Application Note

Use of ThruPLEX® DNA-seq Kit with FFPE Tissues

Kamran Shazand, Komal Kunder & John Paul Jerome, Rubicon Genomics, Inc.

Introduction

Formalin-fixed, paraffin-embedded (FFPE) samples are typically collected and preserved from human biopsies for histological studies. The formalin fixation method allows storage of the material at room temperature for many years by cross-linking the cellular contents of nucleic acids and proteins and encasing in wax. Biobanks and researchers worldwide have millions of samples stored. With advances in library preparation and next-generation sequencing (NGS) techniques, nucleic acids preserved in these samples can now be recovered to yield important genomic information. For this reason, archived FFPE samples have become a rich and precious source of study material. Increasingly, research laboratories are utilizing FFPE samples to profile and understand disease-related genes and mutations.

Multiple factors lead to variability in the quality of the DNA that is isolated from FFPE including chemical alterations that occur to the nucleic acids during the preservation process, the non-standardization of the preservation process, and the age of the archived samples. The low quality of the isolated DNA from FFPE samples is expected to generate low complexity NGS libraries, since the "usable" amount of DNA is low, despite apparent concentrations of isolated DNA.

In order to create libraries with maximum diversity, selecting the right library preparation method is critical. In this application note, we demonstrate the performance and features of the ThruPLEX® DNA-seq Kit, our general purpose library preparation kit, when used with FFPE-derived DNA. The input amount was increased to 100 ng (above the recommended input range suggested by the standard protocol) to compensate for damaged DNA in the FFPE samples. Our results showed that ThruPLEX DNA-seq is an excellent choice for preparing NGS libraries from FFPE-derived DNA for Illumina® platforms. Compared to other library preparation kits, it provides more diverse libraries while minimizing GC bias. We further demonstrate how the ThruPLEX DNA-seq Kit can be successfully integrated with Agilent SureSelect® Enrichment Kits to carry out targeted sequencing of FFPE samples.

Materials & Methods

FFPE tissue blocks were obtained from the Cooperative Human Tissue Network (CHTN) from a variety of tissue types and included tumor tissue along with a matching normal control (Table 1). Six months to one year prior to DNA extraction, tissue had been fixed in 10% neutral buffered formalin for 12–24 hours before embedding in paraffin blocks. From these blocks, sections of 20 µm thickness were cut with a microtome, DNA was extracted using QIAamp® DNA FFPE Tissue Kit (Qiagen), and each sample was visualized on a Bioanalyzer® DNA chip (Agilent Technologies), quantitated by Qubit® (Thermo Fisher Scientific), and sheared to an average size of 200 bp using a Covaris® shearing instrument.

Libraries were prepared with 100 ng in triplicate using the ThruPLEX DNA-seq Kit, the NEBNext® Ultra™ DNA Library Preparation Kit (New England Biolabs), or the KAPA® Hyper Preparation Kit (Kapa Biosystems). All kits were used following manufacturer's instructions with the exception that input DNA was increased to 100 ng for the ThruPLEX kit (recommended input range 50 pg – 50 ng). Pooled libraries were sequenced on an Illumina NextSeq® 500 in 2x76 bp paired-end runs. Sequence data was analyzed with the Picard Pipeline[1] on DNANexus. Briefly, reads were aligned to the human genome, hg19, using Burrows-Wheeler Algorithm[2], BWA-MEM (Li and Durbin), to generate BAM files. Reads were then down-sampled to 2.7M read pairs across all samples. Picard Mark Duplicates (Picard) was used to count duplicate reads and estimate

library diversity (Estimated Library Size) and Picard Collect GC Metrics was used to determine biases based on sequence GC content.

Sample	T/N	Source Tissue
1	NAT	Cervix/Uterus
2	Malignant	Colon/Rectum
3	Malignant	Colon/Rectum
4	NAT	Colon/Rectum
5	Malignant	Ovary
6	Malignant	Kidney
7	NAT	Kidney
8	Malignant	Kidney
9	NAT	Kidney

Table 1. Summary of all FFPE samples and their origin obtained from the Cooperative Human Tissue Network (CHTN). (T/N: Tumor/Normal; NAT: Normal Adjacent Tissue)

ThruPLEX DNA-seq libraries were then subjected to an enrichment step using Agilent ClearSeq® Human DNA Kinome panel (~3.2 Mbp) with Agilent SureSelectXT2 target enrichment kit. Briefly, libraries generated with 100 ng (triplicate) from FFPE samples 4, 6, 7, 8, and 9 (15 total libraries) were pooled and used for capture following our SureSelectXT2 integration protocol[3]. The captured libraries were sequenced on the Illumina NextSeq 500 (2x76 mid output reagent kit). A first low pass run of all samples was downsampled to 2.7M reads. Deeper sequencing of a subset (1, 3 and 4) was downsampled to 30M reads to more accurately measure the key metrics. For kinome enrichment data, Picard CalculateHsMetrics was used to determine capture quality metrics after mapping with BWA-MEM.

