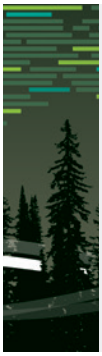


Comprehensive Capture of FFPE RNA for RNA-Seq Using a Continuous Synthesis Chemistry and Post-Library Ribodepletion

Srikanth Perike, Alexander Hoefler, Deanna K. Woo, Xiangdong Meng, Linda Lingelbach, Nish Kumar, Cissy Jiang, Angelica Olcott, Elizabeth Jordan Dreskin
Bio-Rad Laboratories, Inc., 2000 Alfred Nobel Drive, Hercules, CA 94547

Abstract

The most common archived biospecimen in oncology, formalin-fixed paraffin-embedded (FFPE) tissue is considered a key source material for gene expression profiling of tumors using RNA sequencing (RNA-Seq). However, FFPE RNA is highly degraded and is therefore suboptimal for most currently available RNA-Seq library preparation workflows, which also focus on capturing only long (>200 bp) RNAs. In this study, we present a novel RNA-Seq library preparation workflow (Figure 1) that overcomes many of the challenges associated with generating complex, rich data from FFPE RNA and other sample types with scarce and/or degraded RNA. This solution, the SEQuoia Complete Stranded RNA Library Prep Kit, leverages a proprietary retrotransposon enzyme in place of the traditional reverse transcriptase and ligases found in other RNA-Seq library preparation kits, enabling a continuous synthesis reaction that converts RNA to complementary DNA (cDNA) and adds both sequencing adapters in a single step. Our data show that using this kit in conjunction with the SEQuoia RiboDepletion Kit (for post-library ribodepletion) yields a more complex library from FFPE RNA that includes both long (>200 bp) and short (<200 bp) RNA biotypes and is therefore more representative of the complete transcriptome.

Introduction

The ability to reliably investigate the molecular profile of common malignancies has greatly advanced our understanding of disease progression and enabled the development of diagnostic and prognostic clinical tests. To date, much of this research has utilized flash-frozen (FF) tumor specimens, which can provide high-quality yet limited amounts of material to study (The Cancer Genome Atlas Network 2012). FFPE tissues represent an invaluable alternative resource for biomedical research because they are more abundant and readily available compared to FF tissue samples and are often linked to rich clinical and patient outcome data (Pennock et al. 2019). However, FFPE RNA tends to be degraded and possesses chemical modifications that may interfere with downstream enzymatic reactions (Marczyk et al. 2019, Masuda et al. 1999). Moreover, FFPE RNA is often low in yield (Marczyk et al. 2019).

Technological advances, coupled with reduced operational costs, have elevated next-generation sequencing, or massively parallel sequencing, to be the ideal tool for transcriptome analysis. The typical RNA-Seq workflow tends to be a lengthy, multistep process that involves sample RNA extraction, ribosomal RNA (rRNA) removal using either an enrichment or depletion strategy, chemical or enzymatic fragmentation of the residual RNA into smaller fragments, reverse transcription of RNA into cDNA, ligation of sequencing adapters onto the 5' and 3' end of the cDNA fragments, and finally amplification of the library pool to yield a quantity that is sufficient to load onto a sequencing flow cell.

The degraded and heavily modified nature of FFPE RNA presents key limitations for the traditional library preparation strategy in terms of its high RNA quality and yield requirements as well as its complex workflow (Palomares et al. 2019, Sarantopoulou et al. 2019).

Bulk RNA-Seq library protocols typically require 100 ng to 1 µg of input RNA (Naphade et al. 2022). In total, half a dozen unique enzymes may be used in the construction of an RNA-Seq library. The chemical modifications found in FFPE RNA could affect the performance of any one of these enzymes and lead to reduced library quality. In addition, most RNA-Seq library preparation protocols focus on the capture of only long (>200 bp) RNA fragments unless specifically designed to isolate small (<200 bp) RNA fragments (Naphade et al. 2022). Evidence suggesting that both small and long noncoding RNAs (lnRNAs) are key in gene expression regulation (Patil et al. 2014, Ransohoff et al. 2018) encourages the capture of both classes of RNAs in RNA-Seq libraries. However, generating data that represents the complete transcriptome requires the costly and laborious construction, sequencing, and analysis of multiple libraries. Furthermore, most library preparation strategies deplete rRNA before reverse transcription. The smaller RNA fragments that are a hallmark of FFPE RNA samples may be further degraded and lost through handling during the depletion step. In short, obtaining a complete and biologically representative transcriptomic dataset from FFPE RNA is a challenge.

A novel RNA-Seq library preparation workflow (Figure 1) has recently been introduced that may enable researchers to overcome many of the challenges associated with generating complex, rich data from FFPE-derived degraded RNA. The SEQuoia Complete Stranded RNA Library Prep Kit leverages a proprietary retrotransposon enzyme in place of the traditional reverse transcriptase and ligases found in other RNA-Seq library preparation kits. This enzyme possesses greater processivity than a retroviral reverse transcriptase (Bibillo and Eickbush 2002) and readily conducts end-to-end template jumping (Bibillo and Eickbush 2004), enabling a continuous synthesis reaction that converts RNA to cDNA and adds both sequencing adapters in a single step. By first polyadenylating the RNA fragments and then adding a polythymidine sequence to the 3' sequencing adapter to prime the reaction, this enzyme captures long (>200 bp) and short (<200 bp) RNA biotypes as well as degraded RNA fragments in a single library. Here, we demonstrate that using this kit in conjunction with the SEQuoia RiboDepletion Kit (for rRNA-derived fragment depletion) yields a comprehensive library from FFPE RNA that is comprised of both long and short RNA biotypes and therefore more representative of the complete transcriptome.

Fig. 1. Novel RNA-Seq library preparation workflow employing the SEQuoia Complete Stranded RNA Library Prep and SEQuoia RiboDepletion Kits. cDNA, complementary DNA; RNA-Seq, RNA sequencing.

Materials and Methods

Sample Description

Matched infiltrating ductal carcinoma FFPE and FF tissue samples (T2N2aM0) collected in 2018 were obtained from BioIVT (catalog #HMBREASTDC1 [FFPE] and HMBREASTDC2 [FF]). FFPE tissue was cut into 10 µm sections and stored at room temperature until used. FF tissue was stored at –80°C.

RNA Extraction

Total RNA was extracted from FFPE sections with four commonly used commercial kits: RecoverAll Multi-Sample RNA/DNA Isolation Workflow (Thermo Fisher Scientific Inc., #A26135), PureLink FFPE RNA Isolation Kit (Thermo Fisher Scientific, #K156002), AllPrep DNA/RNA FFPE Kit (QIAGEN, #80234); and RNeasy FFPE Kit (QIAGEN, #73504). RNA was purified from the FF sample using PureZOL RNA Isolation Reagent (Bio-Rad™ Laboratories, Inc., #7326880). Three replicate extraction reactions were performed for each kit using equal amounts of starting material.

