Transcriptome Analysis

Aims

In the field of transcriptome analyses, Genomnia is able to offer different services:

- **Total RNA-seq**: this procedure exploits sequencing to analyze, in a quantitative and unbiased manner, the entire coding and non-coding population of transcripts;
- **Small RNA-seq**: the aim of this approach is the selective analysis of the small RNA fraction of a sample;
- **Ampliseq Transcriptome:** this procedure enables to quickly and massively investigate gene expression profiles in human or mouse RNA samples, taking advantage of primer pools specific for RNA targets.

Sequencing transcriptome analysis is characterized by high sensitivity and specificity, and by a dynamic range wider than the one obtained with microarrays. For this reason, such a kind of technology is nowadays considered the method of choice to perform genome-wide RNA expression profile studies.
Thanks to the alternative available strategies, the massive output offered by Ion Torrent technology enables not only the detection and quantification of known transcripts but also the discovery of novel exons, of previously unknown small RNAs, of splice junctions and novel unprocessed transcripts, as well as of those species typically transcribed at very low levels. This technique, coupled with an accurate bioinformatics analysis, also ensure the identification and count of expressed SNPs, leading to the evaluation of allele-specific gene expression and the discovery of fusion transcripts. The library preparation method (*Ligase-Enhanced Genome Detection*) preserves the strandness information of RNA molecules, facilitating the detection of possible partial or total antisense transcripts.

Procedure

The protocol for **Total RNA-seq** requires the initial purification of total RNA and the subsequent polyA isolation or rRNA depletion. The small amount of purified RNA so obtained are sheared in fragment of 25-500 nt, with a mean size of 100 nt, by chemical hydrolysis: this kind of fragmentation method ensures a uniform distribution of sequence reads across the transcripts. Chemical hydrolysis is carried out thanks to the combined action of cations (Mg^{++} or Zn^{++}) and heat (90 - 100°C), producing fragments which need to be phosphorylated using T4 kinase. The length of fragments obtained depends on the reaction time duration.
The procedure to follow for the selective analysis of small RNA (**Small RNA-seq**) is characterized by an initial step of enrichment of the small RNA fraction, subsequently used as template for library preparation.
For both the approaches cited above, RNA samples are converted into a final double-stranded cDNA library through a series of steps. RNA fragments are first hybridized and ligated to an adaptor mix using RNA ligase. This adaptor mix constrains the orientation of RNA molecule in the ligation reaction such that hybridization with the adaptors yields template for sequencing always from the 5' end of the sense strand. The adaptors-linked RNA is then reverse transcribed in a single strand cDNA, subsequently purified using magnetic beads. The cDNA library is PCR-amplified to obtain the amount required for downstream steps of the process and, at the same time, to add the terminal sequences necessary for the sequencing. During this PCR step it is possible to introduce at the 3' end of the molecules specific *barcodes*, short DNA sequences that allow the unambiguous identification of samples in multiplexing sequencing experiments: 96 different barcode sequences are available to date. Finally, libraries are clonally amplified on beads in the emulsion PCR step. These beads represent the support for sequencing reaction.

It is also possible to perform transcriptome analysis using the Ampliseq technology (**Ampliseq Transcriptome**), currently available for both human and murine sequences. The Ampliseq transcriptome allows the evaluation of gene expression levels thanks to a single primer panels designed to investigate in multiplexing over 20,000 genes (corresponding to more than 95% of RefSeq genes for *Homo sapiens* and more than 90% for *Mus musculus*). Primer pairs of these panel are designed to produce small amplicons (one for each gene), guaranteeing optimal results also using RNA extracted from suboptimal and small samples, such as FFPE tissues. In order to avoid the selection of poly(A) RNA or the elimination of ribosomal RNA, this protocol allows the use as input of total RNA, retotranscribed with random primers. cDNA target regions are then amplified in multiplex using specific primer pairs pools. Amplicons are then enzymatically treated to partially remove primer regions and to phosphorylate amplicon ends. Subsequently, specific adaptors (and barcodes, if needed) required for sequencing process are covalently ligated at the amplicon ends.
For more information about Ampliseq technology, please refer to the specific module.

