A long-read whole-genome cancer workflow for integrated genomic, biomarker, and methylation profiling

Zev Kronenberg[1], James M. Holt[1], William J. Rowell[1], Matthew R. Stachowiak[2], Ariel Schwartz[2], Audric Wertz[2], Aaron M. Wenger[1], Zachary D. Stephens[3], Craig A. Struble[1], Akhilesh Pandey[3], Richard Kandasamy[3], Mark J.P. Chaisson[1]

[1] PacBio, Menlo Park, California, USA; [2]GenomOncology, Cleveland, Ohio; [3]Mayo Clinic, Rochester, Minnesota

Introduction

Long-read whole-genome sequencing can capture somatic small variants, structural variants, phasing, methylation, repeat variation, and other tumor biomarkers in a single assay, but analysis is often fragmented across separate tools and workflows. To address this gap, we created secondary analysis workflows that integrate best-in-class long-read tools for cancer genome analysis, integrated with best-in-class tertiary analysis and reporting..

Methods

We developed two PacBio HiFi whole-genome analysis workflows for both *tumor-only* (**Figure 1**) and matched *tumor-normal* samples.

Variant annotation, interpretation, and reporting is fully integrated and available through GenomOncology's Pathology Workbench.

Workflow analyses

Somatic SNV/INDEL	Phasing indicates haplotype origin of somatic mutations.
Somatic structural variants	Complex structural variants hard to resolve with short-reads.
Mutational signatures	Understand mutational processes in cancer.
Allele-specific copy number	Call haplotype-specific copy numbers.
Biomarkers (MSI + TMB)	Common genomic biomarkers for therapy selection.
PGx diplotyping	Understanding drug interactions, and neoantigen prediction.
Mitochondrial analysis	Biomarker and metabolism analyses.
Methylation analysis	Global and gene specific methylation patterns.
GenomOncology	Variant annotation, interpretation, and reporting.

Tumor workflow implementation

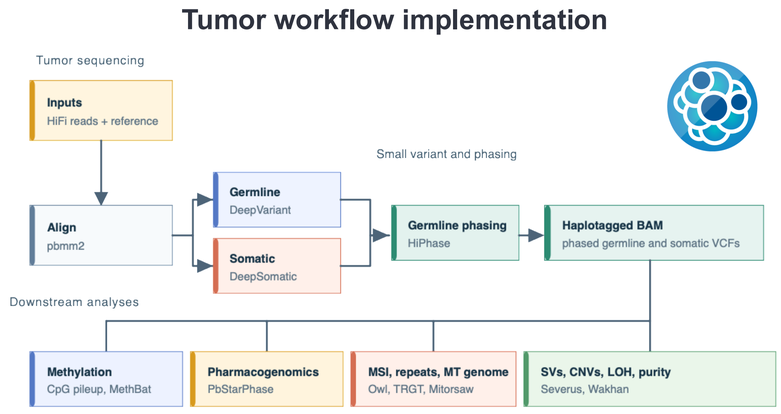


Figure 1. The workflow is implemented in WDL and contains dozens of steps. The first stage aligns, calls small variants, and phases the long reads. Downstream callers operate on phased HiFi data.

Depth of coverage considerations

Somatic variant calling accuracy depends on depth, as shown in **Figure 2**. High-frequency driver mutations (>25%) require ~20-fold coverage to achieve strong recall, whereas low-frequency mutations (<0.5%) require ~60-fold coverage.