Preliminary Library Preparation and Sequencing

To assess the suitability of the different RNA extraction kits for whole transcriptome analysis, 10 ng of total RNA from each extraction was depleted of ribosomal RNA (rRNA) using NEBNext rRNA Depletion Kit v2 (New England Biolabs [NEB], #E7400) according to the manufacturer's instructions. Six microliters of each ribodepletion reaction was subsequently used as input to

prepare libraries using SEQuoia Complete Stranded RNA Library Prep Kit (Bio-Rad, #17005726). Libraries were indexed during the final amplification step using SEQuoia Dual Indexed Primers Plate (Bio-Rad, #12011930). The quality and quantity of each library were evaluated prior to sequencing using the High Sensitivity DNA Kit (Agilent Technologies, Inc., #50674626) run on the Agilent 2100 Bioanalyzer (Agilent, #G2939BA) and the Qubit 1x dsDNA High Sensitivity (HS) Assay Kit (Thermo Fisher Scientific, #Q33231). Libraries were stored at –20ºC prior to sequencing using the Illumina® NextSeq 500 Sequencing System (Illumina, Inc., #SY-415-1001).

Library Construction

Three library construction workflows were evaluated to assess the suitability of pre-library ribodepletion versus post-library ribodepletion and to compare the ability of different library chemistries to construct complex libraries that represent the whole transcriptome (Figure 2). Only FF RNA extracted using the PureZOL Kit and FFPE RNA extracted using PureLink or RecoverAll Kits were used in subsequent experiments, as these samples were most suitable for high-quality library construction. To minimize variation between extraction replicates, RNA samples from each of the extraction kits were pooled and quantified prior to library construction. Libraries were constructed in triplicate using the manufacturers' instructions. All incubation and thermal cycling were performed on a C1000 Touch Thermal Cycler with 96–Deep Well Reaction Module (Bio-Rad, #1851197) using 8-Tube PCR Strips (Bio-Rad, #TLS0851) and Hard-Shell™ PCR Plates (Bio-Rad, #HSP901). Key differences for each of the three workflows are noted as follows:

- **NEBNext Workflow (NEB)** — For each NEB workflow library, 10 ng of total RNA was ribodepleted using NEBNext rRNA Depletion Kit v2 (NEB, #E7400) according to the manufacturer's instructions. Six microliters of each ribodepletion reaction was subsequently used as input to prepare libraries using NEBNext Ultra II Directional RNA Library Prep Kit for Illumina (NEB, #E7760S). Samples were fragmented by incubating at 94°C for 7 min, per the manufacturer's protocol for degraded RNA samples with a RNA integrity number (RIN) value of 2–6. NEBNext Multiplex Oligos for Illumina (Dual Index Primer Set 1) (NEB, #E7600S) were used to index libraries during the final amplification step, which included 14 PCR cycles.
- **SEQuoia Complete Workflow (SEQ)** — For each SEQ workflow library, an input of 10 ng of total RNA was used to construct libraries using SEQuoia Complete Stranded RNA Library Prep Kit (Bio-Rad, #17005726). Total RNA was fragmented by incubating at 70°C for 1 min, per the manufacturer's recommendation for degraded samples. Libraries were amplified using SEQuoia Dual Indexed Primers (Bio-Rad, #12011930) and 12 PCR cycles. Amplified libraries were depleted of cDNA constructs derived from rRNA by adding 0.5 ng of a library to a SEQuoia RiboDepletion Kit (Bio-Rad, #17006487) reaction.
- **Hybrid Workflow (Hybrid**) — For each Hybrid workflow library, 10 ng of total RNA was ribodepleted using NEBNext rRNA Depletion Kit v2 (NEB, #E7400) according to the manufacturer's instructions. Six microliters of each ribodepletion reaction was subsequently used as input to prepare libraries using SEQuoia Complete Stranded RNA Library Prep Kit and SEQuoia Dual Indexed Primers. A 70°C, 1 min fragmentation step was used. The final amplification step included 12 PCR cycles.

The quality and quantity of each library were evaluated by Agilent 2100 Bioanalyzer and the Qubit as described in the Preliminary Library Preparation and Sequencing section.

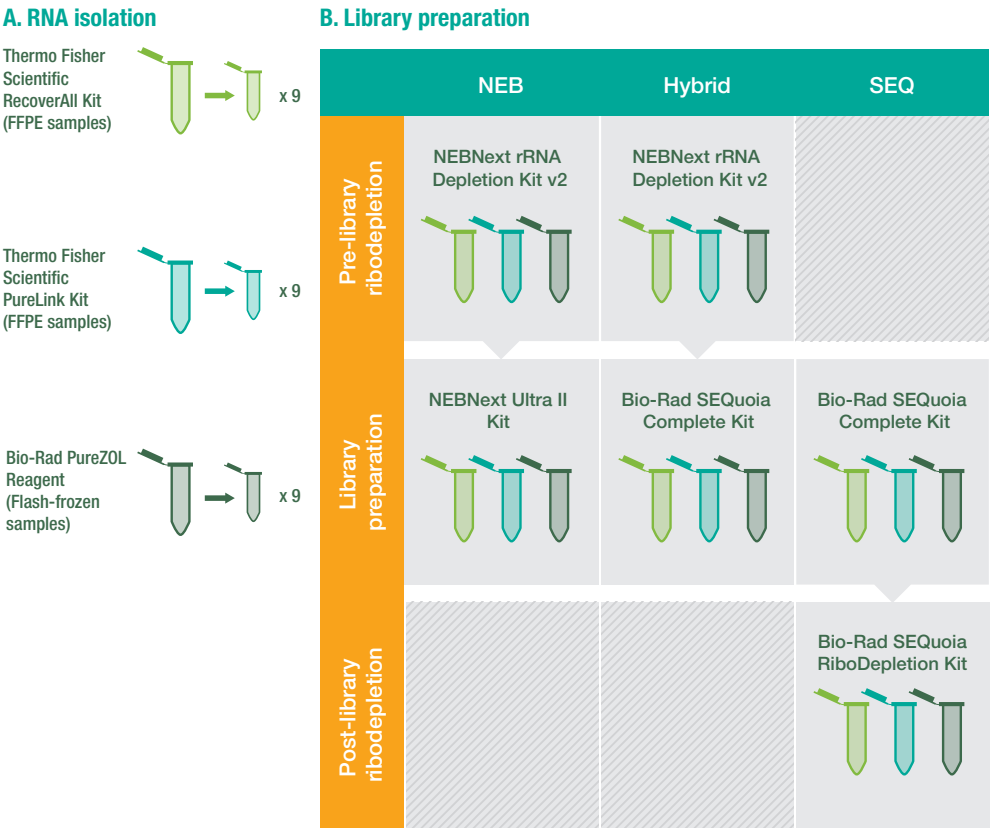
Sequencing and Data Analysis

Individual libraries were diluted to 2 nM using 10 mM Tris (pH 8.0) and 0.1 mM EDTA and then pooled and denatured according to the manufacturer's instructions for the NextSeq 500/550 Mid Output Kit v2.5, 150 Cycles (Illumina, #20024904). Denatured libraries were diluted to 10 pM in HT1 buffer and then further diluted to 1.8 pM by mixing 195 µl of the 10 pM library with 1,100 µl of HT1 buffer and 5 µl of 12.5 pM PhiX Control v3. Libraries were sequenced on Illumina NextSeq 500 Sequencing System with paired-end reads and the following cycle length parameters: read 1, 68; read 2, 8; read 3, 8; read 4, 8.

