BioSystems 107 (2012) 9–17

Contents lists available at SciVerse ScienceDirect

BioSystems

journal homepage: www.elsevier.com/locate/biosystems

Nonlinear response of gene expression to chemical perturbations: A noise-detector model and its predictions

Jose C. Nacher*, Vladimir B. Ryabov*

Department of Complex and Intelligent Systems, Future University Hakodate, 116-2 Kamedanakano-cho, Hakodate, Hokkaido 041-8655, Japan

ARTICLE INFO

Article history:
Received 29 June 2011
Received in revised form 8 August 2011
Accepted 12 August 2011

Keywords:
Systems biology
Gene expression dynamics
Signal processing
Inflammatory response

ABSTRACT

The widespread use of microarrays provided a first glimpse at some simple laws and organizing principles that govern the transcriptome. Previous analyses have shown that the transcriptional organization is very heterogeneous and characterized by a power–law decay for gene expression levels. Moreover, a simple law was unveiled suggesting that gene expression dynamic changes under stress are proportional to their initial expression values. However, to elucidate and assess the underlying governing principles of transcriptional organization, we do not only need to identify them, but also provide theoretical models that are able to faithfully capture and reproduce them. Here we present a method to investigate the gene expression dynamics inspired by the theory of nonlinear transformation of random signals and noise. The model is able to explain not only the well-known power–law decay for abundance of expression levels, but also to reproduce the linear dependence of the standard deviation of gene expression change with respect to the initial expression value (also known as *rich-travels-more dynamics*). To our knowledge, this is the first model applied to gene expression dynamics that is able to simultaneously predict both statistical features. The theoretical framework derives an indicator to measure the coupling between gene expression and specific perturbations. Using genome-wide transcriptional data, this indicator identifies genes strongly coupled to specific inflammatory responses to different pathogens. The novel application of signal and noise theory to study intracellular responses and gene expression changes offers not only a new theoretical avenue to study transcriptional responses to environmental stresses and chemical signals but also provides predictive capability at the genome scale.

1. Introduction

Recent developments in high-throughput data acquisition technologies enable researchers to collect multi-dimensional data and investigate a growing number of organisms, stress conditions and disease stages not only by using the genome sequences but also by means of gene expression profiles (Marton et al., 2001; Brown and Botstein, 1999; Rocke and Durbin, 2001; Okou et al., 2007), protein interactions (Uetz et al., 2000; Giot et al., 2003; Pereira-Leal and Levy, 2008) and metabolic pathways (Raamsdonk et al., 2001; Allen et al., 2003) among many others, offering a novel systems-wide view of cellular processes (Kitano, 2002; Barabási and Oltvai, 2004; Bray, 2003; Barabási et al., 2011).

Many types of cellular networks emerge at different levels from the interaction of fundamental bio-molecules. These networks and pathways also interact with each other painting a huge interwoven pattern of relationships that lead the complex behavior of the cell. The network analysis has shown us that most of these networked structures, from gene regulatory networks and metabolic pathways to protein networks, converge to similar nonrandom architectures with unique features, helping to elucidate the high impact of the topology on systems functionality and behavior (Barabási and Oltvai, 2004; Bray, 2003). However, since the cell is a highly dynamic system, structural analysis alone is not sufficient. We need to reach a global understanding of the behavior of a cell to predict how cells and tissues respond to environmental variations and specific pathologies, or are affected by molecularly targeted treatments (Barabási et al., 2011).

Genome-scale expression analyses have been performed for a wide range of experimental conditions (Ma and Zeng, 2003; De la Fuente et al., 2002) and disease stages, but the relations between the large-scale organization of gene expression and its underlying dynamics have not been sufficiently understood. In particular, we have not found yet a theoretical framework that is able to embed the recently observed universal features of gene expression and offer predictive power.

In recent years, the pioneer analyses of genome-wide gene expression profiles have allowed us to get a first glimpse of some simple laws and organizing principles that govern the transcriptome. Experimental analyses have shown that the transcriptional

* Corresponding authors. Tel.: +81 0138 34 6123.
E-mail addresses: nacher@fun.ac.jp (J.C. Nacher), riabov@fun.ac.jp (V.B. Ryabov).

0303-2647/$ – see front matter © 2011 Elsevier Ireland Ltd. All rights reserved.
doi:10.1016/j.biosystems.2011.08.002

organization is very heterogeneous. The probability that a gene has an amount of expression level k decays as a power–law $f(k) \sim k^{-\gamma}$ (Kuznetsov et al., 2002; Furusawa and Kaneko, 2003; Ueda et al., 2004). This organization seems to be universal across species, from *Escherichia coli* to *Homo sapiens* and even in cell types, tissues and disease states. Although expression levels of individual genes change, the invariance of the power–law suggests a robust property in gene expression dynamics. Moreover, another simple law was unveiled suggesting that gene expression changes are proportional to their initial expression values (*rich-travel-more dynamics*) (Ueda et al., 2004). It was argued that this proportional dynamics explains the power–law organization. To elucidate this finding, a simple model was proposed where the standard deviation of gene expression change increases proportionally to the before-transition gene expression level k_0. In this approach, the temporal gene expression changes are represented by a continuous Markov process (van Kampen, 1992), and the solution of the one-dimensional Kolmogorov forward equation leads to a power–law distribution. However, in spite of its importance, the model is not able to simultaneously predict both observed statistical properties in gene expression profiles. The model only regenerates the power–law distribution once a fluctuation term, proportional to the initial expression value, is inserted. Therefore, the basic mechanism in the transcriptional system that is able to generate the observed response property was not found yet.

