Application Note

Shortened hybridization times with KAPA HyperExome Probes

Applications

Whole-exome sequencing, NGS target enrichment, short hybridization times

Products

KAPA HyperPrep Library Preparation Kit; KAPA HyperExome Probes

KAPA HyperExome Probes yield high-quality results with hybridization times as short as one hour using mechanical DNA fragmentation

The rapid growth of targeted next-generation sequencing (NGS) applications is increasing the demand for workflows with short turnaround times. The KAPA HyperCap Workflow v3 is a high-performance, streamlined target enrichment solution that includes an overnight (16 – 20 hours) hybridization step. This overnight step is standard for hybridization-based workflows, and is one of the key improvement areas in a single-day protocol. This Application Note describes the development of a short-hybridization target enrichment workflow using the KAPA HyperCap Workflow v3 and KAPA HyperExome Probes. Target-enriched libraries were created using hybridization times as short as 15 minutes, and the sequencing metrics were compared to libraries created using the standard 16-hour hybridization process. The results show that the sequencing metrics from a 1-hour hybridization are comparable to the 16-hour hybridization workflow, and that hybridization times as short as 15 minutes had only a slight impact on sequencing metrics.

Introduction

The KAPA HyperCap Workflow v3 is a streamlined NGS target enrichment solution that combines high-efficiency KAPA DNA library preparation kits with high-performance KAPA Target Enrichment probes. The workflow accommodates a variety of sample types and starting input amounts into library preparation, and is compatible with mechanical or enzymatic DNA fragmentation using KAPA HyperPrep Kit or KAPA HyperPlus Kit, respectively. Both kits employ KAPA Universal Adapters and Unique Dual-Indexed (UDI) primers, which are designed to ensure high library construction efficiency and low adapter-dimer formation.

The KAPA HyperCap Target Enrichment portfolio includes KAPA HyperExome Probes, a compact, ~43 Mb (capture target) human whole-exome panel that interrogates exonic regions defined by the CCDS, RefSeq, Ensembl, GENCODE, and ClinVar databases, including medical research relevant variants and 387 sample tracking SNPs to facilitate sample identity tracking through the workflow. The HyperCap Workflow v3 takes approximately 1.5 days to complete, from pre-capture library preparation to amplification and clean-up of enriched libraries. A key area of needed improvement in the development of a single-day target enrichment protocol has been the 16- to 20-hour hybridization step, typically performed overnight, during which the target enrichment probes bind to complementary library molecules.

In this study, we tested the feasibility and performance of shorter hybridization times in the KAPA HyperCap Workflow v3 and demonstrated that the KAPA HyperCap Workflow v3 can deliver high-quality results with the KAPA HyperExome Probes within a single day and with performance comparable to the overnight hybridization time. This is achieved by only reducing the hybridization time while all reagents remain the same. *Note: although the results of this pilot study are promising, this short-hybridization protocol is still in development and has not been completely validated, and thus is not fully technically supported by Roche.*

Methods

Experimental design

Libraries were constructed using the KAPA HyperPrep Kit, which utilizes mechanically fragmented DNA. High-quality human gDNA NA12891 (Coriell Institute) was bulk sheared to a mode size of 200 bp using a Covaris ultrasonicator. The fragmented DNA (100 ng) was used as input into automated library preparation, carried out on a Hamilton NGS STAR platform. Libraries were indexed (barcoded) by ligation to KAPA Universal Adapters and subsequent 8 cycles of pre-capture amplification using KAPA Unique-Dual Indexed (UDI) Primer Mixes. The concentration and size distribution of the resulting libraries were determined with the Qubit dsDNA HS Assay kit (Invitrogen) and the Bioanalyzer DNA High Sensitivity Kit (Agilent).

Three replicate target-enriched libraries were prepared for each hybridization time. Each hybridization was performed in a single-plex reaction using 1 µg of library and KAPA HyperExome probes following instructions in the *KAPA HyperCap Workflow v3.0 User Guide*[1] for a capture target size ≥40 Mb. Hybridization was carried out at 55°C for 15 minutes, 1 hour, 4 hours or 16 hours (16 hours is recommended in the *KAPA HyperCap Workflow v3.0 User Guide, February 2020 version*, Chapter 5, Step 3, Sub-step 7 of page 24). All other steps were carried out as described.

