Authors

Jochen Seggewiss
Staff Scientist

Christian Ruckert
Bioinformatician

Peter Wieacker
Director

Institute for Human Genetics
University of Muenster
Muenster, Germany

Maryke Appel
Technical Director

Sequencing and Life Sciences,
Roche Diagnostics Corporation,
Wilmington, MA, USA

Best of both worlds: a novel, rapid capture protocol that overcomes drawbacks associated with DNA fragmentation in established methods

Rapid capture protocols are an attractive proposition for laboratories that routinely perform targeted next-generation sequencing (NGS), as they enable faster sample-to-result turnaround times. The fragmentation of input DNA for the construction of pre-capture libraries is a bottleneck in established methods. To address this, we integrated the KAPA HyperPlus Kit with enzymatic fragmentation and the SureSelectXT Fast Target Enrichment System from Agilent® Technologies. The new Kapa/Agilent workflow offers the best of both worlds: the speed and convenience of tagmentation-based rapid capture protocols, but the quality and control of ligation-based library construction for targeted sequencing.

Introduction

Mechanical shearing with a focused ultra-sonicator (Covaris®) is currently the gold standard for the fragmentation of input DNA in NGS sample preparation protocols. Despite offering unbiased fragmentation and good control over fragment size distributions, Covaris shearing has two disadvantages: i) it is very time-consuming and laborious when large sample batches have to be processed with a single-tube instrument; and ii) the cost of a high-throughput instrument is prohibitive for most smaller laboratories. "Tagmentation"-based methods (such as the Nextera® Rapid Capture system from Illumina®, or Agilent's SureSelectQXT system) employ transposases for fast and simple library construction. These protocols are, however, associated with significant sequence bias,[1] especially for low-quality FFPE samples; and are extremely sensitive to DNA input—thus requiring meticulous quantification of viscous, high-molecular weight DNA.

In this study, we integrated two industry-leading technologies: the fast, high-performance KAPA HyperPlus Kit (Kapa Biosystems) and the SureSelectXT Fast Target Enrichment System with 90-minute hybridization (Agilent Technologies). The library construction portion of the original XT Fast ("All-Agilent") protocol (which relies on Covaris shearing), was replaced with the single-tube KAPA HyperPlus chemistry with low-bias enzymatic fragmentation. The new method enables robust and reproducible preparation of sequencing-ready capture libraries in a single workday, from both high-quality DNA and challenging FFPE samples.

Materials and methods

DNA samples. High-quality human genomic DNA (gDNA) was isolated from blood using a MagNA Pure® system (Roche®). FFPE DNA was isolated from FFPE tissue sections using a Maxwell® 16 instrument and Maxwell 16 FFPE Plus LEV DNA Purification Kit (Promega). DNA was quantified with a Qubit® dsDNA HS Assay Kit (ThermoFisher Scientific). FFPE DNA quality was assessed with a TapeStation 2200 instrument, using Genomic DNA ScreenTape and Reagents (Agilent®).

Optimization of DNA fragmentation. Two high-quality DNA samples were used to determine the optimal enzymatic fragmentation time for the SureSelectXT Fast workflow, according to the guidelines provided with the KAPA HyperPlus Kit (KK8523).[2] In short, triplicate reactions containing 50 ng, 100 ng, or 250 ng input DNA were set up for each of the two samples. Each set of six reactions was processed with a different fragmentation time (at 37°C), comprising 5 min intervals across the recommended range of 20 – 30 min. After fragmentation, the rest of the library construction protocol (up to final, pre-capture library) was completed in an uninterrupted fashion. All libraries were amplified for 6 cycles. Library size distributions were analyzed using a TapeStation 2200 instrument, with D1000 ScreenTape and Reagents (Agilent).

Library construction and target enrichment. Libraries were prepared from ≤200 ng high-quality or FFPE DNA using the original SureSelectXT Fast ("All-Agilent") protocol,[3], or the newly developed Kapa/Agilent workflow outlined in Figure 1.