Qualitative and quantitative features of RNA samples

For whole transcriptome analysis, two different types of selectively enriched RNA preparations can be used as starting material: poly(A) RNA or rRNA-depleted RNA. For small RNA analysis, this fraction must be previously selected from the total RNA sample.
In both cases, total RNA samples with a RIN value ≥ 7 (better ≥ 8) are required. Total RNA quality must be checked by electrophoresis using the Agilent Bioanalyzer 2100 (RNA 6000 Nano or Pico Kit). This is particularly important for small RNA studies, since RNA degradation products can compete with small RNA during library preparation, decreasing the number of correctly mapped reads.

Poly(A) RNA is isolated from total RNA by one or two rounds of selection using oligo(dT), while rRNA-depleted RNA is obtained using a species-specific kit for the removal rRNA. The small RNA fraction is isolated with size selection from total RNA through a method based on Solid Phase Revers Immobilization (SPRI) technology and on the use of magnetic beads. After the purification step, the absence of 18S and 28S rRNA or the presence of small RNA enriched fraction can be checked with the Bioanalyzer, Agilent RNA 6000 Pico Kit.

The standard amount of RNA from tissues or cells required for these procedures is listed in the following table.
It is important to note that, if possible, higher amount of RNA are preferably, in order to perform an initial quality control. Moreover, samples must be completely free of contaminating DNA.
The lower amount refers to the minimal quantity required according to the specific protocol. However, Genomnia will be glad to evaluate together with the customer the feasibility of the experiment also starting from smaller amount of RNA, trying to optimize the experimental conditions in critical cases.
If possible, an extra amount should be provided to perform initial QC control.

Starting material	Amount (µg)
Total RNA, Poly(A) purification	1 - 5
Total RNA, rRNA depletion	0.5 - 5
Total RNA, small RNA enrichment	1 – 20
	Amount (ng)
RNA poly(A)	10 - 500
rRNA depleted RNA	10 - 500
Small RNA enriched sample	1 - 100
Total RNA, Ampliseq technology	10

Bioinformatic Analysis

The whole transcriptome bioinformatic analysis (available in two levels) aims at investigating the entire set of sample cell or tissue transcripts under both the quantitative (differential analysis of transcripts in two or more different experimental conditions, development phases, tissues and so on) and qualitative (identification and classification of isoforms, non-coding RNAs and so on) aspects. The bioinformatic analysis protocol includes a step of correlation of the transcripts with the reference genome (mapping); a statistical analysis step (differential analysis); a functional analysis step. This protocol is applied to organisms with a well-defined reference genome (even when still incomplete), annotated and available from the international databases and annotation software suites such as NCBI, UCSC Genome Browser, EBI and Ensembl.

In addition to the genome alignment output files, the output of this procedure consists in one or more Excel worksheets including:

- Information on the genome mapping: coverage, mapping statistics and alignment quality values;
- the differential expression values for the requested comparisons;
- the differential splicing results (isoform analysis) and the differential expression evaluated with a second approach (level II);

- the results of the differential analysis elaborated with a second approach and comparisons of the results obtained with the two algorithms (only with level II);
- the functional annotation with Cytoscape "Functional Interactions" module of the over- and under-expressed gene lists obtained by the intersection of the two analytical approaches(only with level II);
- the differential expression analysis of the non-coding RNAs (only with level II).

The microRNA profiling analysis is used to evaluate in a quantitative and qualitative manner the small RNA deep sequencing results with the ION technology. The small RNA expression profiling is performed first with an absolute values quantification and, consequently, with a differential analysis evaluation. The microRNAs are a minority but a functionally important part of the small RNA population. This procedure constitutes the level I analysis and prosecutes with the IsomiR identification procedure, the discovery of putative novel microRNAs and their differential analysis (level II).

