SMRT® Sequencing of DNA and RNA Samples Extracted from Formalin-Fixed and Paraffin Embedded Tissues Using Adaptive Focused Acoustics by Covaris

Primo Baybayan[1], Michael Weiand[1], Kevin Eng[1], Guillaume Durin[2], Brian Packard[2], Anand Sethuraman[1], Lawrence Hon[1] and Steve Kujawa[1]
Pacific Biosciences, 1380 Willow Road, Menlo Park, CA 94025, [2]Covaris Inc., Woburn, Massachusetts 01801

Abstract

Recent advances in next-generation sequencing have led to an increased use of formalin-fixed and paraffin-embedded (FFPE) tissues for medical samples in disease and scientific research. Single Molecule, Real-Time (SMRT®) Sequencing offers a unique advantage for direct analysis of FFPE samples without amplification. However, obtaining ample long-read information from FFPE samples has been a challenge due to the quality and quantity of the extracted DNA. FFPE samples often contain damaged sites, including breaks in the backbone and missing or altered nucleotide bases, which directly impact sequencing and target enrichment. Additionally, the quality and quantity of the recovered DNA vary depending on the extraction methods used.

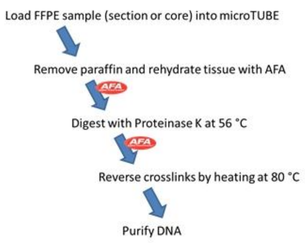
We have evaluated the Covaris® Adaptive Focused Acoustics (AFA) system as a method for obtaining high molecular weight DNA suitable for SMRTbell™ template preparation and subsequent PacBio® RS II sequencing. To test the Covaris system, we extracted DNA from normal kidney FFPE scrolls acquired from the Cooperative Human Tissue Network (CHTN), University of Pennsylvania. Damaged sites in the extracted DNA were repaired using a DNA Damage Repair step, and the treated DNA was constructed into SMRTbell libraries for sequencing on the PacBio System. Using the same repaired DNA, we also tested the efficiency of PCR in amplifying targets of up to 10 kb. The resulting amplicons were also constructed into SMRTbell templates for full-length sequencing on the PacBio System.

We found the Adaptive Focused Acoustics (AFA) system by Covaris to be effective. This system is easy and simple to use, and the resulting DNA is compatible with SMRTbell library preparation for targeted and whole genome SMRT Sequencing. The data presented here demonstrates feasibility of SMRT Sequencing with FFPE samples.

Extraction Workflow

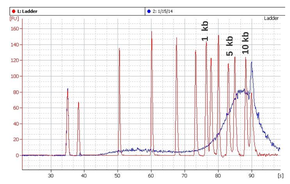
Nucleic acid extraction using the truXTRAC FFPE Kits with Adaptive Focused Acoustics (AFA™) from Covaris, Inc. have made genomic DNA and RNA extractions from FFPE blocks relatively efficient and simple. Multiple samples can be processed to generate high quality nucleic acids suitable for amplification and subsequently, sequencing in the PacBio RS II. The efficiency in recovery is attributed to the effectiveness of the AFA in paraffin removal and tissue hydration, while minimizing further damage to the nucleic acid. The extraction workflows are similar with a minor difference in the incubation time during Proteinase K digestion (figure 1).

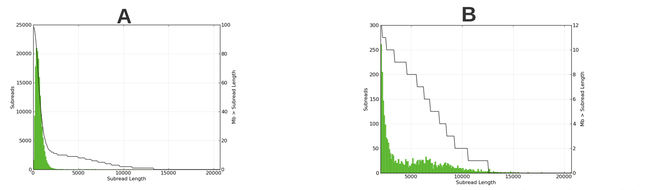
Figure 1. Genomic DNA and RNA extraction workflow. Yield and quality depend on factors such as fixation time, wax to tissue ratio, tissue type and age of FFPE block



Single Molecule, Real-Time Sequencing

Figure 2. Bioanalyzer trace of DNA extracted from normal liver FFPE (CHTN) with a mode of ~ 8 kb, ideal for sequencing in the PacBio RS II. Yield from 5-10 mg scrolls ranges from 1-2 µg DNA.





Mean subread length ~ 700 bp

Mean subread length ~ 5 kb

Figure 3. Sequencing of a FFPE SMRTbell library (A). Re-analysis of data (B) shows subread lengths > 15 kb. A size selection step may be necessary to generate more >5 kb reads.

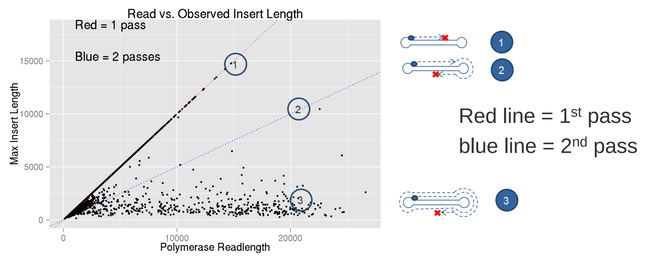
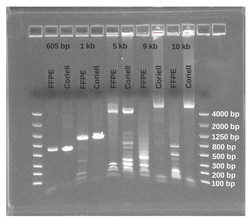


Figure 4. Analysis of sequencing termination events on FFPE SMRTbell templates. Terminations may be caused by residual DNA/Protein crosslinking or possible irreversible damages. Additional optimizations in extraction, library construction and sequencing may be necessary to generate longer subread lengths.