The All-Agilent workflow employs Covaris® shearing; a "conventional", three-step, ligation-based library preparation protocol (~5 hours, inclusive of fragmentation) with Agilent reagents; and 10 cycles of pre-capture amplification with Herculase II Fusion DNA Polymerase (which is also used for post-capture amplification). For the Kapa/Agilent workflow, the streamlined, single-tube KAPA HyperPlus protocol[2] was used for library construction (~1.5 hours, including enzymatic fragmentation), with parameters outlined in Table 1. KAPA HiFi Hotstart ReadyMix and KAPA Library Amplification Primer Mix were used for both pre- and post-capture amplification.

Table 1. Library construction parameters for the Kapa/Agilent workflow

Parameter	High-quality DNA	FFPE DNA
Fragmentation time	22 min	27 min
SureSelectXT Fast Indexed Adapters	Undiluted, 5 µL per reaction	Undiluted, 5 µL per reaction
Pre-capture amplification	4 cycles	9 cycles

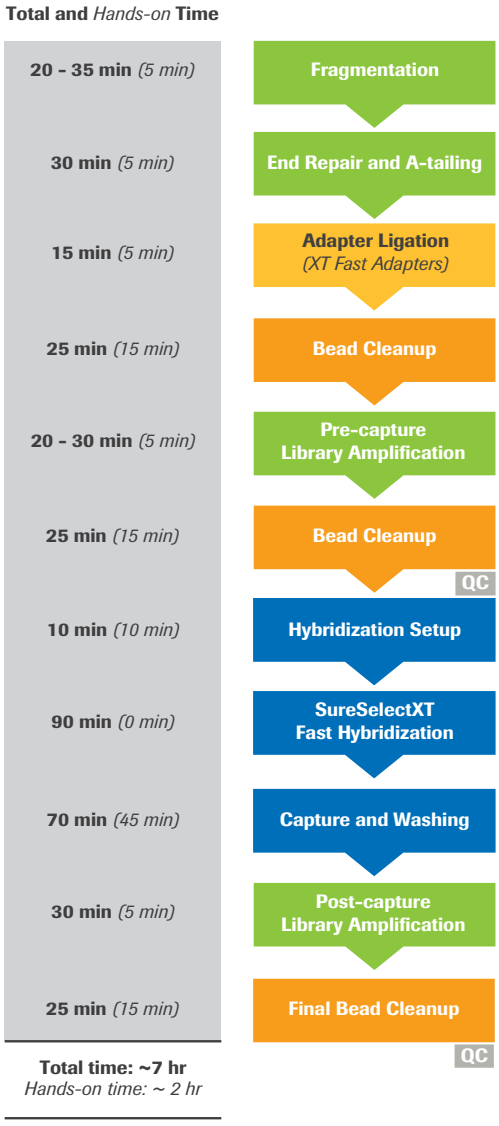

Figure 1. The newly developed Kapa/Agilent workflow. Steps in green are performed with KAPA HyperPlus reagents. Steps in blue are performed with Agilent SureSelectXT Fast reagents. Adapter ligation (yellow) is performed with reagents from the KAPA HyperPlus Kit and SureSelectXT Fast full-length, indexed adapters. All SPRI® bead cleanups are performed with Agencourt® AMPure® XP (Beckman Coulter). Where indicated, QC steps are performed with a Bioanalyzer 2100, or TapeStation 2200 or 4200 instrument (Agilent).

Pre-capture libraries (500 – 750 ng each) were processed individually through target enrichment, performed essentially as described in the original SureSelectXT Fast protocol. For the Kapa/Agilent workflow, a modified streptavidin bead washing step was introduced. After 2.5 min of the 5 min incubation in Wash Buffer 2 (performed at 65°C, in an Agilent SureCycler 8800 thermal cycler), reactions were mixed 10 – 15 times with a multichannel pipette. Reactions were kept in the cycler during this mixing step to avoid temperature fluctuations. After mixing, wells were sealed with fresh strip caps to eliminate the possibility of cross-contamination.

