Choosing between total RNA-Seq and mRNA-Seq

Summary

- The experimental objective should dictate the choice of RNA-Seq method: whole-transcriptome vs. gene expression
- Total RNA-Seq offers the most comprehensive whole-transcriptome analysis
- mRNA-Seq is ideal if the research is focused only on the coding region and limited amounts of starting material are available

Introduction

RNA sequencing (RNA-Seq) is a powerful and comprehensive method for analyzing the cellular transcriptome. This technique enables the qualitative and quantitative examination of multiple types of RNA in biological samples at specific time points. Applications for RNA-Seq are wide ranging, from basic research on cellular structure and function to the analysis of various disease states in clinical samples. For example, gene expression patterns can be compared before and after therapeutic interventions for the presence of a disease. Alternative splicing patterns, post-transcriptional modifications, and exon–intron boundaries can also be determined with RNA-Seq. The data obtained can provide valuable insights into basic cellular mechanisms, genome structures, disease-induced effects, and more [1].

The RNA-Seq process revolves around the construction of a complementary DNA (cDNA) library to be used for sequencing. Library construction begins with the isolation of cellular RNA, followed by quality control measures to determine RNA integrity. Afterwards, a depletion or selection strategy may be chosen to enrich the library for the RNA species of interest. RNA is then reverse-transcribed into cDNA before sequencing. Typically, DNA is the actual molecule being sequenced—not RNA, as the "RNA-Seq" name might imply—although recent advances have made direct sequencing of RNA commercially feasible [2].

Whether total RNA or mRNA sequencing should be used is determined by the objective of the experiment, as several important differences exist between the methods. Total RNA sequencing (also known as whole-transcriptome sequencing) is the most comprehensive approach and typically involves sequencing all of the RNA molecules, both coding and noncoding. However, in this technical note, total RNA-Seq refers to the sequencing of RNA that has been depleted of ribosomal RNA (rRNA), which represents the majority of RNA molecules (Table 1). Total RNA, when originally isolated, is composed of multiple RNA species, including rRNA, precursor messenger RNA (pre-mRNA), messenger RNA (mRNA), and several types of noncoding RNA (ncRNA), such as transfer RNA (tRNA), microRNA (miRNA), and long ncRNA (lncRNA; transcripts longer than 200 nucleotides not translated into protein) [1]. The removal of rRNA in the total RNA-Seq procedure results in improved sequencing data that enables the characterization of these diverse non-rRNA species.

If the research goal is to focus primarily on the coding region, then mRNA-Seq represents the best choice. The mRNA-Seq protocol uses a selection method to enrich for polyadenylated (poly(A)) RNA. mRNA represents only a small percentage of the total RNA molecules (Table 1), so sequencing only mRNA is the most efficient and cost-effective procedure if it meets the overall experimental goals. Table 1 compares different RNA subspecies and offers guidance on whether a total RNA-Seq or mRNA-Seq approach is optimal.

Sample enrichment or depletion approaches

Data quality can be significantly improved through enrichment or depletion steps in both the total RNA-Seq and mRNA-Seq protocols [3]. These steps enable the sequencing of only the RNA molecules of interest and minimize wasted reads. For example, rRNA can account for 80–90% of the total RNA population, and it is usually removed by a depletion method (Table 1). The removal of rRNA transcripts allows for more sequencing reads to be focused on the desired transcripts, resulting in improved sensitivity. This is especially important if the desired transcripts are expressed at low levels. In contrast to the rRNA depletion steps in total RNA-Seq, mRNA-Seq typically utilizes a poly(A) affinity selection method to enrich for mRNA, but both protocols result in samples effectively devoid of rRNA. The selection of an appropriate rRNA removal or poly(A) enrichment method depends on a variety of factors, such as sample quantity and type—for example, prokaryotic, animal, and plant RNA samples will each require different strategies [4]. Figure 1 shows a comparison of rRNA depletion methods using Invitrogen™ Human Brain Reference RNA as input for library preparation, followed by sequencing on an Illumina™ HiSeq™ 4000 system. The lowest percentage of rRNA reads were obtained with the NEBNext™ Ultra™ II RNA Library Prep Kit, followed by the Invitrogen™ Collibri™ Stranded RNA Library Prep Kit.

Table 1. Common types of RNA in typical mammalian cells and the optimal RNA-Seq approaches to study them.

