Optimization of rRNA Depletion by Testing Various Kits

Alyssa Sancio[1] and Dr. Cary Lai[2]

[1]PSM Biotechnology University of San Francisco, [2]Faculty Advisor University of San Francisco

Abstract

RNAseq can be used to compare gene expression between different groups, identify mutations, and confirm gene knockouts. In RNAseq, RNA is extracted and built into a sequenceable library to analyze the transcriptome. Ribosomal RNA (rRNA) makes up roughly 90% of the total RNA extracted and does not provide useful transcriptome information, thus must be removed in order to perform transcriptional analysis. At MBP Titan, the protocol used KAPA's RiboErase Depletion kit using custom methanotroph oligos to specifically bind to rRNA for depletion. On average, the amount of rRNA remaining in the libraries ranged from 10-30%. Several depletion methods and kits were tested to develop a more effective and robust depletion method that can be high-throughput and automated while keeping library preparation costs low and cut down on hands-on time. The KAPA RiboErase kit skipping intermediate purification and the Lucigen RNase H method resulted in an average of 1-5% remaining rRNA.

Introduction

MBP Titan, LLC is a company that synthetically engineers a methanotroph to produce 2,3 butanediol, which is a precursor to synthetic rubber. To validate these gene modifications, RNAseq can be requested from the next-generation sequencing team. As a service, the RNAseq method must be standardized, robust, consistent, high-throughput, automatable, and cost efficient. In developing a protocol, all of these factors contribute to deciding on a method to use. The cost of RNAseq is approximately $300 per sample. rRNA is one of the most expensive portions of the RNAseq workflow. The rRNA depletion using an RNase H based depletion. Single-stranded DNA oligos specifically designed to bind to rRNA of the methanotroph. RNase H removes these rRNA sequences. In order to help drive these costs down, a variety of kits and enzymes were tested: KAPA RiboErase, NEBNext, Takara RiboGone, and Lucigen RNase H enzyme.

Methods

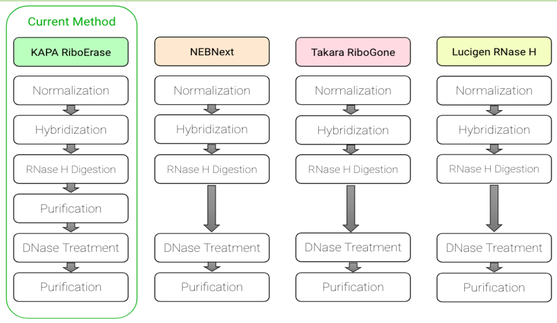


Figure 1: Method comparison across four kits.

Total RNA was extracted from a cell pellet with a biomass of 4 OD using a formamide-EDTA extraction. The total RNA was purified using the Zymo MagBead Clean and Concentrate kit and. The same total RNA was used for all conditions with 500 ng input with 2:1 oligo:RNA ratio. The oligo set was customized from Integrated DNA Technologies to specifically target rRNAs of the methanotroph. The standard KAPA RiboErase protocol was used as the positive control depletion method. The KAPA RiboErase protocol without the intermediate purification was also tested to make the protocol faster. The NEBNext rRNA Depletion kit as well as the enzyme on its own were tested. Buffers were made in house to substitute buffers included in the NEBNext kit. The Takara RiboGone kit was also tested following the standard protocol. The Lucigen RNase H enzyme was tested with buffers made in house. All depletion methods were processed in triplicate. The sample with the highest concentration for each condition was prepped using the Takara SMARTer Stranded RNAseq Library Prep kit. All samples were sequenced on the Illumina MiSeq. Data was analyzed using the Geneious software.

Results and Discussion

Successfully depleted RNA lack the presence of a 16S and 23S ribosomal peak, ~1.5 kilobases and ~2.9 kilobases, that appear at 40 and 45 seconds along the x-axis of the electropherograms (Figure 2). The replicate with the highest concentration for each condition (Table 1) was reverse transcribed to form cDNA and prepped into a sequenceable library (Table 2).

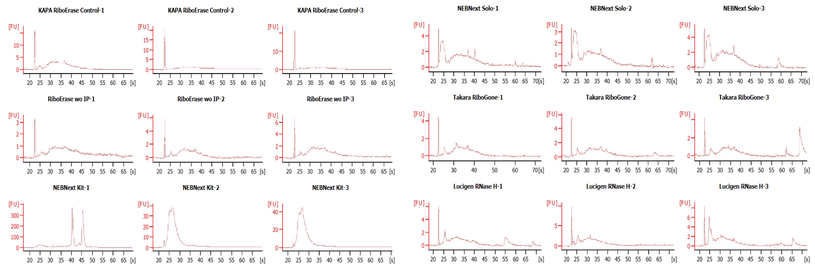


Figure 2: Electropherograms run on the Agilent Bioanalyzer from each condition and replicate using the RNA Pico kit.