Streptavidin beads were resuspended in slightly different volumes prior to post-capture amplification (20 µL for the Kapa/Agilent workflow, vs. 25 µL for the All-Agilent protocol).

A small, custom SureSelectXT capture panel (172.4 kb), targeting 34 genes associated with risk for breast and ovarian cancer, was used for target enrichment of libraries prepared from high-quality DNA; whereas the SureSelectXT Comprehensive Cancer Panel (788 kb) was used for the enrichment of FFPE libraries. Twelve cycles of post-capture library amplification were used for all libraries, irrespective of the nature of the input DNA or capture panel used.

All pre- and post-capture library amplification reactions were performed in an Agilent SureCycler 8800 thermal cycler using the manufacturer's recommended cycling parameters for each enzyme (Herculase II for the All-Agilent protocol and KAPA HiFi for the Kapa/Agilent workflow). SPRI® bead cleanups for both methods were performed with Agencourt® AMPure® XP (Beckman Coulter), as per instructions provided in the All-Agilent[3] or KAPA HyperPlus[2] protocols. For the Kapa/Agilent workflow, DNA was bound to beads for 10 min (vs. 5 min for the All-Agilent protocol). The concentration of ethanol used for bead washes was 70% and 80% for the All-Agilent and Kapa/Agilent methods, respectively. The bead-to-sample ratio for the final cleanup (after post-capture PCR) also differs for the two protocols: a ratio of 1.8X was used for the All-Agilent protocol, and 1X for the Kapa/Agilent workflow.

Enriched, sequencing-ready libraries were analyzed with a TapeStation 2200 instrument using D1000 ScreenTape and Reagents (Agilent) to confirm library size distributions, and quantified using a Qubit® dsDNA HS Assay Kit (ThermoFisher Scientific).

Sequencing and data analysis. Sequencing pools were created by combining eight enriched libraries in an equimolar ratio. Pools were diluted to 4 nM in low-EDTA TE buffer, for paired-end sequencing (2 x 150 bp) on an Illumina® Miseq® instrument, using v2 chemistry kits. Raw data (FASTQ) were processed with an in-house pipeline. BWA v0.7.12 (mem-M) was used for alignment to the human reference genome (GRCh37.75). SAMtools v1.2 was used to convert SAM files to BAM format. Variant calling for libraries generated with the custom 34-gene panel was done with GATK v3.5 (indel realignment and base recalibration according to best practice recommendations), using a BED file containing all RefSeq exons (±20 bp) of the 34 genes.

Results and discussion

Enzymatic fragmentation of high-quality and FFPE DNA

Non-mechanical, commercial fragmentation methods are notoriously sensitive to the amount of input DNA. To evaluate performance of the enzymatic fragmentation reagent supplied in the KAPA HyperPlus Kit, three amounts of high-quality DNA spanning a 5-fold range (50 – 250 ng) were incubated at 37°C for the same period of time. Three fragmentation times (20 min, 25 min, and 30 min) were tested. Electrophoretic analysis of pre-capture libraries (Figure 2) indicated that: i) mode library insert size was independent of input amount; and ii) a fragmentation time of ~25 min was optimal for SureSelectXT Fast target enrichment. For high-quality DNA, a fragmentation time of 22 min was used in all subsequent experiments.

As recommended,[2] a slightly longer fragmentation time (27 min) was employed for the enzymatic fragmentation of FFPE samples. Post-capture libraries—prepared with the Kapa/Agilent workflow from eight FFPE DNA samples of variable concentration and quality, and enriched with Agilent's Comprehensive Cancer Panel (788 kb)—all displayed highly similar and optimal fragment size distributions (Figure 3). This confirmed that—unlike tagmentation-based library construction methods—the KAPA HyperPlus chemistry is capable of yielding robust and reproducible results with both high-quality and FFPE DNA across a range of DNA inputs.

