Roche

Authors

Nancy Nabilsi
Senior Applications Scientist

Drew Cheney
Applications Scientist

Ida van Jaarsveld
*R&D Team Leader
(Bioinformatics)*

Leendert Cloete
Bioinformatics Scientist

Jacqueline Meyer
*Support & Training
Senior Specialist*

Jennifer Pavlica
Applications Manager

Rachel Kasinskas
*Director of Scientific
Support & Applications*

Roche Sequencing & Life Science
Wilmington, MA, USA,
and Cape Town, South Africa

KAPA RNA HyperPrep Workflow: Recommendations and expectations for RNA-sequencing using degraded inputs

High-resolution RNA analysis using next-generation sequencing (RNA-seq) is enabling advances in clinical and molecular diagnostic research. The quality of RNA extracted from biological specimens is highly variable and yields are often low, thus impacting the ability to generate high-quality sequencing libraries. In this study, the effects of RNA type and quality on RNA-seq library construction are assessed and expectations regarding sequencing data quality are addressed. Included are recommendations for informative quality control measures and input-specific modifications that may improve performance.

Introduction

Formalin-fixed paraffin-embedded (FFPE) tissue is a ubiquitous resource for clinical research. The quality of RNA extracted from FFPE tissue can be highly variable due to the damaging nature of the formalin fixation process where crosslinking, chemical modification, and fragmentation can occur. Additional variables that impact RNA quality include the age of the tissue, the extraction method, and the experience of the user. The choice of library construction strategy and the results obtained are influenced by the input amount and quality of the RNA.

RNA–seq workflow selection

Several options are available for RNA-seq library construction. Preparation of total RNA-seq libraries requires the least amount of input material and captures the most comprehensive transcriptome, including ribosomal RNA (rRNA) content (~90% of total RNA content). Removal of rRNA from RNA-seq libraries increases the economy of sequencing and improves coverage of low-abundance transcripts of interest. Two common strategies to remove rRNA during library construction are to enrich for mRNA or to deplete rRNA from the RNA sample.

Enrichment for mRNA typically targets the poly-A tail, either through bead-based captures or selective priming with oligo-dT. Thus, the ability of mRNA enrichment to capture a full-length transcript is dependent on RNA quality. Highly degraded inputs, such as from FFPE-derived RNA, are not suitable for mRNA enrichment, and use of this strategy will invoke a strong bias towards the 3'-ends of transcripts.

Enzymatic depletion of rRNA using complementary DNA oligonucleotides and RNase H is an efficient strategy for upfront removal of rRNA. In contrast to mRNA enrichment, rRNA depletion generates a more comprehensive representation of the transcriptome—as precursor mRNAs and non-coding RNAs are retained—and is highly effective with degraded RNA inputs.

The libraries sequenced in this study were prepared using the KAPA RNA HyperPrep Kit with RiboErase (HMR). This workflow is designed for both manual and automated RNA-seq library construction from 25 ng – 1 µg of total human, mouse, or rat RNA and depletes both cytoplasmic (5S, 5.8S, 18S, and 28S) and mitochondrial (12S and 16S) rRNA species. The protocol is applicable to a wide range of RNA-seq applications, including gene expression analysis and single-nucleotide variation (SNV) discovery, as well as splice junction and gene fusion identification. A workflow summary is provided in Figure 1 with quality control (QC) steps indicated.

Input RNA and library construction QC steps

Input RNA handling and assessment, QC1 and QC2

Efforts should be made to minimize the presence of genomic DNA (gDNA) contamination in the RNA sample. During ribosomal depletion, RNA is incubated under conditions that promote hybridization of rRNA to complementary DNA oligonucleotides, followed by treatment with RNase H to remove rRNA duplexed with DNA. Under these conditions, contaminating gDNA may similarly hybridize to complementary transcripts, leading to undesirable degradation of non-rRNA species. Significant DNA contamination would also lead to an increased number of intergenic reads. Therefore, if not part of the RNA extraction process, total RNA should be subsequently DNase treated (without heat inactivation) and re-purified. In this study, RNA was extracted from tissues using kits that include on-column DNase treatment. See **Appendix A** for more details.

QC1: Fluorometric assays are recommended to quantify RNA samples, as they generally provide a more sensitive measurement of nucleic acid concentration than spectrophotometric methods. For example, the Qubit® RNA HS Assay is highly selective for RNA, will not quantify DNA, protein, or free nucleotides, and is effective on low-abundance RNA samples (designed for RNA sample concentrations between 250 pg/µL and 100 ng/µL).
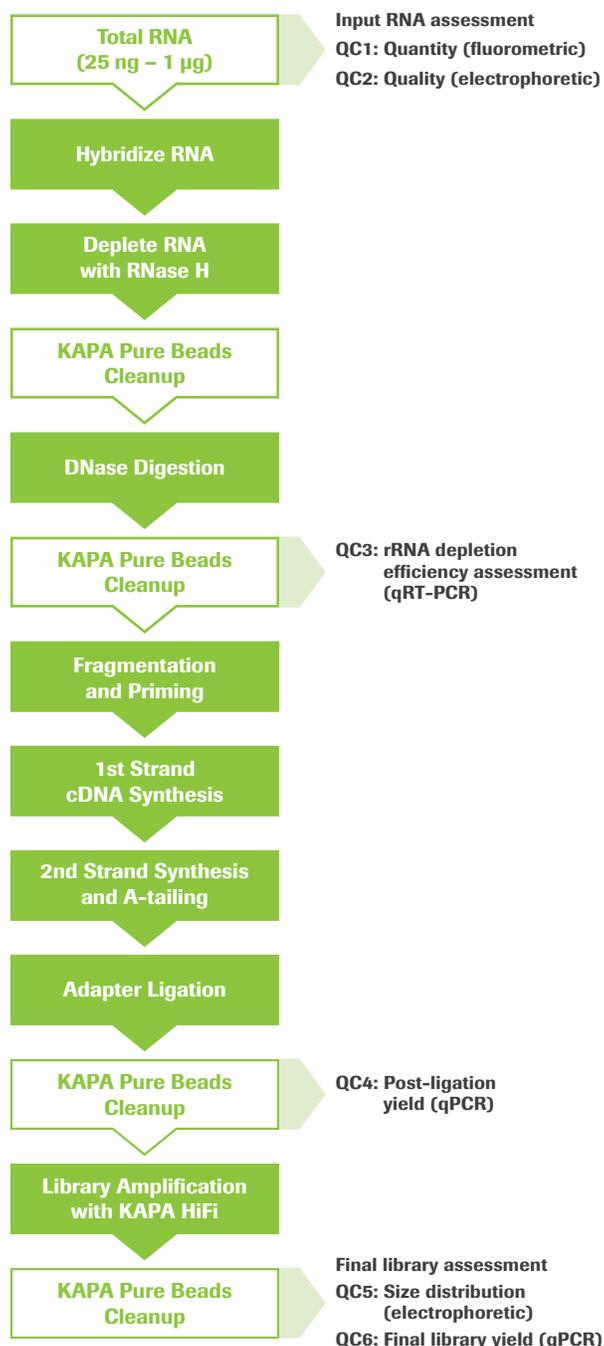

Figure 1: KAPA RNA HyperPrep Kit with RiboErase (HMR) workflow. Quality control (QC) points are indicated.

