Analysis of unreplicated time-course microarray experiments

Luchini A. , Callegaro A. , Bicciato S.

Department of Chemical Engineering Processes – University of Padova – Padova

Motivation

Since transcriptional control is the result of complex networks that interpret a variety of inputs, analyzing dynamical states of gene expression is of paramount importance to detect the multivariate nature of complex biological mechanisms. Although hundreds of studies fully demonstrated the relevancy of microarrays in describing the transcriptional status of different physiological conditions, to access and reconstruct complex interaction pathways it is necessary to analyze the temporal evolution of transcriptional states. However, an appropriate experimental design to accurately identify differentially expressed genes over a meaningful temporal window would require large amounts of microarrays and computational procedures able to assess the correlation structure among data at different time points. Unfortunately, replicates for each time point and experimental condition are not always available, because of cost limitations and/or biological sample scarcity, while common data analysis tools, e.g., ANOVA, do require replicates.

Methods

A method for the identification of differentially expressed genes in unreplicated time-course experiments based on multidimensional permutation tests is proposed. The procedure does not assume any model or distribution function, takes into account the correlation of data, and does not require sample replicates at the various time points, other than the presence of an initial time point t0 for all analyzed conditions (i.e., before the treatment administration). The identification of differentially expressed genes as the result of a system perturbation is formally stated as a hypothesis testing problem in which a defined statistic is used to rank transcripts in order of evidence against the null hypothesis. In general, data obtained from microarray studies does not support the normality assumption for applying standard hypothesis testing. Nevertheless, nominal type I error can be controlled using permutation tests, which are based on empirically constructed null distributions. The only assumption underlying a permutation test is that observations are exchangeable under a true null hypothesis. In this case, i) data are structured so that measurements are correlated in time, within the same biological condition; ii) the null hypothesis is formulated so that changes in expression levels at different time points are equivalent; iii) time point t0 represents the system before the perturbation. Therefore, modulated genes are detected testing the statistical significance of expression differences between physiological states at each time point, once corrected by the variability at t0, and given an empirical null distribution constructed using permutations. Statistical significance is assessed by the q-value, which represents a hypothesis testing error measure for each observed statistic with respect to the proportion of falsely rejected hypotheses (FDR).

Results

The method has been applied to three gene expression databases involving time-course microarray experiments aimed at studying the temporal changes of gene expression in: 1) skeletal muscle cells treated with a histone deacetylase inhibitor (Iezzi et al., 2004), 2) immature

mouse dendritic cells (DC) exposed to larval and egg stages of S. mansoni (Trottein et al., 2004), and 3) human optic nerve head (ONH) astrocytes in response to hydrostatic pressure (Yang et al., 2004). Differentially expressed genes identified using the proposed algorithm were compared and validated with results obtained from ANOVA model and SAM paired test. The biological significance and soundness of selected transcripts was also verified using global functional profiling by means of OntoTools. Results demonstrate that this novel procedure allows the identification of biologically relevant genes using half of the replicates required by standard model-based approaches. References Iezzi et al., Dev Cell, 2004 6(5), 673-684 Trottein et al., J Immunol, 2004 172(5), 3011-3017 Yang et al., Physiol Genomics, 2004 17(2), 157-169

Contact email: silvio.bicciato@unipd.it