ATCGGCA
TTTGGCC
CGCATCG
GACTACG
NGS

KAPA HyperPlus/SeqCap EZ workflow: Improving Data Quality and Turnaround Times for Targeted Next-Generation Sequencing of FFPE DNA

Formalin-fixed paraffin-embedded (FFPE) tissue is an important source of DNA for cancer genomic studies and clinical diagnostics. A major challenge in high-throughput, targeted next-generation sequencing of FFPE DNA on the Illumina® platform is the ability to process samples of variable quality with predictable success rates in competitive turnaround times. The KAPA HyperPlus Kit with integrated enzymatic fragmentation streamlines the SeqCap™ EZ Target Enrichment workflow, and yields libraries of equal or better quality than those produced from Covaris®-sheared DNA.

Authors

Maryke Appel
Technical Director

Kapa Biosystems
Wilmington, MA, USA

Brian Walker
Senior Scientist

Thomas Jones
Scientist

Elisa Izquierdo Delgado
Scientist

Centre for Molecular Pathology
Royal Marsden Hospital
London, UK

Denise Raterman
R & D Scientist

Heidi Rosenbaum
R & D Scientist

Daniel Burgess
Director of Development

Roche NimbleGen
Madison, WI, USA

INTRODUCTION

In clinical settings, improvements to success rates and turnaround times can have a major impact on patient care. The Centre for Molecular Pathology at the Royal Marsden Hospital previously adopted the KAPA Hyper Prep Kit for library construction in their targeted sequencing workflow, which employs custom Roche® NimbleGen™ SeqCap EZ* capture panels. The highly efficient KAPA Hyper Prep chemistry and streamlined protocol reduced library preparation time, and expanded the pool of FFPE samples that could be sequenced with consistent success rates.

Mechanical fragmentation of DNA remained a major bottleneck. With the Centre's M-series Covaris instrument, this is a time-consuming and labor-intensive process that cannot be scaled or automated. To date, enzymatic fragmentation solutions (including tagmentation-based methods) have not proven to be robust with respect to DNA input and quality, and/or suffer from fragmentation and amplification biases that make it difficult to achieve the coverage depth required to reliably detect low-frequency mutations.[1]

In this study, the KAPA HyperPlus Kit from Kapa Biosystems was evaluated. The kit provides for enzymatic DNA fragmentation and library construction in a single tube. The integrated protocol is not only easy to scale and automate, but eliminates the loss of input DNA associated with the physical transfer of fragmented DNA from Covaris tubes—thereby enabling further improvements to library complexity and overall sequence coverage. Results from validation experiments—performed with the FOrMAT assay, which targets regions of 42 genes associated with gastrointestinal (GI) cancer—confirmed that the KAPA HyperPlus workflow with enzymatic fragmentation produces libraries that are equivalent or better than those generated with the KAPA Hyper Prep Kit from Covaris-sheared DNA. Further improvements to the KAPA HyperPlus/SeqCap EZ workflow were achieved through optimization of the capture library used in the GI assay. Turnaround times were reduced significantly by eliminating a second round of hybridization, without sacrificing coverage depth.

MATERIALS AND METHODS

Experimental Design. Three paired DNA samples from FFPE and peripheral blood were processed in parallel, using a previously validated library construction workflow with Covaris® shearing, vs. the new workflow with enzymatic fragmentation. Sequencing data were compared with respect to the following key metrics: (i) % duplicates; (ii) % unique on-target reads; (iii) level of coverage at 100X, 250X and median depth; and (iv) detection of variants at 5% frequency and above.

Library Construction. FFPE DNA (200 ng) or non-tumor blood DNA (50 ng) was sheared to a mode size of 150 – 200 bp using a Covaris M220 instrument. Libraries were prepared with the KAPA Hyper Prep Kit (Kapa Biosystems) and barcoded SeqCap EZ adapters (Roche® NimbleGen™), according to the previously validated protocol[2]—which includes dual-SPRI® size selection after ligation, and pre-capture amplification with KAPA HiFi HotStart ReadyMix (8 cycles for FFPE, 7 cycles for paired normal controls). In parallel, libraries were prepared from the same samples in a similar manner, but using the KAPA HyperPlus Kit with enzymatic fragmentation for 30 min at 37°C.[3,4] Pre-capture amplification was reduced to 6 cycles for both sample types. Positive control libraries were prepared from 50 ng male control DNA (Promega), and negative control libraries without any input DNA.

