Frequently Asked Questions

RNA-Seq

Q1. What is meant by 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, microRNAs and those from ribosomal profiling. 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. Enrichment can also be performed as strand-specific, polyA(+) RNA.
2) Non-polyadenylated and polyadenylated RNAs – NISC can first deplete the abundant interfering 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. Enrichment can also be performed as strand-specific, rRNA depleted.
3) Micro- and other small RNAs – These RNAs are processed from a total RNA sample by selection 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, in order to also examine mRNAs from the same sample, a parallel experiment from a second 1 µg aliquot is needed.
4) Limited amount of RNA – If very limited amounts of material are available, consider amplification of the RNA first. In this case, a minimum of 50 ng of

cDNA in a maximum volume of 5 µl must be sent to NISC. 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 it is severely degraded. In experiments where sequencing data are to be compared between samples, they should have similar RIN values (within a value of 1 of each other).

Q5. Can I determine which strand the transcript is from if the genome is not annotated in the region where the sequence maps ?

A5. Strand-specific RNA-Seq libraries can be constructed at NISC. Generally, this is only recommended when the investigator needs to determine from which strand the RNA originates.

Q6. Can I get sequence information about non-polyA(+) transcripts ?

A6. 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. Please note, approximately twice the number of reads from this library will be required to obtain a comparable level of sensitivity to a polyA(+) -selected library.

Q7. What data are returned by NISC ?

A7. Typically, NISC returns to the investigator a file in “BAM” format containing basecalls and quality scores. The investigator is expected to provide data analyses; this is not offered by NISC. Commercial software is available from both Illumina and third party vendors; e.g., see http://illumina.com/pagesnrn.ilmn?ID=229.

Data files can become quite large; currently, from each mRNA library we generate reads from at least 40 million clusters. 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.

Q8. How many times will a given sequence be found in the total data set (coverage) ?

A8. In the simplest case, 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 36 million 125b single-end reads are needed for 150× 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].

Q9. Is RNA-Seq equivalent to a microarray analysis of the transcriptome ?

A9. A microarray experiment retrieves RNAs of known sequence from a sample through hybridization to defined oligonucleotides bound to a solid surface. Fluorescence signal intensity from the hybrids indicates the relative amount of each of these RNAs in the original sample. This method has a 3-log dynamic range and is faster and cheaper than other methods of measuring gene expression.

An RNA-Seq experiment generates sequence reads for every RNA present in a sample. This method has at least 5-log dynamic range, and shows better quantitation and reproducibility than microarrays. Low abundance molecules are detected; if needed, detection limit is lowered by simply generating more sequencing reads. Unlike microarrays where analysis is limited to the set of known oligonucleotide sequences bound to the solid support, RNA-Seq permits discovery of novel transcripts and splice variants, distinguishes between closely-related sequences and provides for quantitation of expression between alleles. [3]

Q10. How much data can I expect, for example, from a mammalian cell RNA-Seq analysis ?

A10. A transcription profile (relative gene expression) of a sample can typically be made from 40 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.