSMetrics	• NEAR_DISTANCE=5000 • MINIMUM_MAPPING_QUALITY=0 • MINIMUM_BASE_QUALITY=0 • CLIP_OVERLAPPING_READS=false
Picard[1] CollectGcBiasMetrics	• SCAN_WINDOW_SIZE 100
samtools[2] stats	• RL (Read lengths)
bedtools[3] coverage	• –hist • -a CGPregions.bed
DeepVariant[4]	• model_type PACBIO
Sawfish[5]	

Results and Discussion

Figure 2. Mean HiFi read length of 4.6Kb and high proportion of on-target uniquely mapped reads.

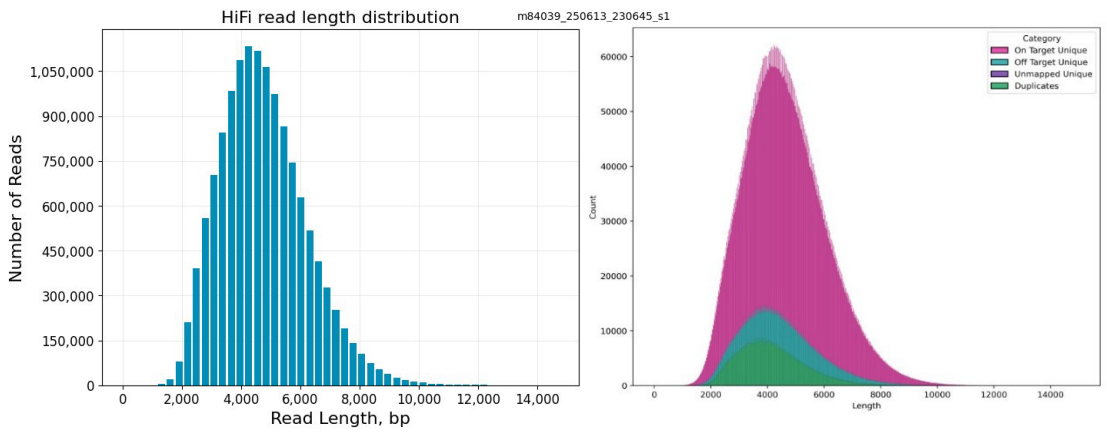


Figure 3. Single plex libraries have slightly lower duplicate rate (and consequently slightly higher base coverage) but similar on-target rate compared to respective multiplex sample.

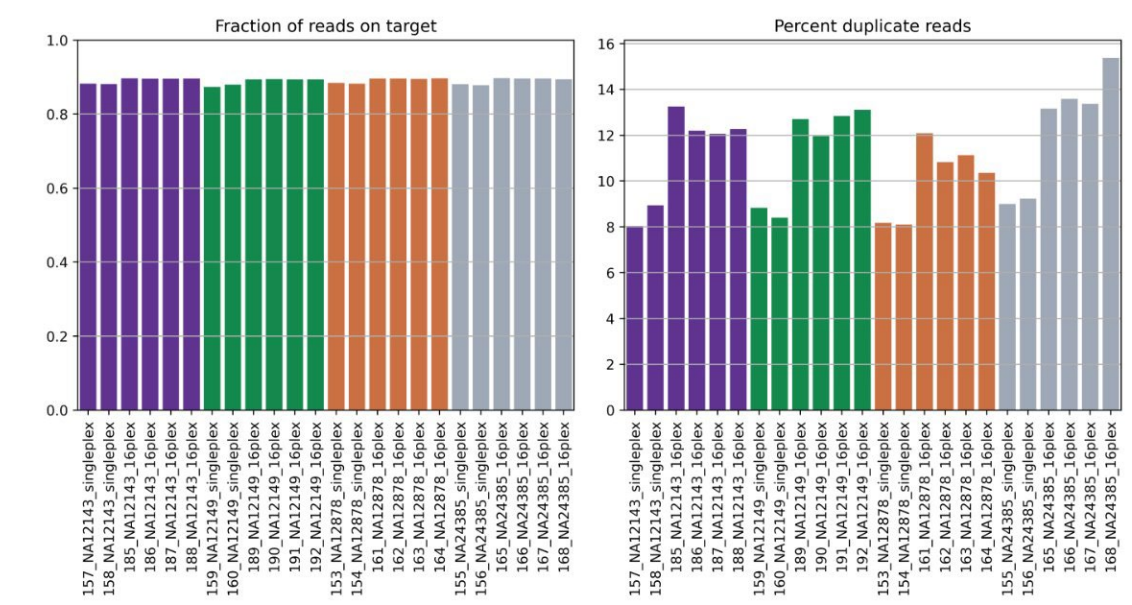


Figure 4. Target bases covered at 20X and 30X are consistent across libraries, with single plex samples showing slightly higher coverage than multiplex samples.

