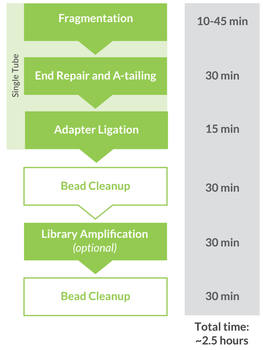
KAPA HYPERPLUS – A SINGLE-TUBE NGS LIBRARY PREP WORKFLOW INTEGRATING ENZYMATIC FRAGMENTATION RESULTS IN HIGH YIELDS AND LOW SEQUENCING BIAS

Bronwen Miller, Maryke Appel, Victoria Van Kets, Beverly van Rooyen, Heather Whitehorn, Martin Ranik, Piet Jones, Adriana Geldart, Rachel Kasinskas and Eric van der Walt

INTRODUCTION

Continuous improvements to library preparation for next-generation sequencing (NGS) are necessary to achieve the highest data quality. One of the crucial steps within library preparation is the initial DNA fragmentation, which can be accomplished through either mechanical or enzymatic processes. Mechanical methods for DNA fragmentation are difficult to scale or automate, and require large investments in expensive instrumentation. Current enzymatic solutions for DNA fragmentation typically exhibit sequence bias, provide poor control over fragment length distribution, and are highly sensitive to input amount.

To address these challenges, we have developed the KAPA HyperPlus Library Preparation Kit by integrating an enzymatic DNA fragmentation technology with fast and efficient library construction to provide a streamlined, easy-to-automate, single-tube solution for preparing NGS libraries from 1 ng – 1 µg of dsDNA.



FRAGMENTATION IS TUNABLE AND REPRODUCIBLE

Existing commercial enzymatic fragmentation solutions for NGS library construction are extremely sensitive to sample type and input amount, but the KAPA HyperPlus Kit provides reproducible results across a range of inputs. Library insert sizes can be adjusted by varying the fragmentation time.

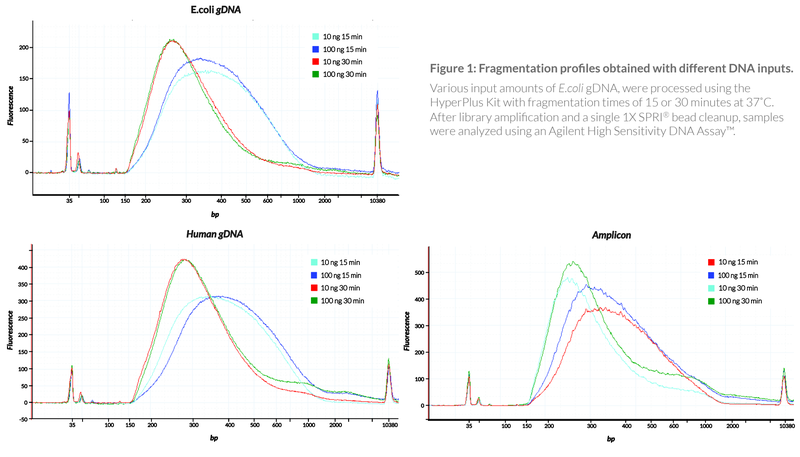
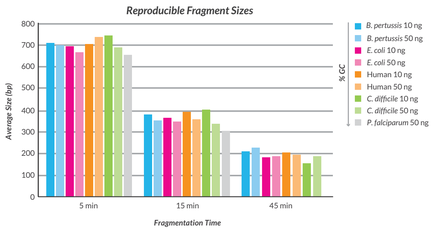


Figure 1: Fragmentation profiles obtained with different DNA inputs.

Various input amounts of *E.coli* gDNA, were processed using the HyperPlus Kit with fragmentation times of 15 or 30 minutes at 37˚C. After library amplification and a single 1X SPRI® bead cleanup, samples were analyzed using an Agilent High Sensitivity DNA Assay™.

Figure 2. Defined fragmentation parameters yield consistent library insert sizes using samples from multiple species across a wide range of GC content.

10 ng or 50 ng of *Bordetella pertussis* (68% GC), *Clostridium difficile* (29% GC), *Escherichia coli* (51% GC), *Plasmodium falciparum* (20% GC) or human gDNA were fragmented for 5, 15 or 45 minutes at 37°C, yielding average insert fragment sizes of ~700, ~350, and~200 bp respectively.

