Guoliang Li, Tze-Yun Leong, Louxin Zhang, Bin Han
Estimating Relative Transcriptional Time Delays from Multiple Microarray Time-Series Data
Appeared in: The 3rd Annual RECOMB Satellite Workshop on Regulatory Genomics (2006) 138-139

Estimating Relative Transcriptional Time Delays from Multiple Microarray Time-Series Data

Guoliang Li[1], Tze-Yun Leong[1], Louxin Zhang[2], Bin Han[3]

[1] Medical Computing Laboratory, School of Computing
[2] Department of Mathematics
National University of Singapore
{ligl, leongty}@comp.nus.edu.sg
matzlx@nus.edu.sg
[3] Clayton School of IT, Monash University, Australia
bhan@csse.monash.edu.au

Keywords: Gene regulatory networks, microarray data, time series, transcriptional time delay, gene expression

1 Introduction

The gene regulatory networks are a type of models to describe a set of genes (possibly all the genes) in a cell and the regulation relationships among them. In the networks, the regulating genes can affect the expression levels of the regulated genes transcribed to mRNAs. Usually, the expression level of one gene is regulated by a set of genes, which are in turn regulated by another set of genes. Such regulation relationships among genes can be represented as a network, called gene regulatory networks. Traditionally the regulation among genes is determined through small scale experiments. Recently the DNA microarray technique can measure the expression levels of thousands of genes simultaneously, which opens the door to explore the regulations among genes in a genomic scale. Currently, elucidating gene regulatory networks from microarray data is a big challenge to understand the mechanisms in biology.

One main factor in all the proposed temporal models for gene regulatory networks is the transcriptional time delay (TTD). The transcriptional time delay is the time from the change in one gene's expression level to its effect on the expression level of another gene. In literature, all the existing work estimated the TTDs from a single microarray time-series data set. Since the sampling time points in one time-series data set are limited, the estimated TTDs may not be so reliable.

In this work, we propose to estimate TTDs from multiple microarray time-series data sets, especially for the cell-cycle-regulated genes. Our idea for estimating TTDs from multiple microarray time-series data sets are as follows. Since the expressions of the cell-cycle-regulated genes usually peak at a certain stage of the cell cycle [2], the TTDs among the cell-cycle-regulated genes are related to the time differences among the cell cycle stages. From Figure 5 in de Lichtenberg *et al*'s paper [1], we know that the time differences among different cell cycle stages are similar in different experiments. Then we postulate that the ratio of the TTD between two cell-cycle-regulated genes to the cell cycle is around a constant value. We named this ratio as the relative TTD and proposed to estimate it from multiple microarray time-series data.

2 Method and Results

The datasets we used in our work are the budding yeast cell cycle microarray data from Spellman et al [2]. First, the data sets are pre-processed to fill in the missing values and smooth the data. Second, the possible relative TTDs are estimated with correlations from individual data sets as stated in Equation 1-2.

$$r_{ij}(\tau) = \frac{\sum_{t=1}^{T-\tau}[x_i(t) - \bar{x}_i][x_j(t+\tau) - \bar{x}_j]}{\sqrt{\sum_{t=1}^{T-\tau}[x_i(t) - \bar{x}_i]^2}\sqrt{\sum_{t=1}^{T-\tau}[x_j(t+\tau) - \bar{x}_j]^2}} \quad (1)$$

$$\bar{x}_i = \frac{1}{T-\tau}\sum_{t=1}^{T-\tau} x_i(t), \quad \bar{x}_j = \frac{1}{T-\tau}\sum_{t=1}^{T-\tau} x_j(t+\tau) \quad (2)$$

where $x_i(t)$ is the expression level of gene *i* at time *t*, τ is the time delay, and *T* is the number of samples in the data set. The cell cycle periods are assumed to be 58 minutes for alpha data, 85 minutes for cdc28 data, and 115 for cdc15 data [3]. Then the relative TTDs from different data sets are aligned based on the cell cycle period, and the relative time in a small range with the high average correlations from different data sets are predicted as the relative TTD.

Experiment results show that 85% of the estimated relative TTDs are the same as those predicted from at least two individual data sets. Table 1 shows the results of estimated relative TTDs and correlations in the 8 known transcriptional factors for the estimated correlation greater than 0.8.

Table 1 Estimated relative TTDs and correlations

Regulator	SWI5		FKH1		SWI4	
Target	ACE2		ACE2		SWI5	
	Relative TTD	correlation	Relative TTD	correlation	Relative TTD	correlation
Cdc15	0.52	-0.92	0.17	0.97	0.09	-0.90
Cdc28	0.00	0.93	0.12	0.92	0.59	0.92
Alpha	0.00	0.80	0.24	0.85	0.60	0.85
Combined	0.00	0.86	0.11	0.85	0.60	0.83
Regulator	FKH1		ACE2		SWI6	
Target	SWI5		SWI5		FKH1	
	Relative TTD	correlation	Relative TTD	correlation	Relative TTD	correlation
Cdc15	0.17	0.88	0.00	0.85	0.70	-0.85
Cdc28	0.24	0.80	0.00	0.93	0.59	-0.91
Alpha	0.12	0.84	0.00	0.80	0.72	-0.77
Combined	0.11	0.80	0.00	0.86	0.71	-0.82

3 Discussion

The relative TTDs from multiple data sets can reveal some consensus transcriptional time delays from different experiments, especially for the relatieve TTDs with high correlations among cell-cycle-regulated genes. Although not all the relative transcriptional time delays are the same under different experimental conditions, this research makes a step toward developing a new approach for discovering the regulations among multiple cell-cycle-regulated genes. The estimated relative transcriptional time delays can be used as a component in building gene regulatory networks in the future research.

References

[1] U. de Lichtenberg, L. J. Jensen, A. Fausboll, T. S. Jensen, P. Bork and S. Brunak, *Comparison of computational methods for the identification of cell cycle-regulated genes*, Bioinformatics, 21 (2005), pp. 1164-1171.

[2] P. T. Spellman, G. Sherlock, M. Q. Zhang, V. R. Iyer, K. Anders, M. B. Eisen, P. O. Brown, D. Botstein and B. Futcher, *Comprehensive identification of cell cycle-regulated genes of the yeast saccharomyces cerevisiae by microarray hybridization*, Molecular Biology of the Cell, 9 (1998), pp. 3273-3297.

[3] L. P. Zhao, R. Prentice and L. Breeden, *Statistical modeling of large microarray data sets to identify stimulus-response profiles*, Proceedings of the National Academy of Sciences of the United States of America, 98 (2001), pp. 5631-5636.