FASTQ files were generated using the BaseSpace Sequence Hub (Illumina). Results were analyzed using the SeqSense Analysis Solution Web App or using the SEQuoia Complete Analysis Toolkit installed on a Linux system. Both analysis options use the same underlying pipeline (Figure 3) with FASTQ files as input for secondary analysis: the quality of raw RNA-Seq reads is assessed using FASTQC. CutAdapt is used to trim the polyA tails and primer and adapter sequences from sequencing contigs. Using STAR aligner, the reads are aligned against the human transcriptome built from the ENSEMBL Genome Reference Consortium Human Build 38 (GRCH38). Picard tools are used as quality control (QC) for the BAM files. The estimated transcripts are summarized to the gene level and the final reports are generated in PDF, HTML, and CVS formats.

Multiplex RiboDepletion

To determine the efficiency of ribodepleting libraries en masse, 48 libraries were constructed using 100 ng Human Placenta Total RNA (Thermo Fisher Scientific, #AM7950) as input for the SEQuoia Complete Stranded RNA Library Prep Kit. Libraries were indexed using SEQuoia Dual Indexed Primers in the final amplification step. An aliquot of each library was either individually depleted (singleplex) using SEQuoia RiboDepletion Kit or pooled by mixing an equal molar ratio of the libraries then depleted (multiplex). Libraries were sequenced on the NextSeq 500 Sequencing System to a read depth of 3 million reads per library.



NEB: NEBNext rRNA Depletion Kit v2 and NEBNext Ultra II Directional RNA Library Prep Kit for Illumina

Hybrid: NEB NEBNext rRNA Depletion Kit v2 and Bio-Rad SEQuoia Complete Stranded RNA Library Prep Kit

SEQ: Bio-Rad SEQuoia Complete Stranded RNA Library Prep Kit and SEQuoia RiboDepletion Kit

Fig. 2. Experimental design to compare RNA extraction, library preparation, and ribosomal depletion strategies. Libraries were constructed in triplicate for each RNA extraction and library preparation/ribodepletion method. **A**, RNA extraction methods; **B**, library preparation/ribodepetion methods. NEB, NEBNext rRNA Depletion Kit v2 and NEBNext Ultra II Directional RNA Library Prep Kit for Illumina; Hybrid, NEB NEBNext rRNA Depletion Kit v2 and Bio-Rad SEQuoia Complete Stranded RNA Library Prep Kit; SEQ, Bio-Rad SEQuoia Complete Stranded RNA Library Prep Kit and SEQuoia RiboDepletion Kit. FFPE, formalin-fixed paraffin-embedded.

Results

RNA Extraction Quality

RNA was extracted from FFPE tissue using four commercially available extraction kits. Despite each extraction using the same amount of starting material (four 10 µm sections), the RNA yield was significantly different between the extraction kits (Table 1). The RNeasy Kit yielded the most RNA (928 ± 159 ng/µl), whereas the RecoverAll and PureLink Kits yielded a modest amount of RNA (518 ± 72 ng/µl and 371 ± 89 ng/µl, respectively). The AllPrep Kit consistently yielded the lowest amount of RNA (280 ± 21 ng/µl). Interestingly, the RIN for all four extraction kits was nearly identical (2.1–2.6), suggesting that the RNA was highly degraded in all FFPE extractions. While the RIN score is commonly used as a relative predictor of successful library preparation, its sensitivity comes into question when measuring RNA from FFPE tissue samples (Illumina 2016). The DV200 score, which reports the percentage of purified RNA with a length greater than or equal to 200 nucleotide bases, tends to be a more reliable determinant of RNA quality and was therefore used to determine the suitability of the RNA extraction methods. RNA extracted from FFPE tissue using the RecoverAll and AllPrep Kits produced the highest DV200 scores (92 ± 0.5 and 83 ± 2, respectively) followed by RNA extracted using the PureLink Kit (60 ± 12). RNA extracted using the RNeasy Kit, on the other hand, consistently had the lowest DV200 score (55 ± 0.2). As suspected, the DV200 and RIN scores did not agree with one another for all samples (Figure 4). The Bioanalyzer electropherogram traces of SEQ workflow libraries using FFPE and FF confirm that the FFPE libraries are comprised of more small RNA fragments than FF libraries, as indicated by the narrower peak in the FFPE library electropherogram (Figure 5).

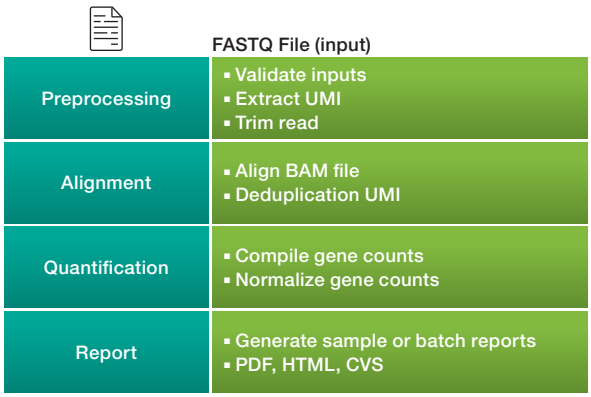


Fig. 3. SEQSense RNA-Seq data analysis pipeline. UMI, unique molecular identifier.

Unexpectedly, the quality of sequencing libraries constructed from RNeasy Kit–extracted RNA was poor. In fact, fewer than 20% of the reads were uniquely aligned. In contrast, libraries constructed from RNA extracted using the other three kits produced high-quality sequencing results (data not shown). Based on the RNA yield, DV200 scores, and sequencing results, we opted to move forward with only RNA extracted using the RecoverAll and PureLink Kits for FFPE. To confirm these two kits yielded RNA with similar quality, we assessed the correlation coefficients of lncRNA and protein-coding RNA transcripts detected at ≥1 RPKM in at least 2 replicates with ≤10% CV for RecoverAll and PureLink library replicates prepared with each workflow (Figure 6). The Pearson correlation coefficients between RecoverAll and PureLink extractions were 0.986–0.999, indicating that the libraries constructed from FFPE RNA extracted using these two methods are very similar.

Table 1. Quantity and quality metrics of RNA extracted from matched FF and FFPE breast cancer tissue. Each extraction was performed in triplicate.

Kit	Sample Type	Replicate	Yield, ng/µl	RIN	DV200	Parameter Quality
RecoverAll	FFPE	1	522	2.5	92	High
		2	589	2.6	93	High
		3	445	2.5	92	High
PureLink	FFPE	1	473	2.5	46	Low
		2	328	2.3	68	Medium
		3	311	2.4	66	Medium
AllPrep	FFPE	1	298	2.1	82	High
		2	257	2.1	83	High
		3	285	2.1	86	High
RNeasy	FFPE	1	995	2.3	55	Medium
		2	1,043	2.5	55	Medium
		3	746	2.3	55	Medium
PureZOL	FF	1	61	7.6	92*	High*
		2	47	7.7		
		3	54	7.5		

* PureZOL DV200 was calculated on a pooled sample due to the low concentrations of the individual replicates.