QC2: The quality and size distribution of input RNA should be assessed prior to library construction by an electrophoretic method (e.g., an Agilent® Bioanalyzer RNA assay). Two metrics frequently used to evaluate RNA quality based on the sample's electrophoretic trace include the RNA Integrity Number (RIN) and the DV200 value. The RIN score is automatically tabulated by the Agilent Expert software which computes the ratio of ribosomal peaks and the presence of degraded product to assign the integrity number. Samples with a RIN ≥7 are considered high quality. FFPE-extracted RNA typically lacks distinctive ribosomal peaks, thus impacting the relevance of the RIN score as a quality metric for such degraded samples.

RNA fragments shorter than 200 nucleotides (nt) are poor substrates for RNA-seq library construction and are likely to be lost following fragmentation, random priming, and subsequent library cleanup steps. The DV200 metric quantifies the percentage of sample RNA greater than 200 nt in length and is expected to be more reflective of RNA quality than the RIN score when applied to lower-quality, degraded inputs. For instructions on how to assess DV200 values using the Agilent® Expert software, see **Appendix B**.

Tissue-derived RNA samples vary widely in their quality metrics. Bioanalyzer traces are shown in Figure 2(A-E) for: one sample of high-quality, intact RNA (Universal Human Reference, UHR); one sample of partially degraded RNA (from fresh-frozen breast tumor tissue); and three FFPE-derived samples (thyroid, duodenum, and breast tumor). Note that the FF and FFPE breast tumor samples were obtained from the same biological specimen. While all four tissue-derived samples are considered of similarly low quality based on their RIN score, the DV200 values are more broadly distributed and suggest that the RNA extracted from fresh-frozen tumor (74%) and FFPE-thyroid (47%) are higher quality inputs than the RNA from FFPE-duodenum (29%). The FFPE breast tumor sample also exhibits a high DV200 value (76%), but this value is likely artificially inflated due to the high molecular weight peak observed in this otherwise highly degraded sample. In some cases, this may reflect material that is cross-linked or was incompletely deparrafinized rather than actual intact RNA.

rRNA depletion efficiency assessment, QC3

QC3: If desired, an optional quality control assessment can be used to confirm successful rRNA depletion of the RNA sample. An aliquot of RNA is retained following rRNA depletion and prior to fragmentation (see Figure 1). qRT-PCR is performed using primers against an rRNA species (in this study, human 28S) and a housekeeping control gene (in this study, GAPDH). Calculation of the delta Ct (dCt) value between the depleted sample and a control input sample reflects the success of rRNA depletion. (See Appendix C for primer sequences and a detailed protocol to assess rRNA depletion efficiency.) In this study, 47 RNA-seq libraries were prepared using either an enzymatic workflow—as employed in the KAPA RiboErase (HMR) module—or a bead-based strategy for removal of rRNA. Samples prepared by enzymatic ribodepletion show a strong correlation between the higher dCt values for 28S rRNA (≥7) and fewer rRNA reads (<5%) in the final RNA-seq data (Figure 3).
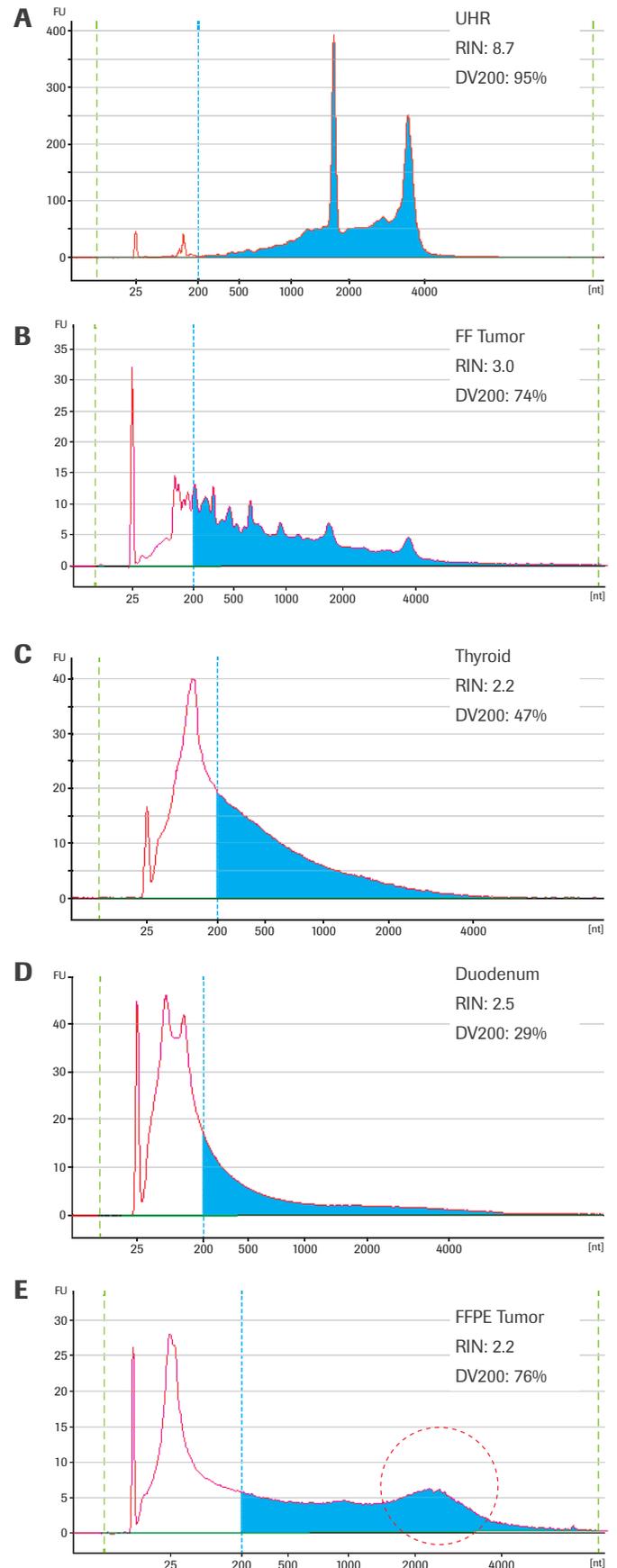

Figure 2: Electrophoretic profiles for representative RNA samples utilized in this study. Samples **C** – **E** were isolated from FFPE tissue; whereas sample **B** originated from fresh-frozen (FF) tissue of the same biological specimen as sample C. High-quality human Universal Human Reference RNA (Agilent Technologies; **A**) is included for comparison. Electropherograms were generated using an Agilent 2100 Bioanalyzer and RNA 6000 Pico Kit. Blue shading highlights RNA fragments >200 nt. The region circled in red in the breast tumor FFPE sample (**E**) designates a high-molecular weight peak that is likely the result of crosslinking or inefficient deparrafinisation (rather than intact transcripts that could be efficiently converted to sequenceable cDNA fragments).

Post-ligation Library Quantification, QC4

QC4: Due to the inherent variability of RNA samples extracted from tissues in general and FFPE-derived samples in particular, quantifying the post-ligation yield of RNA-seq libraries before performing amplification is recommended. This allows the user to estimate the number of cycles necessary to prepare sufficient material for sequencing, quality control, and storage of their sample without over-amplifying. Excessive library amplification can result in unwanted artifacts such as amplification bias, PCR duplicates, chimeric library inserts, and nucleotide substitutions. For this study, cycle numbers were selected based on the sample post-ligation yield to generate ≥10 nM amplified library, a commonly recommended concentration for long-term storage.