Table 1: Depleted RNA Concentrations (ng/µL)

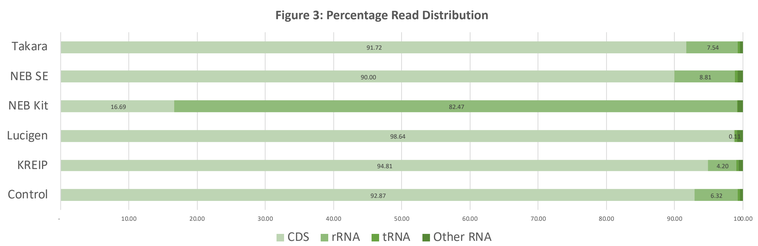
	Control	Condition 1	Condition 2	Condition 3	Condition 4	Condition 5
	KAPA RiboErase	KAPA RiboErase w/o Intermediate Purification	NEBNext Kit	NEBNext-Solo Enzyme	Takara RiboGone	Lucigen RNase H
Rep-1	0.793	0.279	Contaminated	0.957	0.389	0.552
Rep-2	0.283	0.294	4.221	0.756	0.483	0.659
Rep-3	0.279	0.493	5.005	1.084	0.224	0.915

Table 2: Final Libraries

Condition	Concentration (ng/µL)	Avg Size (bp)
Control-1	7.84	314
KREIP-3	7.22	304
NK-3	7.3	276
NSE-3	4.34	345
TRG-2	7.38	310
LRH-3	3.28	304

Table 2: Average size determined from Agilent Bioanalyzer using the HS DNA Kit

Sequenced libraries were aligned to the reference genome. The sample depleted using the standard RiboErase method had 6.32% rRNA remaining (Table 3). The sample depleted using the RiboErase method without the intermediate bead purification step had 4.2% rRNA remaining. The sample depleted using the NEBNext Kit had 82.47% of its reads mapped to rRNA. 8.81% of the reads belonging to the sample depleted using the NEBNext using only the RNase enzyme mapped to rRNA. The sample depleted using Takara RiboGone Kit had 7.54% of sequenced reads mapped to rRNA. Lastly, the sample depleted using the Lucigen RNase H enzyme had 0.11% of its sequences map to rRNA. The TPM values were calculated for every condition and compared against the normalized gene expression of sample depleted using the standard KAPA RiboErase method. With the exception of the sample depleted using the NEB Next kit, all showed a significant correlation with R^2 values >95% (Figure 4). The samples depleted with NEBNext showed the most variation in gene expression (Figure 5).



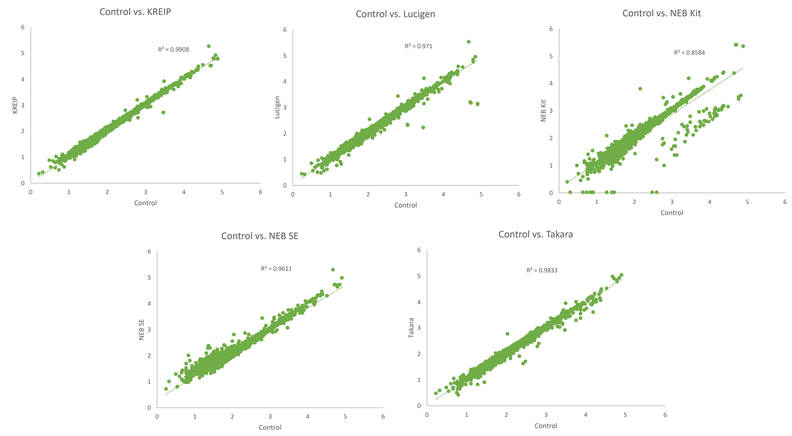


Figure 4: Normalized TPM values graphed for each condition against the standard KAPA RiboErase protocol.

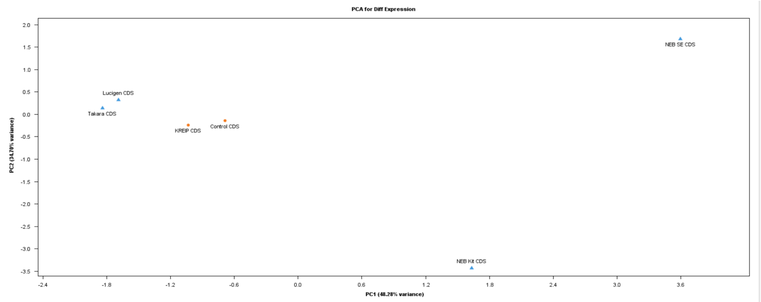


Figure 5: Principal Component Analysis (PCA) demonstrating normalized abundances.

Conclusion

rRNA depletion is a critical step prior to library preparation and sequencing for transcriptome analysis. When scaling up for RNAseq operationalization, it is important that the methods are high-throughput, automatable, robust, and cost efficient. When changing a method, it is also important that the gene expression remain consistent. The Lucigen RNase H method satisfied this criteria. Using that protocol, the mapped sequences that aligned to rRNA was less than 1%. Further, the gene expression remained consistent compared to the sample processed using the standard depletion method. Because this method only requires the purchasing of enzymes and buffers are made in house, the cost of rRNA depletion can be driven down significantly especially as sample processing is scaled up. In order to switch methods with confidence, more experiments need to be performed.

Future Studies

- Repeat Lucigen RNase H protocol with higher samples count
- Repeat with variety of cell pellets
- Deeper data analysis

Acknowledgements

All work was supported by MBP Titan, LLC and the Next-Generation Sequencing Team with assistance from the Data Science Computational Biology team for bioinformatics support. The poster has been reviewed by faculty advisor, Dr. Lai, for the PSM Biotechnology internship.