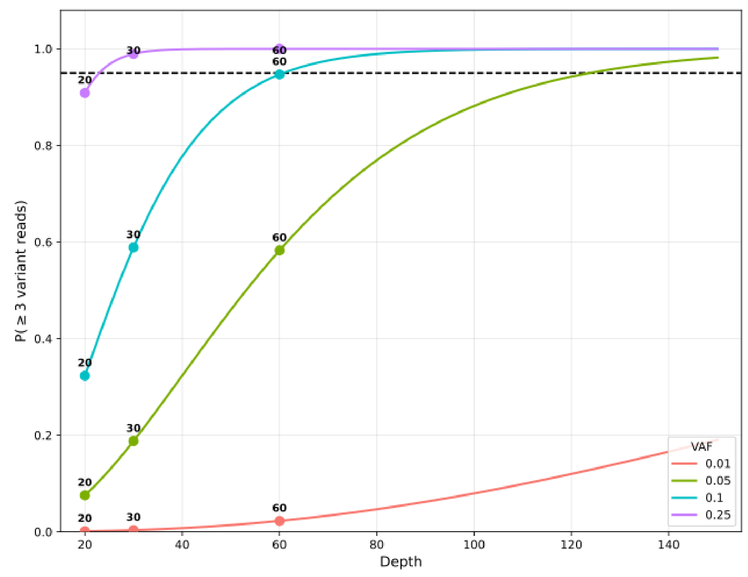


Figure 2. Binomial sampling of reads. The x-axis is sequencing depth, and the y-axis is the probability of reliably calling a variant. Lines are colored by the somatic variant allele frequency.

Variant calling benchmarks

The cancer pipelines achieve strong recall, and precision improves further with a matched normal sample (**Tables 1&2**). The accuracy we observe is consistent with the literature, providing high confidence results.

Small variants

workflow	Recall	Precision
Tumor-only	0.91	0.51
Tumor-normal	0.99	0.89

Table 1. The benchmark truth set was the COLO829 cell line from the SMaHT consortium, sequenced to ~60-fold coverage in both tumor and normal. Variant comparisons were performed using the Aardvark tool (https://github.com/PacificBiosciences/aardvark).

Structural variants

workflow	Recall	Precision
Tumor-only	0.91	0.38
Tumor-normal	0.94	0.61

Table 2. The benchmark truth set was the NIST HG008 cell line, sequenced to ~60-fold coverage. Variant comparisons were performed using the Truvari tool (https://truvari.readthedocs.io/en/latest/).

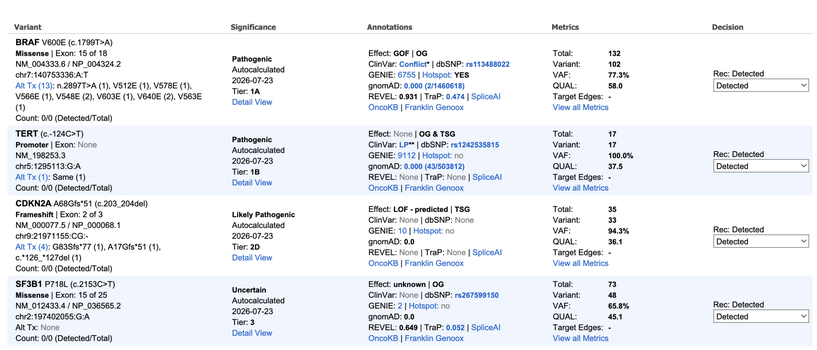
Variant annotation, interpretation, & reporting

Seamless secondary + tertiary analysis

Secondary Analysis can be hosted by GenomOncology: outputs flow directly to tertiary analysis (Pathology Workbench)

