PacBio

Application note

Robust detection of somatic variants from tumor-normal samples with highly accurate long-read whole genome sequencing

Introduction

Highly accurate long-read sequencing has enabled more complete germline variant detection and the completion of the human genome. In contrast to germline mutations, somatic variant detection has presented distinct challenges due to lower variant frequencies, yielding a lower signal-to-noise ratio and requiring higher sequencing depth. However, the exceptional accuracy and long read lengths of PacBio® HiFi sequencing on the Revio™ system is increasingly being applied for the robust detection of complex variants that were previously inaccessible with short reads or less accurate long reads (Vasan et al., 2019; Nattestad et al., 2018), now with the availability of the higher throughput needed to detect variants present at lower allele frequencies.

Short-read sequencing limits the ability to reconstruct important variation in cancer genomes, including complex structural variation and repetitive regions (Cortes-Ciriano et al., 2022). Inaccurate nanopore long-read sequencing faces challenges in the detection of small variants, such as single nucleotide variants (SNVs) (Olson et al., 2022). Paired tumor-normal WGS studies with HiFi **highly accurate long reads** can detect and phase a wide range of cancer-specific genetic variation, including SNVs, structural variants (SVs), deletions and insertions (indels), copy number variations (CNVs) and methylation, in a single assay. This Application note provides the workflow for the detection of somatic small variants, structural variants, and methylation for paired tumor-normal samples with HiFi whole genome sequencing (WGS) (Figure 1).

102-326-582 REV01 OCT2023

PacBio

Experimental results

HiFi WGS of the COLO829 melanoma cell line performed on the Revio system reveals genomic (Figure 1a) and epigenetic (Figure 1b) variation relative to the matched normal cell line.

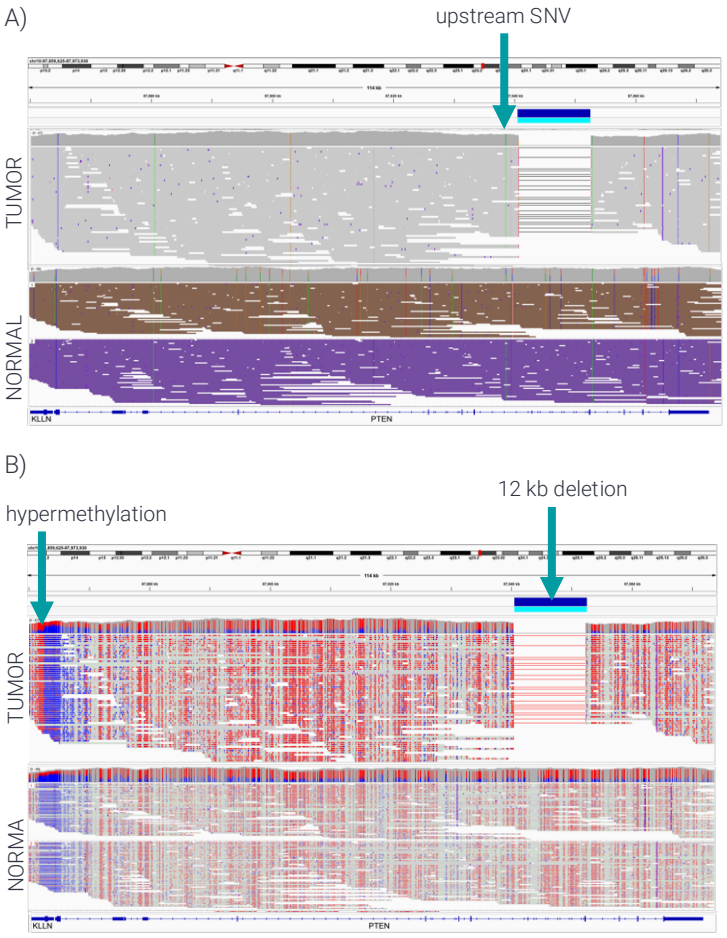


Figure 1. IGV view of a) WGS and (b) methylation data for COLO829 tumor and matched normal COLO829-BL cell lines. Arrows indicate noted tumor variants.

Due to the high accuracy and precision of HiFi sequencing, paired-tumor WGS was able to reveal a 12 kb deletion in the *PTEN* tumor suppressor gene, along with an upstream SNV. Additionally, 5mC methylation calling reveals hypermethylation upstream of *KLLN*, a p53-regulated DNA replication inhibitor. Epigenetics plays a large role in the tumor progression through the regulation of oncogenes and tumor suppressors. Therefore, methylation calling is an important part of this workflow that can reveal valuable insights for cancer research (Hao et al., 2017).

In an analysis using the breast cancer cell line HCC1395, the accuracy of two tumor/normal sequencing depth ratios were compared for SNVs at varying allele frequencies (Figure 2). Both 30X tumor/30X normal and 60X tumor/30X normal coverage for HiFi WGS sequencing approach 0.99 F1 scores down to 0.2 variant allele frequency (VAF). At lower VAF, 60X/30X provided better variant detection, providing guidance for sequencing depth required as a function of VAF.

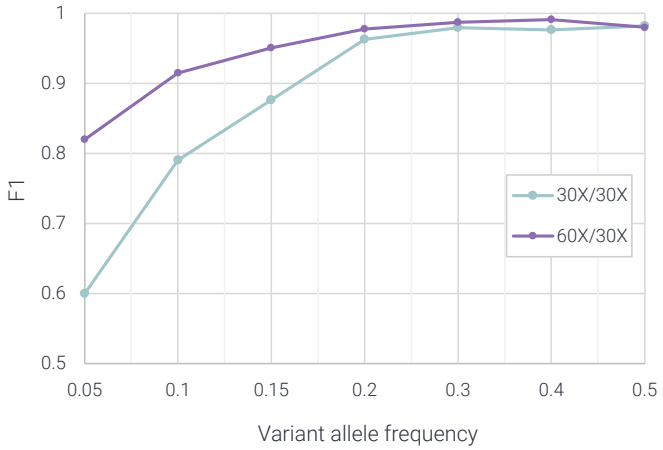


Figure 2. F1 scores for SNV detection in breast cancer cell line HCC1395 at 30X tumor/30X normal and 60X tumor/30X normal coverage.