Samples used in this study were quantified after the post-ligation cleanup and post-amplification cleanup steps (see Figure 1) using the qPCR-based KAPA Library Quantification Kit. This kit includes primers and standards based on Illumina® primer sequences and can be used to quantify libraries that were constructed with full-length adapters. Libraries must be diluted to fall within the dynamic range of the assay. The dilution factor will depend on input quantity and quality. In this study, post-ligation libraries prepared from 25 – 100 ng of partially degraded or FFPE RNA samples were diluted 1:20 prior to quantification. Post-ligation libraries prepared from 25 – 100 ng of high-quality UHR RNA were diluted 1:100 prior to quantification. Data analysis templates for library quantification are available for download at **kapabiosystems.com/support**.

Final library assessment, QC5 and QC6

QC5: The size distribution of final libraries should be confirmed with an electrophoretic method. A LabChip® GX, GXII, or GX Touch (PerkinElmer), Bioanalyzer or TapeStation (Agilent Technologies), Fragment Analyzer (Advanced Analytical) or similar instrument is recommended over conventional gels. Successful library construction is demonstrated by a final library distribution displaying the desired mode or mean fragment size, and minimal adapter-dimer carryover, observed as a sharp 120 – 140 bp peak.

QC6: The KAPA Library Quantification Kit is recommended for high-sensitivity quantification of final library yields. In this study, post-amplification libraries were diluted 1:20,000 prior to

quantification. Libraries were quantified, normalized, and pooled for multiplex sequencing. Final library pools were quantified by qPCR (Table 1) and submitted for sequencing.
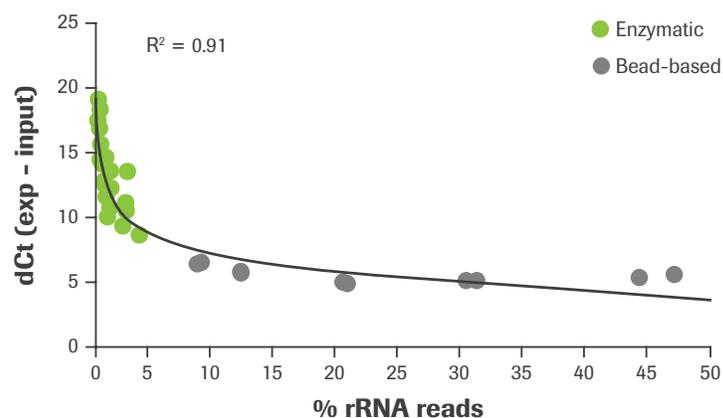

Figure 3: Correlation between rRNA depletion QC3 and RNA-seq data. RNA samples were depleted of rRNA content using either enzymatic (RNase H) or bead-based workflows. Delta Ct values for 28S were measured using qRT-PCR and plotted against the rate of carryover rRNA reads obtained from RNA-seq. For this and all subsequent RNA-seq data, sequencing was performed using an Illumina HiSeq® 2500 in high output mode with v4 chemistry and 2 x 100 bp read length.

Workflow optimization for degraded inputs

FFPE RNA Samples

RNA quality has a significant impact on RNA-seq library construction. For a set input quantity (100 ng), post-ligation yield is decreased and adapter-dimer formation is increased as RNA quality declines (Figure 4). Increasing RNA input quantity generally improves adapter-dimer rates and post-ligation yield. An increase in post-ligation yield indicates that more unique molecules were converted from input RNA to properly adapted library fragments. For samples that are efficiently depleted of rRNA, higher post-ligation yield reflects an increase in library diversity and is associated with lower duplication rate (Figure 5). Notably, the lowest post-ligation yields and highest duplication rates were observed with duodenum FFPE which was the lowest-quality RNA sample assessed.

Figure 6A further illustrates the impact of input RNA quality on library preparation. Bioanalyzer traces for three independent RNA extractions of variable quality from the same FFPE tissue source are shown. Duplicate RNA-seq libraries were prepared from 25 ng of total RNA (without rRNA depletion) using the KAPA RNA HyperPrep Kit. Consistent with results observed for rRNA-depleted libraries, decreased DV200 values are associated with decreased final library yield (Figure 6B). In contrast to rRNA-depleted preparations, adapter-dimer rates are not

appreciably increased. Note that libraries prepared from thyroid sample 3 (DV200: 11%) are shown for illustrative purposes only. It is unclear if useful sequencing data would be obtained from input RNA with such a poor quality score.

When input quantity is limiting and/or sample quality is very low, adapter concentration and post-ligation cleanup ratios may be modified to improve performance. To test the impact of modifying these steps on library quality, titrations of adapter concentration and post-ligation bead clean-up ratios were performed using a highly degraded FFPE input (DV200: 29%). Post-ligation yield improved with increased adapter concentration and less stringent post-ligation cleanups (Figure 7A); however, this was accompanied by increased adapter-dimer formation (Figure 7B).

For this sample, decreasing the adapter concentration to 750 nM (from the standard concentration of 1.5 uM) and relaxing post-ligation cleanup ratios to 0.8X/0.9X (from the standard 0.63X/0.7X ratios) reduced adapter-dimer formation without significantly lowering library yield in comparison to the standard library preparation conditions (Figure 7C). Though this condition was identified as optimal for this sample, ideal conditions are likely to differ with other input amounts and qualities. If workflow optimization is necessary, it is recommended to optimize around the anticipated range of input amounts and qualities for the application. In some cases, adapter-dimers may persist even after workflow optimization. In this case, a second 1X bead cleanup can be performed after amplification to remove small products.
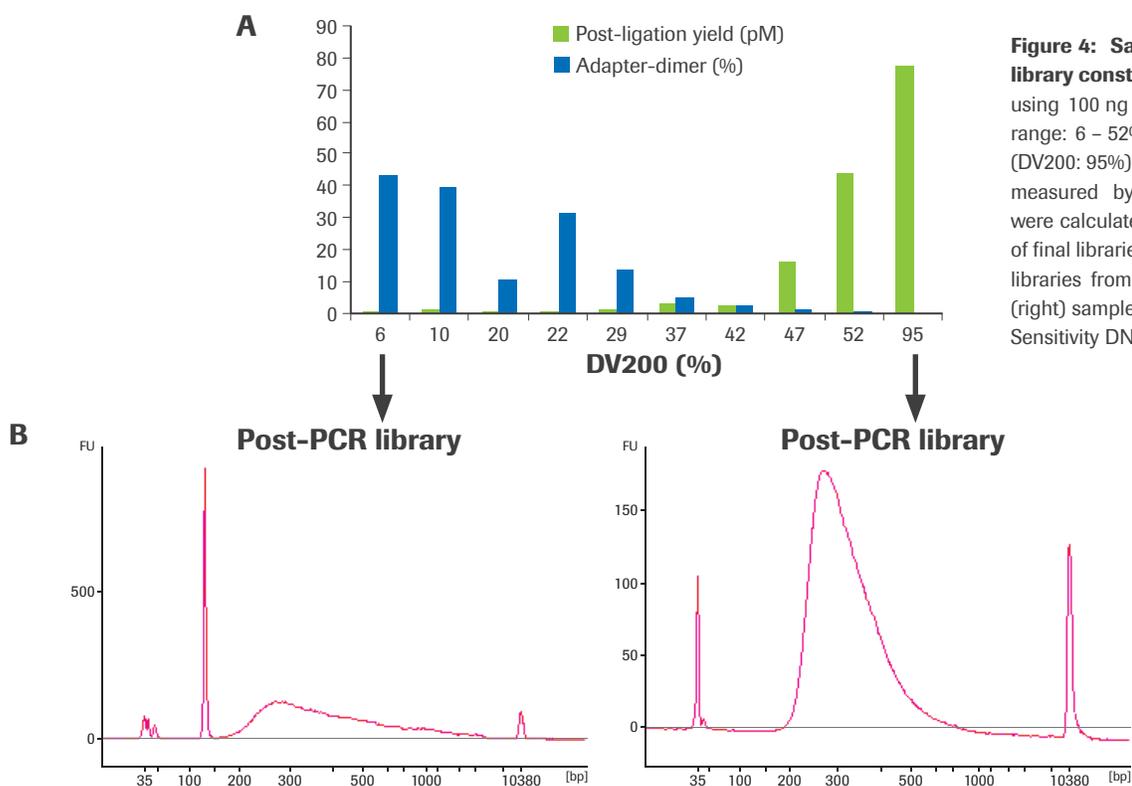

