Poster J-19

Probabilistic comparison of gene sets for detecting key metabolic pathways in the stress response of yeast

Authors:
Jean-Marc Schwartz (*Kyoto University*)
Claire Gaugain (*University of Bordeaux*)
Jose Nacher (*Kyoto University*)
Antoine de Daruvar (*University of Bordeaux*)
Minoru Kanehisa (*Kyoto University*)

Short Abstract: We analyse the flexible response of yeast to different stress conditions, including temperature shocks, nutrient starvation, and exposure to toxins. We show that a combination of correlation-based expression analysis with a probabilistic comparison of sets implemented by the BlastSets tool leads to uncovering key metabolic pathways related to stress response.

Long Abstract:
Functional genomics data should help us to better understand the behaviour of biological cells by uncovering the relationships between molecular mechanisms and cellular functions. Expression data represent a large part of functional genomics information and their analysis has become a major challenge in bioinformatics. The first step in the analysis of expression data is the identification of sets of genes having similar expression profiles. Many methods have been developed to find co-regulated genes in microarray experiments (Eisen et al., 1998; Galitski et al., 1999). Yet a major challenge remains to integrate gene sets derived from expression data with other sources of biological information, such as functional annotations, molecular interactions, cellular localisation, or metabolic pathways.

In this work, we first identified sets of genes that are co-regulated in response to various environmental stress conditions using a correlation-based approach. We combined expression data available from the GEO database (http://www.ncbi.nlm.nih.gov/geo/) corresponding to a number of different stress conditions in Saccharomyces cerevisiae, including temperature shocks, nutrient starvation, and exposure to various toxins from chemical or agricultural products. For each stress condition, we defined a binary signal representing an idealized expression pattern that is equal to one when the stress is present, to zero when it is not. The covariance between experimental expression values and idealized patterns was then computed for each gene. This process led to defining an individual zi(s) score for each gene i and each stress condition s, which is a measure of how strongly the expression pattern of a gene i is coupled to a condition s. Sets were then constructed from the genes which are significantly induced/repressed in each of these conditions.

Then, a bioinformatics tool called BlastSets (Barriot et al., 2004) was used to merge these information with metabolic pathway data. BlastSets makes it possible to bring together heterogeneous data available at the scale of genomes or proteomes. The data are structured in sets of biological entities (genes, proteins) that are stored in a database. These sets are

composed of genes/proteins that are known to have a biological relationship (proteins involved in a complex or a pathway). The database can be queried to find a correlation between any query set (for example, a group of co-regulated genes) and a set from the database (for example, a group of proteins involved in the same metabolic pathway).

We thus queried the sets of induced/repressed genes in stress conditions against the sets of genes built from the metabolic pathways of the KEGG database (http://www.genome.jp/kegg/kegg2.html) to identify pathways whose activity is significantly induced/repressed in response to each stress condition. These results complement the findings presented in the main track of the conference (Nacher, Schwartz et al., 2006), where the objective was to identify connected metabolic units by integrating gene co-expression data with topological information. In contrast, connectivity information is not considered in the present work, but we show that the combination of a correlation-based expression analysis with a probabilistic comparison of sets as implemented by the BlastSets tool leads to uncovering the relationships between the stress response and metabolic pathway activity.

R. Barriot, J. Poix, et al.: New strategy for the representation and the integration of biomolecular knowledge at a cellular scale. Nucleic Acids Res. 32(12), 3581-3589 (2004).

M. B. Eisen, P. T. Spellman, et al.: Cluster analysis and display of genome-wide expression patterns. Proc. Natl. Acad. Sci. U.S.A. 95(25), 14863-14868 (1998).

T. Galitski, A. J. Saldanha, et al.: Ploidy regulation of gene expression. Science 285(5425), 251-254 (1999).

J. C. Nacher, J.-M. Schwartz, et al.: Identification of metabolic units induced by environmental signals. Bioinformatics, accepted for presentation at ISMB 2006.