HIGH-THROUGHPUT PREPARATION OF mRNA CAPTURE LIBRARIES FOR NEXT-GENERATION SEQUENCING

John Palys[1], Zach Smith[2], Sandra Theall[1], Spencer Debenport[1] Eleanor Cowley[1], Jonathan Nowacki[1], Ranjit Kumar[1], Rachel W. Kasinskas[1] and Marsha McMakin[1]
[1]Roche Diagnostics, 200 Ballardvale St #250, Wilmington, MA 01887
[2]Beckman Coulter US, 250 South Kraemer Boulevard, Brea, CA 92821

INTRODUCTION

mRNA-seq is widely applied in studies of differential gene expression, allele-specific expression, and alternative splicing, and is a rapidly-emerging field of interest. Since 2010, the annual number of PubMed documents referencing this topic has increased 12-fold. Despite the utility and applicability of this technique, mRNA-seq Library construction remains a multi-step, time-consuming process, and when undertaken manually, outcomes are dependent upon the skill and accuracy of the operator.

To facilitate high-throughput preparation of mRNA-seq libraries and to alleviate shortcomings associated with manual processing, we have demonstrated an automated protocol for the preparation of mRNA-seq libraries for use on Illumina's® next-generation sequencing platforms using the KAPA mRNA HyperPrep Kit on the Beckman Coulter Biomek i7 Hybrid Automated Workstation.

In this study we compare sequencing outcomes of automated library construction against those derived from manually-prepared libraries. Comparisons are made over a 20-fold range of inputs, from 50 to 1000 ng per preparation.

During this study we encountered higher-than-expected ribosomal RNA carryover (~6%) in the lowest RNA inputs assayed, although carryover at higher inputs was low and uniform. We demonstrate that increasing the mRNA-capture dissociation temperatures reduces carryover to uniformly low levels throughout the range of input RNA assayed.

Finally, data from a high-throughput preparation of constant-input samples are presented, to assess yield and sequencing-metric uniformity across samples.

WORKFLOW, MATERIALS, METHODS

Input RNA: Universal Human Reference RNA (DNase-treated total RNA pooled from 10 human cell line extracts – Stratagene) with average RIN ~ 8 was used as input material.

Sequencing: libraries were quantified, normalized, pooled, and sequenced (2 x 76) on an Illumina NextSeq 550.

Data Analysis: Sequencing reads were downsampled to 6 million read pairs per sample and analyzed using an in-house pipeline, with alignment to hg19 reference and processing with Picard, RNAseq-QC, and kallisto.

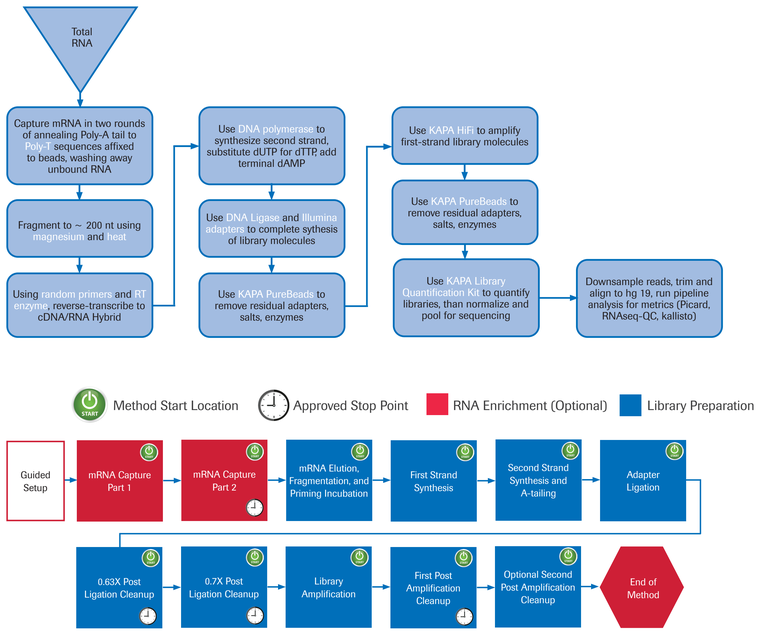

Figure 1. Overview of the mRNA capture, sequencing-library preparation, and bioinformatics analysis employed in manual and automated processing (top), and detail of the library-preparation process as implemented on the Biomek i7 (bottom). *(Detail figure courtesy of Beckman Coulter.)*

ACHIEVE CONSISTENT AND REPRODUCIBLE SEQUENCING RESULTS OVER A RANGE OF INPUTS

Libraries were prepared and analysed in parallel using both the automated method and a manual protocol (n=4, except for the 250-ng automated preparation where n=8). Sequencing metrics (including the percentage of reads mapped to the genome, coverage depth, number of transcripts detected, number of genes detected, and the percentage of residual ribosomal RNA reads) were consistent across inputs and between libraries derived from both methods.