Figure 4: Sample quality affects RNA-seq library construction. Libraries were constructed using 100 ng of RNA from FFPE tissues (DV200 range: 6 – 52%) and a high-quality UHR control (DV200: 95%). **(A)** Post-ligation yield was measured by qPCR, and adapter-dimer rates were calculated from electrophoretic assessment of final libraries. **(B)** Representative traces of final libraries from low-quality (left) and high-quality (right) samples measured using an Agilent® High Sensitivity DNA Kit.
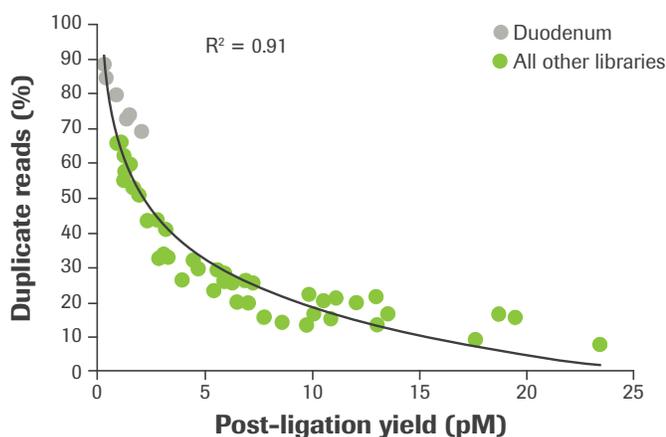

Figure 5: Correlation between post-ligation yield and percent duplicate reads. Post-ligation yield was measured by qPCR and plotted against the percent of duplicate reads obtained from RNA-seq of rRNA-depleted samples. Reads aligning to rRNA were removed and paired reads were randomly subsampled to 14M for comparative analyses.
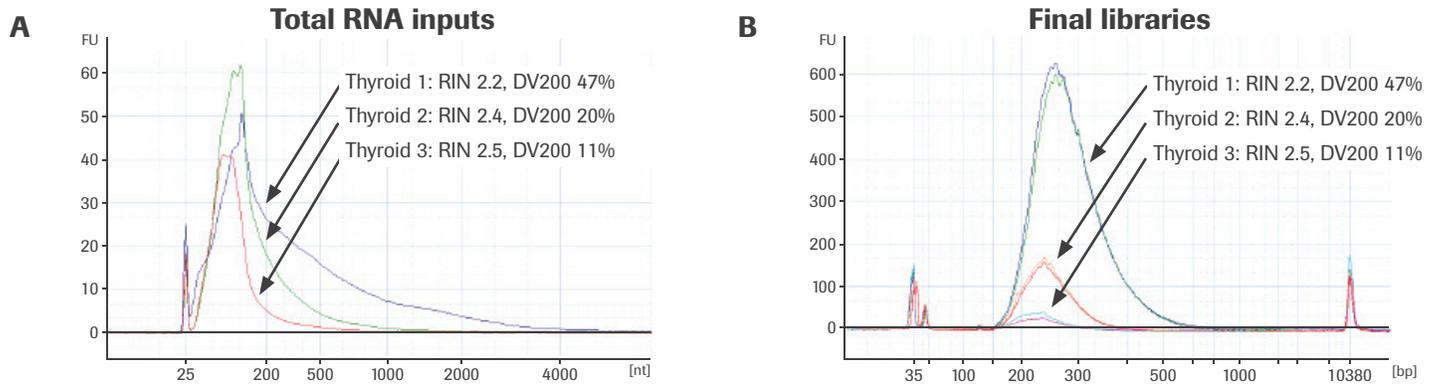
A

Total RNA inputs

Thyroid 1: RIN 2.2, DV200 47%
Thyroid 2: RIN 2.4, DV200 20%
Thyroid 3: RIN 2.5, DV200 11%

B

Final libraries

Thyroid 1: RIN 2.2, DV200 47%
Thyroid 2: RIN 2.4, DV200 20%
Thyroid 3: RIN 2.5, DV200 11%

Figure 6: Effect of RNA quality on the KAPA RNA HyperPrep Kit (without rRNA depletion) workflow. Electrophoretic profiles of **(A)** three separate extractions of RNA from FFPE thyroid tissue and **(B)** final total RNA-seq libraries prepared from each extraction. Final library yield decreases as a function of input RNA quality.
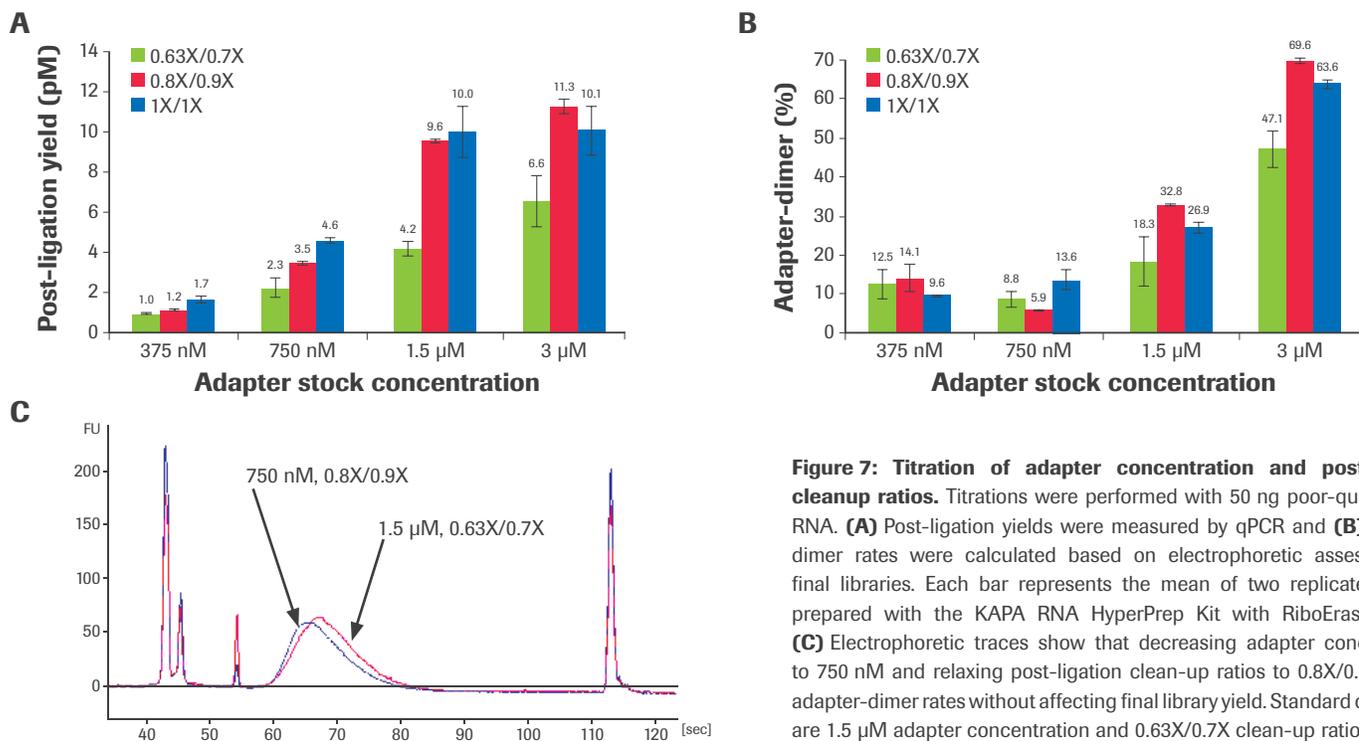
A