On the other hand, systems biology has changed the traditional reductionism approach in molecular biology into a new understanding of the cell as a complex system. In this systemic view, the whole is more than the sum of the parts and the system's functionally emerges from the interactions of its fundamental constituents. This shift has led to the development of simple models that capture the system's main features. Such coarse-grained approach becomes central in statistical physics. It is difficult to predict the behavior of the whole system analyzing the trajectories of single individual water or mRNA molecules, i.e., using a microscopic approach. In contrast, the study of a macroscopic view may often lead to profound understanding of the laws that emerge from underlying complex interactions at the microscopic level.

Statistical data analyses and computational biology provide a variety of techniques, like clustering methods, to identify and classify genes at macroscopic level using transcriptional data (Kerr et al., 2008). However, most of these techniques lack an underlying theoretical model that could explain the mechanism of forming the observed distribution functions from a minimal set of fundamental principles. Unsupervised clustering, for example, does not need at all any biological information of the system. At most, a successful identification of a single or several genes responsible for a given effect can be achieved that does not always lead, however, to an understanding of the governing principles and mechanisms in a cell. In contrast, the coarse-grained approach we present here highlights the importance of universal mechanisms underlying gene expression data that can be successfully utilized in the situation when quantitative information on the complex dynamics of the micro-processes that take place in intra-cellular molecular networks is not available or too complicated.

Here we first perform an analysis on recent microarray time-course experiments from macrophages stimulated with a variety of Toll-like receptor (TLR) ligands that reveals near to proportional dependence between the change in the expression rate and its initial value, as well as close to power–law decay of the inflammatory gene expression response. We then propose a simple probabilistic model that is able to reproduce the both known statistical properties of the gene expression profiles: (1) power–law decay of the expression level distribution function at small k-values and deviations from the power–law at large ones; (2) approximately linear dependence between the temporal variation in the expression level and its initial value. The statistical framework we have developed is based on the theory of random signals and noise (RSN) (Davenport and Root, 1958), and it leads to new insights about the underlying dynamics of the coupling between gene expression and external signals. The model not only allows us to make predictions on the gene expression response to the action of different pathogen-associated molecular patterns but also leads to identification of genes that are most strongly affected by them.

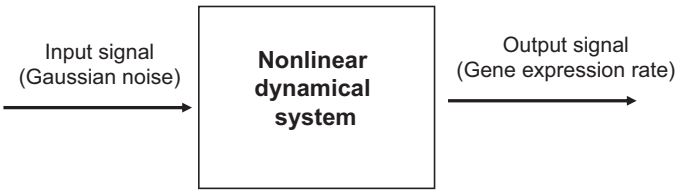

Fig. 1. A schematic of the gene model.

2. Methods

2.1. Datasets

The time-course transcriptional measurement data were downloaded from http://portal.systemsimmunology.org. The dataset contains time-course gene expression data from murine bone marrow-derived macrophages stimulated with six Toll-like receptor agonists (Nykter et al., 2008) (numbered, respectively, with the numbers from 1 to 6 in the subsequent discussion): unmethylated CpG-containing oligodeoxynucleotide (5), lipopolysaccharide (LPS) (15), PAM2 (6), PAM3 (10), Poly(IC) (11) and R848 (8). The numbers in parentheses refer to the number of time point measurements. Gene expression values were annotated every 20 min. These TLR agonists represent different pathogen-associated molecular patterns (PAMPs) that can be detected by the cell surface detectors and process the signals through interwoven cellular pathways that lead to activation and regulation of the inflammatory response. As a result, specific gene expression programs are induced to deal with pathogens. The dataset contains a total 55 time point measurements for each of the 20,202 expressed genes.

2.2. The Model

Below we present a simple probabilistic model that is able to reproduce the known statistical properties of the gene expression rates described in Section 1. We assume that the temporal dynamics of each gene can be approximated with the same model, i.e., each gene can be regarded as a nonlinear input–output system that transforms the fluctuations in the concentration of circumambient chemicals into temporal variations of the expression rate. In the following, we use terminology coming from the theory of random signals and noise (RSN) (Davenport and Root, 1958), since we have found many analogies between the statistics of gene expression levels and similar properties of noise passed through standard radio circuits (Davenport and Root, 1958). Starting from the last century, the RSN-theory has been widely used for building a theory of communication systems (Davenport and Root, 1958), and it constitutes a basis of the modern electronics. Hereinafter, we shall call the fluctuations in the environment as input noise, and the resulting variation in the expression levels – as output noise, or signal, or response (see Fig. 1).

A typical scheme widely used in, e.g., broadcasting radio receivers that transform modulated radio waves into audio signals is shown in Fig. 2. It consists of an input filter/amplifier followed by a detector, a device consisting of a nonlinear element with low-pass filter. The role of the nonlinear element is to "rectify" the high-frequency input signal, i.e., to create a slow oscillation at the output corresponding to the encoded audio signal. This low-frequency signal then can become audible being separated by the low-pass filter from numerous high-frequency components generated by the nonlinear element. Note that such a system works well with virtually any shape of the nonlinear function describing the nonlinear element of course, satisfying certain conditions (Davenport and Root, 1958).

There are several intuitive considerations that justify the analogy between the operation of a receiver and gene dynamics. First, we note that since every gene has a specific function in the genome and occupies a characteristic topological position in the regulatory network, a selective property that takes into account this information should be considered. In terms of RSN-theory, this can be interpreted as different sensitivity of gene to the input signal, therefore, we can assume that there is a mechanism that amplifies (or suppresses) input fluctuations, depending on the presence of certain chemicals nearby. We thus have to include an input element to the model which in the simplest case can be approximated by a linear amplifier or attenuator. Second, the typical time scale of the changes in the expression rate is of the order of minutes, whereas the characteristic time of fluctuations on the molecular level can be of the order of microseconds (Frauenfeldera et al., 2009). In order to transfer the information contained in the very fast processes to slow measurable