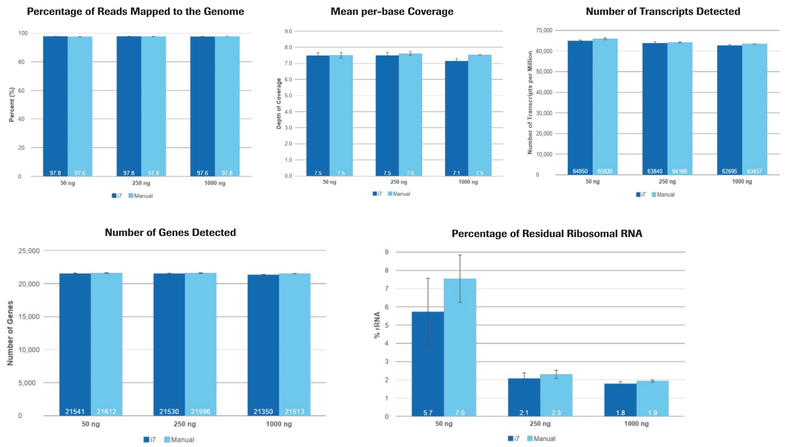

Figure 2. By several sequencing metrics, results for all libraries were consistent across inputs and between libraries prepared using both manual and automated methods. The percentage of residual rRNA was more variable than the other metrics; while values of ~2% are typical for mRNA-capture workflows, the results shown here show ~6% residual rRNA for the lowest-input (50 ng) preparations for both automated and manual preparations. While consistent between methods, this outcome is not ideal. Steps to remediate this result are discussed in the next section.

TEMPERATURE MODIFICATION REDUCES RESIDUAL rRNA

mRNA capture is achieved by annealing the poly-A tail of mature RNA transcripts to poly-T oligos affixed to capture beads, followed by washing away unannealed material including rRNA, tRNA, regulatory RNA, etc. The process is initiated by raising the temperature of the reaction vessel (65°/ 70° in the original protocol) to dissociate all molecules, followed by reducing the temperature to 20° to anneal. Reasoning that increased rRNA contamination in cases of low input samples is a result of promiscuous binding by non-mRNA, we performed a pair of capture experiments using the original temperature profile as well as an elevated-temperature profile (70°/ 75°), for 50 and 100-ng inputs, n=4. Sequencing was performed as described, but in this case downsampled to 11.5 million read pairs per sample. As seen in **Figure 3**, elevated temperatures led to reduced rRNA contamination.

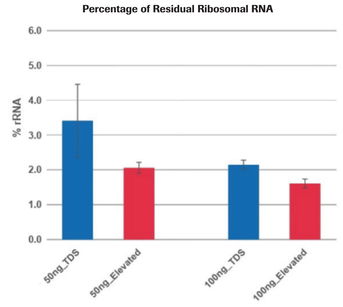

Figure 3. Percentage of residual ribosomal RNA as a function of dissociation temperatures and sample input.

Elevated dissociation temperatures reduce rRNA carryover for both 50 and 100-ng inputs. In the figure, "TDS" refers to *Technical Data Sheet* dissociation temperatures (65°/ 70°), and "Elevated" to increased dissociation temperatures (70°/ 75°).

The effect is more pronounced at the lower (50-ng) input than at the 100-ng input, supporting the hypothesis that simple promiscuous binding at lower-temperature inputs leads to higher rRNA contamination.

Other sequencing metrics are equivalent (not shown).

As a result of these outcomes, the higher dissociation temperatures have been incorporated into the automated protocol.

ACHIEVE HIGH-QUALITY RESULTS AT HIGH THROUGHPUTS

Ninety-Six samples (250 ng input each) were processed using the automated protocol as previously described; 24 libraries were then sequenced and downsampled to 6 million read pairs per sample. In **Figure 5**, outcomes are presented as averages per column, averages per row, average of all samples, and as compared to 250-ng input samples from the experiment described in **Figure 2**.

Figure 4: Concentrations (ng/ul) of 96 libraries resulting from automated processing. All wells yielded usable libraries with an average concentration of ~ 18 ng/uL and concentration CV of 17%.