B

C

Figure 7: Titration of adapter concentration and post-ligation cleanup ratios. Titrations were performed with 50 ng poor-quality input RNA. **(A)** Post-ligation yields were measured by qPCR and **(B)** adapter-dimer rates were calculated based on electrophoretic assessment of final libraries. Each bar represents the mean of two replicate libraries prepared with the KAPA RNA HyperPrep Kit with RiboErase (HMR). **(C)** Electrophoretic traces show that decreasing adapter concentration to 750 nM and relaxing post-ligation clean-up ratios to 0.8X/0.9X lowers adapter-dimer rates without affecting final library yield. Standard conditions are 1.5 µM adapter concentration and 0.63X/0.7X clean-up ratios.

Partially degraded RNA samples

RNA extracted from additional types of biological samples (i.e., fresh-frozen tissues) may present partially degraded electrophoretic profiles. Typically, partially degraded samples have low RIN scores (in comparison to fully intact samples), but relatively high DV200 values (in comparison to FFPE samples). For partially degraded samples, fragmentation conditions may be modified to improve library yield and/or alter fragment size distribution.

Bioanalyzer traces for partially degraded input RNA samples of varying quality are shown in Figures 8A – C, left panels. Fragmentation time and temperature titrations were performed, and the final amplified libraries prepared from each input are shown in Figures 8A – C, right panels.

For all inputs, increasing fragmentation time and/or temperature resulted in reduced post-amplification yield and shorter library fragments that are more tightly distributed. To highlight the impact of sample quality and fragmentation conditions on final library size, mean fragment lengths of the amplified libraries are plotted, grouped by fragmentation condition, in Figure 8D. Note that for a given fragmentation condition, libraries generated from lower-quality samples are shorter than those generated from higher-quality inputs. This trend is particularly evident with decreased fragmentation time and temperature conditions. Thus, when applying a single fragmentation condition across samples of varying qualities, more quality-dependent differences in final library sizes are expected if fragmenting at lower temperatures and/or reduced durations.
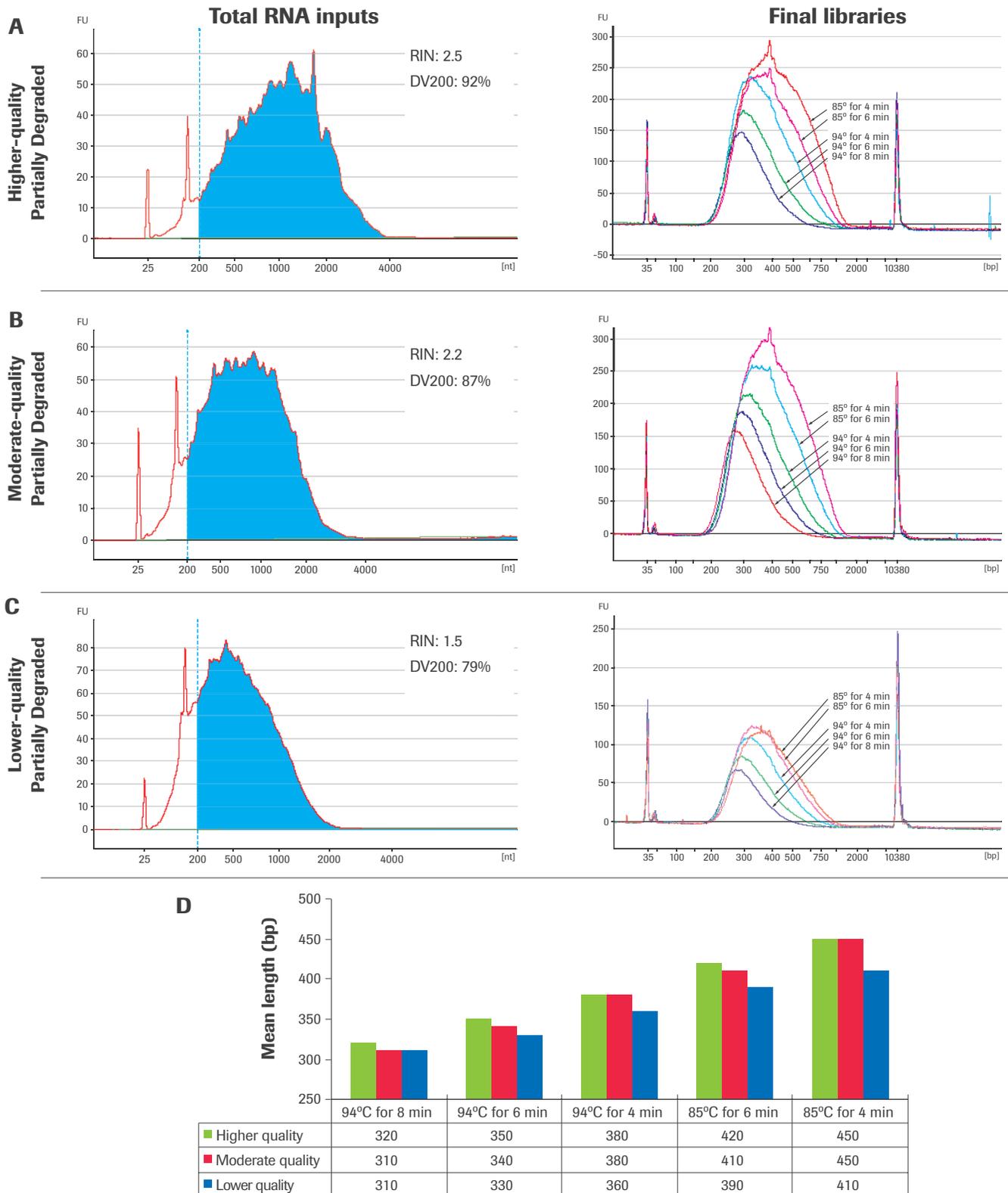
Figure 8: Titration of fragmentation conditions for partially degraded samples. Libraries were constructed using 100 ng of chemically-degraded UHR RNA including **(A)** higher-quality, **(B)** moderate-quality, and **(C)** lower-quality inputs. Total RNA traces were generated with an Agilent® RNA 600 Pico Kit (left panels) and final library traces were generated with an Agilent High Sensitivity DNA Kit (right panels). **(D)** Mean library lengths are grouped by fragmentation condition.

Effect of input quality and quantity on sequencing metrics

RNA samples of varying quality (shown in Figure 2) were used as input into library construction using KAPA RNA HyperPrep Kit with RiboErase (HMR) with 25 and 100 ng inputs.