Tertiary analysis in GenomOncology

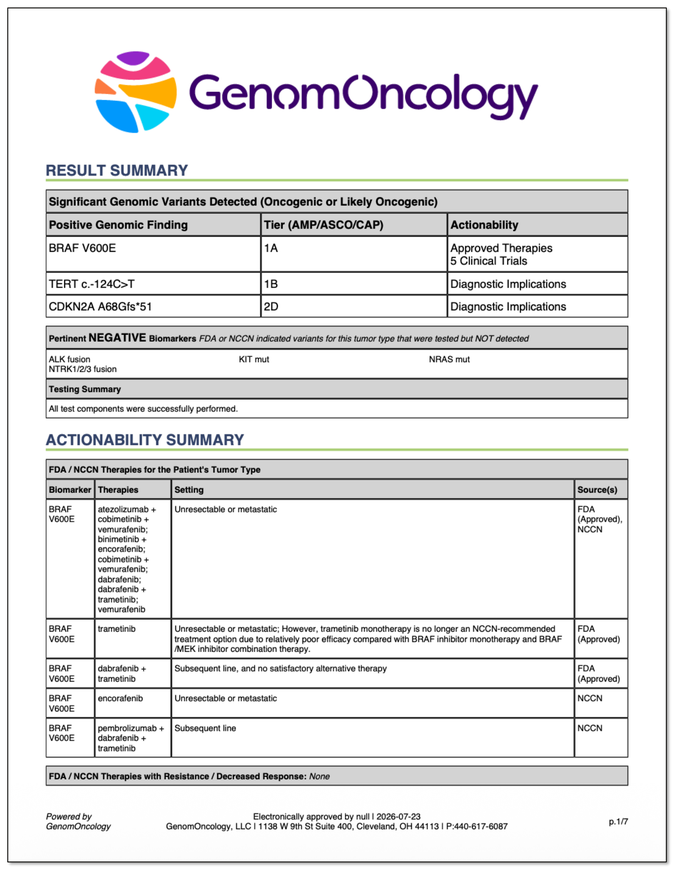
GenomOncology Pathology Workbench identifies, annotates, and auto-classifies the key variants related to cancer including structural variation.



Variant	Significance	Annotations	Metrics	Decision
BRAF V600E (c.1799T>A) **Missense** | Exon: 15 of 18 NM_004333.6 / NP_004324.2 chr7:140753336:A:T Alt Tx (13): n.2897T>A (1), V512E (1), V578E (1), V566E (1), V548E (2), V603E (1), V640E (2), V563E (1) Count: 0/0 (Detected/Total)	**Pathogenic** Autocalculated 2026-07-23 Tier: **1A** Detail View	Effect: **GOF** | **OG** ClinVar: Conflict* | dbSNP: rs113488022 GENIE: 6755 | Hotspot: **YES** gnomAD: 0.000 (2/1460618) REVEL: **0.931** | TraP: **0.474** | SpliceAI OncoKB | Franklin Genoox	Total: **132** Variant: **102** VAF: **77.3%** QUAL: **58.0** Target Edges: **-** View all Metrics	Rec: Detected Detected
TERT (c.-124C>T) **Promoter** | Exon: None NM_198253.3 chr5:1295113:G:A Alt Tx (1): Same (1) Count: 0/0 (Detected/Total)	**Pathogenic** Autocalculated 2026-07-23 Tier: **1B** Detail View	Effect: None | **OG & TSG** ClinVar: LP** | dbSNP: rs1242535815 GENIE: 9112 | Hotspot: no gnomAD: 0.000 (43/503812) REVEL: None | TraP: None | SpliceAI OncoKB | Franklin Genoox	Total: **17** Variant: **17** VAF: **100.0%** QUAL: **37.5** Target Edges: **-** View all Metrics	Rec: Detected Detected
CDKN2A A68Gfs*51 (c.203_204del) **Frameshift** | Exon: 2 of 3 NM_000077.5 / NP_000068.1 chr9:21971155:CG:- Alt Tx (4): G83Sfs*77 (1), A17Gfs*51 (1), c.*126_*127del (1) Count: 0/0 (Detected/Total)	**Likely Pathogenic** Autocalculated 2026-07-23 Tier: **2D** Detail View	Effect: **LOF - predicted** | **TSG** ClinVar: None | dbSNP: None GENIE: 10 | Hotspot: no gnomAD: **0.0** REVEL: None | TraP: None | SpliceAI OncoKB | Franklin Genoox	Total: **35** Variant: **33** VAF: **94.3%** QUAL: **36.1** Target Edges: **-** View all Metrics	Rec: Detected Detected
SF3B1 P718L (c.2153C>T) **Missense** | Exon: 15 of 25 NM_012433.4 / NP_036565.2 chr2:197402055:G:A Alt Tx: None Count: 0/0 (Detected/Total)	**Uncertain** Autocalculated 2026-07-23 Tier: **3** Detail View	Effect: **unknown** | **OG** ClinVar: None | dbSNP: rs267599150 GENIE: 2 | Hotspot: no gnomAD: **0.0** REVEL: **0.649** | TraP: **0.052** | SpliceAI OncoKB | Franklin Genoox	Total: **73** Variant: **48** VAF: **65.8%** QUAL: **45.1** Target Edges: **-** View all Metrics	Rec: Detected Detected

Fully custom reporting

A concise report is generated by Pathology Workbench with AMP/ASCO/CAP tiering. The report format is fully customizable, and sub-panels can easily be created from the Whole Genome Sequencing (WGS) output.