Target Capture. Pre-capture libraries were quantified by Qubit™ (ThermoFisher), after which pools were prepared for multiplexed target enrichment with the FOrMAT panel (~200 kb). Each pool consisted of 1 µg of library DNA, divided more or less equally between five libraries of a single type (FFPE or blood). Positive control (20 ng) and NTC (~0 ng) libraries were included in FFPE pools. Hybridization reactions (one per pool) were set up with SeqCap EZ reagents according to the previously validated protocol.[2] With the original GI panel, two rounds of overnight hybridization (≥16 hours at 47°C) and capture were performed,[5] with 5 cycles of amplification with KAPA HiFi after the first round, and 11 cycles of post-capture amplification.

Sequencing and Data Analysis. Final, enriched library pools where quantified by qPCR using the KAPA Library Quantification Kit (Kapa Biosystems). Each pool was diluted to 4 nM, after which one tumor and one normal pool was combined at a ratio of 4:1, denatured and loaded onto an Illumina® MiSeq® instrument. Paired-end sequencing (2 x 150 bp) was performed using a MiSeq v3 chemistry kit (Illumina). Data were processed using a standardized analysis pipeline, which employs *MiSeq Reporter* for initial data processing and output, and *Picard* for de-duplication and generation of sequencing metrics. Reads were aligned against the hg19 human reference genome. Variant analysis was performed with Illumina *VariantStudio* v2.2, and translocations called with *Manta*.

Sample Preparation and Sequencing with Optimized Capture Panel. Version 2 of the GI panel was created by Roche NimbleGen. Libraries were prepared with the new KAPA HyperPlus workflow.[3] Target enrichment was performed as described above, with the exception that the second round of hybridization and capture (and library amplification after the first round of enrichment) was eliminated—reverting to the single-capture protocol outlined in the SeqCap EZ Library SR User's Guide.[6]

RESULTS

Comparative sequencing (Figure 1) and coverage (Figure 2) metrics for libraries prepared from Covaris-sheared DNA—using the previously validated KAPA Hyper Prep/SeqCap EZ workflow vs. the KAPA HyperPlus/SeqCap EZ workflow with integrated enzymatic DNA fragmentation—are given on the next page. These metrics were chosen as they are reflective of library quality and the utilization of sequencing capacity.

The KAPA HyperPlus workflow outperformed the existing workflow in all five metrics, with notable improvements for the FFPE samples. In addition to eliminating a process bottleneck, the KAPA HyperPlus Kit enabled a significant increase in coverage depth for FFPE samples (from the same amount of sequencing), which is critical for the accurate detection of rare somatic variations.

The final requirement for validation of the KAPA HyperPlus/SeqCap EZ workflow was to compare the detection of sequence variants at a frequency above 5% in libraries prepared with the new workflow vs. the existing protocol with Covaris shearing. The results of this comparison are given in Figure 3. For each of the six libraries, a random number of chromosomal positions were interrogated, and the percentage variant frequency vs. the hg19 human reference genome calculated. Correlations between the two methods ranged between 96.7% and 99.6% for the three FFPE libraries, and between 98.1% and 99.8% for the three libraries prepared from blood DNA. This confirmed that the HyperPlus workflow produced libraries that were functionally equivalent to those produced with the previously validated protocol.