For highly degraded duodenum FFPE RNA, libraries were only prepared with 100 ng input. All samples were prepared using the standard adapter stock concentration (1.5 µM) and post-ligation cleanup ratios (0.63X/0.7X). Relevant input quality and library construction metrics are listed in Table 1. Regardless of quality or input, all samples exhibited efficient depletion of rRNA (<5% rRNA reads). Comparing library construction metrics for the FFPE-derived RNA samples, libraries prepared from the lower-quality duodenum sample showed the lowest post-ligation yield, required additional cycles of amplification to obtain 10 nM final concentration, and exhibited smaller mode fragment sizes and higher adapter-dimer rates compared to libraries prepared from higher-quality thyroid and tumor samples. For thyroid and FF tumor libraries, increasing the input quantity from 25 to 100 ng reduced adapter-dimer rates. As shown in Figure 9A, mapping rates exceeded 90% for all sample types, regardless of sample quality or input. Consistent with exhibiting the lowest post-ligation yield (Table 1), libraries prepared from duodenum showed the highest duplication rates (Figure 9B). For thyroid, tumor and UHR libraries, increasing the inputs from 25 to 100 ng substantially decreased the duplication rate (up to 3.1-fold improvement).

Coverage uniformity is plotted in Figure 9C according to mean CV, or the coefficient of variation in base coverage across transcript length. For this metric, a smaller number is reflective of less variation in base coverage and thus, better coverage uniformity. Lower-quality samples exhibited a higher mean CV (lower coverage uniformity) compared to UHR. Increasing input quantity mildly improved this metric.

Figure 9D shows the number of unique transcripts identified for each sample type. While there appears to be a trend associating lower sample quality with a reduced number of transcripts identified, it is important to note that these samples were extracted from a variety of tissue types, and the expected number of transcripts obtained from each tissue source may not be consistent. Comparing the number of unique transcripts identified for each sample across input quantities, a strong similarity is observed between numbers obtained at 25 and 100 ng. The high-quality UHR samples show the most consistency, with only a 0.18% difference between the number of transcripts identified at 25 and 100 ng inputs. A 0.5%, 0.8%, and 1.5% difference between inputs is observed for FF tumor, thyroid, and FFPE tumor, respectively.

Next, the effect of sample quality on reproducibility in measuring gene expression was examined. Despite the less favorable library construction metrics, the reproducibility between replicates for the duodenum sample was excellent with similar robustness as that observed for the thyroid sample (Figures 10A – C). Additionally, a strong correlation was observed for gene expression data obtained from 25 and 100 ng inputs (Figure 10D). Finally, gene expression data was compared between the matched FF and FFPE tumor samples. Strong agreement was observed between these samples for both 25 and 100 ng inputs (Figures 10E – F).

In summary, RNA extracts of variable quality from tissue samples are suitable inputs for rRNA depletion and RNA-seq library construction using the KAPA RNA HyperPrep workflow. With the exception of coverage balance, library construction from higher-quality FFPE or partially degraded samples yield sequencing metrics that are similar to those obtained with high-quality intact RNA. Increasing the input quantity for partially degraded or FFPE-derived samples improves library diversity, minimizes adapter-dimer formation, and partially improves coverage uniformity. Very low-quality FFPE samples are more challenging,

Table 1: High-level sample, workflow, and data overview

		UHR		FF Tumor		FFPE Tumor		Thyroid		Duodenum
Input RNA	RIN	8.7		3.0		2.2		2.2		2.5
	DV200 (%)	95		74		~76 (inflated)		47		29
	Input (ng)	25	100	25	100	25	100	25	100	100
Library construction conditions	Fragmentation	94°C for 4 min				65°C for 1 min				
	PCR cycles	13	11	15	13	15	13	15	13	17
Library construction QC data	Post-ligation yield (pM)	9.3	77.5	5.9	17.2	2.4	10.3	5.8	16.3	1.4
	Post-PCR yield (nM)	14.3	22.9	36.1	20.9	8.4	13.6	21.0	26.9	13.9
	Mode library size (bp)	333	349	302	295	278	287	312	305	282
Sequencing metrics	rRNA reads (%)	1.6	0.9	4.0	1.5	0.6	0.7	0.4	0.2	1.7
	Adapter-dimer (%)	<1.0	<1.0	1.7	<1.0	3.7	3.7	2.6	<1.0	11.7*

*To facilitate comparison between samples, standard conditions were applied and a second 1X post-amplification clean-up was not employed for the duodenum libraries.

particularly with regards to adapter-dimer formation, duplication rate, and coverage balance. If possible, increasing input (up to 1 μg) is recommended for very-low-quality samples. Reproducibility across input amounts and between replicates is robust, regardless of the input quality; and strong agreement between matched fresh frozen and FFPE tissue lends confidence towards the accuracy of gene expression profiling using RNA derived from FFPE tissue.
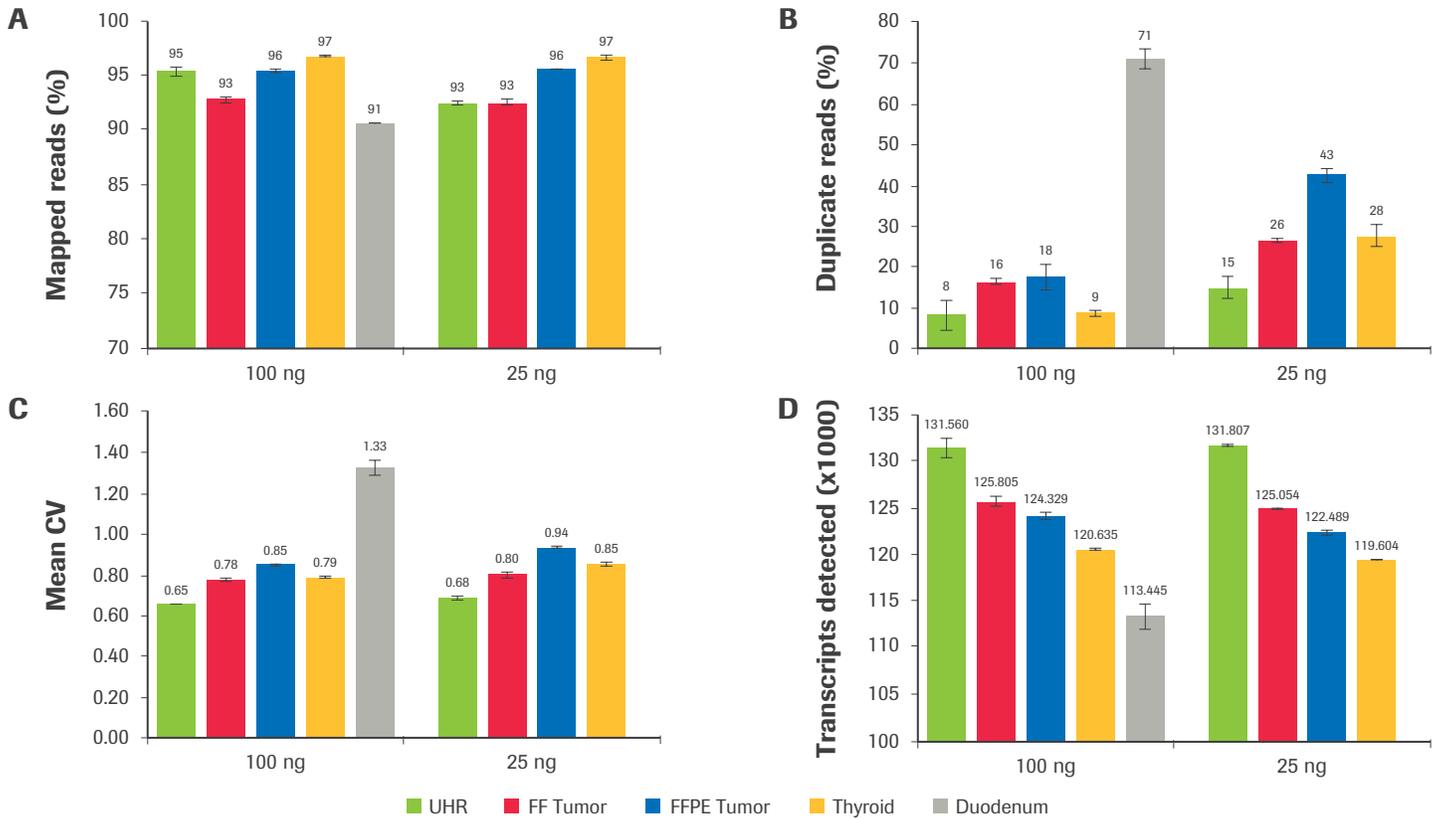

