Poster J-34

Condition transition analysis reveals TF activity related to nutrient-limitation-specific effects of oxygen presence in yeast

Authors:
Theo Knijnenburg (*Mediamatics,TU Delft*)
Lodewyk Wessels (*Mediamatics,TU Delft / Molecular Biology, The Netherlands Cancer Institute*)
Marcel Reinders (*Mediamatics,TU Delft*)

Short Abstract: Differential gene expression between the combinatorial cultivation conditions of a small and homogenous yeast microarray dataset are modeled as transitions. The interrelatedness of the cultivation conditions is exploited to reveal the combinatorial effects on expression behavior and TF activity at these transitions.

Long Abstract:
Regulatory networks are usually presented as graph structures showing the (combinatorial) regulatory effect of transcription factors (TF) on modules of similarly expressed or otherwise related genes. However, from these networks it not clear when and how TF's are activated. The actual conditions or perturbations that trigger a change in the activity of TF's should be a crucial part of the generated regulatory network. Three main reasons for this exclusion can be identified: Firstly, the present inability to directly measure protein levels in vivo prevents direct assessment of the presence of a TF in a particular condition. Secondly, in most cases post-transcriptional and/or post-translational regulation prevent deriving TF activity from gene expression. Thirdly, the trend of employing increasingly large compendia of heterogeneous microarray data, where yeast is grown under a wide variety of very different and unrelated conditions, makes it impossible to incorporate all these conditions in a regulatory program. Hence, the functionality of modules and TF's is assigned based on enrichment in annotation categories (e.g. Gene Ontology). This means that the functionality purely depends on the result of clustering, i.e. the grouping of genes, and not specifically on the cultivation conditions under which the expression behavior is characteristic for a module. This approach can only provide a global overview of TF activity and obstructs novel knowledge discovery, since an existing body of knowledge, i.e. the ontologies, is taken as a golden standard.
Here, we demonstrate the power in uncovering TF activity by focusing on a small, homogeneous, yet very well defined set of chemostat cultivation experiments, where the transcriptional response of yeast grown under four different nutrient limitations, both aerobically as well as anaerobically was measured. We define a condition transition as an instant change in yeast's extracellular environment by comparing two cultivation conditions where either the limited nutrient or the oxygen availability changes. Differential gene expression as a consequence of such a condition transition is represented in a tertiary matrix, where zero indicates no change in expression; 1 and -1 respectively indicate an increase and decrease in expression as a consequence of a condition transition.
Now, by consulting TF binding data (represented in a binary matrix indicating whether a TF can bind the upstream region of a gene or not), a hypergeometric test can be employed to

assess if genes that are up and/or downregulated at a condition transition are bound (upstream) by a TF much more frequently than would be expected by chance. However, the systematic setup of the cultivation conditions in this dataset, allows for selection of more interesting groups of genes to input into the hypergeometric test. For example, genes that are upregulated at an aerobic nutrient limitation transition, yet not upregulated at the same nutrient limitation transition without the presence of oxygen. More specifically, for each of the six nutrient limitation transitions we define nine different groups of genes allowing us to focus on upregulation (1), downregulation (-1) and differential expression (-1 or 1), both specifically for aerobic or anaerobic growth as well as regardless of the oxygen supply.

In an attempt to gain more insight into the dynamics and combinatorial effects within the complete generated regulatory network, in stead of performing stringent tests of individual hypotheses, we add an additional step to our analysis. Here, we aim at modeling the expression behavior at all condition transitions by employing the binding matrix and assess the activity of each TF at a condition transition. This approach is based on the simple biological model that ascribes the change of gene expression levels as observed at a condition transition to changes in TF activity; the means by which the organism adapts to the changed extracellular environment. Our problem finds its mathematical formulation in the orthogonal Procrustes problem, where we explore the possibility that the binding matrix can be rotated into the matrix representing the differential expression at the transitions.

Resultantly, our approach is able to infer TF activity related to very specific changes in combinatorial cultivation parameters. The algorithm that is especially designed for the combinatorial setup of nutrient limitations and oxygen supply in the employed microarray dataset,

not only provides unprecedented detailed insight into the behavior of yeast's metabolism and respiration at the transcriptional level, but also in terms of TF activity. Thus, we do not find many TF's that are globally related to particular nutrients. (These have already been identified in previous studies). More specifically, we identify lots of TF's that are not primarily related to the metabolism of a particular nutrient, yet seem to play a more specific and subtle (and as of yet unknown) regulatory role at these transitions between nutrient limitations. The involvement of these TF's demonstrate the complex and multiple regulatory roles that they exhibit in transcriptional regulation in different processes.

Discussion

Today's main use and strength of bioinformatics tools is generating hypotheses on all types of relationships and functionalities of and between quantifiable parameters inside and outside the cell. Specific biological experiments are, however, still required to validate the automatically generated hypotheses before accepting them as newly discovered knowledge. The common trend of focusing on large compendia of intracellular measurement datasets is often in contrast with the biologist's very specific field of research. These broad approaches are able to recognize global patterns in the data, but miss specific and subtle effects that characterize the complex reality of the cell. In this research we applied a tailor-made informatics approach on a small, well defined dataset. This enabled us to provide the biologist with very detailed hypotheses about the specific biological processes of interest.