GenomOncology

RESULT SUMMARY

Significant Genomic Variants Detected (Oncogenic or Likely Oncogenic)		
Positive Genomic Finding	**Tier (AMP/ASCO/CAP)**	**Actionability**
BRAF V600E	1A	Approved Therapies 5 Clinical Trials
TERT c.-124C>T	1B	Diagnostic Implications
CDKN2A A68Gfs*51	2D	Diagnostic Implications

Pertinent NEGATIVE Biomarkers *FDA or NCCN indicated variants for this tumor type that were tested but NOT detected*

ALK fusion NTRK1/2/3 fusion	KIT mut	NRAS mut

Testing Summary

All test components were successfully performed.

ACTIONABILITY SUMMARY

FDA / NCCN Therapies for the Patient's Tumor Type

Biomarker	Therapies	Setting	Source(s)
BRAF V600E	atezolizumab + cobimetinib + vemurafenib; binimetinib + encorafenib; cobimetinib + vemurafenib; dabrafenib; dabrafenib + trametinib; vemurafenib	Unresectable or metastatic	FDA (Approved), NCCN
BRAF V600E	trametinib	Unresectable or metastatic; However, trametinib monotherapy is no longer an NCCN-recommended treatment option due to relatively poor efficacy compared with BRAF inhibitor monotherapy and BRAF /MEK inhibitor combination therapy.	FDA (Approved)
BRAF V600E	dabrafenib + trametinib	Subsequent line, and no satisfactory alternative therapy	FDA (Approved)
BRAF V600E	encorafenib	Unresectable or metastatic	NCCN
BRAF V600E	pembrolizumab + dabrafenib + trametinib	Subsequent line	NCCN

FDA / NCCN Therapies with Resistance / Decreased Response: *None*

Powered by GenomOncology

Electronically approved by null I 2026-07-23
GenomOncology, LLC I 1138 W 9th St Suite 400, Cleveland, OH 44113 I P:440-617-6087

p.1/7

HiFi data provides phased methylation

Methylation can play a key-role in cancer progression. Global perturbances in methylation are well established in colorectal cancer (CRC) and Multiple Myelomas (MM). We demonstrate how HiFi 5mC methylation calls produced by our workflows can be used to develop biomarkers.

Methylation in a control cell line (HG02) is typically ~70% (**Figure 3A**). However, we find in a colon cancer cell line (LS180) decreased global levels of methylation ~50% (**Figure 3B**).