FF, flash-frozen; FFPE, formalin-fixed paraffin-embedded; RIN, RNA integrity number.

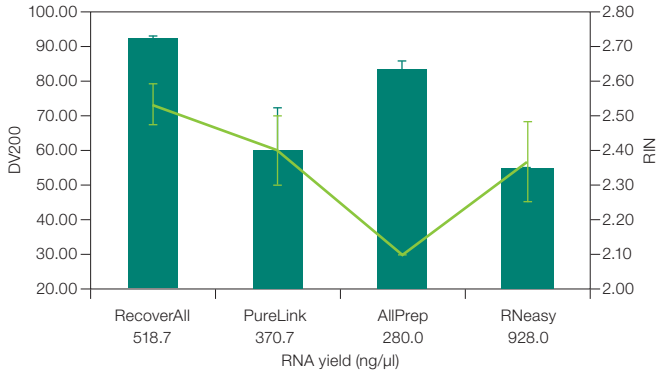


Fig. 4. Poor agreement between DV200 and RIN scores for extracted RNA samples. The DV200 (■) and RIN (—) were the average of three independent extractions using each kit. RNA yield is recorded below the chart. Error bars represent standard deviation (SD). RIN, RNA integrity number.

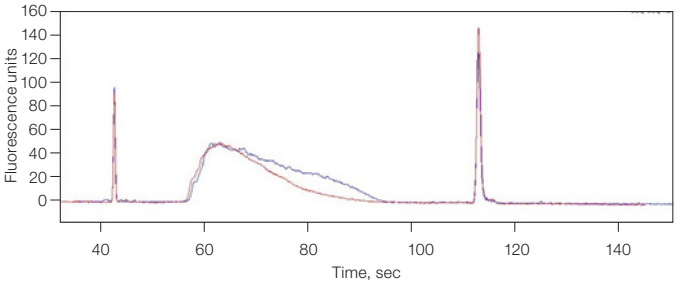


Fig. 5. Example of typical Bioanalyzer electropherogram. SEQ workflow libraries constructed using RNA from FFPE breast cancer tissue (RecoverAll, —) and matched FF tissue (PureZOL, —). FF, flash-frozen; FFPE, formalin-fixed paraffin-embedded.

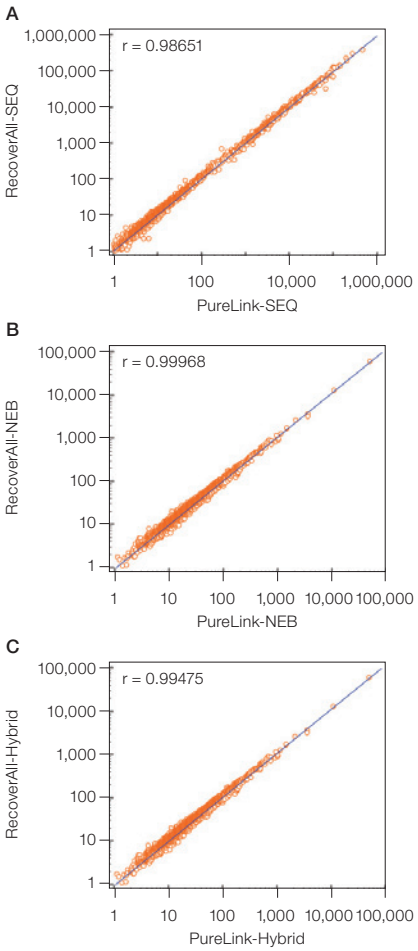


Fig. 6. FFPE RNA extracted using the RecoverAll and PureLink Kits produces similar sequencing results. Correlation coefficients of lncRNA and protein coding genes detected at 1 RPKM in at least two replicates with ≤10% CV between replicates for libraries constructed from FFPE RNAs using RecoverAll and PureLink extraction methods. **A**, SEQ workflow; **B**, NEB workflow; **C**, Hybrid workflow.
CV, coefficient of variation; lncRNA, long noncoding RNA; RPKM, reads per kilobase of transcript per million reads mapped.

Sequencing Library Quality

To evaluate the performance of the three RNA-seq workflows that compare ribodepletion strategy and library chemistry, we constructed libraries using rRNA-depleted RNA from FFPE tissue samples (NEB and Hybrid workflows) or total RNA (SEQ workflow). The libraries constructed using the SEQ workflow underwent enrichment before sequencing by using the SEQuoia RiboDepletion Kit to deplete library constructs containing cDNA inserts that originated from rRNA. After sequencing, the raw data of all three workflows were downsampled to 5 million reads/library. Analytical comparisons focused on several key metrics, including library yield, rRNA depletion efficiency, genome alignment profiles, transcript coverage, and transcript quantification (Table 2).

Library yields were consistently 10-fold greater with the SEQ workflow, which is not entirely unexpected. The SEQ workflow used 10 ng of total RNA for preparing the library. Library constructs were amplified using 12 PCR cycles to increase the amount of library material. Only 0.5 ng of library underwent the post-library ribodepletion process before sequencing. In contrast, the Hybrid and NEB workflows used 10 ng of total RNA that was first ribodepleted. Though RNA concentration was not measured following depletion, typically only 10% of the RNA added to a depletion reaction is recovered. Therefore, we can presume 1 ng of depleted RNA was added to each of the library protocols in the Hybrid and NEB workflows.

The total number of reads for the SEQ (45.5M ± 15) was also greater than the Hybrid (16.3M ± 2.8) and NEB (6.4M ± .2) workflows. It is interesting to note that the Hybrid workflow had more than twice the number of reads as the NEB workflow. Presumably, this reflects the more efficient capture of small RNA biotypes by the SEQuoia Complete Stranded RNA Library Prep Kit that is used in the Hybrid workflow.

As expected, the percentage of reads mapping to rRNA in FFPE libraries was greater in the SEQ workflow (23.5 ± 0.7) and slightly higher in the Hybrid workflow (3.5 ± 0.7) compared to the NEB workflow (0.8 ± 0.6). The small RNA fragments captured by SEQuoia Complete Stranded RNA Library Prep Kit contain a greater percentage of constructs that will not efficiently hybridize to probes in the SEQuoia RiboDepletion kit, resulting in less efficient depletion.

The percentage of uniquely mapped reads was higher in FFPE libraries constructed using the Hybrid workflow (65.5 ± 2) and the NEB workflow (78.5 ± 10.6). The lower percentage of uniquely mapped reads in the SEQ workflow FFPE libraries (39 ± 0.45) may reflect a better capture of small RNA fragments, which tend to map to multiple loci in the genome.

Capture of RNA Biotypes

A closer examination of the RNA biotypes represented in libraries from each workflow reveals another significant distinction between the workflows. Whereas the total number of genes detected was similar across the workflows and between FFPE and FF samples (Figure 7A), significantly more small RNA biotypes were captured in workflows that employed SEQuoia Complete Stranded RNA Library Prep Kit (Figure 7C). Depleting rRNA before library construction resulted in a loss of small RNA biotypes (compare SEQ to Hybrid workflow) but the SEQuoia Complete Stranded RNA Library Prep Kit still captured more small RNA biotypes that remained after depletion (compare Hybrid to NEB workflow). Most notable is the capture of miRNA by the SEQuoia Complete Library Prep Kit from total RNA (Figure 7B), which is significantly more than the number of miRNAs captured by the same method using pre-depleted RNA, indicating loss of valuable material through pre-library prep sample handling.

