that's GOOD science!

TECH NOTE

A streamlined solution for generating high-quality NGS libraries from FFPE DNA samples

- High library diversity
- Balanced GC coverage
- High on-bait coverage

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 the sample in wax. Biobanks and researchers worldwide have millions of FFPE 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 lack of 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 the apparent concentrations of isolated DNA.

In order to create libraries with maximum diversity, selecting the right library preparation method is critical. SMARTer ThruPLEX technology uses stem-loop adaptors that obviate the need for intermediate cleanup 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 1). As an added benefit, the time to create libraries is reduced to around two hours. SMARTer 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.

that's GOOD science!

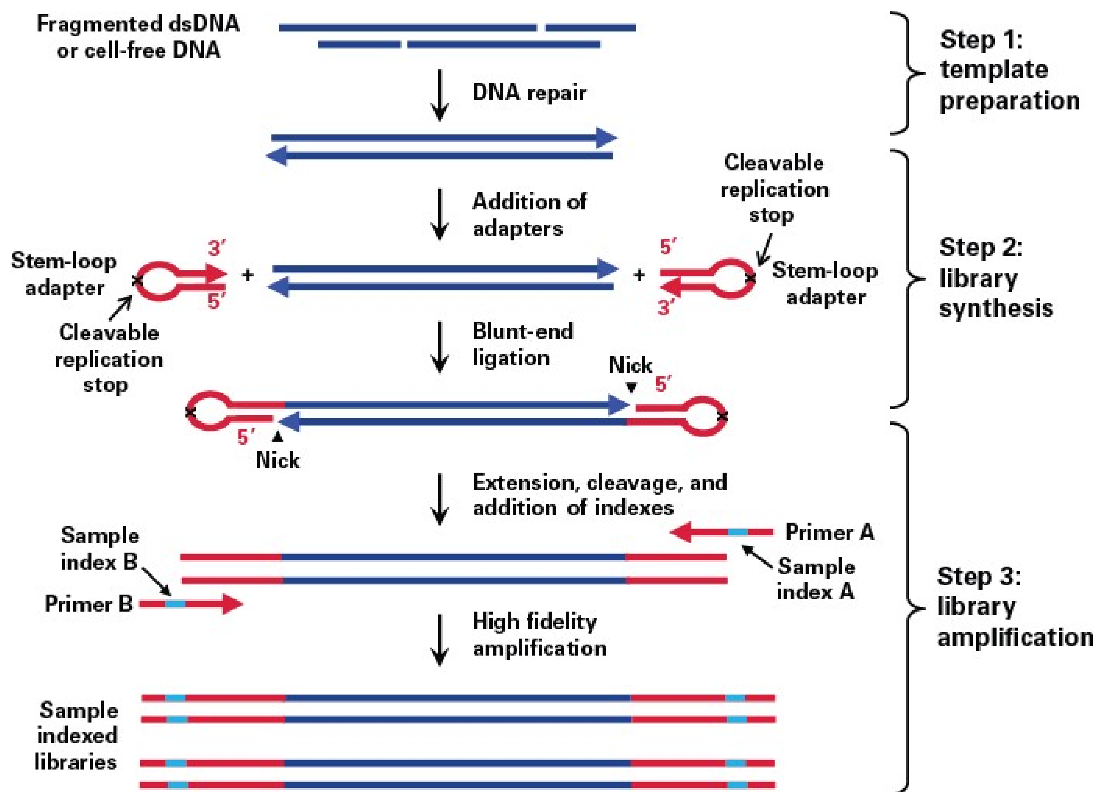

Figure 1. SMARTer ThruPLEX technology. This workflow uses a three-step, single-tube reaction that starts with fragmented double-stranded DNA, which is repaired in a highly efficient process. Next, stem-loop adapters are blunt-end ligated to the repaired input DNA. These molecules are extended, then amplified to include barcodes using a high-fidelity polymerase to yield an indexed Illumina NGS library.

In this technical note, we demonstrate the performance and features of the SMARTer ThruPLEX DNA-seq kit, our general purpose library preparation kit, when used with FFPE-derived DNA. While the recommended input range suggested by the standard protocol for this kit is 0.05–50 ng for non-FFPE samples, here we increased the input amount to 100 ng to compensate for damaged DNA in the FFPE samples. Our results showed that SMARTer ThruPLEX DNA-seq technology 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 SMARTer ThruPLEX DNA-seq kit can be successfully integrated with Agilent SureSelect Enrichment Kits to carry out targeted sequencing of FFPE samples.