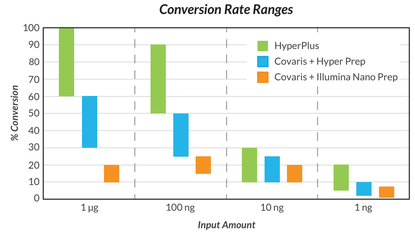


HIGH LIBRARY YIELDS LEAD TO IMPROVED SEQUENCING RESULTS

The HyperPlus workflow results in greater conversion of input DNA to adapter-ligated library. This higher library construction efficiency generates more complex libraries and improved sequencing metrics, including lower duplicate rates.

Figure 3. Conversion rates vary widely for commercial library construction products.

The HyperPlus integrated workflow results in greater conversion of input DNA to adapter-ligated library compared to Covaris-sheared DNA processed using the Illumina® TruSeq™ Nano Prep Kit or the KAPA Hyper Prep Kit. Conversion rates are highest for HyperPlus for both high- and low-input applications.



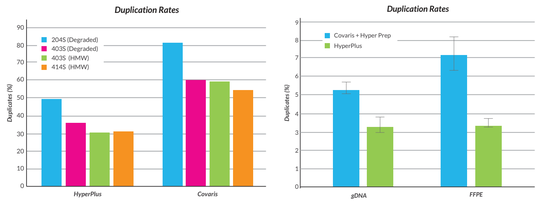


Figure 4. HyperPlus exome capture libraries generate fewer duplicate reads than libraries produced using Covaris.

A. Libraries were prepared from 200 ng of high molecular weight FFPE DNA (403s, 414s) or degraded FFPE DNA (204s, 214s) and captured with the Roche Nimblegen™ double-capture protocol (Data courtesy of Dr. Brian Walker at the Institute of Cancer Research)

B. Libraries were prepared from 50 ng hgDNA or 50 ng FFPE DNA and captured with the Nimblegen SeqCap EZ HGSC VCRome panel.

Roche Nimblegen capture libraries prepared with HyperPlus or Covaris and Hyper Prep were compared to libraries prepared using Nextera™ Rapid Capture. Sequencing metrics show a decrease in off-target reads and an improvement in sensitivity of SNP detection.

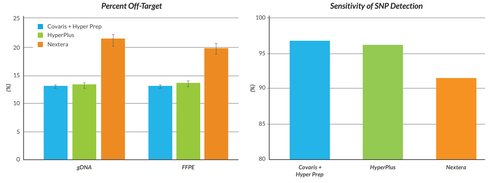


Figure 5: Key sequencing quality metrics demonstrate that HyperPlus performs well for exome capture.

A. Off-target reads were calculated using Picard CalculateHSMetrics.

B. Single nucleotide polymorphisms (SNPs) were called using LoFreq with default settings and a minimum per-base depth of coverage cut-off of 10, and compared to the NA12878 high confidence variant set. Sensitivity of SNP detection (true positive rate) was calculated as true positives divided by the sum of true positives and false negatives. The total number of SNPs analyzed was greater than 25,000.

MINIMAL SEQUENCE COVERAGE BIAS

Bias and coverage uniformity were investigated for libraries prepared for exome sequencing from human DNA.

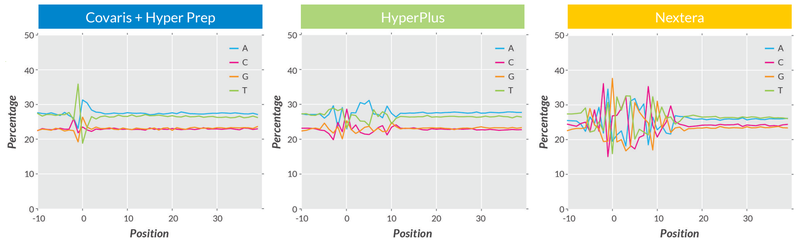


Figure 6: Read start-site bias for human exome libraries.

Nucleotide content over a 30 bp window (-10 to +20 relative to read alignment start) illustrates that start-site complexity of HyperPlus is similar to that of Covaris, and superior to that of Nextera.

