

An equivalence approach to global testing of gene sets based on functional profiles

Alex Sánchez^{1,2}, Miquel Salicrú¹ and Jordi Ocaña¹

¹ Department d'Estadística. Facultat de Biologia. Universitat de Barcelona. Barcelona. Spain

² Statistics and Bioinformatics Unit. Institut de Recerca. Hospital Universitari Vall d'Hebron. Barcelona, Spain

Many tools for the analysis and biological interpretation of the results of high throughput experiments are available ([2]). Most of them rely on the Gene Ontology or similar biological annotation databases to perform some variation Gene Enrichment Analysis (GEA) or the more recent Gene Set Enrichment Analysis (GSEA, [1]).

A type of analysis whose interest is obvious is comparing several lists of genes. This can be interesting “per se” or be the basis for the comparison or the combination of several experiments. In a previous study [3] we developed a framework for the analysis of lists of genes which consists of projecting them into an extended level of the GO, creating “expanded functional profiles” and using a multinomial model as the basis for the analysis. The main characteristic of this approach is that it is performed on all categories at once instead of being done on a category-by-category basis. Once the method has been established and tested the issue has appeared that, in fact, the main interest in many analyses is in establishing the hypothesis of similarity between two lists. In consequence it seems reasonable to adopt a bio-equivalence point of view instead of the classical hypothesis tests setting, designed to reject a null hypothesis of equality. We present an equivalence method which uses a distance-based approach and the confidence interval inclusion principle. Equivalence is declared if a one-sided confidence interval for the distance between two profiles is below a pre-established equivalence limit. We have applied the method to several public data sets obtaining consistent results. Some simulation experiments also prove that the test has an adequate size. A summary of these results will be presented. The method is implemented in a freely available R package (*goProfiles*) available from the Bioconductor (<http://bioconductor.org>) project.

References

- [1] Khatri P, Draghici S (2005) Ontological analysis of gene expression data: current tools, limitations, and open problems. *Bioinformatics* 21:3587-3595.
- [2] Mosquera, J.L. and Sanchez, A. (2008) SerbGO: Searching for the best GO Tool. *Nucleic Acids Research Advance Access* published on May 14, 2008, DOI 10.1093/nar/gkn256
- [3] Sánchez A, Salicrú M, Ocaña J. (2007) Statistical methods for the analysis of high-throughput data based on functional profiles derived from the Gene Ontology. *Journal of Statistical Planning and Inference* 137:3975-3989.