Results

SMARTer ThruPLEX DNA-seq technology produces libraries with greater diversity than either the NEBNext Ultra or the KAPA Hyper kits. On average, SMARTer ThruPLEX libraries yielded an estimated library size that is 80% larger than NEBNext Ultra and 240% larger than KAPA Hyper, indicating the presence of a greater number of unique molecules in ThruPLEX libraries (Figure 2). As expected, SMARTer ThruPLEX DNA-seq generated less than 2% of duplicate reads, comparable to NEBNext Ultra, while KAPA Hyper showed the highest range of duplicate reads (3–6%) (Figure 2). An examination of insert sizes of the library molecules showed that the SMARTer ThruPLEX DNA-seq kit produced inserts of 190 bp, approximately equivalent to the KAPA Hyper libraries, but about 27% longer than the 150-bp inserts from NEBNext Ultra (Figure 2). A longer insert size is often advantageous when sequencing paired-end reads, as having more nonoverlapping reads provides a greater representation of the original sequence information.

that's GOOD science!

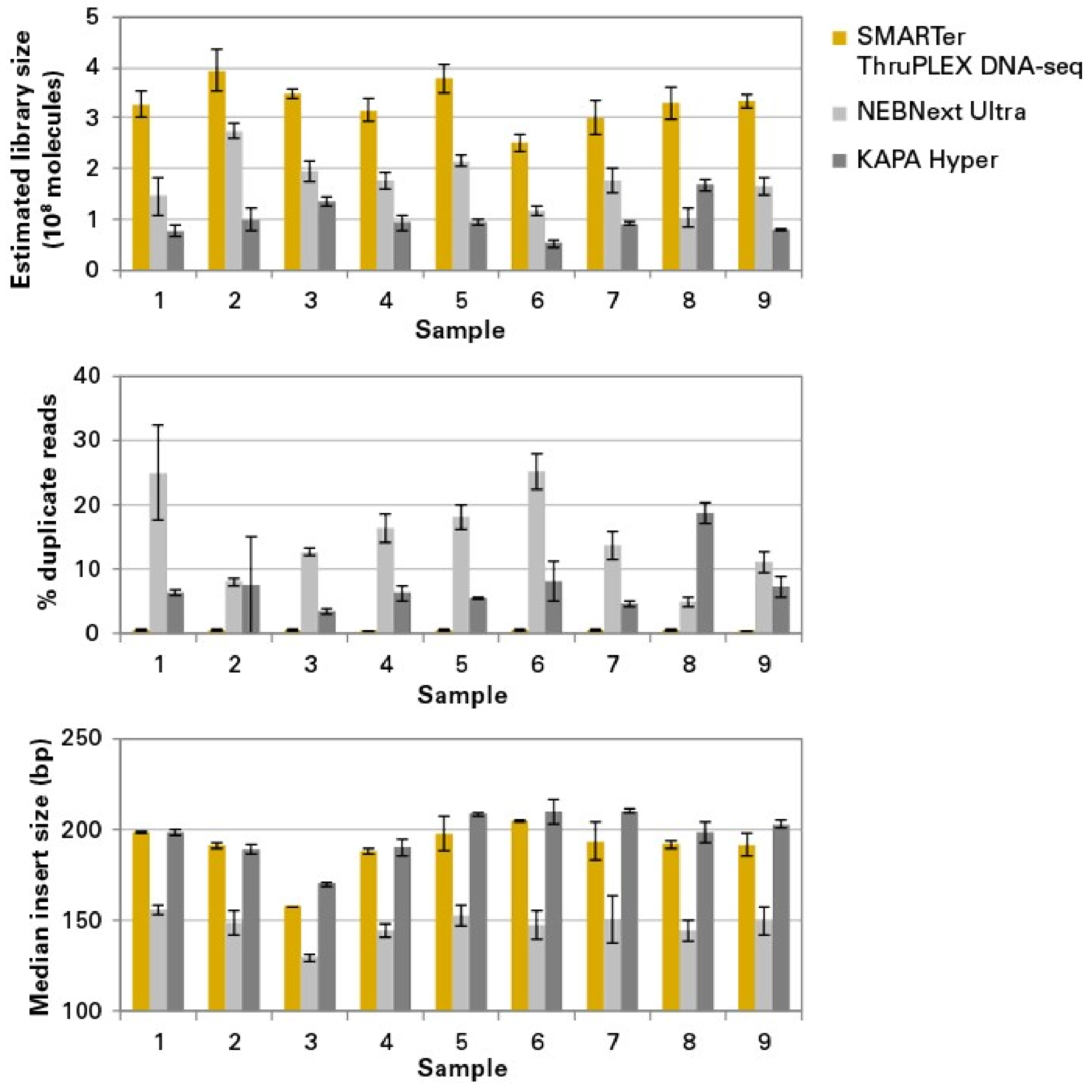

Figure 2. Highly diverse libraries. SMARTer ThruPLEX DNA-seq libraries provide more unique molecules (top) and fewer duplicates (middle) than other tested kits. Bottom: SMARTer ThruPLEX DNA-seq creates molecules with an average insert size that is ~27% longer than NEBNext Ultra.

GC coverage of libraries prepared with each of the three kits was compared. The SMARTer ThruPLEX DNA-seq kit produced the most balanced GC coverage (Figure 3).

that's GOOD science!

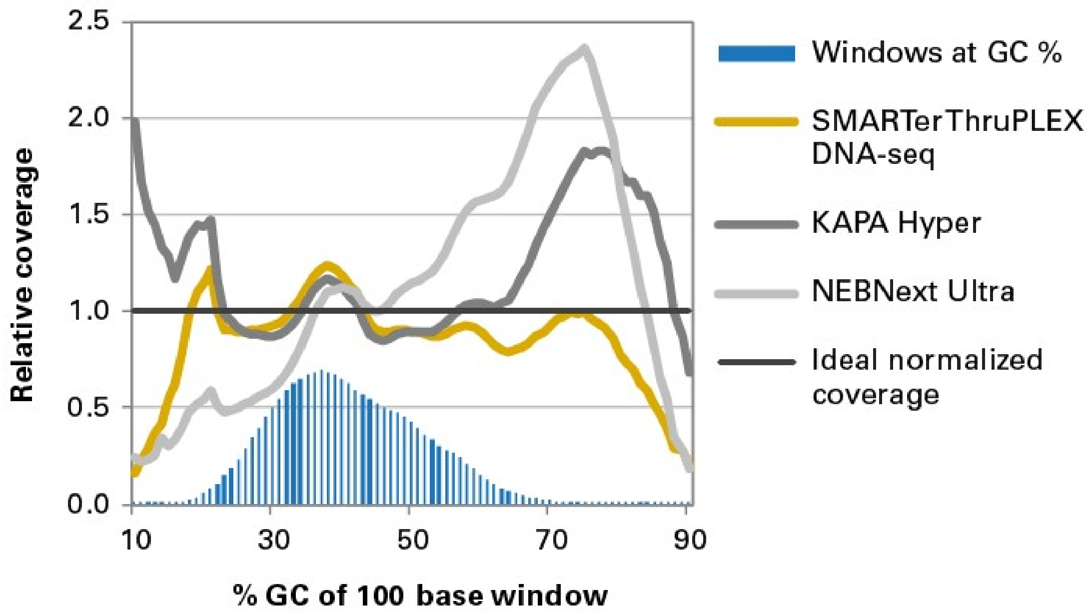

Figure 3. Uniform GC coverage. SMARTer ThruPLEX DNA-seq libraries made from FFPE-derived DNA produced consistent coverage across the human genome.

SMARTer 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 4. 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%.