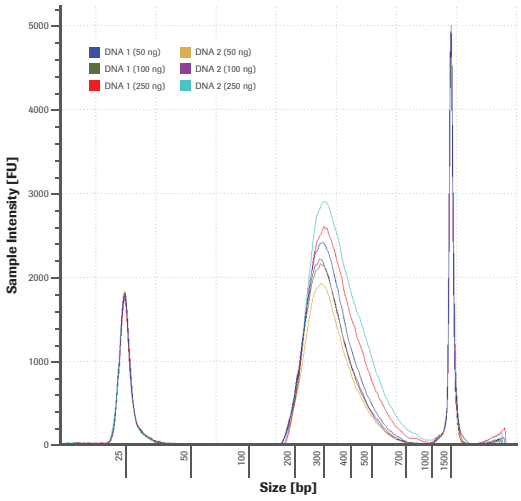

Figure 2. TapeStation profiles of pre-capture libraries prepared from a 5-fold range of input DNA with the KAPA HyperPlus Kit. Three input amounts of two different high-quality DNA samples were fragmented in duplicate, for 20 min, 25 min, or 30 min at 37°C. After completion of the KAPA HyperPlus protocol, all pre-capture libraries generated with a single incubation time displayed a similar fragment size distribution. The 25 min fragmentation time (above) proved to be best for SureSelectXT Fast target enrichment, which requires libraries with a size distribution between ~200 and 700 bp and a peak in the range of 300 – 360 bp.