Following bead cleanup, the quantity and quality of amplified enriched libraries were assessed using the Qubit dsDNA HS Assay, the Bioanalyzer DNA High Sensitivity Assay, and the qPCR-based KAPA Library Quantification Kit. Libraries were normalized, pooled, and sequenced on an Illumina® NextSeq® 500 sequencing instrument using the NextSeq High Output kit (2 x 75 bp).

Sequencing data were analyzed according to our internal standard analysis pipeline based on the published technical note *How To... Evaluate KAPA Target Enrichment Data.*[2] Data was down-sampled to 50X raw coverage. Normalized coverage across GC content was calculated using BEDTools. BEDTools makewindows was used to create 100 bp sliding windows over the length of the capture area. BEDTools nuc and coverage were respectively used to determine GC content and depth of coverage. A combination of R statistical package, Perl and Bash scripts were used to parse the data for graphing purposes.

Results and discussion

Library QC

The quantity and quality of libraries were assessed before and after target enrichment. All pre-capture libraries achieved an average size of 320 bp (data not shown) with a yield ≥1000 ng (Figure 1A), which is the minimum criteria for proceeding to hybridization. All post-capture library yields were sufficient for sequencing (Figure 1B).

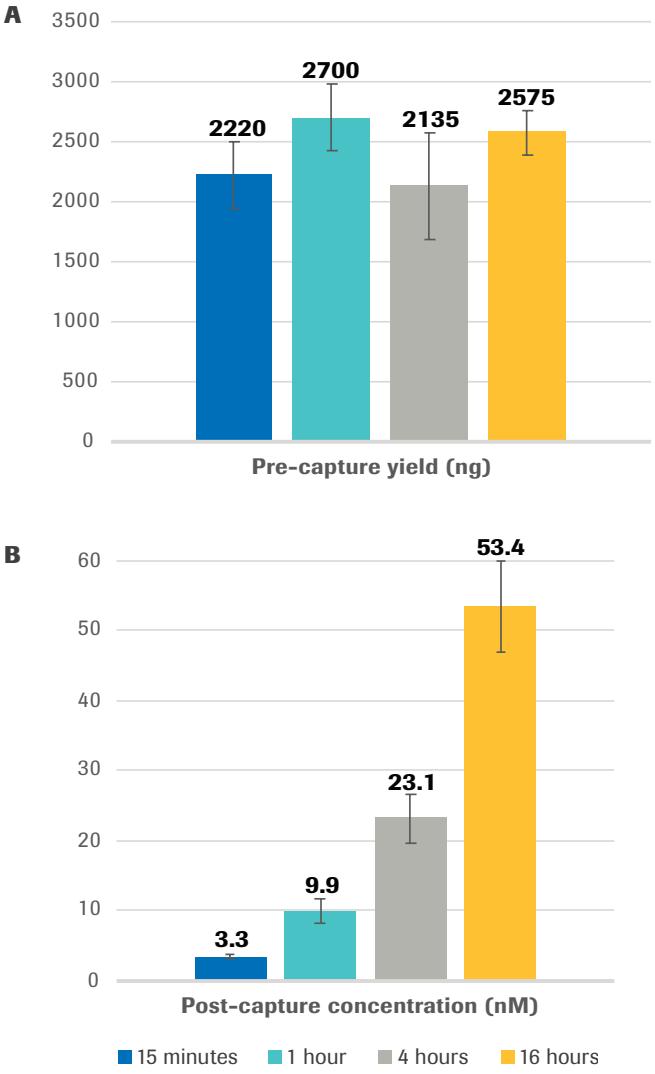

Figure 1. All four hybridization times yielded sufficient library for proceeding to sequencing. (A) Pre-capture yields quantified by Qubit. **(B)** Post-capture concentrations measured by KAPA Library Quantification Kit (qPCR). Bars represent the mean from triplicate libraries and error bars indicate the standard deviation.

Sequencing metrics

The quality of target-enriched libraries generated with each hybridization time was compared using key sequencing metrics: mean coverage, percent of on-target reads, fold-80 base penalty, percent of duplicate reads, and GC coverage uniformity (Figure 2).

