Ontologic Analysis: Challenges for Global Testing

Hummel M[1], Mansmann U[2], Meister R[3]

[1]Institut für Medizinische Informationsverarbeitung, Biometrie und Epidemiologie, Ludwig-Maximilians-Universität München, Deutschland
[2]Institut für Medizinische Informationsverarbeitung, Biometrie und Epidemiologie, Ludwig-Maximilians-Universität München, Deutschland
[3]Technische Fachhochschule Berlin, Deutschland
hummel@ibe.med.uni-muenchen.de

Introduction This talk introduces an ANOVA based global test for differential gene expression between different clinical outcomes. It will be shown how this test offers an alternative for gene set enrichment on the Gene Ontology (GO) graph. Special attention will be given to the adjustment for multiple testing.

Material and Methods The ANOVA based global test for differential gene expression in origin was designed for the comparison of two clinical groups. The approach can easily be extended for the analysis of more complex tasks like differential expression in various groups or differential co-expression of genes.

The method is used to quantify the differential gene expression between phenotype groups for the genes associated with a specific GO category. It is of interest to identify the nodes of the GO tree with differential gene expression between the groups. These nodes may help to deepen insights into biological backgrounds. The strategy of applying global tests on the GO graph offers an alternative for gene set enrichment.

The basic problem is the adjustment for multiplicity within this extensive statistical testing problem. We propose an approach to the multiplicity problem which uses a stochastic process on the GO tree induced by the null hypothesis of no differential expression. The algorithm needs an ordering of GO nodes. We present a first natural ordering idea and discuss its optimality.

Results It will be shown that our approach improves the use of information on changed gene function compared to the strategy of gene set enrichment.

Discussion Various methods of adjustment for multiple testing will be compared and discussed.

Literature

1.   Dudoit S, Shaffer JP, Boldrick JC. Multiple hypothesis testing in microarray experiments. Statistical Science 2003; 18(1): 71-103.
2.   Gentleman R. Using GO for statistical analyses. 2004.
3.   Goeman JJ, de Kort F, van de Geer SA, van Houwelingen JC. A global test for groups of genes: testing association with a clinical outcome. Bioinformatics 2004; 20(1): 93-99.
4.   Khatri P, Draghici S. Ontological analysis of gene expression data: current tools, limitations, and open problems. Bioinformatics 2005.
5.   Mansmann U, Meister R. Testing differential gene expression in functional groups. Methods Inf Med 2005; 44(3).
6.   Subramanian A, Tamayo P, Mootha VK, Mukherjee S, Ebert BL, Gillette MA, Paulovich A, Pomeroy SL, Golub TR, Lander ES, Mesirov JP. Gene set enrichment analysis: a knowledge-based approach for interpreting genome-wide expression profiles. PNAS 2005; 102(43): 15545-15550.
7.   Westfall PH, Young SS. Resampling-based multiple testing – examples and methods for p-value adjustment. New York: Wiley; 1993.