Figure 9: Effect of RNA input on sequencing metrics. After subsampling, data was analyzed to quantify, the **(A)** percent mapped reads, **(B)** percent duplicate reads, **(C)** mean CV, and **(D)** number of unique transcripts identified.
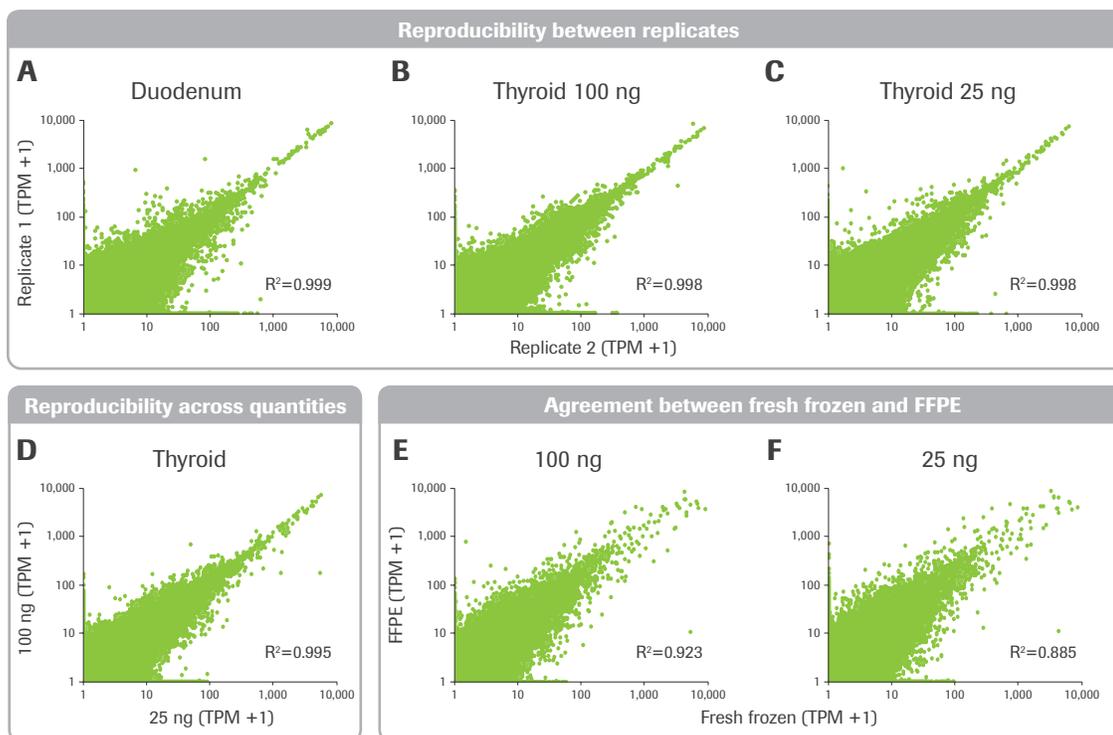

Figure 10: Effect of RNA input on reproducibility and accuracy. Pearson correlation plots are shown for gene expression transcripts per million (TPM) counts across **(A – C)** replicates and **(D)** input quantities. **(E – F)** Agreement between FFPE and matched fresh frozen breast tumor tissue is shown.

Summary

The quick reference tables below summarize library preparation QC points, optimization strategies and recommendations for troubleshooting.

Table 2: Library preparation QC points

QC	Priority	Assay information	Application
QC1: Quantity assessment of input RNA sample	Required	• Fluorometric methods, such as Qubit, are highly preferred • Concentration information from Nanodrop or electrophoretic methods is acceptable	Provides quantitative information regarding the amount of total RNA available for library construction
QC2: Quality assessment of input RNA sample	Strongly recommended	Electrophoretic methods, such as Agilent® Bioanalyzer RNA assays	• Determines if the sample is likely to support successful library construction • Helps to identify appropriate fragmentation conditions and adapter input concentration
QC3: rRNA depletion efficiency	Optional	qRT-PCR, such as the KAPA SYBR® FAST One-Step qRT-PCR Kit	Confirms successful depletion of rRNA prior to sequencing
QC4: Post-ligation yield	Recommended	qPCR, such as the KAPA Library Quantification Kit (only for full-length adapters where P5 and P7 are added during ligation)	• Helps minimize over-amplification • Useful for optimization during the establishment of new workflows and pipelines, or when working with new sample types and input amounts • May also predict library complexity issues (such as high duplication rates) prior to sequencing
QC5: Final library size distribution assessment	Strongly recommended	Electrophoretic methods: LabChip® GX, GXII, or GX Touch (PerkinElmer), Bioanalyzer or TapeStation (Agilent Technologies), Fragment Analyzer (Advanced Analytical), or similar instrument is recommended over conventional gels	Confirms successful library preparation by measuring the final library size distribution and the presence of undesirable products, such as adapter-dimer
QC6: Final library yield	Recommended	qPCR, such as the KAPA Library Quantification Kit	Most accurate method for quantification of NGS libraries prior to sequencing

Table 3: Library preparation optimization strategies

Sample type	Optimization	Standard condition*	Notes
FFPE	Input quantity	25 ng – 1 µg	Increased input amounts, up to 1 µg, can decrease adapter-dimer formation and increase complexity, as well as improve success rates with extremely degraded samples.
	Adapter concentration	1.5 µM	For inputs that are low-quality and low-quantity, decreasing adapter concentration may lower adapter-dimer rates at the expense of post-ligation yield (i.e., library complexity). There may be an optimal balance between adapter concentration and post-ligation cleanup ratios that can reduce adapter-dimer carryover without sacrificing library yield (Figure 7).
	Post-ligation cleanup ratios	0.63X/0.7X	
Partially degraded	Fragmentation	Variable, depending on desired insert size	Increased fragmentation time and temperature results in shorter, more tightly distributed library fragments. Decreased fragmentation time and temperature results in increased yields of longer, more broadly distributed library fragments. Sample quality has a stronger impact on size distribution at decreased times and temperatures (Figure 8).

*From KAPA RNA HyperPrep Kit with RiboErase (HMR) Technical Data Sheet.

