PacBio

Application note

HiFi sequencing of FFPE tumor samples enables comprehensive detection and phasing of somatic variation

Formalin-fixed, paraffin-embedded (FFPE) tissue is a widely used preservation method in clinical pathology that maintains tissue structure and cellular detail for long-term storage, enabling vital histopathological examination and molecular analysis, particularly in cancer research. Despite their value, FFPE samples present significant challenges for genomic analysis. Formalin fixation introduces DNA crosslinks and fragmentation, often resulting in short, chemically damaged molecules. These effects limit the ability of many sequencing technologies, particularly long-read approaches, to recover high-quality genomic information.

PacBio® HiFi sequencing offers distinct advantages for cancer genomics, including highly accurate long reads that enable comprehensive detection of somatic structural variants (SVs), small variants, and haplotype phasing. These capabilities are especially important for resolving complex cancer genomes, where structural rearrangements and allele-specific variation play critical roles. However, the degraded and crosslinked nature of FFPE-derived DNA has historically limited the application of long-read sequencing to these samples.

To address this challenge, PacBio in collaboration with Covaris has developed a workflow that enables extraction, preparation, and sequencing of FFPE-derived DNA for HiFi sequencing. This approach overcomes traditional limitations and unlocks FFPE samples for high-resolution, long-read genomic analysis.

102-326-666 REV01 APR2026

PacBio

Workflow overview for FFPE HiFi sequencing

The workflow integrates optimized DNA extraction, amplification, and library preparation methods to enable high-quality HiFi sequencing from FFPE tissue (Figure 1).

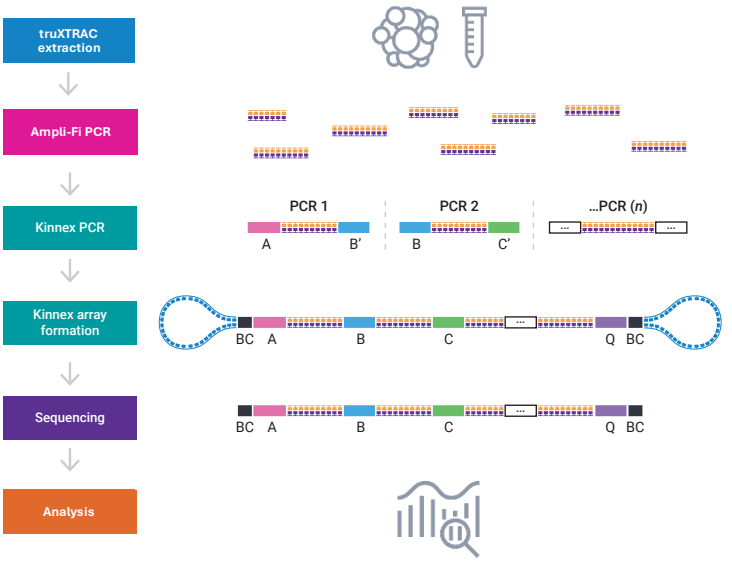

Figure 1. The workflow overview from Covaris extraction, through Ampli-Fi and Kinnex™ PCR, to HiFi sequencing and data analysis. Analysis was conducted in SMRT® Link and variant detection was performed using the HiFi Somatic WDL.

DNA is first extracted from FFPE-preserved tumor samples using Covaris truXTRAC chemistry, which is designed to recover DNA fragments of sufficient length to support long-read sequencing. Extracted DNA is then repaired and prepared for library construction.

To amplify usable DNA, a modified Ampli-Fi workflow is applied. Following amplification, the fragments are processed using the Kinnex library preparation protocol, which concatenates shorter DNA fragments into longer molecules to better utilize HiFi sequencing capacity. These longer constructs are converted into SMRTbell® libraries and sequenced using HiFi sequencing.

Data analysis is performed using the PacBio HiFi somatic WDL workflow and the pb-StarPhase tool, enabling detection of structural variants, small variants, and haplotype phasing. The total workflow can be completed in approximately two days, with ~13 hours of active library preparation time.