For human exome libraries, the HyperPlus workflow shows improved coverage uniformity in comparison to current enzymatic-fragmentation methods and leads to significantly lower AT-dropout, especially for FFPE samples. GC-dropout is similar for all three workflows with high-quality gDNA samples, whereas HyperPlus slightly outperformed Nextera with the FFPE sample.

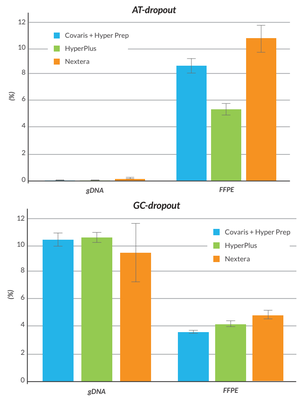


Figure 7: AT- and GC- dropout for gDNA and FFPE samples.

AT- and GC-dropout were calculated using Picard GCBiasSummaryMetrics.

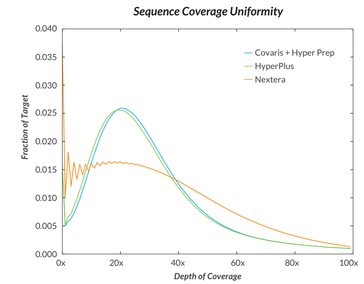


Figure 8: HyperPlus exome sequencing libraries show excellent coverage- depth distribution.

Data for all libraries were down-sampled to equal number of reads. Roche Nimblegen capture libraries prepared with HyperPlus or Covaris and Hyper Prep were compared to libraries prepared with Nextera Rapid Capture. Coverage distribution was similar for Covaris and HyperPlus libraries, while Nextera libraries displayed much broader coverage distribution. A sharper peak and smaller tails indicate more uniform coverage.

GC bias was also investigated in libraries prepared for whole-genome sequencing from microbial DNA. Libraries were prepared from 1 ng of *E. coli*, *B. pertussis* or *C. difficile* gDNA. DNA was either sheared via Covaris to 300 bp and converted to library with the KAPA Hyper Prep kit, or prepared with the KAPA HyperPlus Kit with a 10 min fragmentation time. Additional comparisons were performed with the Nextera kit and the NEBNext® Ultra™ Kit with Fragmentase. HyperPlus shows limited GC bias compared to Covaris, and much less GC bias than Nextera and NEBNext Ultra.

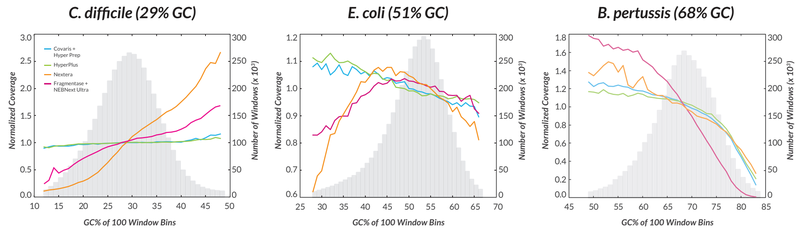


Figure 9: GC bias comparison for microbial samples.

GC bias for *C. difficile* (left), *E.coli* (middle) and *B. pertussis* (right) was assessed by calculating the GC content of the reference in 100 bp bins and plotting normalized coverage across these bins for Covaris and Hyper Prep, HyperPlus, NEBNext Ultra and Nextera workflows using Picard CollectGCBiasMetrics. In the absence of sequencing bias, all bins would be equally represented, indicated by a horizontal distribution centered on a normalized coverage of 1. Distribution of GC content in the genome is indicated by the grey histograms.

CONCLUSION

The KAPA HyperPlus Kit provides a fast and efficient, single-tube enzymatic fragmentation and library preparation solution. This kit combines all the advantages of high-quality, unbiased DNA fragmentation, the speed and scalability of tagmentation, and Kapa's high library yields and low amplification bias in one integrated workflow. This workflow enables the following benefits:

- Automation-friendly DNA fragmentation and library prep in as little as 2.5 hours
- Tunable and reproducible fragmentation that is flexible for DNA sample inputs from 1 ng – 1 µg
- Industry-leading library construction efficiency resulting in reduced bias and maximum sequence coverage
- High-quality results from challenging sample types and applications such as FFPE DNA