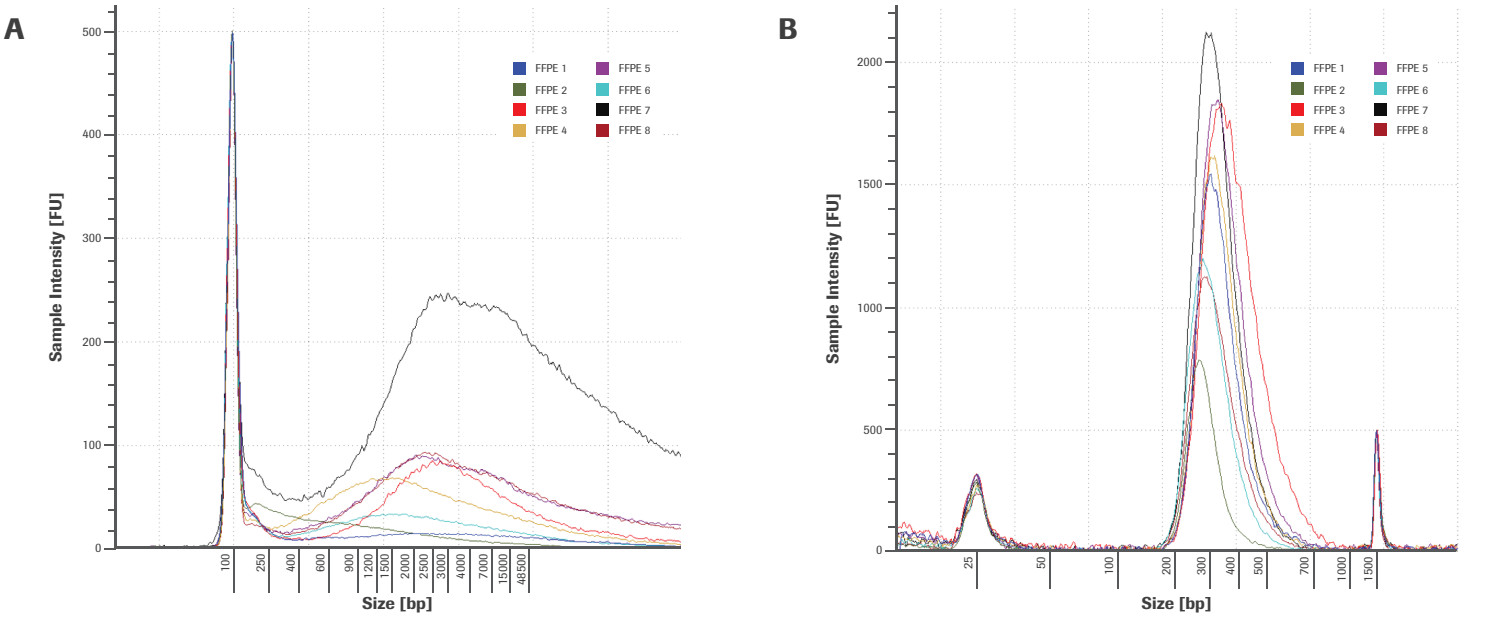

Figure 3: TapeStation profiles for input DNA (A) and sequencing-ready libraries (B), prepared from FFPE DNA of varying quality with the Kapa/Agilent workflow. Libraries were prepared from eight FFPE DNA samples of varying concentrations and degrees of degradation, as indicated in the electropherogram (generated with a Genomic DNA Assay) on the left. Of each sample, ≤200 ng was used for library construction and target enrichment using the Kapa/Agilent workflow and Agilent's Comprehensive Cancer Panel (788 kb). All eight samples yielded sequencing-ready libraries (right) with a similar and optimal fragment size distribution and a concentration sufficient for sequencing. The FFPE sample with the worst quality (FFPE_2; dark green line) previously failed Sanger sequencing.

Comparison of the All-Agilent and Kapa/Agilent workflows

To compare the performance of the original All-Agilent and new Kapa/Agilent workflows, libraries were prepared from four high-quality DNA samples with both workflows, and enriched with the custom 34-gene panel (172.4 kb). Workflows were compared with respect to two key sequencing metrics, namely percentage duplicates and on-target rates. (Figure 4). A high percentage of unique reads (or low % duplicates) is an indicator of high library quality (complexity). On-target rates reflect how efficiently sequencing capacity is utilized, within the context of a specific capture panel and experimental design.

The KAPA HyperPlus Kit employed in the Kapa/Agilent workflow yielded more complex libraries than the original All-Agilent protocol; as a result of higher library construction efficiency, combined with low-bias fragmentation and library amplification. This translated to a 5-fold reduction in duplication rate over the original All-Agilent protocol (Figure 4A). Achieving high on-target rates is a challenge when working with small, custom capture panels. Improvements to capture and wash steps in the Kapa/Agilent workflow resulted in a >1.5-fold improvement in on-target rates (Figure 4B). Improvements in library complexity and on-target rates contribute to higher overall sequence coverage, which is important for the detection of low-frequency mutations.

Performance with FFPE DNA

Hybridization capture methods—such as the SureSelectXT Fast system—have many advantages over amplicon-based strategies for targeted NGS. Amplicon-based strategies are preferred for FFPE samples in cases where available DNA is limited. However, results are often poor and/or inconsistent as DNA degradation/damage affects primer binding sites in a stochastic manner.

Selected mapping and capture statistics for the eight FFPE libraries described earlier (Figure 3) are given in Table 2 and Figure 4 (light green bars). As expected, duplication rates were higher and more variable for the FFPE samples, but on-target rates were better due to the larger capture size.

Table 2: Mapping statistics and duplication rates for FFPE libraries prepared with the Kapa/Agilent protocol and Comprehensive Cancer Panel

FFPE sample #	% mapped reads	% mapped reads inside target	% duplication inside target
4	99.7	64.3	7.4
6	99.8	66.0	10.2
8	99.7	66.5	8.4
5	99.8	70.6	2.8
1	99.8	64.2	12.7
3	99.8	69.8	9.8
7	99.7	69.8	7.4
2	99.7	65.0	28.2

Samples are arranged in order of decreasing quality, based on electrophoretic profiles as shown in Figure 3A.