Table 4: Recommendations for troubleshooting

Category	Metric	Effect of sample quality	Solution
Library preparation metrics	QC3: rRNA depletion efficiency	None observed in this study	No troubleshooting needed in this study
	QC4: Post-ligation yield	Decreases as sample quality declines	Increase input quantity
	QC5: Adapter-dimer formation	Increases as sample quality declines	To prevent: ▪ Increase input quantity ▪ Decrease adapter concentration and/or modify post-ligation cleanup ratios (Figure 7) To mitigate in a final library: ▪ Implement a second 1X cleanup post-amplification
	QC6: Final library yield	Decreases as sample quality decreases (for a set number of amplification cycles)	▪ Increase input quantity ▪ Increase the number of amplification cycles, if necessary, but note that this may impact sequencing metrics such as duplication rate
Sequencing metrics	Mapping rate	None observed in this study	No troubleshooting needed in this study
	Reproducibility	None observed in this study	No troubleshooting needed in this study
	Duplication rate	Increases as sample quality declines	Increase input quantity
	Coverage balance	Decreases as sample quality declines	Increasing sample input quantity partially improves coverage balance. For very highly degraded samples, coverage balance will likely remain lower than observed for higher-quality samples.

Appendix A: Materials and methods

Experimental design

High-quality UHR RNA, partially-degraded UHR RNA, and RNA of variable qualities extracted from fresh frozen and FFPE tissues were processed using the KAPA RNA HyperPrep Kit with RiboErase (HMR). Library construction QC data were compared for rRNA depletion efficiency, library yield, and adapter-dimer formation. Sequencing data were compared with respect to % mapped reads, % rRNA reads, % duplicate reads, coverage balance, and number of unique transcripts identified.

RNA preparation

RNA extraction from both sample types included on-column DNase treatment. High-quality UHR RNA was purchased (Agilent Technologies). For preparation of partially degraded samples (Figure 8), UHR RNA was chemically fragmented using heat and $Mg2+$. RNA samples were quantified using the Qubit® RNA HS Assay and quality was assessed using a 2100 Bioanalyzer instrument and an Agilent® RNA 6000 Pico Kit.

Library construction and QC

Replicate libraries were prepared from each sample using the KAPA RNA HyperPrep Kit with RiboErase (HMR). Input quantities, fragmentation conditions, and number of PCR cycles used are indicated in Table 1. Negative (no template) control reactions were included during PCR. Aliquots of RNA were retained following rRNA depletion with RiboErase and depletion efficiency was assayed as described in Appendix C. Post-ligation and post-amplification library concentrations were measured using the KAPA Library Quantification Kit.

Sequencing and Data Analysis

For libraries submitted for sequencing, final libraries were pooled and quantified by qPCR using the KAPA Library Quantification Kit. Each pool was diluted, denatured and loaded onto an Illumina® HiSeq®2500 instrument according to manufacturer instruction. Paired-end sequencing (2 x 100 bp) was performed using a HiSeq v4 chemistry kit (Illumina). Adapter and quality trimming was performed using cutadapt and trimmomatic, respectively. Reads were aligned to a hard-masked version of human reference GRCh38, filtered to remove rRNA reads, and down-sampled to the lowest common number of paired reads (14M). Gene expression was normalized and quantified using Kallisto (0.42.4).

Appendix B: Calculate DV200 value using Agilent Expert software

To assess the DV200 value of a sample in the Agilent® Expert software:

1. Under the Global tab, change Normal to Advanced.

2. Select the box for Smear Analysis.

3. Double click on Table, add a region, and enter 200 – 10,000 nt in the popup window. (An error message about choosing the appropriate upper limit may appear—modify the upper limit accordingly.)

4. The results are displayed in the Region Table tab.

5. The results are displayed in the trace window as % of Total for the selected region.

Appendix C: Confirm successful rRNA depletion using qRT-PCR

Materials:

1. KAPA SYBR® FAST One-Step qRT-PCR Kit

2. Primer set targeting a housekeeping gene, such as GAPDH (shown below):
   - GAPDH_RNA_1F 5'- ACCATCTTCCAGGAGCGAGA
   - GAPDH_RNA_1R 5'- ATGGTGGTGAAGACGCCAGT

3. Primer set targeting a ribosomal region, such as 28S \(shown below):
   - 28S_RNA_1F 5'- TACCGGCACGAGACCGATAG
   - 28S_RNA_1R 5'- TTAACGGTTTCACGCCCTCTT

   Note: primers are designed against a human target reference and have not been tested against mouse and rat RNA. Primer BLAST indicates that the 28S primer sequences are 100% conserved in mouse and rat. GAPDH primer sequences are also 100% conserved in rat however, the reverse GAPDH primer contains 1 nt mismatch against the mouse reference.

Procedure:

Take a 2 µL sample aliquot after rRNA-depletion by modifying the following protocol steps in the KAPA RNA HyperPrep Kit with RiboErase (HMR) Technical Data Sheet:

   Step 6.2: Resuspend beads in 24 µL of 1X Fragment/Prime/
   Elute (FPE) Buffer.

   Step 6.5: Carefully transfer 22 µL of supernatant into a
   new tube.

Remove 2 µL from the collected sample and store at -20°C for QC3 at a later point.

When ready to perform the qRT-PCR assessment:

1. Thaw and briefly centrifuge depleted RNA sample aliquots from step 6.5.

2. Dilute rRNA-depleted samples by adding 3 µL RNase-free water to the tube and mix thoroughly by gently pipetting up and down several times. Never vortex RNA samples.

3. Calculate the concentrations of the 5 µL RNA-depleted samples, assuming no RNA enrichment occurred. Please see the following example for a sample prepared using 100 ng total RNA:

   a. 100 ng into 24 µL of FPE buffer = 4.17 ng/µL

   b. 4.17 ng/µL x 2 µL (initial aliquot volume) = 8.34 ng

   c. 8.34 ng into 5 µL (diluted volume) = 1.67 ng/µL

   Note: this dilution volume supports measuring each sample once (i.e., no technical replicates) for two assays (28S and GAPDH). Adjust volumes and calculations accordingly if adding technical replicates or additional assays, but do not dilute samples to <50 pg/µL.

4. Using remaining total RNA (i.e., material that has not been rRNA depleted), dilute to the same concentrations as the depleted samples to be measured on the same plate.

5. According to the KAPA SYBR FAST One-Step qRT-PCR Kit Technical Data Sheet, make an individual master mix for each primer set (28S and GAPDH), assuming 2 µL of sample (either depleted or total RNA) will be added to each well.

6. Follow the thermocycling protocol outlined in the KAPA SYBR FAST One-Step qRT-PCR Kit Technical Data Sheet.

   Note: dissociation/melt curve analysis is recommended to verify primer specificity.

Data analysis

1. Plot Ct scores for depleted and total RNA for both primer sets for comparative purposes.

   a. Ideally, GAPDH Ct scores will be similar between the control (total RNA) and sample (depleted RNA). This indicates conservation of GAPDH and minimal nonspecific material loss during rRNA depletion.

   b. Ideally, the 28S Ct scores for the depleted samples will be significantly higher than the control. This indicates a lower concentration of rRNA present in the depleted sample as compared to total RNA.

2. Calculate the delta Ct (dCt) score: subtract the Ct scores (Ct depleted sample – Ct control sample) for each sample and assay.

   a. Ideally, the GAPDH dCt value should be close to 0. For high-quality RNA, this value may range from -1 to 1. For FFPE-extracted RNA, this value may reach up to 3.

   b. Ideally, the 28S dCt value should be ≥10. Efficient depletion (>95%) has been observed for samples with dCt ≥7. Samples with a 28S dCt <7 may exhibit rRNA carryover >5%.

For more information about Roche RNA-seq products and solutions, please visit: **sequencing.roche.com/RNA-seq**