Table 2. Sequencing metrics of libraries created with the different library and ribosomal RNA depletion protocols.

	SEQ Workflow			Hybrid Workflow			NEB Workflow		
	RecoverAll (FFPE)	PureLink (FFPE)	PureZOL (FF)	RecoverAll (FFPE)	PureLink (FFPE)	PureZOL (FF)	RecoverAll (FFPE)	PureLink (FFPE)	PureZOL (FF)
PCR cycles	12	12	12	12	12	12	14	14	14
Library yield, nm	44.47	42.17	34.97	1.00	3.15	3.41	1.66	3.42	6.11
Mean library size, bp	300	293	453	370	360	396	322	321	413
Total reads	28,600,594	47,913,879	60,122,647	14,809,271	19,629,284	14,618,192	6,637,246	6,199,090	7,459,959
Downsampled to	5,000,000	5,000,000	5,000,000	5,000,000	5,000,000	5,000,000	5,000,000	5,000,000	5,000,000
rRNA, %	24	23	13	4	3	1	1.2	0.4	0.2
Uniquely mapped reads, %	39.59	40.22	56.93	64.08	66.96	81.57	71.00	86.00	82.00
Uniquely mapped reads	1,979,210	2,011,183	2,846,326	3,203,960	3,347,894	4,078,173	3,527,169	4,276,429	4,120,035
Transcripts RPKM ≥ 0.3	16,039	16,846	17,137	18,487	18,698	17,540	17,674	18,192	17,574
Transcripts RPKM ≥ 1.0	13,187	13,314	13,996	15,583	15,701	14,923	14,925	15,245	14,589
Transcript integrity score	42.9461	39.9025	54.1157	44.2536	52.6077	55.2504	42.4337	48.0525	44.8903

FF, flash-frozen; FFPE, formalin-fixed paraffin-embedded; RPKM, reads per kilobase of transcript per million reads mapped; rRNA, ribosomal RNA.

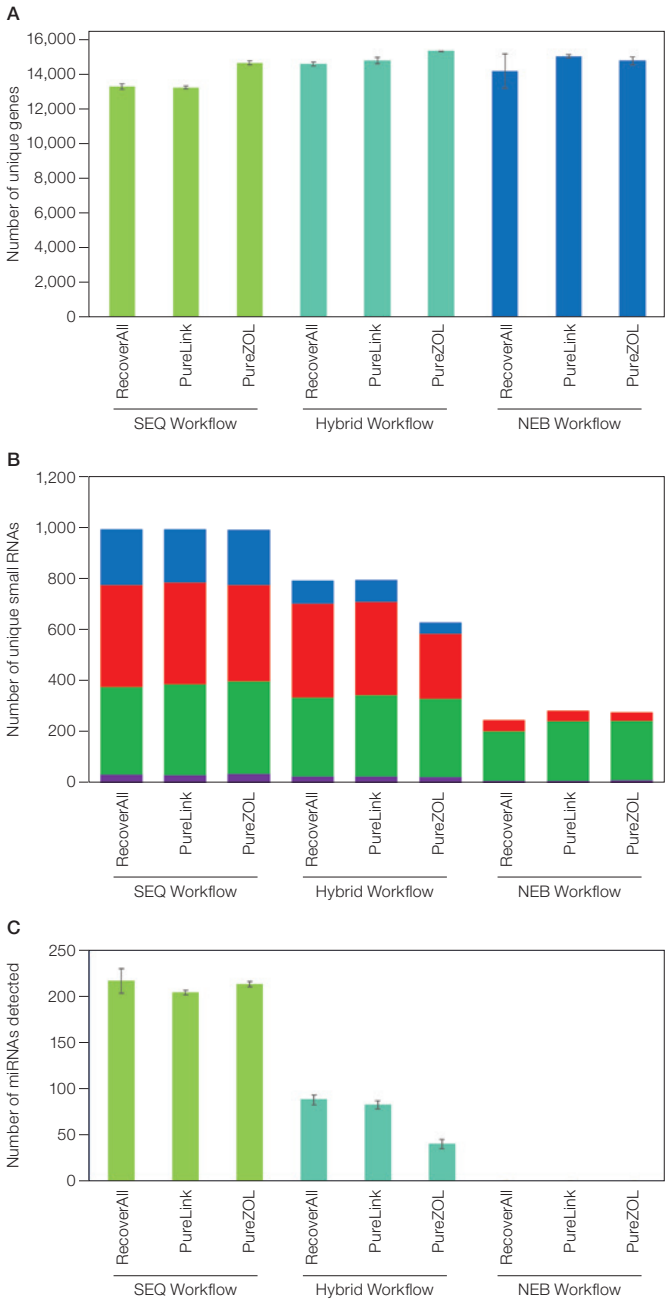


Fig. 7. More small RNAs were detected with post-library ribodepletion than pre-library ribodepletion. The number of unique transcripts per RNA biotype detected in libraries prepared using each of the workflows are shown in the following panels: **A**, unique long RNAs (protein coding and lncRNAs) detected at ≥5 reads; **B**, combined unique small RNAs, including miRNA (■), tRNA (■), snoRNA (■), and snRNA (■), detected with ≥5 reads each; **C**, unique miRNAs detected at ≥5 reads. Error bars represent SD. lncRNA, long noncoding RNA; miRNA, microRNA; RPKM, reads per kilobase of transcript per million reads mapped; SD, standard deviation; snRNA, small nuclear RNA; snoRNA, small nucleolar RNA; tRNA, transfer RNA.

Uniformity and Continuity of Coverage

To assess the uniformity of transcript coverage, we calculated the median coefficient of variation (CV) for the top 1,000 highly expressed transcripts from the libraries constructed using the SEQ, Hybrid, and NEB workflows (Figure 8). We observed that both SEQ workflow (CV = 0.71) and NEB workflow (CV = 0.68) libraries have an increased median CV compared to Hybrid workflow libraries (CV = 0.63, $P < 0.05$) but this difference is not statistically significant. At 10M read depth, the SEQ workflow had an equivalent CV (CV = 0.56) to the Hybrid workflow (CV = 0.58).

