Methodologies for Modeling Gene Regulatory Networks

Shruti Mishra
Siksha 'O' Anusandhan University, India

Debahuti Mishra
Siksha 'O' Anusandhan University, India

INTRODUCTION

One of the major breakthroughs in the field of molecular biology is the development of a computational branch: famously stated as *Bioinformatics*. It comprises of two major fields that is molecular biology and information technology. The recent development in this field includes understanding the human genome, new forensic techniques (Gomaa, 2011), as well as the discovery of new medicines of the future (Claverie & Nortredame, 2007). One of the scientific aspects of bioinformatics is that it integrates different fields like machine learning, biology, statistical modeling etc.

Genes being the basic units of heredity in the living organisms plays a major role in the control of cellular processes (Li, Lam, & Shu, 2010).They are the codes of proteins and is transcribed into the corresponding mRNA in the transcription process and then is translated into a protein. With the advancement in the techniques of molecular biology it is possible to measure the gene expression level. Present microarray technology (Schena, Shalon, Davis, Brown, 1995; Shalon, Smith, & Brown, 1996; Ramsay, 1998; Lockhart & Winzeler, 2000) allows measuring the expression levels of thousands of genes in a particular cell at a particular time in a specific condition (Lin, Yeh, Cheng, & Soo, 2007). Hence, they provide information about the gene networks.

Bioinformatics being such an important advancement in the recent development process has one of the most challenging problems: that is to discover relationships and interactions among genes (i.e. discovering gene regulatory networks (Alleyne, 2009)). The network usually represents the regulations among the genes and the basic objective is to extract the expression features, activations and inhibitions from the gene expressions changes of the gene that are present in the microarray data. As the expression of the gene that encodes the regulator, is also regulated by the functional products of some other genes, this forms many complicated regulatory interactions that constitute the structures of underlying Gene Regulatory Networks (GRNs) (Patrick & Keith, 2008).

An intense study of computational biology and the analysis of gene expression networks and metabolic pathways have given rise to various types of GRN (Kauffman, 2007; Savageau, 1996; Arkin,Ross, & McAdams 1998; D'haeseleer, Wen, Fuhrman, & Somogyi, 1999) models. The modeling criterion varies in terms of the details of biochemical interactions, discrete or continuous gene expression level used, deterministic or stochastic approach that has been applied, and so forth (D'haeseleer, Liang, Fuhman, & Somogyi, 2000; Noman & Iba, 2007). These criterions also define how closely the model can represent the genetic interactions. Usually the detailed modeling is very useful for acquiring the precise mechanism in common regulatory pathways. But as we try to simply the network, the complexity of the model increases accordingly. In addition, with the increase in model complexity, the data requirement for learning the model parameters also increases. Despite of various advancement in technology, the microarray technologies (including oligonucleotide arrays and complementary DNA (cDNA) microarrays) were not able to acquire the quality and quantity of data that is required for the accurate reconstruction of genetic networks (Schena, 1999). Therefore, depending on the characteristics of the model used, the availability of the gene expression data, and the level of noise present in the data, the concept of reverse engineering of a genetic network is possible and successful. Thorough review of various integration

DOI: 10.4018/978-1-4666-5888-2.ch041

strategies to support gene regulatory construction was introduced (Chen & VanBuren, 2012) which talked a lot of different distinctive strategies for data integration in gene regulatory network construction.

A massive scale gene construction and testing was done (Gibson, Ficklin, Isaacson, Feltus, & Smith, 2013) using random matrix theory, that showed a huge decrease in network construction time and computational requirement.

Gene networks inferred solely based on the microarray data are often not sufficient for rigorous analysis. In order to overcome the problem, combination of the biological knowledge with the modeling process becomes necessary (Segal, Barash, Simon, Freidman, & Koller, 2002). In molecular biology, transcription factors are important proteins that facilitate the transcription of DNA into mRNA, and they play a significant role in the gene regulatory mechanism (Lin et al., 2007). The transcription factors always control the regulation of the genes that bind to the promoter region of the DNA sequence, located in the upstream regions of the genes. Scientists believe that genes are regulated by the transcription factors at the corresponding transcription factor binding sites in the promoter region that usually leads to gene expression change that is observed in microarray data. Therefore, networks between the transcription factors and their target genes can help biologists understand the complex regulatory mechanisms in a cell. A Markov logic network learning was proposed (Brouard, Vrain, Dubois, Castel, & Debily, 2013) with some set of rules that can be used to predict some new regulations.

Understanding of regulatory networks has always been a crucial task for the understanding of fundamental cellular processes that involves growth, development, hormone secretion, and cellular communication (Mitra, Das, & Hayashi, 2011). The determination of transcriptional factors which control gene expression, and their targets, basically offer a clear view into the dysregulated expressions that is common in many human diseases. This is plays a crucial role for the applications in drug development, medicine, nutrition, and other therapeutic activities.

The objective is of this article is to review all the information available about the Gene Regulation.

BACKGROUND

The prime interest of the researchers is the identification of gene regulatory network (GRN) as it provides an insight to activation or inhibition of a gene in living organisms. GRN (Ibrahim,Ngom, & Tawfik, 2011) is basically collection of genes that influence or regulate other genes or themselves. Technically, a GRN is a collection of DNA segments in a cell which interact with each other indirectly through their RNA and protein expression products with different substances in the cell, governing the rates at which genes in the network are transcribed into mRNA. The DNA microarrays have provided a simple way for collecting the expression levels of thousands of genes. So the ultimate goal of much analysis is learning gene interaction networks to describe and simulate how genes interact with each other (Yavari,Towhidkhah, Gharibzadeh, 2011). The network aims to represent relationships that govern the rates at which genes in the network are transcribed into mRNA.

Genes are stated to be the *on-off switch* (Kanterakis, Kafetzopoulos, Moustakis, & Potamias, 2008) of a cell that operates at the gene level. Regulation of gene being one of the most activating process in living cells, must decide which genes to express at a particular time. As the development process advances, different needs and functions entail an efficient mechanisms to turn the required genes on while leaving the others off. Cells can also activate new genes to respond effectively to environmental changes and the knowledge of which gene triggers a particular genetic condition can help ward off the potential harmful effects by turning that gene off. For example, cancer can be controlled by deactivating the gene that causes it.

Figure 1. shows a generalized model of gene regulatory model where nodes represent genes and the edges represent interactions among them. Figure 2 (Barker, 2011) depicts a genetic network for part of the phage λ decision circuit. This network is constructed from DNA that includes the genes *c*I and *c*II that are blueprints for producing proteins, CI and CII.

A vital question is how to build a GRN model. There are many simplified and complex models (Farina, Santis, Morelli, & Ruberti, 2007) available for