Targeted Sequencing



	Forward Sequence	Reverse Sequence	Size (bp)
BRCA2_9169	AGCCTCTTAACTCTAATTCAAGGAC	ACCATACCTATAGAGGGAGAACAGA	605
BRCA1_729436	TGAGTTCCATCAAGGTGCTTACA	TGTCTAAGAACACAGAGGAGAATTTTA	1,003
HLA-C_F2	TGCTTAGATGTGCATAGTTCCGGAA	TGGACCCAATTTTTACAAACAAATA	5,000
PTGS2_01a_F	GCTTGCAAACCTAACCCCATCAACAACAGAAGAAC	CGCAGTACAGAAGTATCACAGGCTTCCATTGAC	92,88
ALOX5_04a	GCAAAGTTGGCCCGAGATGACCAAATTCAC	GCTTGTACCGTGGGTATCATCATCCCTGTC	10,178

Primer Sequences

Figure 5. FFPE vs Coriell DNA, Amplifications of 605 bp (BRCA) 1 kb (BRCA), 5 kb (HLA), 9 kb (Alox5) and 10 kb (PTGS2), using Clontech's PrimeSTAR® GXL DNA Polymerase.

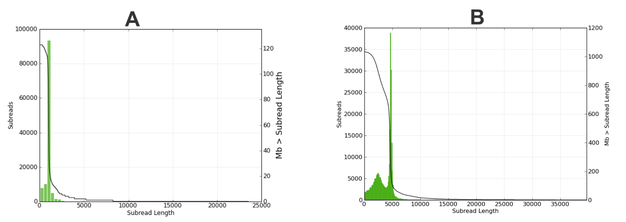


Figure 6. Full-length sequencing of 1 kb and 5 kb PCR products using P6 polymerase and C4 chemistry. The major peaks represent full-length reads of 1 kb (BRCA) and 5 kb (HLA class II) amplicons.

FFPE RNA and Iso-Seq™ Analysis

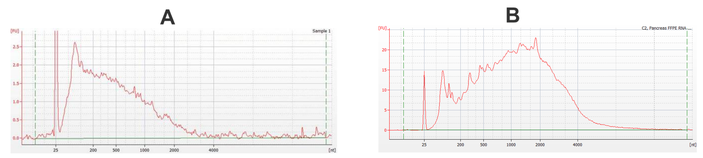


Figure 7. FFPE RNA isolated from Prostate (A, 28 months old) and Pancreas (B, 22 months old) FFPE from BioServe Biotechnologies.

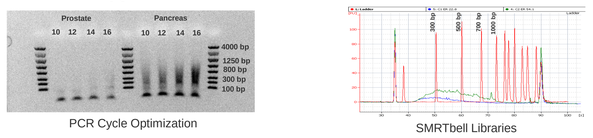


Figure 8. cDNA Library Distribution. Transcripts were captured using a poly-A based method (Clontech® SMARTer® PCR cDNA Synthesis Kit) and amplified using KAPA™ HiFi PCR Kit, and subsequently constructed to SMRTbell libraries.

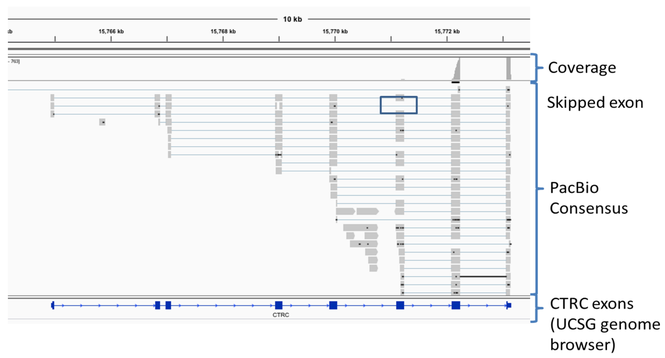


Figure 9. Consensus of a *de novo* assembly of PacBio data from pancreas FFPE RNA. One of the full-length isoforms detected (CTRC, chronic pancreatis) shows a skipped exon. The partial length isoforms show 3' ends but missing 5' ends presumably due to damage during the fixation process.

Conclusions

- Genomic sequencing of DNA from FFPE using truXTRAC and Adaptive Focused Acoustics by Covaris, Inc. is compatible with single molecule sequencing.
- Up to 15 kb subreads can be generated from the CHTN normal liver FFPE DNA. Read length highly depends on the quality of genomic DNA .
- For targeted sequencing, the extraction method generates good quality and large DNA fragments suitable for amplification of targets up to 5 kb for full-length sequencing in the PacBio RS II system.
- The Iso-Seq method allows for the detection of novel isoforms from FFPE RNA with no assembly required. Detection of full-length isoforms may be impacted by the degree of degradation.