

Assignment of Direct Binding in Gene Regulatory Systems from Expression Profiles

Sachiyo Aburatani¹

sachiyo@grt.kyushu-u.ac.jp

Satoru Kuhara¹

kuhara@grt.kyushu-u.ac.jp

Hiroyuki Toh²

toh@kuicr.kyoto-u.ac.jp

Katsuhisa Horimoto³

khorimot@ims.u-tokyo.ac.jp

¹ Laboratory of Molecular Gene Technics, Graduate School of Genetic Resources Technology, Kyushu University, Hakozaki 6-10-1, Higashi-ku, Fukuoka 812-8581, Japan.

² Laboratory of Genome Informatics, Bioinformatics Center, Institute of Chemical Research, Kyoto University, Uji, Kyoto 611-0011, Japan

³ Laboratory of Biostatistics, Human Genome Center, Institute of Medical Science, University of Tokyo, 4-6-1 Shirokane-dai, Minato-ku, Tokyo 108-8639, Japan

Keywords: direct DNA binding, expression profiles, graphical Gaussian modeling, mutant

1 Introduction

Expression analysis with DNA microarrays allows monitoring changes in mRNA levels in living cells. The monitoring of changes in a genomic scale reveals new findings of gene regulatory systems. However, there are some limits of microarray monitoring. For example, microarray monitoring has no ability to discriminate direct from indirect effects of the genes that are controlled by specific regulatory factors.

In this work, we present an approach to assign direct binding effects in gene regulatory systems by combination with DNA microarray analyses in mutant cells and a computational method that we recently developed [1, 3, 4, 5].

2 Method and Results

2.1 Expression Data and Computational Method

The expression profiles of 521 genes that are directly and indirectly controlled by 94 transcription factors (TFs) were monitored in 107 mutant cells in our laboratory. The relationships between genes are assigned by ASIAN [1, 3, 4, 5] that is a system for inferring a network by application of graphical Gaussian modeling (GGM) [2], which is a statistical method to discriminate conditionally independent variables among the correlated variables. The ASIAN, at first, automatically classify the genes into some clusters, and then infer the relationship between the clusters.

2.2 Classification and Network Inference

The results obtained by the analysis of the present data by ASIAN are as follows:

- i) The 521 genes are classified into 37 clusters whose numbers of members are ranged from 1 to 81.
- ii) In a network between 37 clusters, 294 (44.1%) of 666 connections are broken off, while 373 (55.9%) are established. The numbers of connections between the clusters are ranged from 4 to 27, and the average is 20.1.

2.3 Correspondence of Inferred Network with Known Regulatory Systems

With reference to the knowledge in *Saccharomyces cerevisiae* Proteome Database, we found 893 gene pairs with the relationship between a transcriptional factor and the regulated gene in our data. Then, we survey whether a transcriptional factor and the corresponding regulated gene(s) belong to the connected or the same cluster in the network.

Overall, the inferred network corresponded well with the known gene regulatory relationship. Among them, 590 of 893 gene pairs with the regulatory relationship were found in the connected or the same clusters. The present correctness for the inference of gene pairs in the known regulatory systems (66.1%) is clearly larger than the fraction of the established connections to all connections (55.9%). Indeed, the correctness is verified from a statistical viewpoint. In a simulation, the chance probability of the present correctness of the inference was estimated to be *ca.* 0.042, indicating that present inference is well significant.

2.4 Assignment of Direct Binding

Among 94TFs, 9 TFs whose sequence motifs for DNA binding are known were mutated, and their 208 regulated genes were monitored in the present data. Then, the 208 genes are divided into four classes: the genes with the binding motifs and with no motifs in the connected clusters and those in the non-connected clusters. In Table 1, the correctness is 74.1% with the probability of *ca.* 0.002 by the simulation. Furthermore, the ratio of the genes with the binding motifs to those with no motifs in the connected clusters (8.79) is statistically larger than the ratio of the genes with the motifs to those with no motifs (2.09), by χ^2 test ($\chi^2 = 15.725$ and $P < 0.0001$). These results indicate that present connection in inferred network reveals the gene regulatory relationship by the direct binding.

Table 1: Divisions of regulated genes in terms of the connection and the existence of binding motif.

	motif	no motif
connected	123	14
not-connected	48	23

3 Discussion

In the statistical sense, GGM discriminates between direct and indirect correlations between the variables. In the present application of GGM to expression profiles, the direct bindings of TF to their regulated genes are assigned, although the present work is performed in a limited set of TFs and their regulated genes. Thus, the present approach serves to reveal an initial step in a whole gene network in a cell. Furthermore, our approach may serve to identify the unknown regulated genes, apart from usual approach in which the combination of the clustering techniques for expression profiles with the sequence analyses of upstream regions in the clustered sequences is adopted.

References

- [1] Aburatani, S., Kuhara, S., Toh, H., and Horimoto, K., Deduction of a gene regulatory relationship framework from gene expression data by the application of graphical Gaussian modeling, *Signal Processing*, in press.
- [2] Edwards, D., *Introduction to Graphical Modelling*, Springer, 2000.
- [3] Horimoto, K. and Toh, H., Statistical estimation of cluster boundaries in gene expression profile data, *Bioinformatics*, 17:1143–1151, 2001.
- [4] Horimoto, K., Aburatani, S., Kuhara, S., and Toh, H., ASIAN — automatic system for inferring a network from gene expression profiles, *Res. Commun. Biochem. Cell Mol. Biol.*, in press.
- [5] Toh, H. and Horimoto, K., Inference of a genetic network by a combined approach of cluster analysis and graphical Gaussian modeling, *Bioinformatics*, 18:287–297, 2002.