Recovery of FFPE DNA suitable for long-read sequencing

DNA extracted from FFPE tumor samples using the Covaris workflow demonstrated sufficient yield and integrity to support downstream HiFi sequencing. Across multiple tissue types, including brain, kidney, and uterine tumors, DNA yields exceeded 10 µg per sample from two to four 10-micron scrolls, far exceeding the protocol input requirements of 1–20 ng per sample. DNA integrity numbers (DIN) ranged from 4 to greater than 6.5, indicating moderate preservation despite fixation. Ampli-Fi library size distribution was further validated via Fragment Analyzer (FA) and showed size distributions primarily in the 500–1,000 bp range, which are suitable for downstream concatenation using the Kinnex workflow.

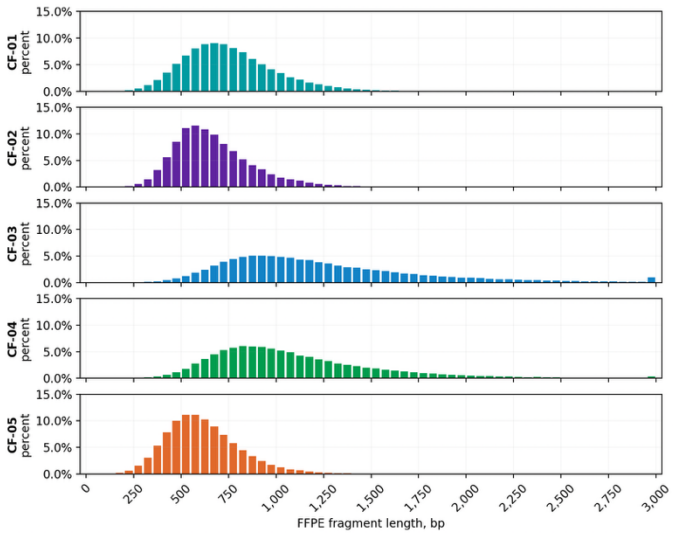

Figure 2. Read lengths for FFPE libraries of five distinct samples (CF01-CF05). Mean lengths between 500–1,000 bp are of adequate size for concatenation with the Kinnex workflow.

These results demonstrate that, even with degraded FFPE material, it is possible to recover DNA fragments of sufficient quality and length to enable HiFi long-read sequencing when paired with appropriate library preparation strategies.

Robust HiFi sequencing performance across FFPE tumor samples

Application of the Ampli-Fi and Kinnex workflows enabled consistent and high-quality HiFi sequencing across diverse FFPE tumor samples. Sequencing generated more than 100 million HiFi reads per sample, with mean

Sample ID	Tissue	Loading concentration	Kinnex HiFi reads			Segmented reads		
			Reads	Yield	Length	Reads	Length	Duplicates
CF-01	Kidney	150 pM	9.6 M	134.4 Gb	14.0 kb	152.0 M	863 bp	31%
CF-02	Kidney	150 pM	10.3 M	129.9 Gb	12.6 kb	162.4 M	779 bp	61%
CF-03	Uterus	150 pM	7.1 M	147.5 Gb	20.7 kb	110.5 M	1,306 bp	10%
CF-04	Uterus	150 pM	7.6 M	142.5 Gb	18.7 kb	118.3 M	1,179 bp	12%
CF-05	Brain	63 pM	8.9 M	104.2 Gb	11.7 kb	139.0 M	729 bp	34%

Table 1. HiFi sequencing metrics for FFPE-preserved tumor tissues. Yield and read lengths met or exceeded specifications, with moderate duplicate rates for segmented reads.

read lengths ranging from approximately 11 kb to over 20 kb (Table 1).

Importantly, sequencing performance was consistent across multiple tissue types and varying levels of DNA quality, demonstrating the robustness of the workflow across heterogenous FFPE tumor samples.

Comprehensive detection of somatic variants in cancer-associated genes

HiFi sequencing of FFPE samples enabled detection of both structural and small variants across the genome, including in clinically relevant cancer genes.

Results summarized in Table 2 highlight genome-wide and gene-level coverage across FFPE samples following alignment to GRCh38. Across all samples, ≥99.7% of the 21,502 protein-coding genes achieved at least 1x coverage, indicating near-complete representation. Between 78.0–93.4% of genes are captured at or above 20x coverage, demonstrating the ability to provide both broad and sufficiently deep coverage of protein-coding regions to support reliable downstream variant analysis.

On average, approximately 11,000 structural variants and 5.1 million small variants were identified per sample. These variants were detected with high confidence and included alterations in key cancer-associated loci such as *TP53* and *EGFR*, demonstrating the applicability of the workflow to biologically and clinically meaningful targets.