RNA species	Biological function	Estimated percent of total RNA*	Total RNA-Seq	mRNA-Seq
Ribosomal RNA (rRNA)	Protein synthesis	80–90	– (ribosomal reduction)	– (ribosomal reduction)
Transfer RNA (tRNA)	Protein synthesis	10–15	–	–
Messenger RNA (mRNA)	Translation	3–7	Yes	Yes
Long intergenic noncoding RNA (lincRNA)	Various (cell identity and differentiation)	<0.2	Yes	No
Small nucleolar RNA (snoRNA)	Post-transcriptional modification of RNA	0.04–0.2	Yes	No
Pseudogenes	Various (transcriptional regulation)	Variable**	Yes	No
MicroRNA (miRNA)	RNA silencing, post-transcriptional regulation	0.02	Yes	No

* Percentage by mass [7].

** Pseudogene transcripts can be counted in both coding and noncoding RNA.

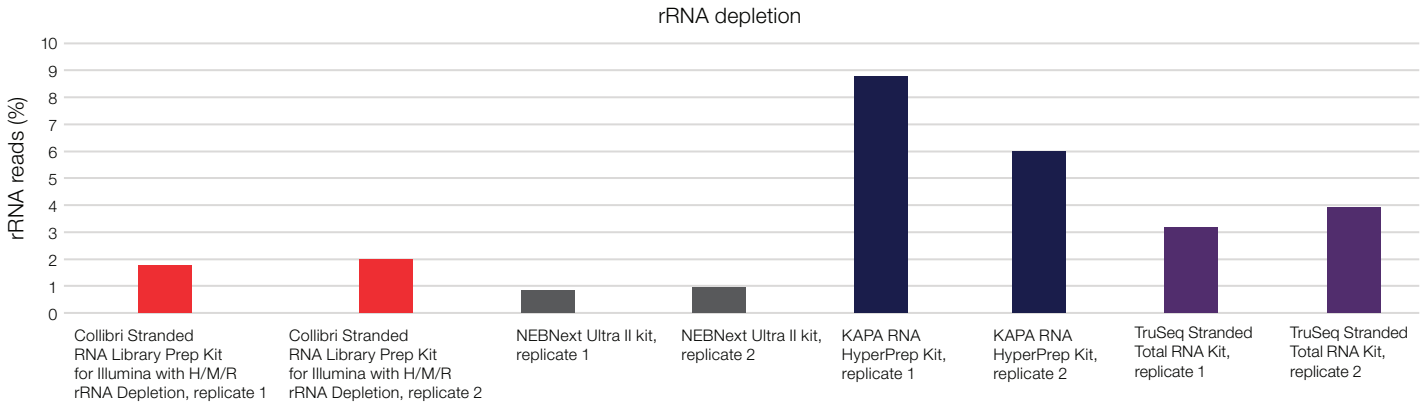

Figure 1. Comparison of rRNA depletion for RNA-Seq library preparation. The Collibri Stranded RNA Library Prep Kit for Illumina with H/M/R rRNA Depletion Kit provides superior depletion of rRNA. Libraries were prepared from 100 ng of Human Brain Reference RNA using the standard protocol from each manufacturer and sequenced on a HiSeq 4000 system with an average normalized read count of 30 million reads per sample. The residual rRNA carryover was measured and plotted as a percentage of total reads.

Factors to consider when choosing between total RNA-Seq and mRNA-Seq

Total RNA-Seq yields the most comprehensive transcriptome analysis, as it provides information on both coding RNA, and noncoding RNA such as lncRNA and miRNA (Table 1). This method can be used to examine regulatory regions, global expression levels of transcripts, splicing patterns, and to identify exons and introns, and their boundaries.

Many of the lncRNA reads will overlap with mRNA [5]. It is estimated that in the human genome, nearly 20% of genes are overlapping and are transcribed from opposite strands [6]. Thus, stranded protocols for total RNA-Seq are needed to differentiate among them. A stranded protocol can identify the specific DNA strand (coding or template) from which an RNA transcript was derived. In addition, it can provide more accurate gene expression data by improving the annotation of reads and increasing the number of reads that can be aligned.

It is imperative to initially determine what information is desired from the RNA-Seq data, so that other types of RNA that may not be relevant are not sequenced. mRNA-Seq should be selected if the coding region is of primary interest. Focusing only on mRNA will result in much better gene expression data, because mRNA accounts for only about 3–7% of the mammalian transcriptome. Library preparation can be performed with smaller sample sizes for mRNA-Seq compared to total RNA-Seq, and sequencing depth can be increased.

