Experiment Specific Expression Patterns

Tobias Petri, Robert Küffner, and Ralf Zimmer

LMU Munich, Department of Informatics,
Amalienstr. 17, 80333 Munich, Germany
{petri,kueffner,zimmer}@bio.ifi.lmu.de
http://www.bio.ifi.lmu.de

The differential analysis of genes between microarrays from several experimental conditions or treatments routinely estimates which genes change significantly between groups. As genes are never regulated individually observed behavior may be a consequence of changes in other genes. Existing approaches like co-expression analysis aim to resolve such patterns from a wide range of experiments. The knowledge of such a background set of experiments can be used to compute expected gene behavior based on known links. It is particularly interesting to detect previously unseen specific effects in other experiments. Here, a new method to spot genes deviating from expected behavior (PAttern DEviation SCOring – *Padesco*) is devised. It uses linear regression models learned from a background set to arrive at gene specific prediction accuracy distributions. For a given experiment it is then decided whether each gene is predicted better or worse than expected. This provides a novel way to estimate the experiment specificity of each gene. We propose a validation procedure to estimate the detection of such specific candidates and show that these can be identified with an average accuracy of about 85 percent.

The extend of differential expression alone does not indicate experiment *specific* involvement of genes. Based on the prediction performance we identified specific candidates genes that exhibit experiment *specific* expression, i.e. expression changes that cannot be explained (predicted) by our models. This analysis is related to co-expression studies and complements differential expression analysis. It enables to focus on concise candidate lists for follow-up studies that consist of experiment-specific candidates only. We screened for filter thresholds and estimated *Padesco*'s performance from permutation tests as comprehensive gold standards for the experiment specific expression of genes are not available. This newly devised simulation approach suggests that *specific* candidates are identified reliably by *Padesco* ($> 85\%$ precision at $padscore > 1.5$) even if they show only marginal levels of differential expression. On the other hand, more than 90% of the genes selected by differential expression alone exhibit only generic expression patterns and could thus be excluded from further studies. *Specific* candidates are likely to represent characteristic features of the corresponding experimental conditions.

We evaluated *Padesco* selected genes for two data sets on prostate cancer. Besides interesting new candidates, we found several genes with a known involvement in the disease. Some of them, such as IL-2RB, have already been reported as promising drug targets. We demonstrated that such examples are more difficult to detect by differential expression analysis alone. Instead, differential expression tends to pick up genes that act similarly in other, biologically unrelated experiments. Thus, in combination with differential expression analysis, *Padesco* is a promising protocol for the detection and analysis of particularly distinctive features of microarray experiments.