SMRT® Sequencing of DNA and RNA Samples Extracted from Formalin-Fixed and Paraffin-Embedded Tissues

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

Introduction

Recent advances in next-generation sequencing have led to the 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 in that it allows 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. DNA samples extracted from FFPE often contain damaged sites, including breaks in the backbone and missing or altered nucleotide bases, which directly impact sequencing and amplification. Additionally, the quality and quantity of the recovered DNA also vary depending on the extraction methods used.

We have evaluated the Adaptive Focused Acoustics (AFA™) system by Covaris as a method for obtaining high molecular weight DNA suitable for SMRTbell template preparation and subsequent single molecule sequencing. Using this method, genomic DNA was extracted from normal kidney FFPE scrolls acquired from Cooperative Human Tissue Network (CHTN), University of Pennsylvania. Damaged sites present in the extracted DNA were repaired using a DNA Damage Repair step, and the treated DNA was constructed into SMRTbell libraries suitable for sequencing on the PacBio RS II System. Using the same repaired DNA, we also tested PCR efficiency of target gene regions of up to 5 kb. The resulting amplicons were constructed into SMRTbell templates for full-length sequencing on the PacBio RS II System. We found the Adaptive Focused Acoustics (AFA) system combined with truXTRAC™ by Covaris to be effective and efficient. 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 single molecule sequencing of DNA samples extracted from tissues embedded in FFPE.

Extraction Workflow

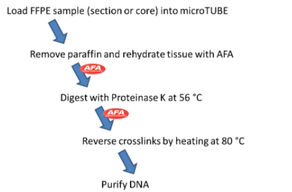

Figure 1. DNA and RNA extraction workflow using truXTRAC. Quality and yield depend on factors such as FFPE fixation time, wax to tissue ratio, tissue type and age of FFPE block.

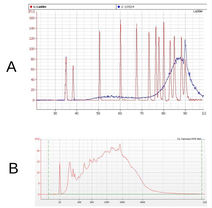

Figure 2. Bioanalyzer traces of genomic DNA (A) and total RNA (B) from FFPE. The genomic DNA shows large fragments ideal for amplification. Yield from a 5-10 mg scroll ranges from 1-2 µg DNA. The total RNA was extracted from 22 month-old Pancreas FFPE (BioServe Biotechnologies).

Amplification of FFPE DNA

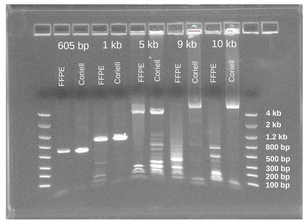

Figure 3. Examples of genes amplified from FFPE DNA. 1 kb targets can be amplified routinely using PrimeSTAR® GXL Kit (Clontech). There is evidence that 5 kb targets can also be amplified successfully with adjustments to the total input and amplification cycling parameters.

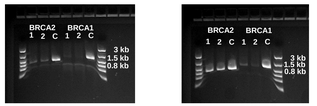

Figure 4. Treatment of DNA with damage repair enzymes (preCR® Repair mix, NEB) prior to amplification is highly recommended for increased yield and efficiency. A is untreated, B is treated with repair enzymes.

	Forward Sequence	Reverse Sequence	Size (bp)
BRCA2_9169	AGCCTCTTAACTCTCAATTCAAGGAC	ACCATACCTATAGAGGGAGGAGACAGA	605
BRCA1_729436	TGAGTTCCATCAAGGTGCTTACA	TGTCTAAGAACACAGAGGAGAATTTTA	1,003
HLA-C_f2	TGCTTAGATGTGCATAGTTCCGGAA	TGGACCCAATTTTACAAACAAATA	5,000
PTGS2_01a_F	GCTTGCAAACCTAACCCCATCAACAAGAAGAAC	CGCAGTACAGAAGTATCACAGGCTTCCATTGAC	92,88
ALOX5_04a	GCAAAGTTGGCCCGAGATGACCAAATTCAC	GCTTGTACCGTGGGTATCATCATCCCTGTC	10,178

Table 1. Primer pairs used in the evaluation

Sequencing of PCR Products

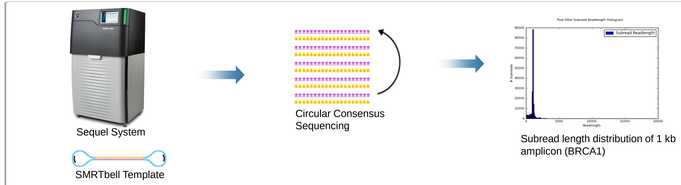

Figure 5. SMRT Sequencing workflow. Amplified products are constructed to SMRTbell templates and sequenced to generate full-length, high-accurate consensus sequence.

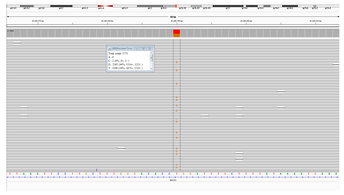

Figure 6. Representative PacBio CCS reads aligned to hg19 showing a G/T SNP from BRCA1. SMRT Sequencing provides full-length sequencing of PCR products without the need for assembly. Long reads also allow SNP phasing that short-read technologies can't.

Sequencing of Genomic DNA

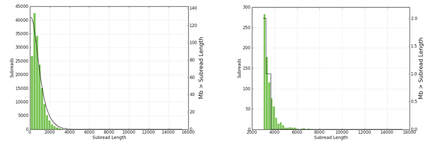

Figure 7. Mean Subread length of a SMRTbell library constructed from FFPE DNA. In this example, the mean subread length is 700 bp (A). When subreads <2 kb are excluded from the analysis, mean subread length increases to 5 kb (B) suggesting presence of large DNA fragments extracted from the Adaptive Focused Acoustics (AFA) system / truXTRAC method.

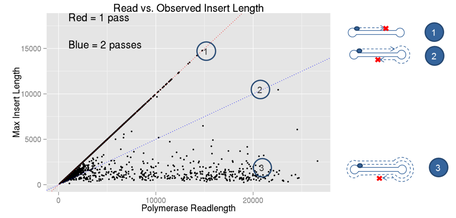

Figure 8. Analysis of sequencing terminations in FFPE DNA. There is evidence of terminations in the SMRTbell backbone, possibly due to residual DNA/protein crosslinking and/or damages to the DNA backbone, irreversible by DNA repair treatment. Additional optimizations in the extraction method may be necessary to minimize sequencing terminations.

Sequencing of RNA

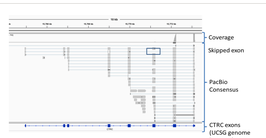

Figure 9. A cDNA library was generated from total RNA extracted from pancreas tissue embedded in FFPE and subsequently constructed to a SMRTbell library for SMRT sequencing (Iso-Seq Method). PacBio reads were assembled *de novo*, showing isoforms of CTRC (Chronic Pancreatis) gene, with one of the isoforms showing a skipped exon.

Summary

- Sequencing of DNA and RNA from FFPE using truEXTRAC and Adaptive Focused Acoustics is compatible with SMRT Sequencing. Read length depends on quality of the DNA and RNA.
- This method extracts large DNA fragments allowing good amplification of > 1 kb targets and possibly up to 5 kb.
- Additional optimizations in the extraction method may be necessary to minimize sequencing terminations due to damage to the DNA.