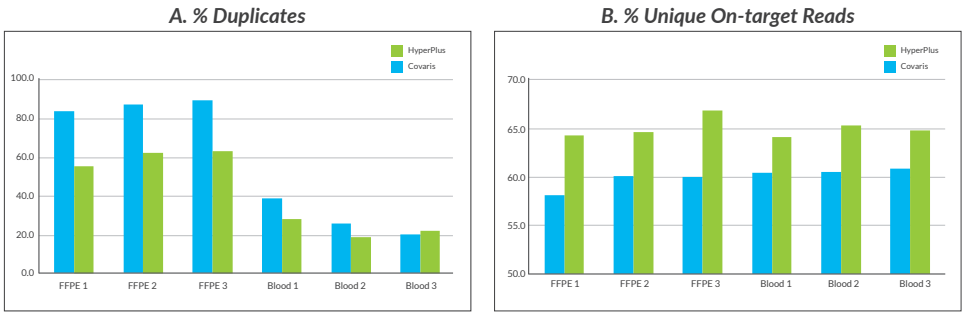

Figure 1. Comparative sequencing metrics. A) Duplication rates were lower for the majority of libraries prepared with the HyperPlus workflow. The average % duplicates for FFPE libraries (60.0%) represented a 26.3% decrease compared to the original protocol. As expected, the improvement in average duplication rate (23.0% vs. 27.8%) was smaller for high-quality blood samples, which have a lower intrinsic duplication rate. **B**) Consistent on-target rates were achieved for all libraries, and across both sample types. With the HyperPlus workflow, the average % unique on-target reads was 65.3% and 64.7% for FFPE and paired normal libraries, respectively. This equates to a 5.9% and 4.1% improvement over the original protocol.

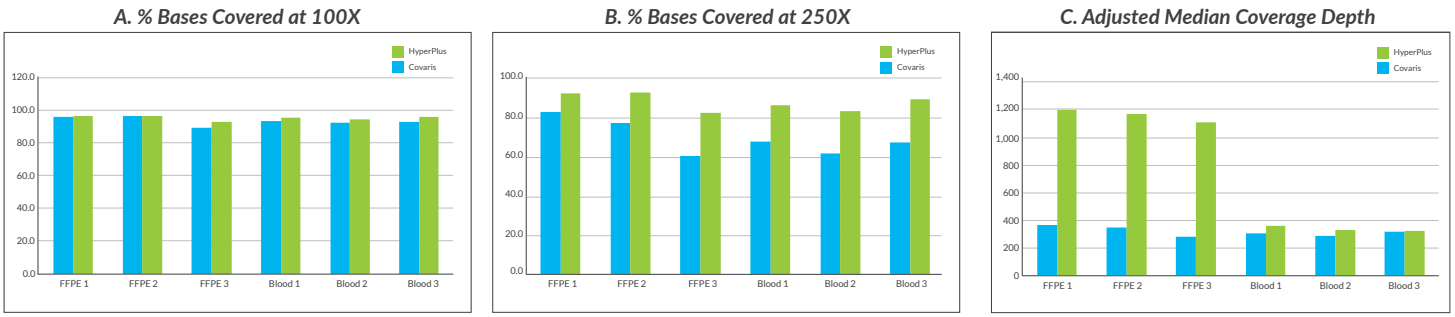

Figure 2. Comparative coverage metrics. A) The level of coverage at 100X coverage depth was high and similar for both protocols and sample types. **B**) At higher coverage depth (250X), the KAPA HyperPlus workflow outperformed the original protocol. The average HyperPlus coverage at 250X was 89.2% and 86.2% for FFPE and paired controls, respectively—as opposed to 73.7% and 65.7% for the original protocol with Covaris® shearing. **C**) Since coverage is impacted by the number of reads obtained for each library in the sequencing pool, the calculated value for each sample was adjusted to reflect the difference in median coverage between the two protocols, should the number of reads have been equal for all libraries. The average median coverage achieved with the HyperPlus workflow for the FFPE samples (1,154X) represents a 3.5-fold improvement over the existing protocol (average median coverage = 328X). The difference in average median coverage for the paired blood samples was much smaller (333X for HyperPlus vs. 302X for the Covaris protocol, or a 1.1-fold improvement).

