

*Use of GO Terms to Understand the Biological Significance of Microarray Differential Gene Expression Data*

Diaz-Uriarte, R., Dopazo, J.

Centro Nacional de Investigaciones Oncologicas, Madrid, Spain

This paper shows how to use Gene Ontology (GO) to understand the biological relevance of statistical differences in gene expression data from microarray experiments. We have used, as an example to illustrate our methodology, the data from [8]. Our proposed approach involves three sequential steps: 1) we analyze the microarray data to sort genes according to how much they differ between/among organs using, for example, a linear regression model or an Analysis of Variance (ANOVA); 2) using the sorting information, we divide the genes in groups based on "how much or how strongly" they differ (e.g., we separate those that differ vs. those that do not, or those more expressed in one organ vs. those more expressed in the other organ); 3) finally, to examine the relative frequency of GO terms in the two groups, we use Fisher's exact test, with correction for multiple testing, to assess which of the GO terms differ significantly between the groups of genes. By using the GO terms, we obtain biological information on, for instance, the predominant biological processes or molecular functions of the genes that are differentially expressed between organs, and thus it is easier to evaluate the biological relevance of inter-organ differences. Moreover, this method, when applied to novel situations (e.g., comparing difference cancer conditions) can provide important hints on the biologically relevant aspects and characteristics of the differences between conditions. Finally, this proposed method is of easy application and interpretation.