	1	2	3	4	5	6	7	8	9	10	11	12
A	22.8	18.7	22.1	18.6	15.1	21.8	20.6	17.1	18.5	17.7	18.3	17.6
B	19.4	19.40	13.8	20.0	15.8	16.60	16.5	17.6	11.9	15.30	18.8	19.0
C	18.7	18.0	20.2	20.50	17.4	14.9	15.6	20.50	15.4	15.6	13.5	20.8
D	21.7	17.30	16.8	26.1	15.0	16.60	21.4	16.2	19.3	12.80	22.7	22.7
E	23.7	15.6	20.9	20.10	15.4	14.4	16.6	15.80	15.4	17.1	11.8	25.0
F	21.8	21.40	20.7	21.4	16.6	16.50	19.6	18.9	15.1	14.60	17.3	18.0
G	21.0	22.9	21.3	23.90	16.9	16.7	17.8	20.20	17.9	18.6	18.2	21.3
H	22.4	22.3	24.9	26.3	14.6	16.3	20.8	21.7	15.9	20.6	21.0	22.7

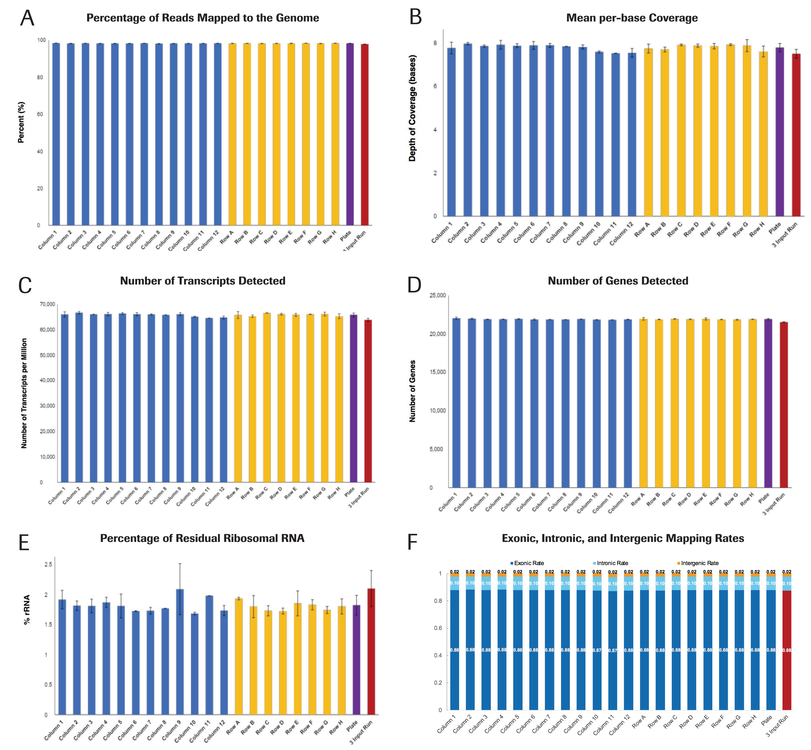

Figure 5. Comparison of sequencing metrics reveals that this high-throughput automated library preparation method yields high-quality libraries. (A) Percentage of reads mapped to the genome. Mapping rate is ~98% and is consistent across rows, columns, and with previous run. **(B)** Mean per-base coverage. ~7.7 x coverage was attained throughout, slightly higher than but comparable to the previous run. **(C)** Number of transcripts detected with count > 1 per million, as reported by kallisto. **(D)** Number of genes detected. **(E)** Percentage of residual ribosomal RNA. Results were consistent throughout, slightly lower than previous run for the 250-ng inputs and within expectations for mRNA capture workflows. **(F)** Exonic, Intronic, and Intergenic Mapping Rates. Consistent throughout, comparable to previous run and within expectations for mRNA capture workflows.

CONCLUSION

Sequencing metrics—including the percentage of reads mapped to the genome, coverage depth, number of transcripts detected, number of genes detected, and the percentage of residual ribosomal RNA reads—are consistent across all inputs and are comparable to libraries derived from manual processing. This establishes the automated protocol as a powerful tool to facilitate reliable, higher-throughput NGS analyses of mRNA and its role in a wide range of biological processes, and establishes the KAPA mRNA HyperPrep Kit as a flexible and automatable tool for use in this dynamic, emerging workflow.

MC-US-08561 A602