Whole Transcriptome Analysis (total RNA-seq) or microRNA Analysis (small RNA-seq)

Aims

This procedure is aimed at exploring by sequencing in a quantitative and unbiased manner the entire coding and non-coding population of RNA transcripts, or the entire small RNA population, in a given sample. With its high sensitivity, specificity and larger dynamic range than the microarray technology, this technology is considered the standard method for doing genome-wide RNA expression profiles studies.

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, previously unknown small RNAs, splice junctions and novel unprocessed transcripts as well as of those typically transcribed at a very low level. This technique, coupled with bioinformatics analysis, permits also the identification and count of expressed SNPs, the evaluation of allele-specific gene expression and the discovery of fusion transcripts. The library preparation method (Ligase-Enhanced Genome Detection) preserves the information about molecules strandeness facilitating the detection of opposing and overlapping (antisense) transcripts.

Procedure

The purification of total RNA and the subsequent polyA isolation or rRNA depletion are the first step for the whole transcriptome analysis.

Little amounts of purified RNA are sheared in fragment of 25-500bp, with a mean size of 100 bp, by chemical hydrolysis. 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. This kind of fragmentation method produces a uniform distribution of sequence tags across the transcripts.

The initial step for the construction of a small RNA library is the selection of this small fraction from total RNA sample. The fragmented RNA (small RNA or fragments derived from long RNAs) is converted into a final double-stranded cDNA library. At the beginning, RNA fragments are hybridized and ligated to an adapter mixture using the RNA ligase. The adaptor mix constrains the orientation of the RNA in the ligation reaction such that hybridization with the adaptors yields template for sequencing from the 5' end of the sense strand. The adaptors-linked RNA is converted in single strand cDNA using reverse transcriptase and purified using magnetic beads. The cDNA library is finally amplified using a variable number of PCR cycles, in order to meet the sample quantity requirements for ION TORRENT sequencing. During the PCR step it is also possible to introduce at the 3' end of the molecules, specific short DNA sequences acting as barcode used to identify different samples. There are 96 barcode sequences available. These libraries are clonally amplified on beads in the emulsion PCR step and then sequenced on ION TORRENT platform.

RNA sample type and amount

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

Poly(A) RNA is isolated from total RNA by one or two rounds of selection using oligo(dT). rRNA-depleted RNA is obtained using a species-specific kit for binding rRNA. The small RNA fraction is isolated with size selection from total RNA through a magnetic bead-based method. 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 and samples must be completely free of contaminating DNA. The minimal amounts are reported just as an indication of

the lower limits of the technique and as a basis for discussion to meet the requirements of particular experimental conditions.

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

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.

The output of this procedure consists in one or more Excel worksheets including:

- the genome mapping, coverage, mapping statistics and alignment quality values files (both analytical levels);
- the differential expression values for the requested comparisons (level I);
- the differential splicing values (isoform analysis) and the differential expression evaluated with a second approach (level II);
- the intersection of LogFC values evaluated with the two different algorithms (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(level II);
- the differential expression analysis of the non-coding RNAs (RefSeq and/or EnsEMBL) obtained with a direct comparison of the files obtained from the selection of the reads from the alignment files against a specialized ncRNA transcripts database.

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). The conceptual workflow of this analytical procedure is detailed in the scientific paper “*Deep-sequencing of endothelial cells exposed to hypoxia reveals the complexity of known and novel microRNAs”* (ref.)

Bibliographic References (in bold Genomnia coauthors)

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 (transcriptome analysis)	RNA-BF01
Bioinformatic Analysis II: RNA (transcriptome advanced analysis)	RNA-BF02
Bioinformatic Analysis III: RNA (custom analysis)	RNA-BF03
Bioinformatic Analysis I: smallRNA	Small-BF01
Bioinformatic Analysis II: smallRNA (advanced)	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