The mean coverage over the entire target region was consistent for all hybridization times. Libraries created with shorter hybridization times displayed a lower percentage of bases covered at 20X and 30X. At 50X coverage, performance was similar for all libraries.

The overall efficiency of targeted sequencing is described by two sequencing metrics: on-target rate and coverage uniformity (often reported as fold-80 base penalty). The on-target rate refers to the proportion of reads that fall into the target regions, or the genomic regions covered by the panel. All four hybridization conditions tested in this study yielded a similar percent of reads on-target, with a slightly lower rate for libraries hybridized for only 15 minutes (Figure 3A).

Coverage uniformity describes how evenly the target regions are represented in the sequencing data, and is expressed here as the fold-80 base penalty. This metric is defined as the fold, or amount, of additional sequencing required to achieve the mean coverage level for 80% of the sequenced bases. A score of 1.0 indicates that no further sequencing is required; lower values close to 1.0 indicate better uniformity. All four hybridization conditions tested in this study exhibited low (below 2.0) fold-80 base penalty scores (Figure 3B), with similar scores for all libraries. These results indicate that shorter hybridization times do not substantially impair target enrichment efficiency of the KAPA HyperCap Workflow v3 using KAPA HyperExome.

The level of PCR duplicates in sequencing data indicates library complexity, with low values indicating greater complexity and fewer wasted sequencing reads. In this experiment, the level of PCR duplicates was consistent with the lower post-capture library concentrations observed with samples hybridized for shorter times. (Figure 3C), reflecting a decrease in library complexity as the hybridization time is decreasing. As expected the impact was higher for the 15 min hybridization. However, for 1 or 4 hours the duplication was very similar to the overnight (16 hours) hybridization time.

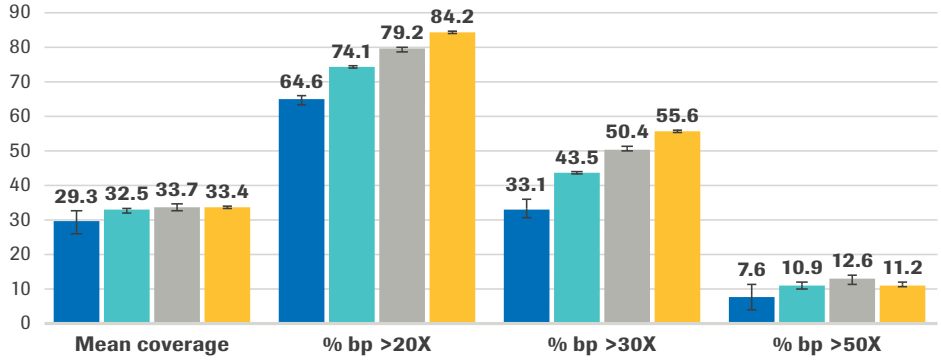

Figure 2. Shorter hybridization times resulted in reduced sequencing coverage at 20X and 30X coverage, but similar coverage at 50X. Bars represent the mean from triplicate libraries and error bars indicate the standard deviation.

