A Linear Systems Analysis of Expression Time Series

T. Gregory Dewey+ and Ashish Bhan‡, +Keck Graduate Institute of Applied Life Sciences and ‡Department of Mathematics, Claremont Graduate University, Claremont, CA 91711.

A common approach to tackling complex phenomena is to first establish the extent and limitations of the linear response domain of the system. While there is no a priori reason to assume that the time series of a gene expression profile can be well related via a simple linear model, it is a reasonable starting point and can provide important clues to the origin of nonlinearity in the system. Linear models, in themselves, can often lead to surprisingly complicated responses, especially for multivariate systems. One way to establish the applicability of such models is to examine the time correlation functions of the system. While it is possible to have oscillatory time correlation functions, a hallmark of linear systems is exponential decaying correlation functions. Accordingly, we have examined the time dependence of the cross correlation function between different gene profiles in the cell-cycle of yeast. These correlation functions obtained for the cell cycle data in yeast are surprisingly simple and many of them appear as decaying exponential functions. This provides a strong motivation for an in-depth exploration of linear models. Additionally, an important feature of a systems analysis based on linear response is that it can be used to ascertain the match between the time scales of the system and the time scale of a given experiment. This traditionally has been used to separate fast and slow variables from each other. Thus, one can identify and remove genes that respond either too fast or too slow during the time course of the experiment. This leads to a reduction in complexity of the problem. In this work, a dynamic, linear response model for analyzing time series of whole genome expression is presented. The simplest assumption about expression profile data is that the expression state represented in the data from one time point determines the expression state seen at the next time point. This assumption is equivalent to modeling the data by a first-order Markov process. The linear version of this model is described by a single transition matrix, L, defining the transitions from one state to the next. The expression levels of each of m genes in the two points in time can be described by column vectors, a(t) and a(t-1), each of length m. The transition between the two states is modeled by: a(i,t)=sum{L(i,j)a(j,t-1)} where a(i,t) is the expression level of the ith gene at time t after some exposure or treatment. The transition coefficients are L(i,j) which are the respective elements of the transition matrix. The matrix elements represent the influence of the expression level of the jth gene on that of the ith gene. Using this model, we calculate a model state transition matrix for both cell-cycle and diauxic shift data is yeast. These models are statistically robust and lead naturally to a network of interactions reflected in the data. This provides a direct method of classifying genes according to their place in the resulting network and offers an alternative to traditional clustering approaches. These network groupings compare favorably with previously used methods like cluster analysis. The network derived by this method shows a hierarchical structure that is dominated by a collection of central hubs. These hubs are interconnected and have a cascade of tree-like structures attached to them. Non-linear and higher order Markov behavior of the network can also be included by a self-consistent method. Our dynamic method appears to give a broad and general framework for data analysis and modeling of gene expression arrays.