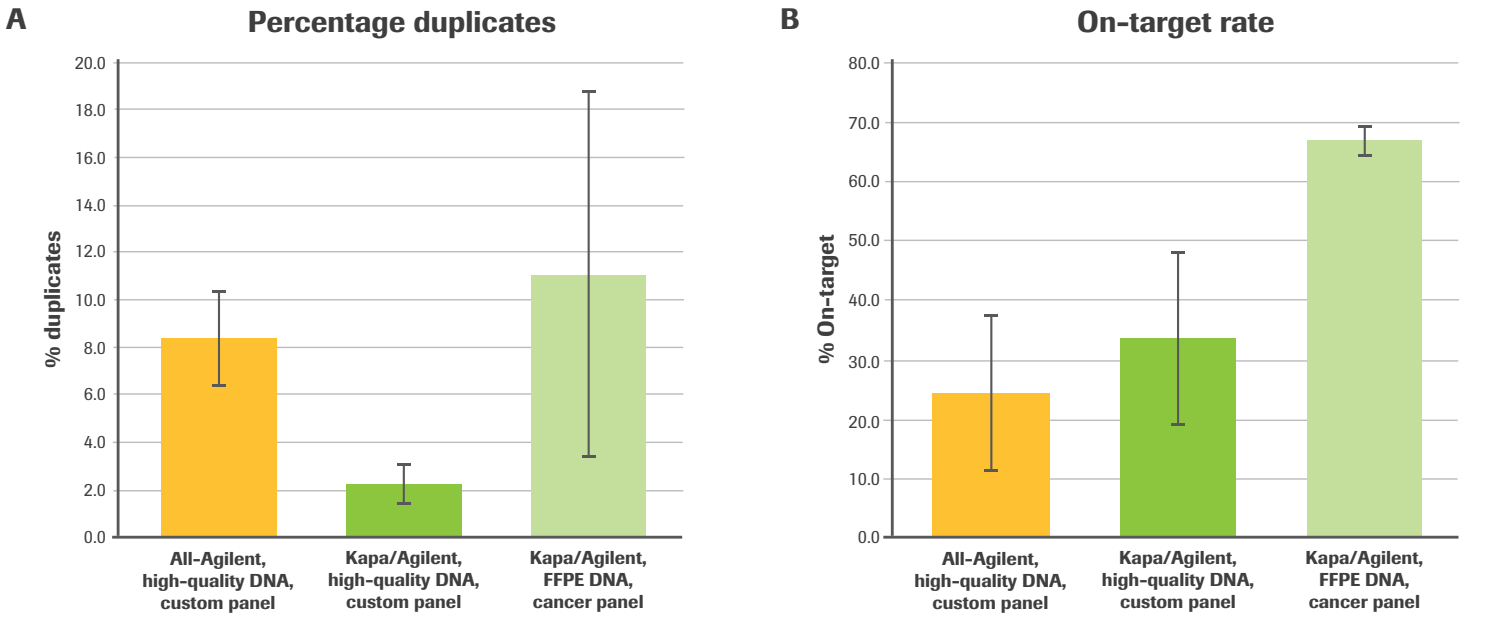

Figure 4. Percentage duplicates (A) and on-target rates (B) for libraries prepared from high-quality and FFPE DNA, using different capture panels. Libraries were prepared from four high-quality DNA samples using the All-Agilent protocol (yellow) or new Kapa/Agilent workflow (green), and the 34-gene custom SureSelectXT capture panel (172.4 kb). Light green bars represent data for libraries prepared from eight FFPE samples of variable quality, captured with Agilent's Comprehensive Cancer Panel (788 kb). The new Kapa/Agilent protocol achieved significantly lower duplication rates than the original All-Agilent workflow. Optimization of capture and wash steps resulted in higher on-target rates. The new Kapa/Agilent workflow achieved excellent results with FFPE DNA.