The ability to simultaneously detect structural variants and small variants using a single sequencing assay provides a more complete view of tumor genome architecture compared to traditional short-read approaches.

Sample ID	Mean coverage	≥1x	≥2x	≥5x	≥10x	≥20x	≥30x
CF-01	22.1x	99.9%	99.8%	99.0%	96.1%	86.7%	76.0%
CF-02	12.0x	99.9%	99.8%	98.8%	93.4%	78.0%	62.1%
CF-03	37.2x	99.7%	99.7%	99.4%	98.0%	92.4%	85.6%
CF-04	34.7x	99.7%	99.7%	99.5%	98.3%	93.4%	87.4%
CF-05	15.9x	99.9%	99.9%	99.7%	98.6%	91.8%	83.8%

Table 2. Percent of genes covered at each depth. Genome-wide coverage and coverage of 21,502 protein-coding genes after aligning deduplicated FFPE samples to GRCh38.

Sample ID	Read length (bp)	HLA-A	HLA-B	HLA-C
CF-01	863	./.	*40:431:02/*40:431:02	*07:551N/*07:551N
CF-02	779	./.	./.	./.
CF-03	1,306	*24:02:01:53/*24:02:01:53	*35:01:01:07/*35:01:01:07	*04:526Q/*06:02:01:108
CF-04	1,179	*01:01:118/*01:01:118	*08:01:01:26/*35:03:29	*04:01:01:179/*07:01:01:92
CF-05	729	./.	*08:329Q/*08:329Q	./.

Table 3. HLA diplotyping using pb-StarPhase demonstrates accurate allele resolution at read lengths sufficient to span the defined HLA regions.

Haplotype phasing of somatic variation

In addition to variant detection, HiFi sequencing enabled phasing of approximately 60% of variants into haplotypes. This capability allows researchers to determine the allelic context of mutations, providing insight into compound heterozygosity, allele-specific expression, and clonal architecture in tumors.

HiFi sequencing also enabled accurate diplotyping of HLA genes using the pb-StarPhase tool. Successful diplotyping required reads spanning at least 50% of the target region, and improved results were observed in samples with longer mean read lengths.

These results highlight the importance of long, accurate reads for resolving complex and clinically relevant genomic regions in cancer

Discussion

FFPE samples have long represented an underutilized resource for long-read sequencing due to DNA damage and fragmentation introduced during fixation. The workflow presented here overcomes these limitations by combining optimized extraction with amplification and concatenation-based library preparation.

The Covaris extraction workflow recovers DNA fragments of sufficient length for downstream processing, and the Ampli-Fi and Kinnex workflows transform fragmented DNA into long molecules compatible with HiFi sequencing. Together, these approaches enable generation of high-yield, high-quality sequencing data from challenging FFPE samples.

This integrated workflow supports comprehensive genomic analysis, including structural variant detection, small variant calling, and haplotype phasing, all from archival tumor material. By enabling HiFi sequencing of FFPE samples, this approach expands access to previously inaccessible datasets and opens new opportunities for cancer research.

Conclusion

A combined Covaris and PacBio workflow enables high-quality HiFi sequencing from FFPE tumor samples, overcoming traditional limitations associated with DNA damage and fragmentation.

Key advantages of the workflow include:

- Recovery of DNA fragments suitable for long-read sequencing from FFPE tissue
- Robust HiFi sequencing performance across diverse tumor types and DNA quality levels
- Comprehensive detection of structural and small variants in a single assay
- Phasing of somatic variants into haplotypes for deeper biological insight
- Compatibility with clinically relevant targets and complex genomic regions

This approach unlocks the full potential of FFPE samples for long-read sequencing, providing a more complete and accurate view of cancer genomes and enabling new discoveries from archival tumor collections.

102-326-666 REV01 APR2026

Resources

Amirault, K., et al. (2025). Fully automated extraction of high-quality total nucleic acids from FFPE specimens for comprehensive genomic profiling of solid tumors. *SLAS Technology*, *31*, 100252. https://doi.org/10.1016/j.slast.2025.100252

Covaris FFPE Whitepaper

Covaris truXTRAC FFPE SMART Solutions

PacBio AmpliFi protocol

PacBio Kinnex full-length RNA library prep protocol

PacBio HiFi somatic WDL Application note

PacBio pb-StarPhase GitHub

102-326-666 REV01 APR2026