Integrated Gene Set Analysis for microRNA Studies

Francisco García-García[1], Joaquín Panadero[2], Joaquín Dopazo[1,2,3,4] , David Montaner[1,2]

1 Systems Genomics Lab, Computational Genomics, Centro de Investigación Príncipe Felipe (CIPF), Valencia, Spain.
2 Genómetra, Valencia, Spain. 3 Functional Genomics Node (INB). 4 BIER, CIBER de Enfermedades Raras (CIBERER)

Aim

From a systems biology perspective, gene set analysis (GSA) allow us to understand the molecular basis of a genome-scale experiments. Gene set methods are much more sensitive than single enrichment methods in detecting gene sets (defined as sets of genes with a common annotation) with a joint implication in a genomic experiment. But currently there are not GSA methods tailored for the miRNA context. In this work we present a novel approach to the functional interpretation of miRNA studies which keeps the advantages of the GSA.

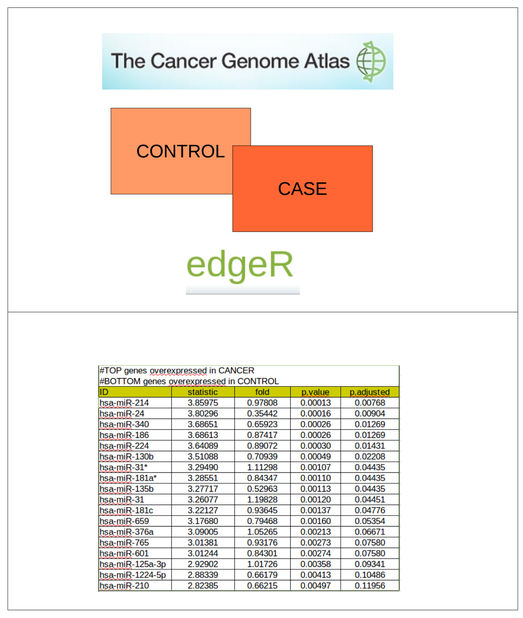
Methods

Data. We downloaded 20 datasets from The Cancer Genome Atlas (http://cancergenome.nih.gov/), containing tumoral and normal samples from HiSeq techonology.

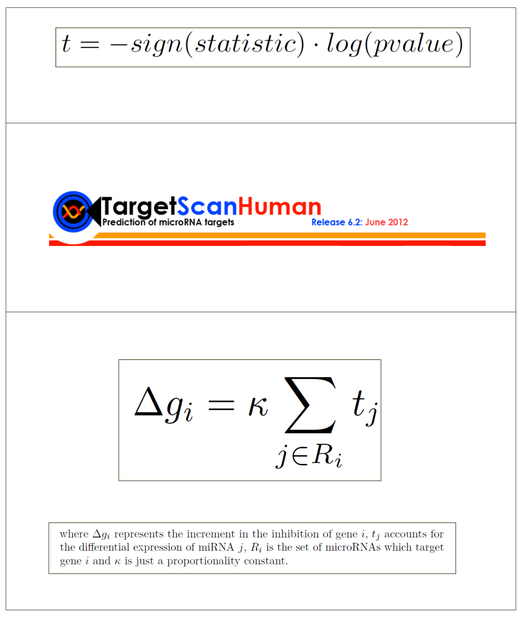
- Differential expression analysis was carried out for mRNA and miRNA levels (Bioconductor library edgeR) (1).
- Information from miRNA was transferred to gene level by adding its effects and generating a new index which ranks genes according to their differential inhibition by miRNA activity across biological conditions
- Given such ranking statistics of the genes for each functional class, we apply the logistic regression models for GSA (2,3). P-values were corrected for multiple testing using the method Benjamini and Yekutieli (4).

Results

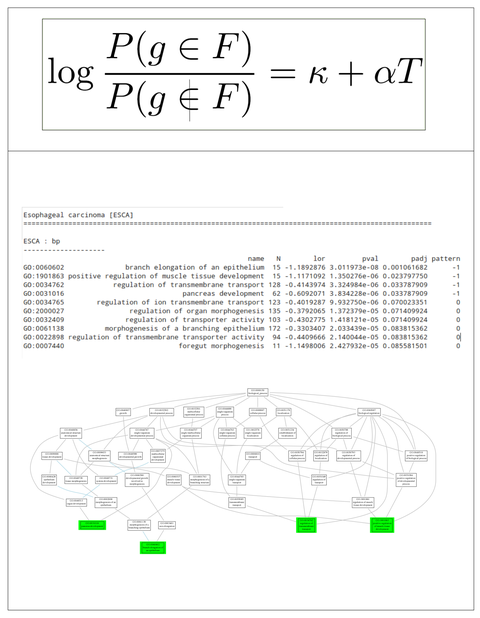
- This new approach has allowed to obtain a genomic functional profiling for different cancers when using miRNA data.
- In our study we used Gene Ontology terms (http://www.geneontology.org/) to define gene sets, obtaining detailed functional results for each ontology (biological process, cellular component and molecular function).



Step 1. Differential miRNA Expression Analysis



Step 2. Transferring information from miRNA to mRNA



Step 3. GSA from logistic regression models

Conclusions

- This method may be successfully applied in genomic functional profiling, transferring miRNA data to gene level so that GSA can be properly applied.
- Functional results take advantage of the knowledge already available in biological databases and can help to understand large-scale experiments from a systems biology perspective.

References

(1) M. D. Robinson, D. J. McCarthy, and G. K. Smyth. edgeR: a Bioconductor package for differential expression analysis of digital gene expression data. Bioinformatics, 26(1):139–140, Jan 2010
(2) D. Montaner and J. Dopazo. Multidimensional gene set analysis of genomic data. PLoS ONE, 5(4):e10348, 2010.
(3) M. A. Sartor, G. D. Leikauf, and M. Medvedovic. LRpath: a logistic regression approach for identifying enriched biological groups in gene expression data. Bioinformatics, 25(2):211–217, Jan 2009.
(4) Yoav Benjamini and Daniel Yekutieli. The control of the false discovery rate in multiple testing under dependency. The Annals of Statistics, 29(4):1165–1188, 08 2001. doi: 10.1214/aos/1013699998.