

# INTEGRATIVE GENE SET ANALYSIS OF GENE AND MIRNA EXPRESSION DATA

FRANCISCO GARCIA-GARCIA<sup>1</sup>, JOAQUIN DOPAZO<sup>1,2,3</sup>, DAVID MONTANER<sup>1</sup>

<sup>1</sup>Computational Genomics Department, Centro de Investigación Príncipe Felipe (CIPF)

<sup>2</sup>Spain Bioinformatics of Rare Diseases (BIER), CIBER de Enfermedades Raras (CIBERER)

<sup>3</sup>Spain Functional Genomics Node, (INB) at CIPF

**Introduction:** From a systems biology perspective, gene set analysis (GSA) allows us to understand the molecular basis of a genome-scale experiments. Employing a systems biology approach that includes several genome-scale measurements gets a better functional interpretation. In this work we present a multidimensional method to the functional profile of mRNA and miRNA studies which integrates both expression data.

**Methods:** We downloaded 20 datasets from The Cancer Genome Atlas (<http://cancergenome.nih.gov/>), containing tumoral and normal samples. Differential expression analysis was carried out for mRNA and miRNA levels (Bioconductor library edgeR). 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 both ranking statistics of mRNA and miRNA, for each functional class, we apply the logistic regression models for GSA. P-values were corrected for multiple testing using the method Benjamini and Yekutieli.

**Results:** This new approach has allowed to obtain a genomic functional profiling for different cancers when using an integrated approach with mRNA and 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).

**Discussion:** Integrative Gene Set Analysis of mRNA and miRNA expression data constitutes a novel approach of functional profiling which allows us to detect interactions between gene and miRNA that account for functional roles dependent on several genomic properties or measurements. From this method, we can differentiate several patterns for functional modules to understand and discover of new cell functionalities with complex dependences.

**Conclusion:** This method may be successfully applied in genomic functional profiling, transferring miRNA data to gene level and integrating mRNA and microRNA data at the same level, so that GSA can be properly used. 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:

- Benjamini, Y., Yekutieli, D. (2001). *The control of the false discovery rate in multiple testing under dependency*. The Annals of Statistics, 29(4), 1165–1188.
- Montaner, D., Dopazo, J. (2010). *Multidimensional gene set analysis of genomic data*. PLoS One. Apr 27;5(4):e10348. doi: 10.1371/journal.pone.0010348.
- Montaner, D., Minguez, P., Al-Shahrour, F., Dopazo, J., (2009) *Gene set internal coherence in the context of functional profiling*. BMC Genomics 10: 197.
- Robinson, MD., McCarthy, DJ., Smyth, G K. (2010). *edgeR: a Bioconductor package for differential expression analysis of digital gene expression data*. Bioinformatics, 26(1), 139–140.
- Sartor, MA., Leikauf, GD., Medvedovic, M. (2009) *LRpath: a logistic regression approach for identifying enriched biological groups in gene expression data*. Bioinformatics, 25(2):211–217.
- Smyth, GK. (2004). *Linear models and empirical bayes methods for assessing differential expression in microarray experiments*. Stat Appl Genet Mol Biol 3: Article3.