A comparative analysis of library preparation technologies for RNA sequencing from FFPE samples

Giulia Corbet, Jen Pavlica, Kailee Reed, Thomas Harrison, Travis Sanders, David Gelagay, Josh Haimes, Ariele Hanek, Kristina Giorda, Brian Kudlow – Watchmaker Genomics

WATCHMAKER GENOMICS

Abstract #2940

Introduction

Formalin-fixed, paraffin-embedded (FFPE) samples are an invaluable resource in the oncology space, providing access to a vast library of archived diseased tissue samples paired with relevant donor information. Despite the broad utility of these samples, RNA extracted from FFPE tissue is typically difficult to process due to the presence of residual crosslinks and its degraded nature. Further, these samples often vary widely in performance as the fixation process, block age, block storage, and extraction method can impart large impacts on resulting template quality. As a result, robust and reproducible RNA sequencing from FFPE-derived RNA remains a challenge with unpredictable and high failure rates.

We evaluated four commercially available Whole Transcriptome Analysis (WTA) solutions to determine which performed best with this challenging sample type with respect to library complexity, inter-input and intra-sample concordance, and overall reproducibility. Matched fresh frozen and FFPE liver samples were used such that the fresh frozen data set serves as a comparative truth set for the FFPE data.

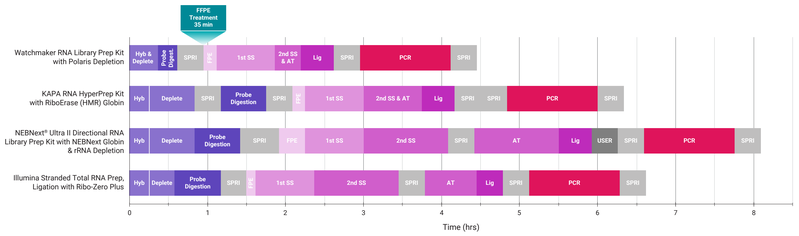

Figure 1. Whole transcriptome library prep workflow of four commercially available kits. The same fundamental steps are present in all workflows. Some workflows are more streamlined — combining steps, shortening incubations, and bypassing the need for additional SPRI cleanups. Only the Watchmaker chemistry includes an optional FFPE treatment step. Workflow times for the different kits vary, ranging from 4.5 to 8 hours.

Materials and Methods

Samples: RNA samples from matched fresh frozen and FFPE liver tissues were purchased from Biochain. FFPE RNA quality was assessed by Agilent Tapestation to obtain both a DV200 (64%) and RIN (1.2) score.

Library preparation and sequencing: 100 ng and 10 ng of total RNA derived from FFPE or matched fresh frozen tissue was used as input into depletion following each kit's available protocol. Following depletion, library preparation was performed according to manufacturer instructions. The Watchmaker libraries followed the stubby adapter library preparation recommendations. Libraries were pooled to be equimolar and sequenced on an Illumina NextSeq2000 using the P2, 200 cycle chemistry with 75 nt paired-end reads.

Data analysis: For analysis, all libraries were subsampled to 10.8M read pairs per library. Standard RNA-seq metrics such as strandedness, duplication rate, and sequence composition were generated using Picard Tools RNASeqMetrics. Unique genes in each library were detected using featureCounts with an applied deduplicated readcount cutoff of 5. Differential expression analysis utilized DESeq2 which employs a variance stabilizing transformation to stabilize variance for downstream statistical analysis

Library Complexity and Gene Detection Sensitivity

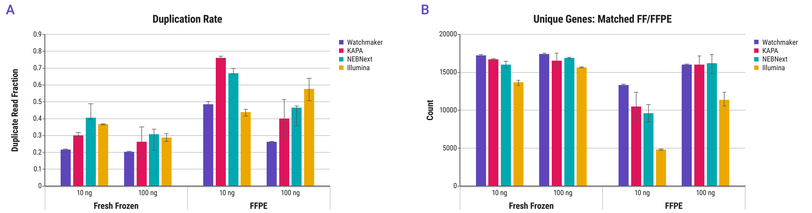

Figure 2. Improve library complexity with lower duplication rate and more unique genes detected. (A) Library duplication rate is impacted by the quantity and quality of the RNA input for all chemistries. **(B)** With high-input (100 ng) and high-quality (fresh frozen) RNA, all chemistries show similar gene detection sensitivity with only slight differences in the number of unique genes detected. As inputs lower and quality decreases, clear differences emerge. The Watchmaker chemistry detects the most unique genes, followed by KAPA and NEBNext with Illumina performing poorly with this sample type and input.