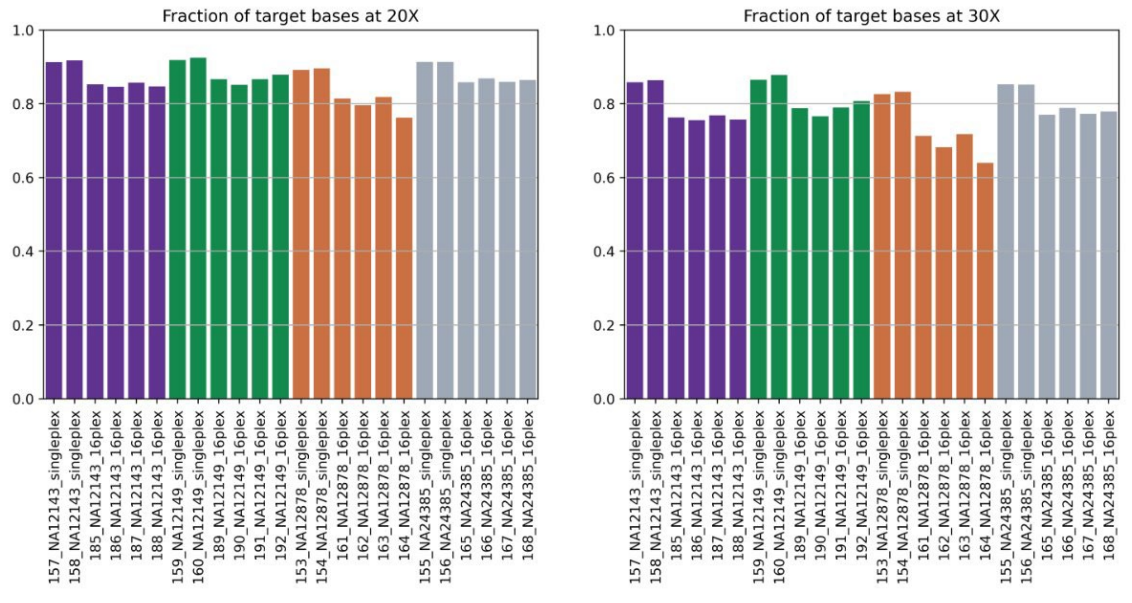
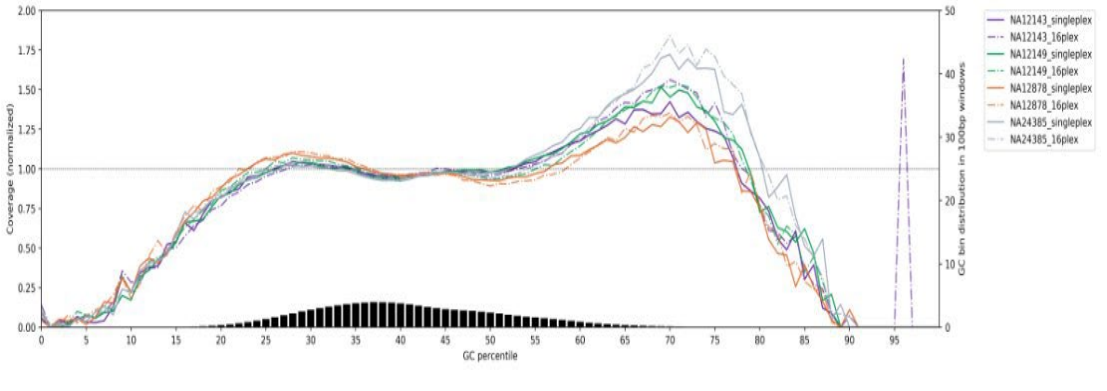


Figure 5. Single plex and multiplex libraries show similar coverage across all GC content with some sample-to-sample variation.



Results and Discussion

Figure 6. Only reporting variant calling results for NA12878 libraries. Single plex samples on average have more true positives, fewer false positives, and fewer false negatives than multiplex samples for SNPs and InDels. Precision and recall metrics are based on NIST v4.2.1 benchmark regions overlapping the CGP panel – 2,809 SNPs and 110 InDels.

