Frequently Asked Questions

Bulk RNA-Seq

Q1. What is meant by bulk RNA-Seq ?

A1. The collection of RNA transcripts, the Transcriptome, is dynamic and constantly changing. RNA-Seq is a powerful next generation sequencing method that can deliver a detailed snapshot of RNA transcripts present in a sample. Due to the variety and very large number of molecules sequenced, this represents the most advanced method for discovery of novel transcripts and isoforms, for identification of alternative splice sites and allele-specific expression, as well as for monitoring changes in gene expression over time or under differing conditions. RNA-Seq has a 5-log dynamic range, providing exceptional relative quantitation of transcripts and making detection of rare transcripts possible. A wide variety of RNA types can be sequenced, including polyadenylated [polyA(+)] and non-polyadenylated mRNAs, non-coding RNAs, and microRNAs. Due to their differences in size and structure, various preparation methods are required for sequencing different RNA species and, therefore, may not all appear in the same data set [1].

Q2. What material should I send to be analyzed by RNA-Seq ?

A2. Generally, we start with 1 µg of total RNA from which we create a "library" of molecules ready for sequencing. Since a typical experiment will focus on a particular RNA species, we perform one of several enrichment techniques to increase the efficiency of recovering from the population sequence reads of interest.

1) Polyadenylated mRNA – PolyA(+)-selected RNA is prepared by oligo(dT) hybridization of the total RNA sample; 1 μg is required in a maximum volume of 50 μl. If instead, polyA(+)-selected RNA is delivered to NISC, then the investigator should provide at least 50 ng in a maximum volume of 5 μl. Libraries are strand-specific.
2) Non-polyadenylated and polyadenylated RNAs – NISC can first deplete the abundant ribosomal RNAs through subtractive hybridization and then follow with sequencing the pool of residual RNAs. One µg of total RNA is required in a maximum volume of 10 μl. Libraries are strand-specific.
3) Micro- and other small RNAs – These RNAs are processed from a total RNA sample by selecting specifically for their small size; 1 μg is required in a maximum volume of 5 μl. Be sure to use an RNA isolation procedure designed specifically to retain microRNAs. Since this library procedure excludes larger molecules, a parallel experiment from a second 1 μg aliquot is needed to also examine mRNAs from the same sample.
4) Limited amount of RNA – If very limited amounts of material are available, an RNA amplification option can be employed. Please consult NISC for recommendations for amplification.

Q3. How should samples be shipped ?

A3. RNA samples should be submitted in 1.5-1.7 ml microfuge tubes (example: VWR cat. no.89000-028) or 2 ml screw cap tubes (example: Sarstedt cat. no. 72.694.007). Please DO NOT send samples in 0.5 or 0.2 ml tubes. RNA should be shipped on ample dry ice.

Q4. How should the RNA be qualified ?

A4. NISC requires a Bioanalyzer trace for each total RNA sample submitted. A Bioanalyzer analysis includes a Relative Integrity Number (RIN) for each sample, where a value of 10 indicates a perfectly intact sample and a value of 1 indicates a severely degraded sample. In experiments where sequencing data are to be compared between samples and poly(A)+ enrichment is desired, samples should have similar RIN values (within a value of 1 of each other).

Q5. Can I get sequence information about non-polyA(+) transcripts ?

A5. When non-polyadenylated transcripts are of interest, rRNA can be depleted from the sample to eliminate wasteful sequencing of these molecules. Besides non-polyA(+) and polyA(+) RNAs, additional species will be present in the library, such as non-coding, incompletely spliced, and fragmented RNAs. This library type can be of value when the RNA sample is somewhat degraded. If the RIN is between 4 and 7, it is recommended to use this rRNA depletion method rather than polyA(+) selection, which will recover all RNA fragments rather than just those attached to a polyA tail.

Q6. What data are returned by NISC ?

A6. Typically, NISC returns to the investigator fastq files containing basecalls and quality scores. The investigator is expected to provide data analyses; this is not offered by NISC.

Data files can become quite large; currently, from each mRNA library we typically generate at least 50 million read-pairs. For efficiency, a sequencing lane typically will contain a pool of barcoded samples, so demultiplexing is part of our data processing. Sequence data for each sample will be in a separate file. Please note that quality of input nucleic acid will greatly influence the actual amount of quality sequences recovered. Also, poorly annotated genomes can make data analysis significantly more difficult.

Q7. How many times will a given sequence be found in the total data set (coverage) ?

A7. This will depend on the relative level of expression of the transcript of interest, but in general, the depth of coverage is a ratio of bases generated (number of fragments read times length of read) divided by bases of sequence targeted. If the entire transcriptome (human coding potential) were targeted (30 Mb), then 50 million pairs of 150 b reads will yield 500× coverage. This is a good starting point, but as with any experimental procedure, the read depth for any given sequence will vary from the average. One can determine empirically through a pilot experiment whether more or fewer reads are needed to sufficiently cover a region of interest [2].

Q8. How much data can I expect from an RNA-Seq analysis of a mammalian RNA sample ?

A8. A transcription profile (relative gene expression) of a sample can typically be made from 50 million clusters of reads. For detection of splice variants or low-expressing genes, then up to 150 million clusters of reads may be desirable.

References :

1. Mutz, K-O., *et al*. (2013) “Transcriptome analysis using next-generation sequencing.” *Current Opinions in Biotech*. **24**: 22-30.
2. Sims, D., *et al*. (2014) “Sequencing depth and coverage: key considerations in genomic analyses.” *Nature Rev*. *Genetics* **15**: 121-132.
3. Shendure, J. (2008). "The beginning of the end for microarrays?". *Nature Methods* **5**: 585-587.