Additional considerations with regard to an overall budget should be made prior to method selection. Total RNA-Seq requires more sequencing data (typically 100–200 million reads per sample), which will increase the cost compared to mRNA-Seq. If only mRNA information is required, then mRNA-Seq offers greater read depth at lower cost than total RNA-Seq. This is because sequencing reads (typically 25–50 million reads per sample) are focused on poly(A)-enriched RNA molecules.

mRNA-Seq is the most viable approach for limited amounts of starting material. Enrichment for mRNA results in better sequencing read data, lower cost, and requires less starting material. Table 1 highlights the different RNA species that can be sequenced with total RNA-Seq compared to mRNA Seq.

Conclusions

Total RNA-Seq and mRNA-Seq approaches have technical advantages and disadvantages, and selecting between the two requires careful consideration of the overall experimental objective, the underlying biological question, and the specific technical limitations of each method.

The most comprehensive transcriptome analysis can be achieved with total RNA-Seq, and the best coding-region data can be obtained using mRNA-Seq. However, other factors—such as the type of sample, amount of starting material, and project budget—need to be considered as well.

invitrogen

Ordering information

Product	Quantity	Cat. No.
RNA-Seq kits		
Collibri Stranded RNA Library Prep Kit for Illumina*	24 preps	A38994024
	96 preps	A38994096
Collibri Stranded RNA Library Prep Kit for Illumina with H/M/R rRNA Depletion Kit*	24 preps	A39003024
	96 preps	A39003096
ERCC RNA Spike-In Mix	1 kit	4456740
ERCC ExFold RNA Spike-In Mixes	1 kit	4456739
Library quantification		
Collibri Library Quantification Kit*	100 rxns	A38524100
	500 rxns	A38524500
Qubit 4 Fluorometer	1 instrument	Q33238
Qubit 1X dsDNA HS Assay Kit	100 assays	Q33230
Qubit 4 NGS Starter Kit	1 kit	Q33240
Library amplification		
Platinum SuperFi Library Amplification Master Mix	50 rxns	A38539050
	250 rxns	A38539250
Platinum SuperFi Library Amplification Master Mix with Primer Mix*	50 rxns	A38540050
	250 rxns	A38540250
Purification		
PureLink RNA Mini Kit	25 preps	12183020
	250 preps	12183025
MagMAX *mir*Vana Total RNA Isolation Kit	96 preps	A27828
Thermo Scientific accessories		
KingFisher Flex Purification System with 96 Deep-Well Head	1 system	5400630
Applied Biosystems accessories		
Veriti 96-Well Thermal Cycler	1 instrument	4375786
ProFlex 96-Well PCR System	1 instrument	4484075
MicroAmp EnduraPlate Optical 96-Well Clear Reaction Plates with Barcode	20 plates	4483354
MicroAmp Optical 96-Well Reaction Plate	10 plates	N8010560
MicroAmp Clear Adhesive Film	100 films	4306311
MicroAmp 8-Tube Strip with Attached Domed Caps, 0.2 mL	125 strips	A30589

H = human, M = mouse, R = rat.
* Not all kits are available in all countries.

References

1. Hrdlickova R, Toloue M, Tian B (2017) RNA-Seq methods for transcriptome analysis. Wiley Interdiscip Rev RNA 8(1).
2. Garalde DR, Snell EA, Jachimowicz D et al. (2018) Highly parallel direct RNA sequencing on an array of nanopores. *Nat Methods* 15(3):201–206.
3. Conesa A, Madrigal P, Tarazona S et al. (2016) A survey of best practices for RNA-Seq data analysis. *Genome Biol* 17:13.
4. Guo Y, Zhao S, Sheng Q et al. (2015) RNA-Seq by total RNA library Identifies additional RNAs compared to poly(A) RNA library. *Biomed Res Intl* 2015:862130.
5. Ning Q, Li Y, Wang Z et al. (2017) The evolution and expression pattern of human overlapping lncRNA and protein-coding gene pairs. *Sci Rep* 7:42775.
6. Zhao S, Zhang Y, Gordon W et al. (2015) Comparison of stranded and non-stranded RNA-Seq transcriptome profiling and investigation of gene overlap. *BMC Genomics* 16(1):675.
7. Palazzo AF, Lee ES (2015) Non-coding RNA: what is functional and what is junk? *Front Genet* 6(2).

Find out more at **thermofisher.com/collibri**

ThermoFisher SCIENTIFIC