Results

ThruPLEX DNA-seq produced libraries with greater diversity than either the NEBNext Ultra or the KAPA Hyper kits. On average, ThruPLEX libraries yielded 80% more unique molecules than NEBNext Ultra and 240% more than KAPA Hyper (Figure 1). As expected, ThruPLEX generated less than 2% of duplicate reads, comparable to NEBNext Ultra, while KAPA Hyper showed the highest range of duplicate reads (3-6%) (Figure 1). The insert size of the library molecules were also examined and showed that ThruPLEX DNA-seq produced inserts of 190 bp, approximately equivalent to the KAPA Hyper libraries, but about 27% longer than the 150 bp for NEBNext Ultra (Figure 1). Longer insert size is often advantageous when sequencing paired-end reads, as the non-overlapping reads provide more unique sequence information for a given cluster.

GC coverage of each of the three kits was compared and found to be quite different from those normally observed with non-fixed DNA. Of the three library preparation kits examined, ThruPLEX DNA-seq produced the most balanced GC coverage (Figure 2) from FFPE-isolated DNA.

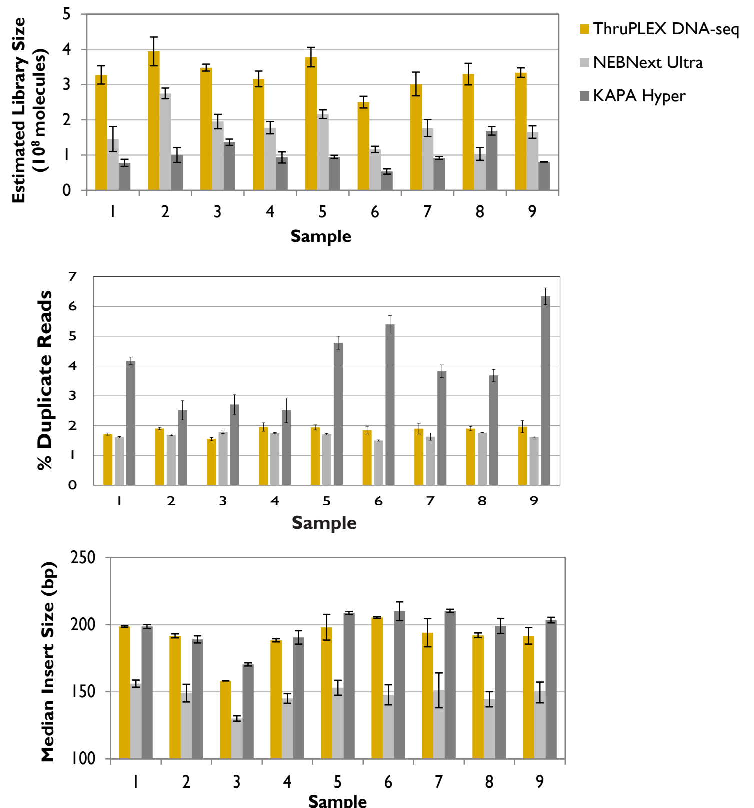

Figure 1. Highly diverse libraries. ThruPLEX DNA-seq libraries provide more unique molecules (top) and fewer duplicates (middle) than other tested kits. Bottom: ThruPLEX creates molecules with an average insert size ~27% longer than NEBNext Ultra.

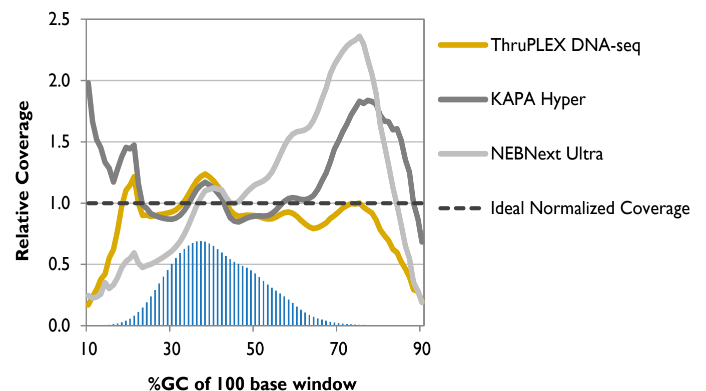

Figure 2. Uniform GC Coverage. ThruPLEX DNA-seq libraries made from FFPE-derived DNA produced consistent coverage across the human genome.

ThruPLEX DNA-seq libraries were then enriched using the Agilent ClearSeq® Human DNA Kinome panel. In each case, the libraries were shown to be of high quality as evidenced by the data shown in Figure 3. Kinome coverage as measured by the depth of sequencing was sample dependent and sequenced at 40-50x coverage of the targeted panels. All samples had on-bait coverage greater than 70%.