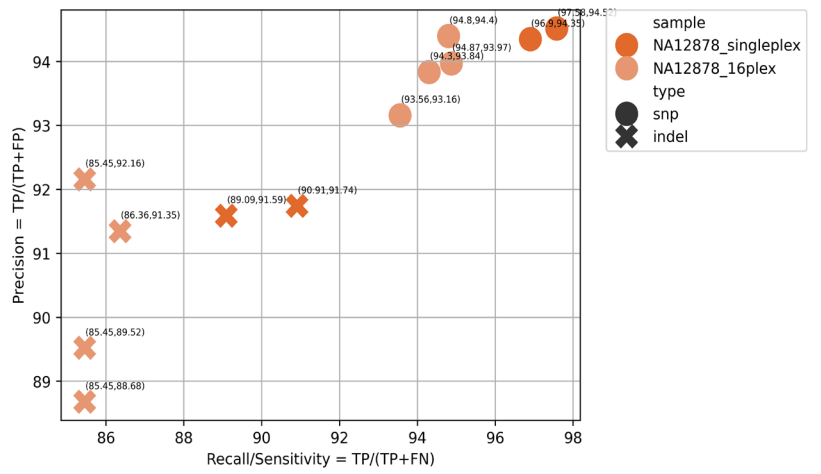


Table 1. SV calling results for HG002 samples for PacBio using Sawfish v2.3.0 and Illumina short read using Manta[6] v1.6.0. Benchmark file from T2T HG002 Q100 project. Data for SV calling was normalized to 66x of the CGP design, accounting for read lengths.

Genotype, Allele Depth & Depth per SV	PacBio155 single plex	PacBio156 single plex	short read replicate1 single plex	short read replicate 2 single plex
chr10:88,132,299 INS96bp	0/1; 15, 32; 47	0/1; 11,23; 34	*FN*	*FN*
chr13:19,173,946 DEL772bp	*FN*	*FN*	*FN*	*FN*
chr19:1,278,175 INS63bp	1/1; 0,4; 4	1/1; 0,2; 2	*FN*	*FN*
chr19:19,147,209 DEL274bp	1/1; 0,24; 24	1/1; 0,26; 26	*FN*	*FN*
chr20:56388,093 INS61bp	1/1; 0,36; 36	1/1; 0,57; 57	*FN*	*FN*
chr22:20,989,693 DEL37bp	0/1; 29,35; 64	0/1; 22,34; 56	0/1; 15,33; 48	0/1; 64,43; 107
chr22:29,299,960 DEL70bp	1/1; 0,32; 32	1/1; 0,36; 36	*FN*	*FN*

Figure 7. IGV screenshot of homozygous 274bp deletion on chr19. Both PacBio single plex samples (top panels) accurately detect this SV, while both short read single plex samples (bottom panels) fail to identify it. The longer read lengths of the PacBio reads span across the deletion, while short read samples are incapable due to their smaller read lengths. The LINE repeat on the 3' end of the SV further complicates alignment for short reads in this region.

Results and Discussion

Figure 8. IGV screenshot of homozygous 70bp deletion on chr22. PacBio single plex samples (top panels) accurately detect the deletion, as its longer read lengths span the SV. The bottom two panels of short read data fail to detect the variant, likely due to insufficient read length and fragmented coverage that does not fully span the SV.

Conclusions

- Off-the-shelf SSEL Max library prep reagent kit and cancer CGP panel optimized for short read sequencing generated robust performance with no panel optimization, no additional enzymes or beads, and minimal workflow optimization
- Samples were successfully pooled into capture following standard pooling recommendations– up to 16 samples for panels smaller than exome
- Demultiplexing was successfully applied using SSEL Max barcodes; PacBio's standard analysis pipeline through SMRT link can be used to analyze the data

References

1. Broad Institute. Picard Toolkit. Broad Institute, GitHub Repository. https://broadinstitute.github.io/picard/.
2. Danecek, P. et al. Twelve years of SAMtools and BCFtools. *GigaScience***10**, giab008 (2021). https://doi.org/10.1093/gigascience/giab008
3. Quilan, A.R. & Hall, I.M. BEDTools: a flexible suite of utilities for comparing genomic features. Bioinformatics**26**, 841-842 (2010). https://doi.org/10.1093/bioinformatics/btq033
4. Poplin, R. et al. A universal SNP and small-indel variant caller using deep neural networks. *Nat. Biotechnol.***36**, 983–987 (2018). https://doi.org/10.1038/nbt.4235
5. Saunders, C.T. et al. Sawfish: improving long-read structural variant discovery and genotyping with local haplotype modeling. *Bioinformatics***41**, btaf136(2025). https://doi.org/10.1093/bioinformatics/btaf136
6. Chen. et al. (2016) Manta: rapid detection of structural variants and indels for germline and cancer sequencing applications. Bioinformatics, 32, 1220-1222. doi:10.1093/bioinformatics/btv710

https://www.agilent.com/en/promotions/asms

Research Use Only. Not for Use in Diagnostic Procedures

This information is subject to change without notice.

DE59599796

© Agilent Technologies, Inc. 2024

Published in USA, May 31,2024

Agilent

Trusted Answers