A new workflow for FFPE tumor samples enables a streamlined solution for structural variant detection and phasing of somatic mutations through long-read sequencing

Poster # 3253

Camille Conner[1], Ian McLaughlin[1], Juniper Lake[1], Davy Lee[1], Heather Ferrao[1], Greg Endress[2], Ulrich Thomann[2], Martina Werner[2], Luca Beker[2]
[1]PacBio, 1305 O'Brien Drive, Menlo Park, CA 94025 [2]Covaris, 14 Gill St, Woburn, MA 01801

Introduction

HiFi long-read sequencing provides highly accurate reads of fragments from 500 bp to 25,000 bp in length. These long reads offer an agnostic measure of sequence length and enable structural variant detection, phasing, and direct methylation analysis. However, FFPE tissue has been difficult for HiFi sequencing because extraction often yields short, damaged fragments and fixation crosslinks DNA which disrupts long-read sequencing. This poster demonstrates a workflow that can be used to extract longer DNA fragments from FFPE-preserved tissue to realized the benefits of HiFi sequencing for FFPE DNA.

Methods

- DNA was extracted from FFPE-preserved kidney, uterus, and brain tissue using Covaris truXTRAC® chemistry to recover longer fragments on the Hamilton Sonication Star™
- Input DNA was repaired and A-tailed prior to adapter ligation
- A custom Ampli-Fi workflow was used to amplify FFPE DNA and improve usable input for library preparation
- Kinnex library preparation concatenates shorter FFPE fragments into longer molecules via PCR-based assembly
- Libraries were converted to SMRTbell templates and sequenced using HiFi sequencing
- Data was analyzed using the HiFi WGS variant analysis pipeline for structural variants, small variants, and haplotype phasing.

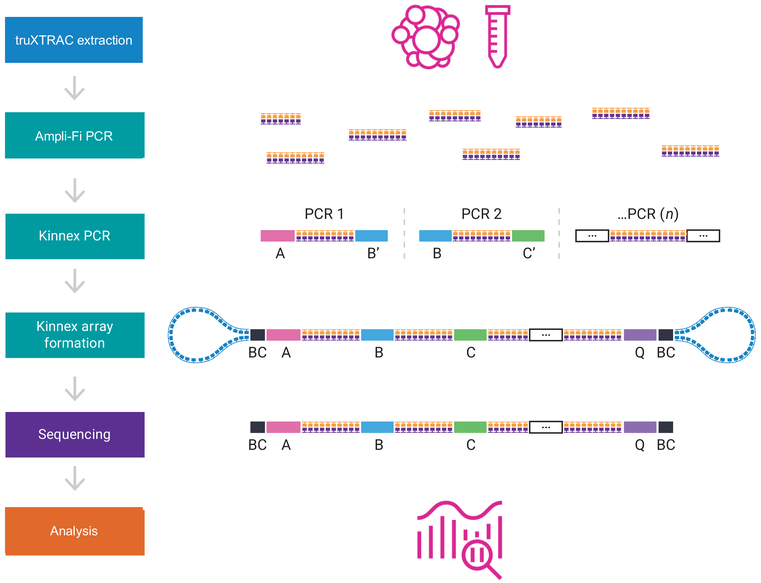

Figure 1. The workflow overview from Covaris extraction, through Ampli-Fi and Kinnex PCR, to HiFi sequencing and data analysis. Analysis was conducted in SMRT Link and variant detection was performed using the WGS variant calling pipeline and the HiFi Somatic WDL.

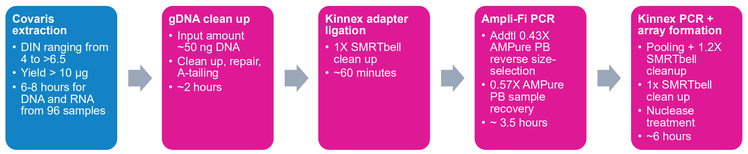

Figure 2. Detailed wet lab workflow from Covaris (blue) and PacBio (magenta) with timing estimates for each step. Starting material was 2-4 scrolls of FFPE-preserved tissue. Total run time is about 2 days.

