Chapter 4
Inferring Gene Regulatory Networks from Genetical Genomics Data

Bing Liu
Monsanto Co., USA

Ina Hoeschele
Virginia Polytechnic Institute and State University, USA

Alberto de la Fuente
CRS4 Bioinformatica, Italy

ABSTRACT

In this chapter, we review the current state of Gene Regulatory Network inference based on 'Genetical Genomics' experiments (Brem & Kruglyak, 2005; Brem, Yvert, Clinton & Kruglyak, 2002; Jansen, 2003; Jansen & Nap, 2001; Schadt et al., 2003) as a special case of causal network inference in 'Systems Genetics' (Threadgill, 2006). In a Genetical Genomics experiment, a segregating or genetically randomized population is DNA marker genotyped and gene-expression profiled on a genomewide scale. The genotypes are regarded as natural, multifactorial perturbations resulting in different gene-expression 'phenotypes', and causal relationships can therefore be established between the measured genotypes and the gene-expression phenotypes. In this chapter, we review different computational approaches to Gene Regulatory Network inference based on the joint analysis of DNA marker and expression data and additionally of DNA sequence information if available. This includes different methods for expression QTL mapping, selection of regulator-target pairs, construction of an encompassing network, which strongly constrains the network search space, and pairwise and multivariate methods for Gene Regulatory Network inference, such as Bayesian Networks and Structural Equation Modeling.

DOI: 10.4018/978-1-60566-685-3.ch004

INTRODUCTION

A fruitful abstraction of biochemical systems is that of 'networks' (Barabasi & Oltvai, 2004; Dorogovtsev & Mendes, 2003; Newman, 2003; Pieroni et al., 2008; Watts & Strogatz, 1998). Such networks include Transcription Regulatory Networks (TRNs) (Lee et al., 2002; Luscombe et al., 2004; Shen-Orr, Milo, Mangan & Alon, 2002), Protein Interaction Networks (Pieroni et al., 2008; Schwikowski, Uetz & Fields, 2000), Metabolic Networks (Jeong, Tombor, Albert, Oltvai & Barabasi, 2000; Wagner & Fell, 2001), Gene Regulatory Networks (GRNs) (Brazhnik, de la Fuente & Mendes, 2002; D'Haeseleer, Liang & Somogyi, 2000) (see also A. de la Fuente – this book}, and Phenotype Networks (Nadeau et al., 2003). Inferring, or 'reverse engineering', such biological networks is therefore currently an area of research receiving a lot of interest and attention. It advances our knowledge about the integrated biochemical machinery of living cells (systems biology) and our understanding of general features of complex traits (complex trait biology). Constructing phenotype networks provides information about the functionality of complex systems (such as cardiovascular function) at the organismal level, and constructing GRNs furthers our understanding of the molecular basis of complex traits and diseases (Chen et al., 2008; Lum et al., 2006; Schadt et al., 2005). GRNs have other applications (Brazhnik, de la Fuente & Mendes, 2002), including the discovery of direct drug targets (di Bernardo et al., 2005; Gardner, di Bernardo, Lorenz & Collins, 2003). It has been shown that classical concepts from genetics, such as dominance and epistasis, can be readily understood in terms of networks and their properties (Kacser & Burns, 1981; Omholt, Plahte, Oyehaug & Xiang, 2000).

Many different experimental and computational approaches to GRN inference have been proposed. Data from experiments without targeted perturbations, or data from observational studies, only allow for inference of undirected Co-Expression Networks that are based on a measure of association between the expression profiles of pairs of genes (e.g.de la Fuente, Bing, Hoeschele & Mendes, 2004; Ghazalpour et al., 2006; Schäfer & Strimmer, 2005a,, 2005b; Wille & Buhlmann, 2006; Wille et al., 2004; Zhang & Horvath, 2005). In particular, one can construct an Undirected Dependency Graph (UDG), which contains edges only between those genes that interact directly, and which can be estimated based on partial correlations (de la Fuente, Bing, Hoeschele & Mendes, 2004; Shipley, 2002). The construction of a UDG can be a first step in a regulatory network analysis of a Genetical Genomics or Systems Genetics experiment.

A strategy of targeted perturbation is required to enable causal inference needed for the identification of the directed structure of GRNs. In such a strategy, targeted perturbations are created and responses of the gene-expression levels to the perturbations are measured. It has been shown that this approach can provide a reliable identification of GRNs (Brazhnik, de la Fuente & Mendes, 2002; de la Fuente, Brazhnik & Mendes, 2002; Gardner, di Bernardo, Lorenz & Collins, 2003; Wagner, 2001). There are two major types of targeted perturbation experiments. One approach uses one-at-a-time, specific perturbations in the expression of individual genes (e.g.Hughes et al., 2000; Mnaimneh et al., 2004). These experimental perturbations are relatively expensive and difficult to perform, especially in quantities required for comprehensive GRNs identification. Such perturbations (knock-outs, over-expressions) also tend to have strong biological effects, making it potentially difficult to distinguish between 'normal' functional relationships and relationships that emerge when the 'normal' functionality of a system is compromised.

The second type of targeted perturbation experiments, Genetical Genomics and Systems Genetics, uses naturally occurring, multi-factorial perturbations in segregating or genetically randomized populations