Similarly, both sequencing depth ratios exhibit high recall (>83%) of a set of 62 structural variants from the Valle-Inclan et al., 2022 truth dataset using the Severus caller (Figure 3).

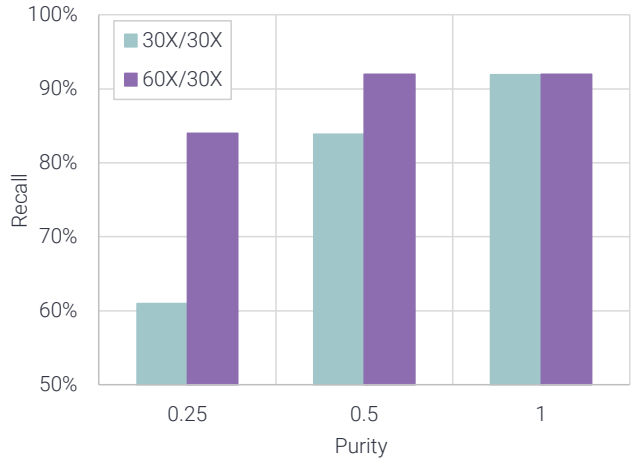


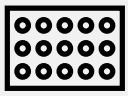
Figure 3. Recall of 62 structural variants improve with increasing COLO829 tumor purity.

Tumor-normal WGS workflow overview

Extract DNA from blood and/or tissue

Because paired tumor-normal sequencing typically requires sampling from both tissue and blood, it is recommended to use extraction protocols according to the Nanobind® tissue kit. The protocols have been validated for many tissue types including heart, liver, spleen, kidney, and colon and require 15 mg of tumor sample (input material quantity may vary depending on tissue). Extracting HMW DNA from whole blood samples using this kit requires 200 µL of whole blood.

Library prep



It is recommended to construct single WGS libraries from extracted gDNA using the SMRTbell® prep kit 3.0 and with 2 µg of DNA input per Revio SMRT® Cell (lower DNA input has been shown to yield high quality libraries with minimal decline in sequencing depth). High-quality genomic DNA will maximize sequencing coverage per SMRTbell library and reduce sequencing costs. The recommended Genome quality number (GQN) of the gDNA prior to shearing is 9.0 or higher at 10 kb. SRE or gel-based size selection methods can be used to remove short DNA and improve read lengths on lower quality samples.

HiFi sequencing

The resulting libraries can be prepared for sequencing on the Revio system following instructions in the Sample Setup module of SMRT® Link.

Analysis

The Revio system provides on-board 5mC methylation calling, and third-party bioinformatic callers are available for detecting somatic small and structural variants. ClairS is a deep-learning method for long-read somatic small variant calling and Severus and Sniffles2 are recommended for detecting structural variants.

Conclusion

These results demonstrate the ability of the PacBio somatic tumor-normal WGS workflow to accurately detect biologically relevant variants, including methylation, across sample and sequencing conditions. This variant-calling performance is driven by the high accuracy of HiFi sequencing and is particularly important in cancers where allele frequencies are much lower. As such, this workflow represents the achievement of a more complete interrogation of cancer genomes, including the detection of both small and complex structural variants.

Resources and references

Resources

Guide & overview – Nanobind tissue kit

Procedure & checklist – Preparing whole genome and metagenome libraries using SMRTbell prep kit 3.0

Dataset – HCC1395 tumor and matched normal WGS on the Revio system

Dataset – COLO829 tumor and matched normal WGS on the Revio system

References

Cortés-Ciriano, I., Gulhan, D. C., Lee, J. J. K., Melloni, G. E., & Park, P. J. (2022). Computational analysis of cancer genome sequencing data. *Nature Reviews Genetics*, 23(5), 298-314. https://doi.org/10.1038/s41576-021-00431-y

Hao, X., et al. (2017). DNA methylation markers for diagnosis and prognosis of common cancers. *Proceedings of the National Academy of Sciences*, 114(28), 7414-7419. https://doi.org/10.1073/pnas.1703577114

Mandelker, D., & Ceyhan-Birsoy, O. (2020). Evolving significance of tumor-normal sequencing in cancer care. *Trends in Cancer*, 6(1), 31-39. https://doi.org/10.1016/j.trecan.2019.11.006

Nattestad, M., et al. (2018). Complex rearrangements and oncogene amplifications revealed by long-read DNA and RNA sequencing of a breast cancer cell line. *Genome Research*, 28(8), 1126-1135. https://doi.org/10.1101/gr.231100.117

Olson, N. D., et al., (2022). PrecisionFDA Truth Challenge V2: Calling variants from short and long reads in difficult-to-map regions. *Cell Genomics*, 2(5). https://doi.org/10.1016/j.xgen.2022.100129

Valle-Inclan, J. E., et al. (2022). A multi-platform reference for somatic structural variation detection. *Cell Genomics*, 2(6). https://doi.org/10.1016/j.xgen.2022.100139

Vasan, N., et al. (2019). Double PIK3CA mutations in cis increase oncogenicity and sensitivity to PI3Kα inhibitors. *Science*, 366(6466), 7. https://doi.org/10.1126/science.aaw9032