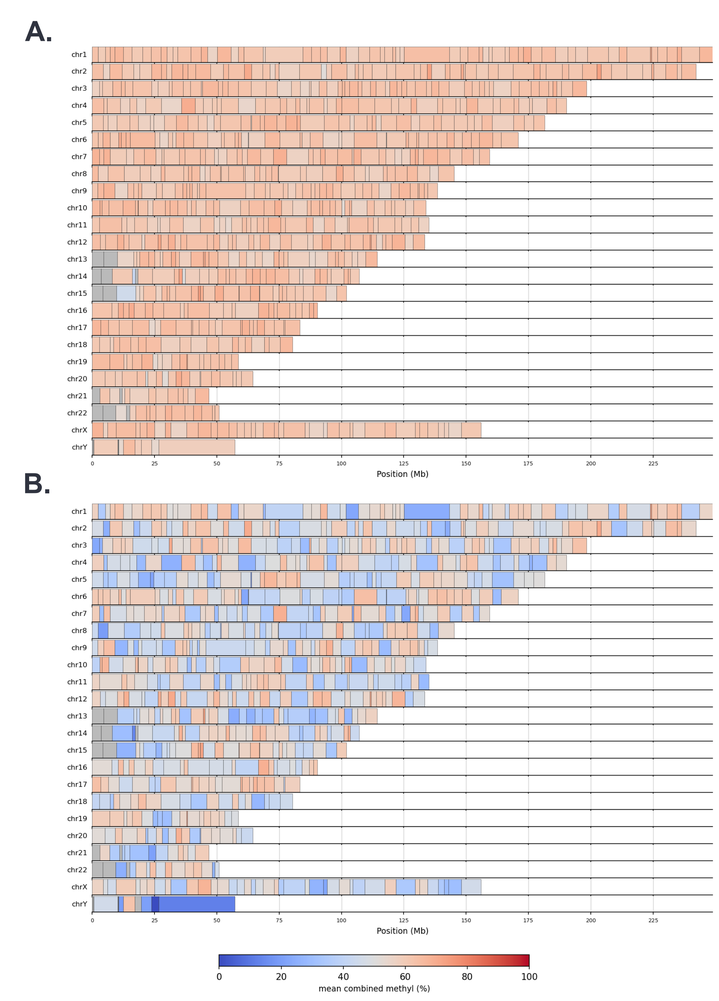


Figure 3. Ideograms of 5mC methylation levels averaged within cytobands. **A)** HG002 a normal control. **B)** LS180 CRC shows a global decrease in methylation, and methylation heterogeneity.

We explored how methylation rates in MM cell lines correlate with driver mutation (**Figure 4**). Analysis reveals that increased methylation correlate with *myc* amplification positive cell lines. Relative to controls, the MM cell lines have lower levels of methylation.

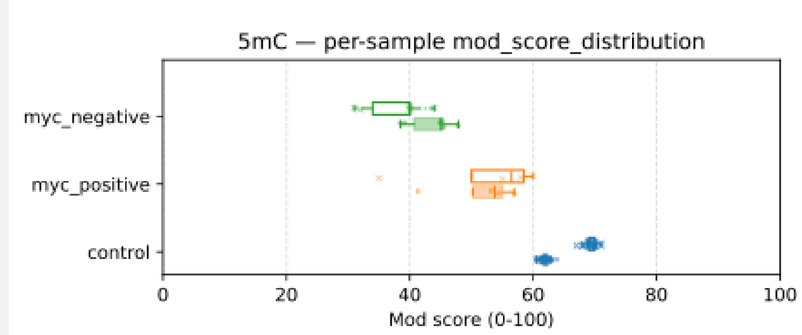


Figure X. Correlation between *myc* status and methylation patterns in multiple myeloma. The top panel displays the distribution of methylation for cell lines without *myc* amplification, with *myc* amplification, and unaffected controls. The open boxes are mean methylation values, and the shaded boxes are median.

Conclusions

- The cancer workflows deliver a comprehensive suite of analyses, providing key insights into driver mutations, biomarkers, and methylation patterns, all from HiFi data.
- The GenomOncology workbench effectively integrates with the secondary analysis, classifies and reports actionable variants.
- An early version of the pipeline is available at: https://github.com/PacificBiosciences/HiFi-somatic-WDL.
- Questions can be directed to Zev Kronenberg at zev@pacbio.com or Audric Wertz audric@genomoncology.com.

Research use only. Not for use in diagnostic procedures. ©2026 Pacific Biosciences of California, Inc. ("PacBio"). All rights reserved. Information in this document is subject to change without notice. PacBio assumes no responsibility for any errors or omissions in this document. Certain notices, terms, conditions and/or use restrictions may pertain to your use of PacBio products and/or third-party products. Refer to the applicable PacBio terms and conditions of sale and to the applicable license terms at pacb.com/license. Pacific Biosciences, the PacBio logo, PacBio, Circulomics, Omniome, SMRT, SMRTbell, Iso-Seq, Sequel, Nanobind, Revio, Revio SPRQ, Kinnex, PureTarget, SPRQ, Vega and Vega SPRQ are trademarks of PacBio. Reference to third-party companies, products, or services are for informational purposes only and do not imply endorsement, sponsorship, or affiliation.