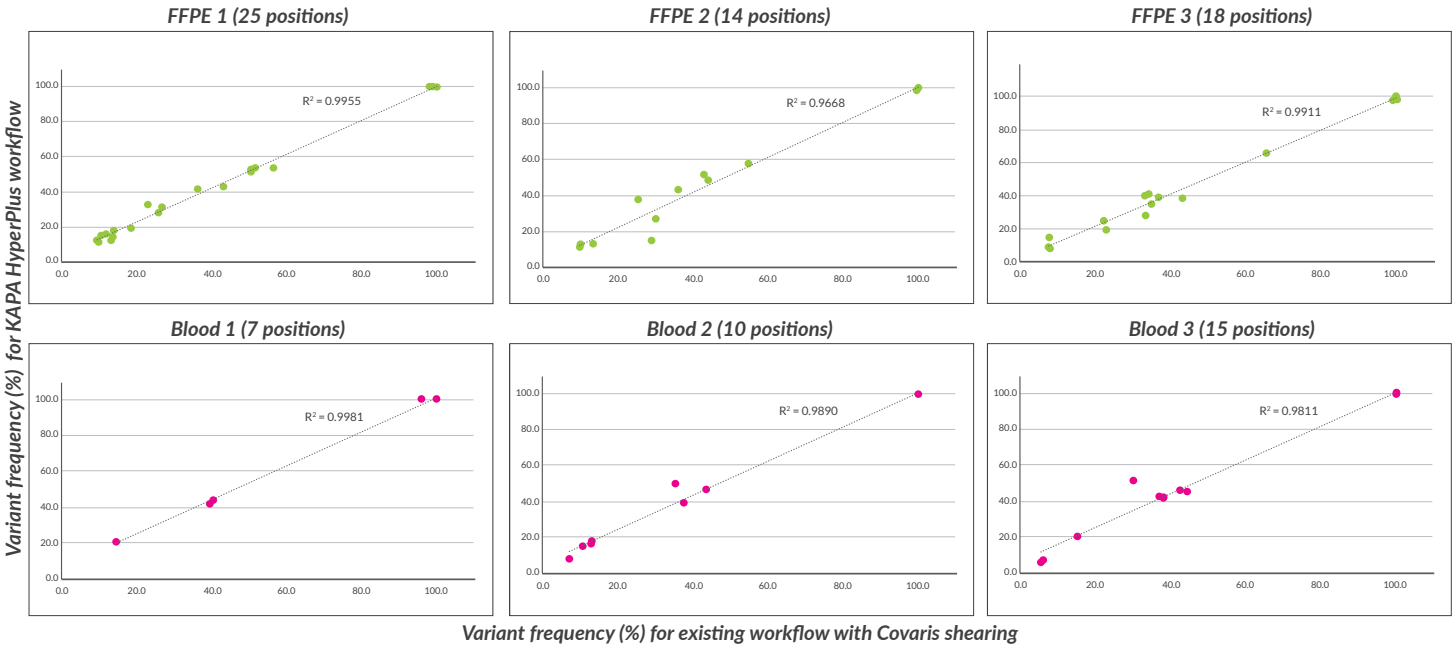

Figure 3. Variant detection correlations. For each of the three FFPE (tumor) libraries (top row) and its normal paired library (prepared from peripheral blood DNA; bottom row), the variant frequency (vs. the hg19 human reference genome) was calculated for a select number of chromosomal positions with expected variant frequencies between 5% and 100%. Each data point represents the calculated variant frequency (%) obtained with the existing workflow (x-axis) vs. the KAPA HyperPlus workflow (y-axis). Regression analysis was performed in Microsoft® Excel.

Elimination of Covaris® shearing reduced library preparation time (fragmentation to pre-capture library) by ~30%. However, the majority (48 hours) of the total turnaround time from tissue to sequence data (~76 hours, excluding data analysis) was still dedicated to the "double-capture" protocol—previously shown to be advantageous for small and challenging capture panels.[5] Optimization of the gastrointestinal cancer target library—through the addition of probes in areas that were historically poorly covered, and removal of probes from non-unique regions—enabled elimination of the second round of hybridization and capture, and the 5 cycles of amplification between the two rounds. This translated to a significant time saving (1 day) with a further decrease in duplication rates (Figure 4A). As expected, a single round of capture with a small panel resulted in lower on-target rates (not shown), but mean coverage depth remained similar as a result of more uniform coverage achieved with the optimized panel (Figure 4B).