Covaris truXTRAC extraction recovered fragments of adequate length to take advantage of HiFi long read sequencing

- **Yield:** DNA concentration was determined via fluorometric quantitation. All selected tissues yielded total DNA masses exceeding the requested 10 µg threshold.
- **Integrity:** Quality was evaluated using DNA Integrity Numbers (DIN), with values ranging from 4 to > 6.5 as measured with TapeStation.
- **Sizing:** Genomic DNA size distribution was further validated via Fragment Analyzer (FA) to confirm adequate length for Kinnex concatenation.

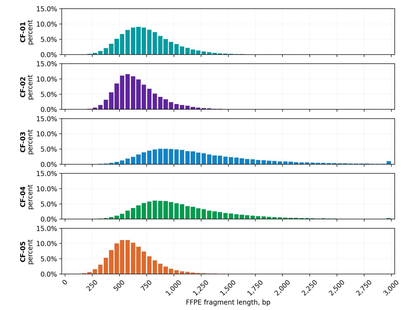

Figure 3. Fragments extracted from FFPE. The CF01-CF05 are library traces prior to Kinnex concatenation. DIN is after extraction but before Ampli-Fi and the traces shown are after Ampli-FI but before Kinnex. Mean fragment lengths between 500-1,000 bp in length are adequate size for concatenation with the Kinnex workflow.

Robust sequencing yield and read length achieved across multiple types of FFPE-preserved tumors

- Sequencing generated >100 million HiFi reads per sample
- Mean Kinnex HiFi read lengths ranged from 11 kb to over 20 kb
- Consistent performance observed across brain, kidney, and uterine FFPE tumor samples
- Workflow supports sequencing across variable DNA quality and tissue types

			Kinnex HiFi reads			Segmented reads		
Sample ID	Tissue	Loading conc.	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 low duplicate rates for segmented reads.

Structural and small variants detected and phased into haplotypes comprehensively across samples

- 11,000 structural variants detected per sample
- 5.1 million small variants identified genome-wide
- ~60% of variants phased into haplotypes
- HiFi sequencing enables integrated detection of structural and small variants in FFPE samples

Detection of clinically relevant variants in cancer genes

- High-quality variant detection observed in clinically relevant loci
- Structural and small variants identified in key cancer-associated genes
- Demonstrates applicability of the workflow to real tumor biology

		Percent of Genes Covered at Each Depth					
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. Genome-wide coverage and coverage of 21,502 protein-coding genes after aligning FFPE samples to GRCh38.

Figure 4. Cancer associated gene, TP53, sequenced with sufficient coverage for detection and analysis.

- HLA genes were diplotyped using pb-StarPhase, which uses small variant calls and aligned reads to identify the most likely set of star alleles.
- pb-StarPhase requires reads to span at least 50% of the defined HLA regions, samples with higher mean read length produced more complete HLA diplotyping.

		Diplotypes of HLA Genes		
Sample ID	Read Length	HLA-A	HLA-B	HLA-C
CF-01	863 bp	./.	*40:431:02/*40:431:02	*07:551N/*07:551N
CF-02	779 bp	./.	./.	./.
CF-03	1,306 bp	*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 bp	*01:01:118/*01:01:118	*08:01:01:26/*35:03:29	*04:01:01:179/*07:01:01:92
CF-05	729 bp	./.	*08:329Q/*08:329Q	./.

Table 3. HLA diplotyping is possible when read lengths are high enough to phase the defined HLA region with pb-StarPhase (https://github.com/PacificBiosciences/pb-StarPhase).

Conclusion

Covaris extraction combined with PacBio Ampli-Fi and Kinnex library preparation enables high-quality HiFi sequencing from FFPE samples

- Kinnex library construction substantially improves sequencing performance compared to standard approaches
- The workflow generates high-yield, high-quality data across diverse tumor types and DNA quality levels
- Enables comprehensive detection of structural variants, small variants, and haplotype phasing from archival FFPE tumor samples
- Unlocks FFPE tissue for long-read genomic analysis in cancer research
- FFPE samples represent the largest repository of clinically annotated tumor material
- This workflow enables long-read sequencing of previously inaccessible samples, providing a more complete view of cancer genome structure and variation

Questions and comments about the PacBio workflow can be directed to **cconner@pacbio.com**
Questions and comments about the Covaris workflow can be directed to **gendress@covaris.com**

Acknowledgements

The authors thank the PacBio and Covaris teams for sample preparation, sequencing, and analysis support.