

# Lag Analysis of Genetic Networks in the Cell Cycle of Budding Yeast

Mamoru Kato<sup>1</sup>

mkato@ims.u-tokyo.ac.jp

Tatsuhiko Tsunoda<sup>2</sup>

tatsu@ims.u-tokyo.ac.jp

Toshihisa Takagi<sup>1</sup>

takagi@ims.u-tokyo.ac.jp

<sup>1</sup> Human Genome Center, Institute of Medical Science, University of Tokyo, 4-6-1 Shirokane-dai Minato-ku, Tokyo 108-8639, Japan

<sup>2</sup> SNP Research Center, RIKEN, 4-6-1 Shirokane-dai Minato-ku, Tokyo 108-8639, Japan

**Keywords:** DNA chip, cell cycle, genetic networks, time series analysis

## 1 Introduction

The recent emergence of whole-genome expression data requires a novel computational method to efficiently extract biological information from the large-scale data. The typical method for extracting information about genetic networks is cluster analysis and Fourier analysis of gene-expression patterns [2, 3, 4]. These analyses have uncovered some of the genetic networks. However, in order to comprehend all of the genetic networks in a variety of biological processes, it is necessary to develop computational methods from various viewpoints. Here we propose a computational time-series analysis to infer gene-regulatory networks from whole-genome expression data.

## 2 Method

The lag analysis we propose is a method of analyzing a cross-correlation function between mRNA expression levels to screen for the genes whose expression levels have a time-lagged correlation with that of a known transcription factor. The cross-correlation function is defined as:

$$r_{xy}(k) = \frac{\sum_{t=1}^{N-k} (x_t - \bar{x})(y_{k+t} - \bar{y})}{\sqrt{\sum_{t=1}^{N-k} (x_t - \bar{x})^2} \sqrt{\sum_{t=1}^{N-k} (y_{k+t} - \bar{y})^2}}$$

where  $x_t$  is the expression level of a known transcription factor at time  $t$ , and  $\bar{x}$  is the mean of  $x_t$  over  $t$ .  $y_{k+t}$  is the expression level of a gene at time  $k + t$ , and  $\bar{y}$  is the mean of  $y_{k+t}$  over  $t$ .  $N$  is the number of time points and  $k$  is the time delay. We calculate a cross-correlation function and test a no-correlation hypothesis of the cross-correlation function. The statistic for the no-correlation test ( $r = 0$ ) is defined as:

$$t = r \sqrt{\frac{n-2}{1-r^2}},$$

where  $n$  is the number of samples and equals  $N - k$  in the formula of the cross-correlation function. The statistic follows the  $t$ -distribution with  $n - 2$  degree of freedom, and the statistical significance of the statistic equals the value of the cumulative distribution function of the  $t$ -distribution in a one-sided test.

We calculate and compare the statistical significance values at all delays within one cell-cycle, and define the delay with the largest significance value as a “lag”. If a gene has a significance value of more than 0.99 at a lag, we screen for the gene. This means we select the gene for which the no-correlation

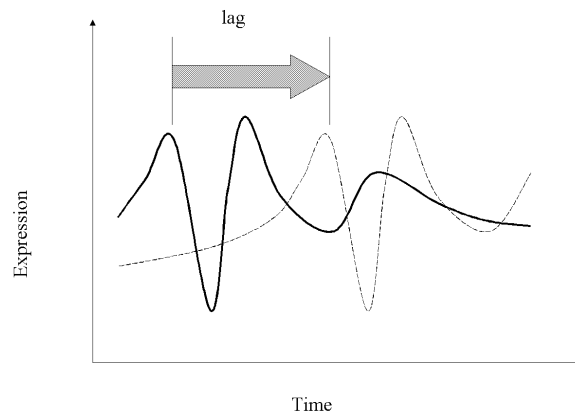


Figure 1: A lag is the time required until a transcription factor begins to influence a gene. The expression level of a transcription factor is depicted as a solid line, and the expression level of a gene is depicted as a broken line.

hypothesis is rejected at a lag at a 0.01 significance level. When naturally interpreted, a lag means the time needed from the time a transcription factor is expressed until its acting on a gene (Figure 1).

Moreover, we undertake upstream sequence analysis to screen the genes selected by lag analysis for those that are directly controlled by a known transcription factor. We examine the number of already known consensus regulatory elements in a 600-base upstream region of the genes screened by lag analysis. For the reliability of analysis, we screen for the genes that have two or more consensus regulatory elements in the upstream region.

### 3 Results

We applied our analysis to the data gathered by Cho et al. [1], who measured the expression levels of 6220 mRNA species of budding yeast. These data provided us with 17 time points taken at 10-minute intervals. We make our results available at [5]. The results include many genes that are known targets of MBF (Swi6 + Mbp1), SBF (Swi6 + Swi4), and Swi5, and include those that are unknown targets of them in previous studies. Interestingly, the results show a “central lag”, an “exceptional lag”, and “different lags.” A central lag is the time at which a lot of genes are influenced concentrically. Some genes are influenced at an exceptional lag, which is the time far from a central lag. Different lags indicate two transcription factors (Swi6 and Swi4) have quite different times in influence on a gene at mRNA level, although they form into a protein complex (SBF), which is a final transcriptional regulator.

### References

- [1] Cho, R.J., Campbell, M.J., Winzeler, E.A., Steinmetz, L., Conway, A., Wodicka, L., Wolfsberg, T.G., Gabrielian, A.E., Landsman, D., Lockhart, D.J., and Davis, R.W., A genome-wide transcriptional analysis of the mitotic cell cycle, *Mol. Cell*, 2:65–73, 1998.
- [2] Eisen, M.B., Spellman, P.T., Brown, P.O., and Botstein, D., Cluster analysis and display of genome-wide expression patterns, *Proc. Natl. Acad. Sci. USA*, 95:14863–14868, 1998.
- [3] Spellman, P.T., Sherlock, G., Zhang, M.Q., Iyer, V.R., Anders, K., Eisen, M.B., Brown, P.O., Botstein, D., and Futcher, B., Comprehensive identification of cell cycle-regulated genes of the yeast *Saccharomyces cerevisiae* by microarray hybridization, *Mol. Biol. Cell*, 9:3273–3297, 1998.
- [4] Tavazoie, S., Hughes, J.D., Campbell, M.J., Cho, R.J., and Church, G.M., Systematic determination of genetic network architecture, *Nat. Genet.*, 22:281–285, 1999.
- [5] [http://www.hgc.ims.u-tokyo.ac.jp/mkato/Research/Ycell\\_cycle/](http://www.hgc.ims.u-tokyo.ac.jp/mkato/Research/Ycell_cycle/)