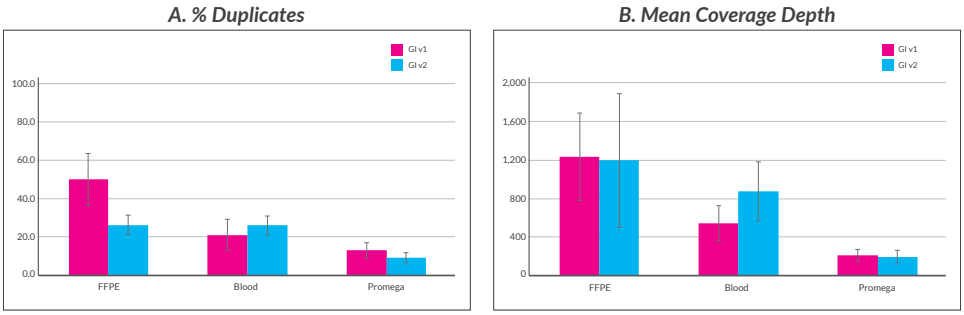

Figure 4. Selected metrics for libraries prepared with an optimized capture panel and capture protocol. Duplication rates **A**) were generally lower with the optimized panel and protocol with less library amplification. Elimination of the second round of hybridization and capture had no significant impact on mean coverage depth **B**) for FFPE samples.

Data represent the averages from seven MiSeq® runs with the original FOrMAT GI cancer panel (GI v1) and the double-capture protocol; and three MiSeq runs with the optimized panel (GI v2) and single-capture protocol. The total number of individual samples in each data set are as follows: GI v1–FFPE (tumor) = 37; Blood (normal) = 28, and Promega positive controls = 7 (1 per MiSeq run); GI v2–FFPE (tumor) = 14; Blood (normal) = 4, and Promega positive controls = 3.

CONCLUSIONS

Implementation of the KAPA HyperPlus Kit in a targeted sequencing assay employing a small, custom SeqCap EZ Target Enrichment panel enabled significant improvements to both turnaround time and FFPE library quality. The robust, single-tube, low-bias HyperPlus chemistry eliminates mechanical shearing, and enables streamlining of the capture workflow through improved library construction efficiency. Overall, sample preparation time (tissue to raw sequence data; analysis excluded) was reduced by ~30%, whilst improving library complexity and coverage depth.

REFERENCES

1. Miller B, et al. (2015). *KAPA HyperPlus—A Single-tube NGS Library Prep Workflow Integrating Enzymatic Fragmentation Results in High Yields and Low Sequencing Bias*. Kapa Biosystems Poster A064.
2. Walker B (2015). *Covaris Targeted Capture of DNA Using KAPA Hyper Prep Kit and NimbleGen™ Baits*, version 1.0. Royal Marsden index number SMD72.
3. Walker B (2015). *Targeted Capture of DNA Using KAPA HyperPlus Kit and NimbleGen Baits*, version 2.0. Royal Marsden index number SMD74.
4. KAPA HyperPlus Technical Data Sheet. Kapa Biosystems document number KR1145.
5. Burgess D, et al. (2012). *Double Capture: An Alternative Protocol for Sequence Capture of Difficult Targets*. Roche® NimbleGen Technical Note 06870406 001.
6. SeqCap EZ Library SR User's Guide, v5.0 (2014). Roche NimbleGen document number 06588786 001.

KAPABIOSYSTEMS

Headquarters, United States
Wilmington, Massachusetts
Tel: 781.497.2933
Fax: 781.497.2934
sales@kapabiosystems.com

International Office
Cape Town, South Africa
Tel: +27.21.448.8200
Fax: +27.21.448.6503
sales@kapabiosystems.com

United Kingdom Office
London, England
Tel: +44.845.512.0641
Fax: +44.203.745.5862
uksales@kapabiosystems.com

*** For life science research only. Not for use in diagnostic procedures.**