Bioinformatics Advance Access published January 17, 2006

BIOINFORMATICS

Inference of Gene Regulatory Networks and Compound Mode of Action from Time Course Gene Expression Profiles

Mukesh Bansal [1,2], Giusy Della Gatta [1,3], Diego di Bernardo [1,2]

[1]Telethon Institute of Genetics and Medicine, Via P. Castellino 111, 80131 Naples, Italy, [2]European School of Molecular Medicine (SEMM), Naples, Italy, [3]Seconda Universita degli Studi di Napoli, Naples, Italy

Associate Editor: Satoru Miyano

ABSTRACT

Motivation: Time series expression experiments are an increasingly popular method for studying a wide range of biological systems. Here we developed an algorithm that can infer the local network of gene-gene interactions surrounding a gene of interest. This is achieved by a perturbation of the gene of interest and subsequently measuring the gene expression profiles at multiple time points. We applied this algorithm to computer simulated data and to experimental data on a 9 gene network in *Escherichia coli*.

Results: In this paper we show that it is possible to recover the gene regulatory network from a time series data of gene expression following a perturbation to the cell. We show this both on simulated data and on a nine genes subnetwork part of the DNA-damage response pathway (SOS pathway) in the bacteria *Escherichia coli*.

Contact: dibernardo@tigem.it

1 INTRODUCTION

Recent developments in large-scale genomic technologies, such as DNA microarrays and mass spectroscopy have made the analysis of gene networks more feasible. However, it is not obvious how the data acquired through such methods can be assembled into unambiguous and predictive models of these networks. Different experimental and computational methods have been proposed to tackle the network identification problem (Tong *et al.* (2002); Lee *et al.* (2002); Ideker *et al.* (2001); Davidson *et al.* (2002); Arkin *et al.* (1997); Yeung *et al.* (2002)). Although implemented with some success, they are data intensive and they may require a certain degree of *a priori* information.

A variety of mathematical models can be used to describe genetic networks (de Jong (2002); Savageau (2001); ALevchenko and Iglesias (2002)), including Boolean logic (Shmulevich *et al.* (2002); Liang *et al.* (1998)), Bayesian networks (Hartemink *et al.* (2002)), graph theory (Wagner (2001)) and ordinary differential equations (Tegner *et al.* (2003)). We concentrated our efforts on the last method as it offers a description of the network as a continuous time dynamical system that can be used to infer the genes with the major regulatory functions in the network.

In a recent study (Gardner *et al.* (2003)), we developed an algorithm (Network Identification by multiple regression −NIR) that used a series of steady state RNA expression measurements, following transcriptional perturbations, to construct a model of a 9 gene network that is a part of the larger SOS network in *E. coli*(Gardner *et al.* (2003)). Though the NIR method proved highly effective in inferring small microbial gene networks, it requires prior knowledge of which genes are involved in the network of interest, and the perturbation of all the genes in the network via the construction of appropriate episomal plasmids. In addition, it requires the measurement of gene expressions at steady state (i.e., constant physiological conditions) after the perturbation. This experimental setup is challenging for large networks, it is not easily applicable to higher organisms, and, most importantly, it is not applicable if there is no prior knowledge of the genes belonging to the network.

In this paper we are presenting an algorithm **TSNI** (Time Series Network Identification) that can infer the local network of gene-gene interactions surrounding a gene of interest by perturbing only one of the genes in the network. To this end, we need to measure gene expression profiles at multiple time points following perturbation of the gene, or genes, of interest.

We investigated the effect of noise and a limited number of data points on the performance of the algorithm, and we devised techniques to overcome these problems.

Our algorithm is illustrated and tested *in silico* on computer simulated gene expression data and applied to an experimental gene expression data set obtained by perturbing the SOS system in the bacteria *Escherichia coli*.

The novelty of our approach is in the idea of a gene-centric inference method that can be applied to infer the regulatory interactions of a gene of interest. State-of-the-art inference algorithms start from the assumption that a gene network is unknown and experiments are performed to perturb it. Gene expression data are then used to reconstruct the network. In a real life situation, large scale gene expression data from a given cell type involve thousands of responsive genes and there are many different regulatory networks activated at the same time by the pertrubations. In this case, inference methods can be successful but only on a subset of the genes (i.e. a specific network) (Basso *et al.* (2005)). This subset of genes (network), however, cannot be defined a priori but depends on the data set. Networks involving genes that never change in the data set, for example, cannot be inferred. We aim at developing an integrated experimental and computational approach to infer the network of a specific gene of interest.