To assess the continuity of transcript coverage in SEQ and NEB workflow libraries using pre– and post–rRNA depletion methods, we calculated the percentage of gaps (gap %) without read coverage across the top, middle, and bottom 1,000 expressed transcripts (Figure 9) using RNA-Seq QC tools from the Broad Institute (DeLuca 2012). The gap percentage is defined as the total cumulative gap length divided by the total cumulative transcript length in each library (RNA-SEQ QC, Broad Institute). Each comparison shows the percentage of transcript coverage in each replicate library. In the top 1,000 expressed transcripts, the gap % for SEQ workflow libraries was 15.8% compared to 8.9% in the Hybrid workflow and 9.9% in the NEB workflow libraries (Figure 9A). We have also detected 18.2% of unique small RNAs in the top 1,000 expressed transcripts of the SEQ workflow libraries but not in the Hybrid or NEB workflow libraries. This suggests that the increased gap % in the top 1,000 expressed transcripts of SEQ workflow libraries might be due to the presence of more unique small RNA biotypes in addition to the key lncRNA and protein-coding genes. In bottom expressed transcripts (Figure 9C), no small RNAs were detected in any library preparation. However, in the middle 1,000 expressed transcripts (Figure 9B), SEQ workflow libraries detected 0.2% of small RNAs but none were detected in the NEB or Hybrid workflow libraries. It is notable that in the middle and bottom 1,000 expressed transcripts, the SEQ workflow library preparation exhibited the smallest gap % compared with Hybrid and NEB workflow libraries. Furthermore, the Hybrid workflow libraries showed less gap % in these same categories than the NEB workflow libraries, indicating that the SEQuoia Complete Stranded RNA Library Prep Kit results in better continuity of transcript coverage than the other workflows.

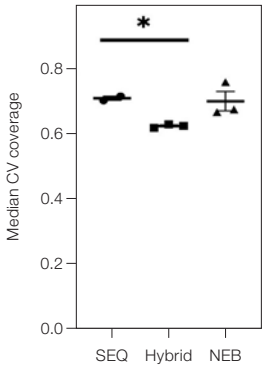


Fig. 8. The median CV of coverage of the top 1,000 expressed transcripts. RecoverAll RNA was used in the SEQuoia Complete (SEQ), NEB, and Hybrid workflows. Results were reported by the CollectRnaSeqMetrics utility from Picard tools. Data represents mean and SEM from SEQ (n = 2), NEB (n = 3), and Hybrid (n = 3) workflow libraries. * $P < 0.05$ (Student's t-test). CV, coefficient of variation; SEM, standard error of the mean.

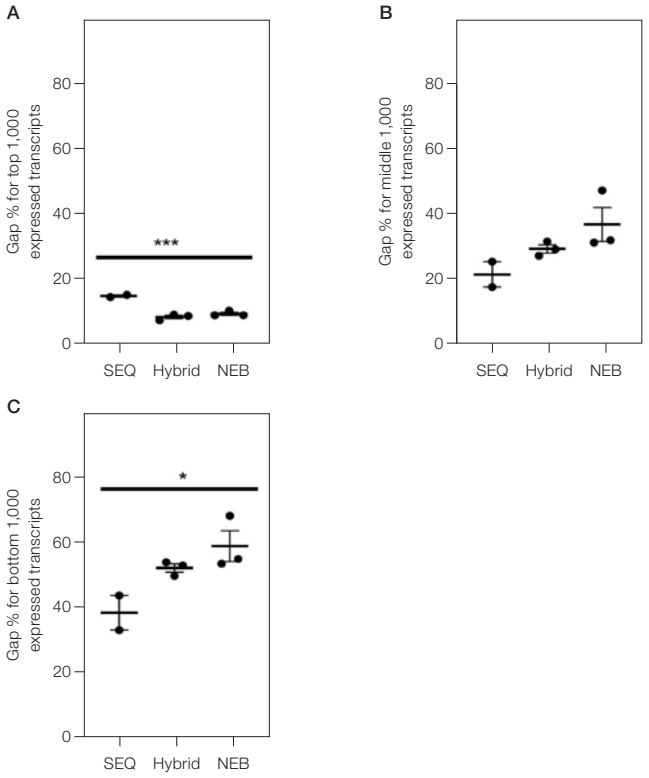


Fig. 9. Percentage of gaps in transcript for the top, middle, and bottom 1,000 expressed genes. A, top 1,000 expressed genes; **B**, middle 1,000 expressed genes; **C**, bottom 1,000 expressed genes. In the top 1,000 expressed transcripts, 18.2% of unique small RNAs were detected in SEQ workflow but not in the NEB or Hybrid workflows. Data represents mean and SEM from SEQ (n = 2), NEB (n = 3), and Hybrid (n = 3) workflow libraries. * $P < 0.05$, *** $P < 0.0005$ (one-way ANOVA). ANOVA, analysis of variance; SEM, standard error of the mean.

Comprehensive Transcriptome Coverage

There is a significant difference in the RNA biotypes captured by each of the workflows. Whereas only slightly different subsets of long RNA biotypes were observed between the different workflows (Figure 10A), the SEQ workflow libraries were much more complex in terms of the small RNA biotypes. Notably, over 500 unique small RNAs including miRNA, tRNA, snRNA, and snoRNA subtypes were detected in SEQ workflow libraries (Figure 10B), compared with only seven small RNAs in the NEB workflow libraries.

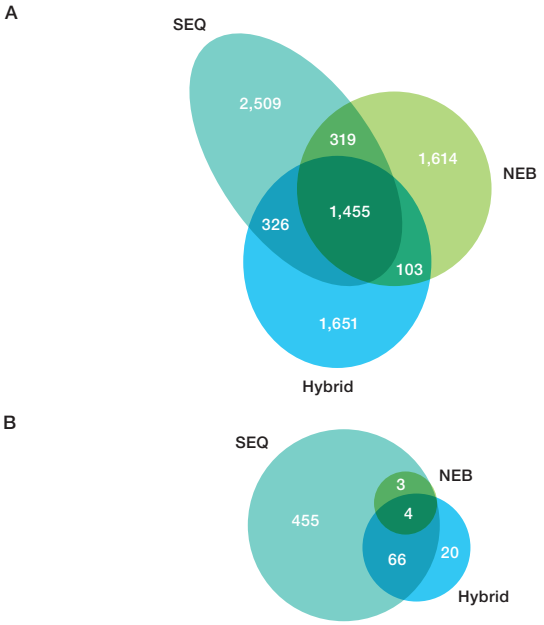


Fig. 10. SEQuoia library workflow provides more complete transcriptome coverage of small RNAs in addition to long RNAs. Venn diagrams demonstrating number of genes detected at ≥1 RPKM and ≤10% CV in libraries constructed from FFPE RNA extracted using the RecoverAll Kit with the same library prep/ribodepletion. **A**, lncRNA and protein coding genes; **B**, small RNAs (miRNA, tRNA, snoRNA, snRNA). CV, coefficient of variation; lncRNA, long noncoding RNA; miRNA, microRNA; RPKM, reads per kilobase of transcript per million reads mapped; snRNA, small nuclear RNA; snoRNA, small nucleolar RNA; tRNA, transfer RNA.

Multiplexing Ribodepletion Reactions

In a separate experiment, the multiplexing capability of the SEQ workflow was also evaluated. Forty-eight RNA-Seq libraries were constructed and indexed using SEQuoia Complete Stranded RNA Library Prep Kit. Each library was split into two; half of each library was ribodepleted individually using SEQuoia RiboDepletion Kit, and the other half of each library was pooled by mixing an equimolar ratio of libraries and then depleted en masse using the SEQuoia RiboDepletion Kit. The percentage of reads mapping to rRNA in libraries that were depleted en masse was consistently within 5% of the same library depleted in a singleplex reaction (Figure 11A). In addition, the transcriptome profile is preserved when libraries are depleted en masse, as illustrated in Figure 11B where the Pearson correlation coefficient of transcripts between the multiplexed and single ribodepletion reactions ranged from 0.977 to 0.996, with an average of 0.992.