Figure 3. RNA Seq chemistry used changes library composition. The distribution of sequencing reads across exonic, intronic, intergenic, and ribosomal RNA varies by chemistry, sample, and — in some cases — input amount. NEBNext and Illumina libraries generated with FFPE demonstrate high residual ribosomal RNA reads indicating poor depletion and reducing sequencing efficiency. KAPA chemistry demonstrates the lowest residual rRNA reads but shows variance across inputs for both high-quality and FFPE samples. Watchmaker chemistry is consistent between 100 ng and 10 ng inputs for both high-quality and FFPE samples, indicating confident representation even for limited sample input.

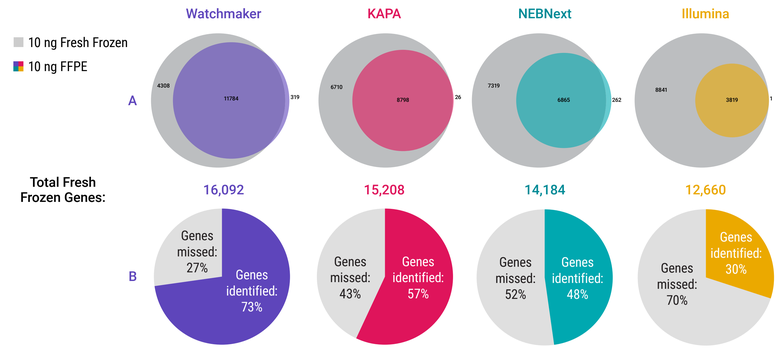

Figure 4. Improved gene detection with FFPE samples. (A) Gene overlap analysis of 10 ng fresh frozen and 10 ng FFPE. The fresh frozen data set is used as the truth set to determine how much potential information can be accessed with FFPE samples for a given chemistry. Only genes detected in both technical replicates were included in the analysis. Unique genes were identified using featureCounts with a cutoff of 5 deduplicated raw reads. **(B)** Pie charts demonstrate the amount of data retained for FFPE analysis for each chemistry. For all chemistries, most genes detected in the FFPE sample are also present in the fresh frozen libraries. The Watchmaker chemistry detects more genes overall, more genes with FFPE samples, and demonstrates the highest overlap (73%) in genes detected in both fresh frozen and FFPE samples.

Capture More with Challenging Samples

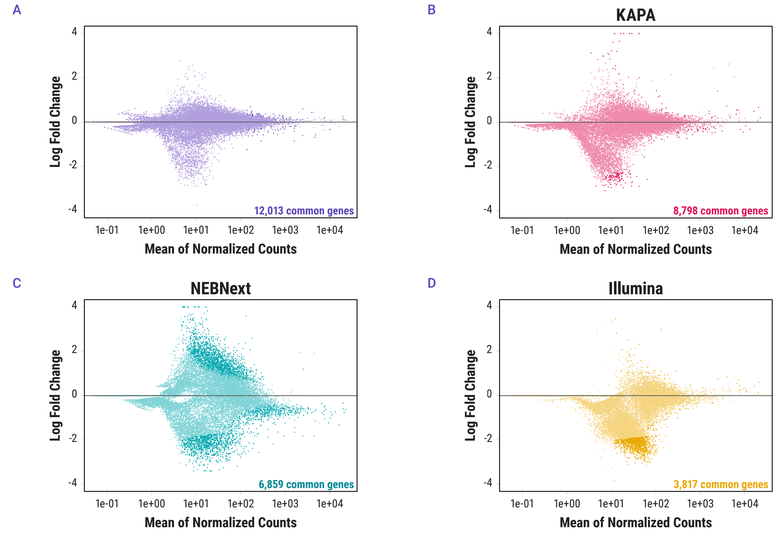

Figure 5. Reduce data distortion at low inputs. Differential expression analysis via DESeq2 between averaged 100 ng FFPE (control) and 10 ng FFPE (experimental) libraries for each of **(A)** Watchmaker, **(B)** KAPA, **(C)** NEBNext and **(D)** Illumina. Only commonly identified genes were included in the analysis. The number of common genes between the 100 ng and 10 ng data sets is indicated in the lower right of each plot. Dark data points indicate genes identified as differentially expressed. The Watchmaker chemistry outperforms in terms of more common genes detected and fewer genes identified as differentially expressed.

Conclusions

- All chemistries tested generate quality libraries when RNA inputs are of high quality and quantity. Chemistries are differentiated in performance as RNA quality or input decreases. Overall, the Watchmaker chemistry acceses more information with challenging samples.
- Libraries prepared from challenging, low-input, and degraded RNA samples demonstrate increased complexity with the Watchmaker chemistry due to core features of the kit, including:
  - Novel FFPE decrosslinking step
  - A reverse transcriptase specifically engineered for improved conversion of RNA to cDNA
  - Fewer bead purification steps, thereby preventing sample loss
- The Watchmaker solution demonstrates higher performance and promotes confidence with low-input and degraded samples as evidenced by:
  - Greater gene detection overlap of Fresh Frozen and FFPE samples
  - Less data distortion with differential expression analysis for low-input FFPE samples