

# Identifying Genetic Network Using Experimental Time Series Data by Boolean Algorithm

**Kazumi Hakamada**<sup>1</sup>

h011416m@mbbox.media.nagoya-u.ac.jp

**Taizo Hanai**<sup>1,2</sup>

taizo@brs.kyushu-u.ac.jp

**Hiroyuki Honda**<sup>1</sup>

honda@nubio.nagoya-u.ac.jp

**Takeshi Kobayashi**<sup>1</sup>

takeshi@nubio.nagoya-u.ac.jp

<sup>1</sup> Department of Biotechnology, Graduate School of Engineering, Nagoya University, Furo-cho, Chikusa-ku, Nagoya 464-8603, Japan

<sup>2</sup> Present address: Laboratory for Applied Biological Regulation Technology, Graduate School of Bioresource and Bioenvironmental Science, Kyushu University, 6-10-1 Hakozaki, Higashi-ku, Fukuoka 812-8581, Japan

**Keywords:** gene expression analysis, Boolean algorithm

## 1 Introduction

Recently, a variety of experimental techniques for biological field have been developed. These technologies have made it possible to observe the expression of many genes simultaneously and to accumulate a vast amount data. One of the challenging research areas is to extract the genetic networks from these large data. A lot of methods for this problem proposed; quantitative model [2], statistic model [3], hybrid model [7], and Boolean model [3]. However, a lot of these papers analyzed only artificial data [5] and deletion strain data [6]. In this paper, we applied Boolean algorithm for extraction genetic network from experimental time series data. Using binary data that was made from real data, our system was achieved to categorize genes to some equivalence classes and discover genetic interactions.

## 2 Method and Results

### 2.1 Data Processing

In this study, we used expression data from the study of Chu *et al.* [4]. *Saccharomyces cerevisiae* was synchronized by transferred them to sporulation medium (SPM) at  $t = 0$  to maximize the synchrony of sporulation. RNA was harvested at time  $t = 0, 0.5, 2, 5, 7, 9$  and 11.5 hours after transfer to SPM. Polyadenylated RNA was prepared by purification with oligo(dT) cellulose column. Each gene's mRNA expression level just before transfer to SPM was used as control. About 6100 genes of expression profiles are included in their data. Using them, we followed the same method [4] to extract 512 genes that showed significant increase of mRNA levels during sporulation. Twenty one [1] genes of which function had been well known biologically were used for the following analysis.

### 2.2 Extraction of Gene Interactions

Boolean model can only use binary data so experimental data was needed data preprocessing to convert to the binary data. We combine our data preprocessing methods that were used differential value and quantified value of gene expression data with Boolean model. At first Boolean algorithm extracts reflexive transitive closure from binary data, and then it makes some equivalence classes from reflexive transitive closure, at last it searches interaction in the equivalence classes by means of depth first searching algorithm. Using this result, network was drawn. Figures show the flow chart of network extraction from time series data using Boolean algorithm. The following sentences also show the way of extracting genetic network from time series data.

1. Convert time course data (Fig. 1) to differential data and quantified data (Fig. 2)

2. Binary data made from differentiation data using threshold and quantified data (Fig. 3)
3. Categorize binary data into some equivalence class and search equivalence interaction.
4. Drawing interaction into graph (Fig. 4)

### 3 Result and Discussions

The results of extracting genetic networks using 21 genes by the Boolean Network model are given in Figure 4. These genes were categorized in 21 equivalence classes. This result shows each of equivalence classes has only one gene in the class. And this suggests that each gene is under individual regulation.

Figure 4 (a) shows the extracted network using early time stage data. And this network contains some genes that were known to express in the early phase. Figure 4 (b) shows the extracted network using late time stage data. And this network contains some equivalence classes that were expressed in the late phase. These two networks are reasonable from the point of view of biological knowledge. This result suggests that our combination of data preprocessing methods with Boolean algorithm is able to extract genetic network from experimental time series data.

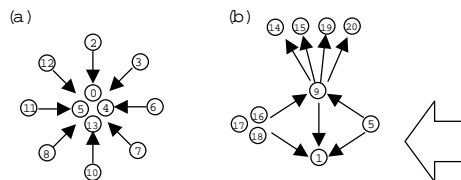
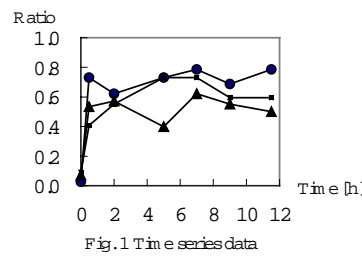


Fig. 4 Gene interactions  
 (a) Extracted network using early time stage data.  
 (b) Extracted network using late time stage data.  
 absent classes have no interactions to any

	0 - 0.5	0.5 - 2	2 - 5
IM E2	1.407	-0.072	0.036
HOP1	0.925	0.026	0.058
ZIP1	0.632	0.096	0.059

	0	0.5	2
IM E2	0.026	0.730	0.621
HOP1	0.071	0.533	0.573
ZIP1	0.092	0.408	0.552

Fig. 2 Data from time series  
 (a) differential data.  
 (b) quantified data.

	IM E2	HOP1	ZIP1
IM E2	0	0	0
HOP1	0	0	0
ZIP1	1	1	0

Fig. 3 Binary data

### References

- [1] Aaron, P.M., *Microbiological Review*, 58:56–70,1994
- [2] Akutsu, T., Miyano, S., and Kuhara, S., Inferring qualitative relations in genetic networks and metabolic pathway, *Bioinformatics*, 16:727–734, 2000
- [3] Akutsu, T., Miyano, S., and Kuhara, S., Algorithm for identifying Boolean network and related biological networks based on matrix multiplication and fingerprint function, *Journal of Computational Biology*, 7:331–343, 2000
- [4] Arkin, A., Shen, P., and Ross, J., A test case of correlation metric construction of a reaction pathway from measurements, *Science*, 277:1275–1279, 1997
- [5] Chu, S., Derisi, J., Eisen, M., Mulholland, J., Botstein, D., Brown, P.O., and Herskowitz, I., The transcriptional program of sporulation in budding yeast, *Science*, 282:699–705, 1998
- [6] Maki, Y., Tominaga, D., Okamoto, M., Watanabe, S., and Eguchi, Y., Development of a system for the inference of large scale genetic networks, *Pacific Symposium on Biocomputing 2001*, 446–458, 2001
- [7] Trey, I., Vestein, T., Jeffrey, A.R., Rowan, C., Jeremy, B., Jimmy, K.E., Roger, C., David, R.G., Ruedi, A., and Leroy, H., Integrated genomic and protein analyses of a systematically perturbed metabolic network, *Science*, 292:929–934, 2001
- [8] Yhu, C.-H., Bolouri, H., and Dacidson, E.H., Genomic cis-regulatory logic: experimental and computational analysis of a sea urchin gene, *Science*, 279:1896–1902, 1998