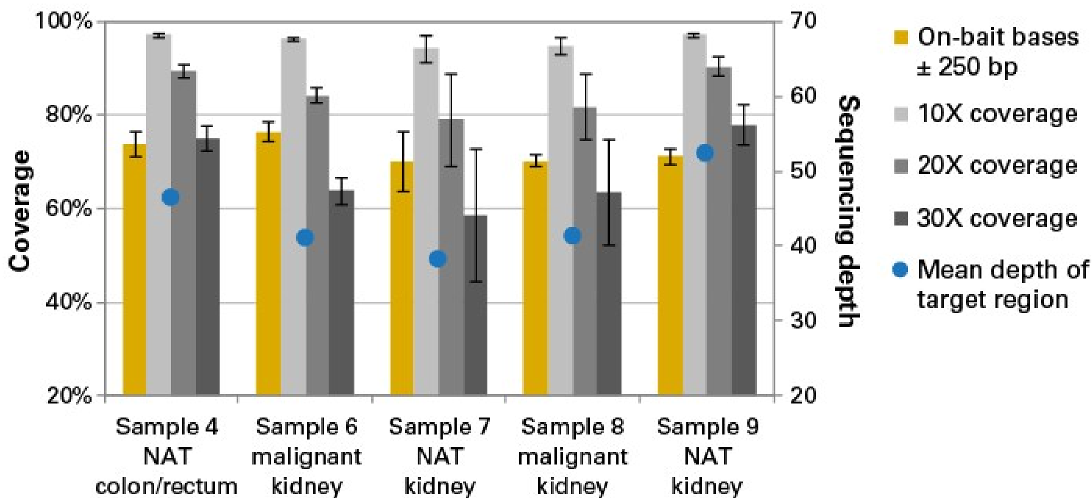

Figure 4. Targeted sequencing metrics of SMARTer ThruPLEX DNA-seq libraries enriched using the Agilent ClearSeq Human DNA Kinome panel.

Conclusions

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 SMARTer 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% higher library complexity) and contained fewer duplicates than other library preparation kits.

Methods

FFPE tissue blocks were obtained from the Cooperative Human Tissue Network (CHTN) from a variety of tissues (see table below) and included tumor tissue (Malignant) along with a matching normal control (normal tissue adjacent to the tumor, or NAT). Six months to one year prior to DNA extraction, the tissue had been fixed in 10% neutral buffered formalin for 12–24 hours before embedding in paraffin blocks. From these blocks, 20-µm thick sections 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.

Sample	Type	Source tissue
1	NAT	Colon/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

Libraries were prepared with 100 ng in triplicate using the SMARTer 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 the manufacturer's instructions with the exception that input DNA was increased to 100 ng for the SMARTer ThruPLEX kit (recommended input range 0.05–50 ng). Pooled libraries were sequenced on an Illumina NextSeq® 500 in 2 x 76 bp paired-end runs. Sequence data was analyzed with the Picard Pipeline1 on DNAnexus. Briefly, reads were aligned to the human genome, hg19, using Burrows-Wheeler Algorithm2, 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.

SMARTer ThruPLEX DNA-seq libraries were then subjected to an enrichment step using Agilent ClearSeq Human DNA Kinome panel (~3.2 Mbp) with the Agilent SureSelect^{XT2} 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 SureSelect^{XT2} integration protocol. The captured libraries were sequenced on the Illumina NextSeq 500 (2 x 76 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.

References

Broad Institute. Picard Tools - A set of command line tools (in Java) for manipulating high-throughput sequencing (HTS) data and formats such as SAM/BAM/CRAM and VCF. <http://broadinstitute.github.io/picard/>

Li, H. & Durbin, R. Fast and accurate short read alignment with Burrows-Wheeler transform. *Bioinformatics* 25, 1754–1760 (2009).

Related Products

Cat. #	Product	Size	License	Quantity	Details
R400406	SMARTer® ThruPLEX® DNA-seq 48D Kit	48 Rxns			
R400427	SMARTer® ThruPLEX® DNA-seq 48S Kit	48 Rxns			
R400523	SMARTer® ThruPLEX® DNA-seq 6S (12) Kit	12 Rxns			
R400407	SMARTer® ThruPLEX® DNA-seq 96D Kit	96 Rxns			
R400428	ThruPLEX® DNA-seq 12S (48) Kit	48 Rxns			

Add to Cart

Takara Bio USA, Inc.
United States/Canada: +1.800.662.2566 • Asia Pacific: +1.650.919.7300 • Europe: +33.(0)1.3904.6880 • Japan: +81.(0)77.565.6999
FOR RESEARCH USE ONLY. NOT FOR USE IN DIAGNOSTIC PROCEDURES. © 2018 Takara Bio Inc. All Rights Reserved. All trademarks are the property of Takara Bio Inc. or its affiliate(s) in the U.S. and/or other countries or their respective owners. Certain trademarks may not be registered in all jurisdictions. Additional product, intellectual property, and restricted use information is available at takarabio.com.