Bibliographic References (in bold Genomnia coauthors)

Megiorni F, **Colaiacovo M**, Cialfi S, McDowell HP, **Guffanti A**, Camero S, **Felsani A**, Losty PD, Pizer B, Shukla R, Cappelli C, Ferrara E, Pizzuti A, **Moles A**, Dominici C. A sketch of known MYCN-associated miRNA networks in neuroblastoma. Oncology Reports. 2017;38(1):3-20.

Callari M, **Guffanti A**, Solda G, Merlino G, Fina E, **Brini E**, et al. In-depth characterization of breast cancer tumor-promoting cell transcriptome by RNA sequencing and microarrays. Oncotarget. 2016;7(1):976-94.

Soreq L, Salomonis N, **Guffanti A**, Bergman H, Israel Z, Soreq H. Whole transcriptome RNA sequencing data from blood leukocytes derived from Parkinson's disease patients prior to and following deep brain stimulation treatment. Genom Data. 2015;3:57-60.

Fazi B, **Felsani A**, Grassi L, **Moles A**, D'Andrea D, Toschi N, et al. The transcriptome and miRNome profiling of glioblastoma tissues and peritumoral regions highlights molecular pathways shared by tumors and surrounding areas and reveals differences between short-term and long-term survivors. Oncotarget. 2015;6(26):22526-52

Soreq L, **Guffanti A**, Salomonis N, Simchovitz A, Israel Z, Bergman H, et al. Long non-coding RNA and alternative splicing modulations in Parkinson's leukocytes identified by RNA sequencing. PLoS Comput Biol. 2014;10(3):e1003517.

Megiorni F, Cialfi S, McDowell HP, **Felsani A**, Camero S, **Guffanti A**, et al. Deep Sequencing the microRNA profile in rhabdomyosarcoma reveals down-regulation of miR-378 family members. BMC cancer. 2014;14:880.

Guffanti A, Simchovitz A, Soreq H. Emerging bioinformatics approaches for analysis of NGS-derived coding and non-coding RNAs in neurodegenerative diseases. Front Cell Neurosci. 2014;8:89.

Papait R, Cattaneo P, Kunderfranco P, Greco C, Carullo P, **Guffanti A, Viganò V**, Stirparo GG, Latronico MV, Hasenfuss G, Chen J, Condorelli G. Genome-wide analysis of histone marks identifying an epigenetic signature of promoters and enhancers underlying cardiac hypertrophy. Proc Natl Acad Sci U S A. 2013, 110:20164-9

Voellenkle C, Van Rooij J, **Guffanti A**, **Brini E**, Fasanaro P, Isaia E, et al. Deep-sequencing of endothelial cells exposed to hypoxia reveals the complexity of known and novel microRNAs. RNA. 2012;18(3):472-84.

Guffanti A., Iacono M., Pelucchi P *, Kim N., Soldà G., Croft L.J., Taft R.J., Rizzi E., Askarian-Amiri M., Bonnal R.J., Callari M., Mignone F., Pesole G., Bertalot G., Rossi Bernardi L., Albertini A., Lee C., Mattick J.S., Zucchi I., de Bellis G. A transcriptional sketch of a primary human breast cancer by 454 deep sequencing. BMC Genomics 2009, 10:163.

Ordering information

Item	Catalog N.
QC: Quality control of total RNA preparations	RNA03
Poly(A)$^+$ RNA purification from total RNA	RNA06
Enrichment in Small RNA fraction	RNA07
rRNA depletion of total RNA	RNA10
Barcoded RNA library preparation	LRb
Forward sequencing 200 bp tags with barcode	SEQI200B
Bioinformatic Analysis I: RNA	RNA-BF01
Bioinformatic Analysis II: RNA	RNA-BF02
Bioinformatic Analysis I: smallRNA (known microRNAs)	Small-BF01
Bioinformatic Analysis II: smallRNA (novel microRNAs, isomiRs and targets)	Small-BF02

Genomnia srl
Via L. Ariosto, 21 - 20091 Bresso (MI) - tel +39 02 93305700 - fax +39 02 93305777
www.genomnia.it - info@genomnia.it