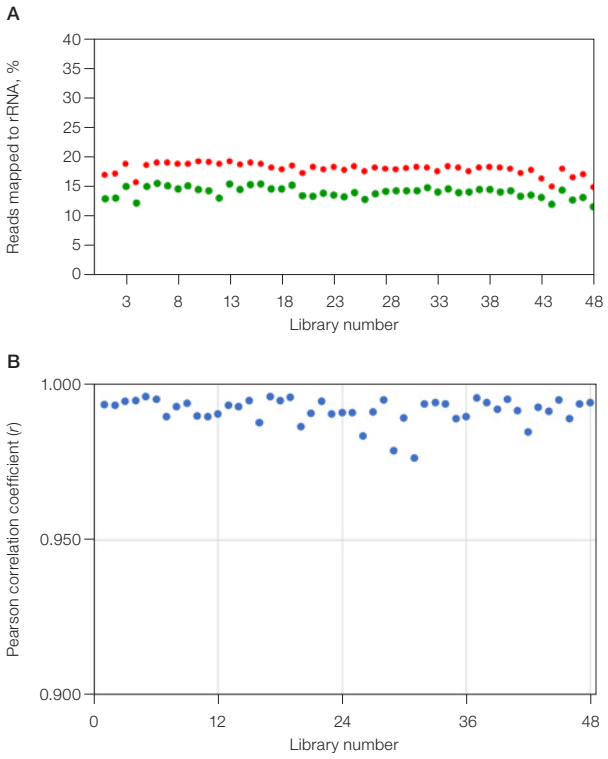


Fig. 11. Multiplexing capability of SEQuoia RiboDepletion Kit. Forty-eight libraries were constructed using the SEQuoia Complete Stranded RNA Library Prep Kit. An aliquot of each library was either individually depleted using the SEQuoia RiboDepletion Kit or pooled by mixing an equal molar ratio of the libraries and then depleted. **A**, the percentage of reads mapped to rRNA for each library are plotted (●, multiplexed depletion; ●, singleplex depletion). **B**, the number of total unique transcripts with RPKM > 0.1 in the libraries depleted in the pooled reaction was compared to the number in the individually depleted libraries. The Pearson correlation coefficient was calculated between the individually and multiplexed ribodepleted libraries. RPKM, reads per kilobase of transcript per million reads mapped; rRNA, ribosomal RNA.

Discussion

RNA-Seq can be a powerful tool for elucidating the molecular profile of RNA isolated from FFPE tumor samples. However, to realize the full potential of this sample type many challenges must be overcome. One significant challenge is the degradation and chemical modifications to nucleic acids caused by the fixation and storage process, as these interfere with the chemistry involved in preparing and sequencing samples and thus may confound the fidelity of RNA expression profiles (Groelz et al. 2013). In addition to the low RNA quality, low RNA yield (<100 ng) is also common in FFPE samples and can lead to low complexity and suboptimal results during RNA-Seq (Adiconis et al. 2013). Another challenge is the decision of which RNA biotypes to study. Biomarker discovery research using FFPE tissue samples has concentrated on elucidating differential expression patterns of long RNA biotypes, due mainly to the preponderance of pre-existing data obtained from FF tissue samples using PCR techniques (Spencer et al. 2013). In recent years, there is growing interest in exploring alterations in the expression of small RNA biotypes as potential biomarkers of cancer diagnosis and progression (Marino et al. 2014). Interrogating the whole transcriptome for both long and short RNA biotypes requires the construction of multiple libraries, which is costly and labor intensive both from the bench and bioinformatics perspective. The results of this study demonstrate a novel approach to the library preparation and ribodepletion that captures more of the transcriptome in a single library, enabling a researcher to interrogate both long and short RNA biotypes in a single library.

Using the SEQuoia Complete Stranded RNA Library Prep Kit greatly improves the number of molecular targets captured even from modest amounts of degraded RNA from FFPE tissues. The number of unique long transcripts captured by the SEQ and Hybrid workflows was similar to the NEB workflow, consistent with what others have reported when comparing capture efficiencies between library prep kits (Ribarska et al. 2022). In contrast, the number of small RNA biotypes captured in the SEQ workflow was significantly greater than in the Hybrid and NEB workflows. Importantly, over 100 unique miRNAs were consistently detected in the SEQ workflow from both FFPE and FF RNA samples. Deregulation of miRNA is thought to be a major determinant for the initiation and progression of breast cancer and may represent novel therapeutic targets and biomarkers of early disease (van Schooneveld et al. 2015). Correlating the dysregulation of miRNA expression profiles to changes in coding RNA expression has been a challenge up until now because multiple libraries are required: one that captures long RNA biotypes and one that captures only miRNAs. Each library is constructed, sequenced, and analyzed individually. The SEQ workflow represents an opportunity to study the expression profile of miRNA in parallel with long coding and noncoding RNAs using data obtained from a single library, which may reveal new insights.

Depleting rRNA-derived library constructs to enrich a library, as opposed to depleting total RNA of rRNA prior to library construction, is a strategy well-suited to highly fragmented or low-abundance RNA samples. Depletion of rRNA from total RNA prior to library construction requires multiple pipetting steps and bead-based cleanup steps — both considerable risks for material loss. A post-library depletion strategy allows RNA, which is highly susceptible to degradation, to be converted into more stable cDNA early in the workflow. The addition of sequencing adapters to short RNA fragments lengthens the fragments and reduces the risk of material loss during pipetting and cleanup steps. Our results support this notion: the SEQ workflow contained considerably more unique small RNA transcripts than the NEB and Hybrid workflows. The difference between the SEQ workflow and the Hybrid workflow, where pre-depleted RNA was the input for the SEQuoia Complete library preparation, is most striking. These results clearly show that small RNAs are lost through the ribodepletion process when conducted prior to library preparation.

We have also shown that the SEQ workflow more efficiently captures scarce RNA biotypes. Depleting RNA from total RNA prior to library construction resulted in bigger gaps in the transcript profile of lowly expressed genes, consistent with the idea that converting RNA to cDNA and increasing the fragment length by adding adapters helps retain more of the transcriptome.

Of course, depletion efficiency will be lower when constructing a library with small RNA biotypes or when using highly degraded RNA, which can be low quality and low yield. The SEQuoia RiboDepletion Kit utilizes a biotinylated labeled probe and streptavidin-bead capture technique to remove rRNA-derived constructs. Short library constructs from highly degraded and small RNAs may not bind to the ribodepletion probes as efficiently as long constructs, resulting in lower depletion efficiency and a greater percentage of sequencing reads mapping to rRNA (Lin et al. 2019, Herbert et al. 2018). While the potential shortcoming of high residual rRNA-derived library constructs may be seen as a negative in that it decreases the amount of relevant sequencing reads and increases the sequencing cost, one should weigh this against the gain: a more complex dataset that represents more of the transcriptome, even from highly degraded and low RNA yield samples.