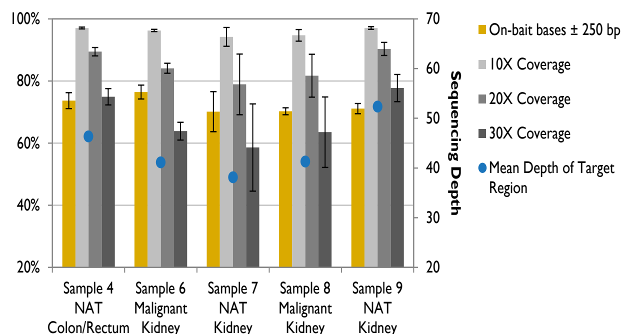

Figure 3. Targeted sequencing metrics of ThruPLEX DNA-seq libraries enriched using the Agilent ClearSeq Human DNA Kinome panel.

Discussion

FFPE samples can be extremely challenging to sequence. The variable degree of chemical damage introduced during the fixation process reduces the amount of "quality" material available to construct a library. To account for this, the standard ThruPLEX DNA-seq protocol was modified by increasing the input amount to 100 ng in this series of experiments. This allowed us to create libraries that were more diverse (80 – 240% superior) with fewer duplicates than other library preparation kits.

ThruPLEX technology uses stem-loop adaptors that obviate the need for intermediate clean-up steps, maintain a single-tube workflow, and preserve the complexity of the libraries. The final amplification makes use of all of the repaired and ligated molecules, resulting in the high number of unique molecules found in the library (Figure 4). As an added benefit, the time to create libraries is reduced to around 2 hours. ThruPLEX DNA-seq libraries can also be easily integrated with other applications such as target enrichment. The high diversity of the input material ensures that capture will yield libraries with excellent coverage throughout the targeted regions of the genome.

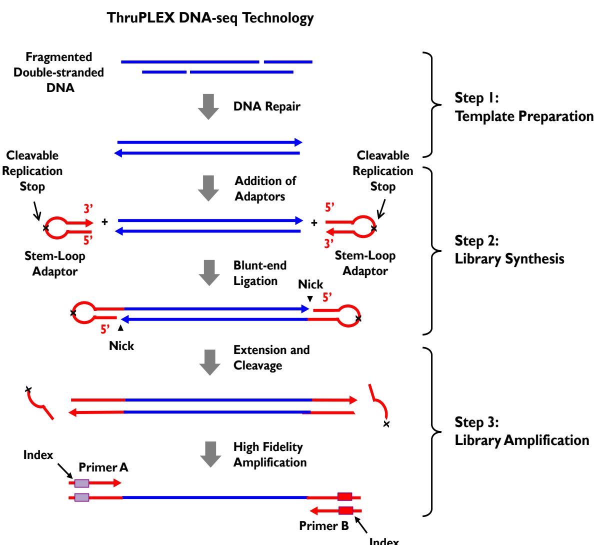

Figure 4. ThruPLEX technology. A three-step, single-tube reaction that starts with fragmented double-stranded DNA (0.05 ng to 50 ng) which is repaired in a highly efficient process. Next, stem-loop adapters are blunt-end ligated to the repaired input DNA. These molecules are extended, then amplified to include barcodes using a high fidelity polymerase to yield an indexed Illumina NGS library.

References

1. Picard - A set of tools (in Java) for working with next generation sequencing data in the BAM format. http://broadinstitute.github.io/picard.

2. Li H, Durbin R. Fast and accurate short read alignment with Burrows-Wheeler transform. Bioinformatics. 2009, 25, 1754-1760.

3. Exome Capture of ThruPLEX® Libraries with Agilent SureSelect® XT2 Target Enrichment system. (153002)

Trademarks

ThruPLEX is a registered trademark of Rubicon Genomics, Inc.
Covaris is a registered trademark of Covaris, Inc.
Qubit is a registered trademark of Thermo Fisher Scientific, Inc.
Bioanalyzer, ClearSeq, and SureSelect are registered trademarks of Agilent Technologies, Inc.
Illumina and NextSeq are registered trademarks of Illumina, Inc.
QIAamp is a registered trademark of Qiagen N.V.
KAPA is a registered trademark of Kapa Biosystems.
NEBNext and Ultra are registered trademarks of New England Biolabs, Inc.

For research use only. ThruPLEX DNA-seq may not be used for any other purposes including, but not limited to, use in diagnostics, forensics, therapeutics, or in humans. It may not be transferred to third parties, resold, modified for resale or used to manufacture commercial products without prior written approval of Rubicon Genomics, Inc. ThruPLEX DNA-seq Kit is protected by U.S. Patents 7,803,550; 8,071,312; 8,399,199; 8,728,737 and corresponding foreign patents. Additional patents are pending.