

# Chapter 19

## A Linear Programming Framework for Inferring Gene Regulatory Networks by Integrating Heterogeneous Data

**Yong Wang**

*Academy of Mathematics and Systems Science, China*

**Rui-Sheng Wang**

*Renmin University, China*

**Trupti Joshi**

*University of Missouri, USA*

**Dong Xu**

*University of Missouri, USA*

**Xiang-Sun Zhang**

*Academy of Mathematics and Systems Science, China*

**Luonan Chen**

*Osaka Sangyo University, Japan*

**Yu Xia**

*Boston University, USA*

### ABSTRACT

*There exist many heterogeneous data sources that are closely related to gene regulatory networks. These data sources provide rich information for depicting complex biological processes at different levels and from different aspects. Here, we introduce a linear programming framework to infer the gene regulatory networks. Within this framework, we extensively integrate the available information derived from multiple time-course expression datasets, ChIP-chip data, regulatory motif-binding patterns, protein-protein*

DOI: 10.4018/978-1-60566-685-3.ch019

*interaction data, protein-small molecule interaction data, and documented regulatory relationships in literature and databases. Results on synthetic and real experimental data both demonstrate that the linear programming framework allows us to recover gene regulations in a more robust and reliable manner.*

## **INTRODUCTION**

Cells efficiently carry out molecular synthesis, energy transduction, and signal processing across a range of environmental conditions by gene networks, which we define broadly as networks of interacting genes, proteins, and metabolites. Microarray technologies enable the simultaneous measurement of all RNA transcripts in a cell, producing tremendous amounts of gene expression data from different research groups. For instance, the Stanford Microarray Database (SMD) has deposited data for 70,113 experiments, from 341 labs and 56 organisms, as of 2007 (Demeter et al., 2007). Thus there is a pressing need for the development of sophisticated algorithms for reverse-engineering gene networks. So far, many computational algorithms have been developed to analyze gene expression profiles to detect dependencies among genes over different conditions.

Generally speaking, there are two strategies for studying the relationships among genes. The “physical (direct) interaction” approach seeks to identify true physical interactions between regulatory proteins and their binding promoters to reconstruct the so-called transcriptional regulatory network (R. S. Wang, Wang, Zhang, & Chen, 2007). The second strategy, the “genetic (indirect) interaction” approach seeks to identify regulatory influences between RNA transcripts to reconstruct the so-called gene regulatory network (Y. Wang, Joshi, Zhang, Xu, & Chen, 2006). Thus, in general, the regulator transcripts may exert their effects indirectly through the action of proteins, non-coding RNA, metabolites, and the cell environmental factors. An advantage of the influence strategy is that the model can implicitly capture regulatory mechanisms at the protein and metabolite level that are not physically measured (Gardner & Faith, 2005). In this study we focus on the inference problem for gene regulatory networks. The detailed descriptions on the first strategy, i.e. inferring transcriptional regulatory networks, can be found in (R. S. Wang et al., 2007).

So far, a wide variety of approaches have been proposed to infer gene regulatory networks from time-course data or perturbation experiments (De Hoon, Imoto, Kobayashi, Ogasawara, & Miyano, 2003; Dewey & Galas, 2001; Friedman, 2004; Gardner, di Bernardo, Lorenz, & Collins, 2003; Holter, Maritan, Cieplak, Fedoroff, & Banavar, 2001; Husmeier, 2003; Nachman, Regev, & Friedman, 2004; Tegner, Yeung, Hasty, & Collins, 2003). These approaches include discrete models of Boolean networks and Bayesian networks, and continuous models of neural networks and difference/differential equations. A common challenge for all these models is the scarcity of the data, since a typical gene expression dataset consists of relatively few time points (often less than 20) with respect to a large number of genes (generally over thousands). In other words, the number of genes far exceeds the number of time points for which data are available, making the problem of determining gene regulatory network structure a difficult and ill-posed one (D’Haeseleer, Liang, & Somogyi, 2000).

On the other hand, there are many heterogeneous data sources closely related to gene regulatory networks. These data sources provide rich information for depicting complex biological processes in cellular systems at different levels and from different aspects. It is necessary and important to understand gene expression and regulation through mining these data sources. Currently high-throughput microar-

24 more pages are available in the full version of this document, which may be purchased using the "Add to Cart" button on the publisher's webpage:

[www.igi-global.com/chapter/linear-programming-framework-inferring-gene/38247](http://www.igi-global.com/chapter/linear-programming-framework-inferring-gene/38247)

## Related Content

---

### Ambient Assisted Living (AAL) Systems to Help Older People

Rita Komalasari (2023). *Exploring Future Opportunities of Brain-Inspired Artificial Intelligence* (pp. 84-99).

[www.irma-international.org/chapter/ambient-assisted-living-aal-systems-to-help-older-people/320611](http://www.irma-international.org/chapter/ambient-assisted-living-aal-systems-to-help-older-people/320611)

### An Incremental Evolutionary Approach to Tabu Search

Yoshinori Suzuki and Juan David Cortes (2017). *Recent Developments in Intelligent Nature-Inspired Computing* (pp. 133-151).

[www.irma-international.org/chapter/an-incremental-evolutionary-approach-to-tabu-search/179367](http://www.irma-international.org/chapter/an-incremental-evolutionary-approach-to-tabu-search/179367)

### Genetic-Algorithm-Based Optimization of Fragile Watermarking in Discrete Hartley Transform Domain

Sudipta Kr Ghosal and Jyotsna Kumar Mandal (2016). *Handbook of Research on Natural Computing for Optimization Problems* (pp. 103-127).

[www.irma-international.org/chapter/genetic-algorithm-based-optimization-of-fragile-watermarking-in-discrete-hartley-transform-domain/153810](http://www.irma-international.org/chapter/genetic-algorithm-based-optimization-of-fragile-watermarking-in-discrete-hartley-transform-domain/153810)

### The Grand Challenges in Natural Computing Research: The Quest for a New Science

Leandro Nunes de Castro, Rafael Silveira Xavier, Rodrigo Pasti, Renato Dourado Maia, Alexandre Szabo and Daniel Gomes Ferrari (2014). *Natural Computing for Simulation and Knowledge Discovery* (pp. 237-250).

[www.irma-international.org/chapter/the-grand-challenges-in-natural-computing-research/80069](http://www.irma-international.org/chapter/the-grand-challenges-in-natural-computing-research/80069)

### Existence of Positive Solutions of Nonlinear Second-Order M-Point Boundary Value Problem

F. H. Wong, C. J. Chyan and S. W. Lin (2011). *International Journal of Artificial Life Research* (pp. 28-33).

[www.irma-international.org/article/existence-positive-solutions-nonlinear-second/52976](http://www.irma-international.org/article/existence-positive-solutions-nonlinear-second/52976)