In addition, post-library ribodepletion allows multiplexing of the ribodepletion reaction, which can reduce costs considerably. Whereas depleting rRNA from total RNA before library construction requires each sample to be depleted separately, one may consider pooling multiple libraries when using a post-library depletion strategy. The libraries can be indexed during the library amplification step prior to depletion. We demonstrate in this study high concordance in the number of total unique transcripts found in libraries that are depleted either in bulk (48 pooled libraries) or as single reactions. Moreover, the depletion efficiency was relatively similar; the percentage of reads mapping to rRNA in libraries that were depleted in bulk was only slightly higher than in the same libraries depleted in individual reactions. In sum, to construct a more complex library that includes both long and small RNA biotypes, a post-library ribodepletion strategy is ideal.

Conclusions

The SEQuoia Complete Stranded RNA Library Prep Kit, a novel library preparation chemistry that leverages a T4 polynucleotide kinase and a proprietary retrotransposon instead of a retroviral reverse transcriptase and ligases, comprehensively captures the transcriptome, enabling the interrogation of both long and short RNA biotypes in a single library. Furthermore, depleting rRNA after library construction using SEQuoia RiboDepletion Kit preserves small RNA biotypes and fragments from FFPE samples, resulting in a richer, more complex dataset that more accurately represents the whole transcriptome. The ability to use an input RNA as low as 10 ng for RNA-Seq library preparation expands the opportunities for RNA-Seq in oncology research using limited and/or degraded tissue samples. Based on the results of this study, we suggest translational researchers carefully consider the library preparation strategy employed to ensure successful biomarker discovery from FFPE tumor samples.

References

Adiconis X et al. (2013). Comparative analysis of RNA sequencing methods for degraded or low-input samples. Nat Methods 10, 623–629.

Bibillo A and Eickbush TH (2002). High processivity of the reverse transcriptase from a non-long terminal repeat retrotransposon. J Biol Chem 277, 34,836–34,845.

Bibillo A and Eickbush TH (2004). End-to-end template jumping by the reverse transcriptase encoded by the R2 retrotransposon. J Biol Chem 279, 14,945–14,953.

DeLuca DS et al. (2012). RNA-SeQC: RNA-seq metrics for quality control and process optimization. Bioinformatics 28, 1,530–1,532.

Groelz D et al. (2013). Non-formalin fixative versus formalin-fixed tissue: A comparison of histology and RNA quality. Exp Mol Pathol 94, 188–194.

Herbert ZT et al. (2018). Cross-site comparison of ribosomal depletion kits for Illumina RNAseq library construction. BMC Genomics 19, 199.

Illumina, Inc. (2016). Evaluating RNA quality from FFPE samples. https://www.illumina.com/content/dam/illumina-marketing/documents/products/technotes/evaluating-rna-quality-from-ffpe-samples-technical-note-470-2014-001.pdf, accessed October 18, 2022.

Lin X et al. (2019). A comparative analysis of RNA sequencing methods with ribosome RNA depletion for degraded and low-input total RNA from formalin-fixed and paraffin-embedded samples. BMC Genomics 20, 831.

Marczyk M et al. (2019). The impact of RNA extraction method on accurate RNA sequencing from formalin-fixed paraffin-embedded tissues. BMC Cancer 19, 1,189.

Marino ALF et al. (2014). MicroRNA expression as risk biomarker of breast cancer metastasis: A pilot retrospective case-cohort study. BMC Cancer 14, 739.

Masuda N et al. (1999). Analysis of chemical modification of RNA from formalin-fixed samples and optimization of molecular biology applications for such samples. Nucleic Acids Res 27, 4,436–4,443.

Naphade S et al. (2022). Systematic comparative analysis of strand-specific RNA-seq library preparation methods for low input samples. Sci Rep 12, 1,789.

Palomares MA et al. (2019). Systematic analysis of TruSeq, SMARTer and SMARTer Ultra-Low RNA-Seq kits for standard, low and ultra-low quantity samples. Sci Rep 9, 7,550.

Patil VS et al. (2014). Gene regulation by non-coding RNAs. Crit Rev Biochem Mol Biol 49, 16–32.

Pennock ND et al. (2019). RNA-Seq from archival FFPE breast cancer samples: Molecular pathway fidelity and novel discovery. BMC Med Genomics 12, 195.

Ransohoff JD et al. (2018). The functions and unique features of long intergenic non-coding RNA. Nat Rev Mol Cell Biol 19, 143–157.

Ribarska T et al. (2022). Optimization of enzymatic fragmentation is crucial to maximize genome coverage: A comparison of library preparation methods for Illumina sequencing. BMC Genomics 23, 92.

Sarantopoulou D et al. (2019). Comparative evaluation of RNA-Seq library preparation methods for strand-specificity and low input. Sci Rep 9, 13,477.

Spencer DH et al. (2013). Comparison of clinical targeted next-generation sequence data from formalin-fixed and fresh-frozen tissue specimens. J Mol Diagn 15, 623–633.

The Cancer Genome Atlas Network. (2012). Comprehensive molecular portraits of human breast tumours. Nature 490, 61–70.

van Schooneveld E et al. (2015). Dysregulation of microRNAs in breast cancer and their potential role as prognostic and predictive biomarkers in patient management. Breast Cancer Res 17, 21.

Visit **bio-rad.com/NGS** for more information.

Bio-Rad Laboratories, Inc.

Life Science Group

Website *bio-rad.com* ***USA*** *1 800 424 6723* ***Australia*** *61 2 9914 2800* ***Austria*** *00 800 00 24 67 23* ***Belgium*** *00 800 00 24 67 23* ***Brazil*** *4003 0399*
Canada *1 905 364 3435* ***China*** *86 21 6169 8500* ***Czech Republic*** *00 800 00 24 67 23* ***Denmark*** *00 800 00 24 67 23* ***Finland*** *00 800 00 24 67 23*
France *00 800 00 24 67 23* ***Germany*** *00 800 00 24 67 23* ***Hong Kong*** *852 2789 3300* ***Hungary*** *00 800 00 24 67 23* ***India*** *91 124 4029300* ***Israel*** *0 3 9636050*
Italy *00 800 00 24 67 23* ***Japan*** *81 3 6361 7000* ***Korea*** *82 2 3473 4460* ***Luxembourg*** *00 800 00 24 67 23* ***Mexico*** *52 555 488 7670*
The Netherlands *00 800 00 24 67 23* ***New Zealand*** *64 9 415 2280* ***Norway*** *00 800 00 24 67 23* ***Poland*** *00 800 00 24 67 23* ***Portugal*** *00 800 00 24 67 23*
Russian Federation *00 800 00 24 67 23* ***Singapore*** *65 6415 3188* ***South Africa*** *00 800 00 24 67 23* ***Spain*** *00 800 00 24 67 23* ***Sweden*** *00 800 00 24 67 23*
Switzerland *00 800 00 24 67 23* ***Taiwan*** *886 2 2578 7189* ***Thailand*** *66 2 651 8311* ***United Arab Emirates*** *36 1 459 6150* ***United Kingdom*** *00 800 00 24 67 23*

Bulletin 3421 Ver A US/EG *22-0657 1022 Sig 0122*