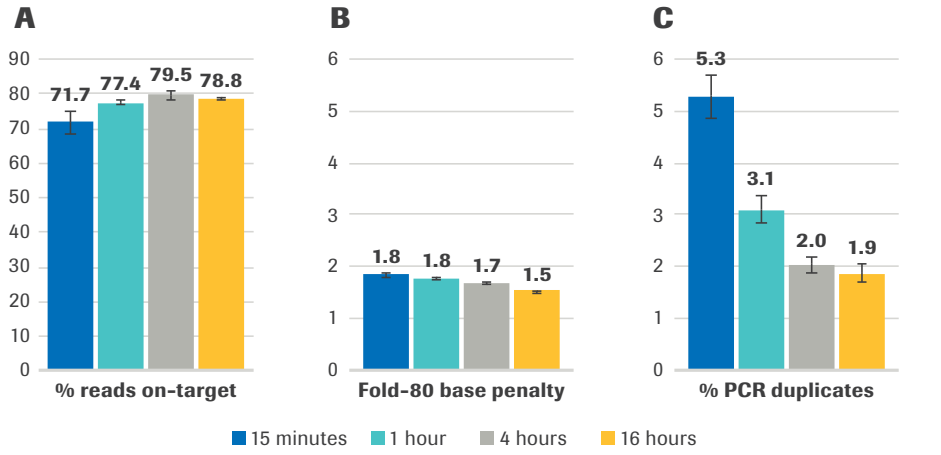

Figure 3. Shorter hybridization times have minimal impact on capture efficiency but can affect library complexity. (A) The percent of on-target reads refers to the percent of mapped, non-duplicate reads overlapping a target region by at least 1 base. **(B)** Coverage uniformity expressed as fold-80 base penalty. **(C)** Percent of duplicate reads, associated to the library complexity. Bars represent the mean from triplicate libraries and error bars indicate the standard deviation.

The KAPA HyperExome Probes yielded highly uniform enriched libraries with minimal GC% bias regardless of the duration of the hybridization step (Figure 4). GC bias describes the relationship between the actual GC content of a region and read coverage across that region; when all regions are equally represented regardless of GC content then the normalized coverage across all regions is 1.0 (Figure 4), and when GC-poor and/or GC-rich regions are under- or over-represented in the sequencing data the normalized coverage is greater or less than 1.0. The normalized coverage of KAPA HyperExome libraries prepared with shorter hybridization times—even as short as 15 minutes—was very similar to the standard 16-hour hybridization over the entire target region.

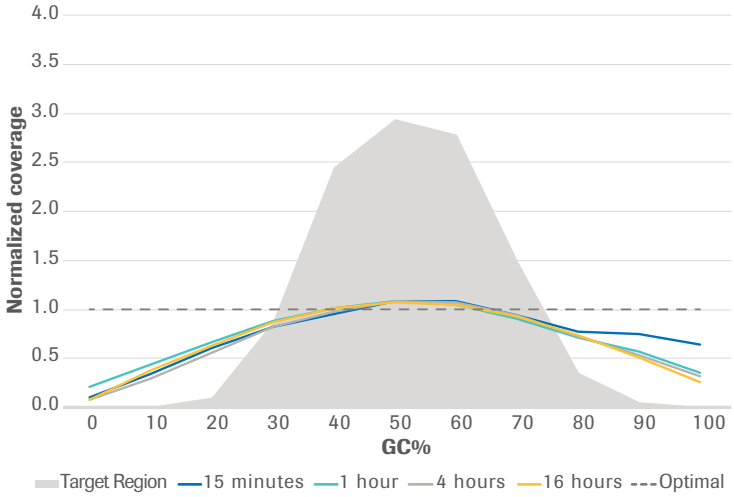

Figure 4. Library uniformity across the GC% spectrum is not affected by reduced hybridization duration. Each curve represents the average of normalized coverage across GC% bins for triplicate libraries. The horizontal line represents the optimal normalized coverage, which would be 1.0 if all sample-to-data steps were completely unbiased. The shaded plot is the GC% distribution of 100 bp windows in the target region.

Conclusion

Target-enriched NGS workflows with shorter hybridization offer the potential for same-day sample-to-sequencer turnaround times—a valuable time-saver for laboratories that routinely perform targeted NGS. The shorter hybridization method described here enables the generation of sequencing-ready target-enriched libraries in a single workday using the KAPA HyperCap Workflow v3 while reagents remain the same. Target-enriched libraries generated with all four hybridization times tested (15 minutes, 1 hour, 4 hours and the standard 16 hours) yielded high-quality sequencing metrics while libraries prepared with shorter hybridization times demonstrated only slightly lower uniformity, reduced coverage and higher duplicate rates than libraries with 16-hour hybridization. Through our initial efforts to reduce the KAPA HyperCap Workflow v3 turnaround time, we have shown comparable performance for the KAPA HyperExome Probes between the standard overnight and shortened hybridization times. This enables the workflow to complete in less than a single day.

References

1. Roche Sequencing Solutions. Instructions for Use. KAPA HyperCap Workflow v3.0. February 2020.
2. Roche Sequencing Solutions. Technical Note. How To Evaluate KAPA Target Enrichment Data. March 2020.

Published by:

Roche Sequencing Solutions, Inc.
4300 Hacienda Drive
Pleasanton, CA 94588

sequencing.roche.com

MC--06374 01/23