With the Kapa/Agilent workflow, duplication rates for the FFPE samples were in the same range as those obtained with the original All-Agilent protocol for high-quality DNA (average=8.4% when the lowest quality FFPE sample (#2) is excluded). These improvements to library quality and capture efficiency with the Kapa/Agilent workflow reduce the input and quality requirements for successful sample preparation from FFPE DNA.

Variant calling

Eleven high-quality DNA samples were processed with the All-Agilent and/or Kapa/Agilent workflows, and sequenced across two runs on an Illumina® MiSeq® instrument. Data were processed as described in Materials and Methods. Variant calling statistics for the three samples processed in one run with both protocols are given in Table 3. Correlations between methods were high, indicating they produced libraries that are functionally equivalent.

Table 3. Variant calling statistics for high-quality DNA libraries prepared with both workflows and the 34-gene custom panel

Sample #	9		10		11	
Run Protocol	*B* All-Agilent	*B* Kapa/Agilent	*B* All-Agilent	*B* Kapa/Agilent	*B* All-Agilent	*B* Kapa/Agilent
Mapped reads	99.89%	99.84%	99.99%	99.86%	99.77%	99.86%
Mapped inside target +200 bp	9.11%	44.67%	32.63%	40.39%	30.99%	17.30%
Variants	111	117	102	107	113	121
SNPs	86.49%	82.05%	84.31%	80.37%	84.07%	81.82%
Insertions	1.80%	5.98%	5.88%	9.35%	3.54%	6.61%
Deletions	11.71%	11.97%	9.80%	10.28%	12.39%	11.57%
dbSNP entry	94.59%	91.45%	97.06%	94.39%	96.46%	92.56%
Ti/Tv	2.43	2.43	2.31	2.31	2.39	2.30
Heterozygous	61.26%	64.10%	67.65%	69.16%	69.03%	71.90%
Homozygous	38.74%	35.90%	32.35%	30.84%	30.97%	28.10%
Missense	24.32%	23.08%	15.69%	14.95%	18.58%	19.01%
Synonymous	19.82%	18.80%	21.57%	20.56%	15.04%	14.05%

Conclusions

The new, streamlined KAPA HyperPlus/SureSelectXT Fast workflow developed in this study represents a significant improvement for routine and rapid targeted sequencing on the Illumina® platform. The new workflow:

- Allows for the preparation of sequencing-ready, SureSelect capture libraries in a standard 8-hour working day
- Does not require expensive fragmentation equipment
- Obviates the need for careful quantification and normalization of input DNA
- Yields robust and reproducible results with high-quality DNA and FFPE samples of variable quality
- Yields more complex libraries than the original All-Agilent SureSelectXT Fast method through a combination of low fragmentation bias, very efficient conversion of input DNA to adapter-ligated library, and low-bias library amplification (typically employing fewer cycles of pre-capture PCR)
- Achieves lower duplication rates than established protocols as a result of higher library complexity
- Offers higher on-target rates for both small, custom panels and larger, catalog panels through improvements to capture and wash procedures
- Does not come at any cost to the sensitivity and specificity of variant detection
- Enables sequencing of low-quality DNA not appropriate for Sanger and/or amplicon sequencing

Acknowledgements

The authors wish to thank Alex Siebold (Agilent Technologies) for significant scientific contributions to the development of the new protocol. Operational and technical contributions from Anniek de Witte and Mary Napier (Agilent) are acknowledged with gratitude. We also thank Carolin Dreier (University of Muenster) for expert technical assistance.

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. KAPA HyperPlus Kit for SureSelectXT Fast Target Enrichment Technical Data Sheet, v1.16 (2016). Kapa Biosystems document number KR1323.
3. SureSelectXT Fast Target Enrichment System for Illumina Paired-End Multiplexed Sequencing Library, vA.0 (2015). Agilent Technologies document number G9701-90000.

Published by:

Roche Sequencing and Life Science
9115 Hague Road
Indianapolis, IN 46256
sequencing.roche.com

MC-US-10932 APP109004 A520 07/22