Identification of Gene Regulatory Networks by Strategic Gene Disruptions and Gene Overexpressions

Tatsuya Akutsu [1]
takutsu@ims.u-tokyo.ac.jp
Satoru Kuhara [2]
kuhara@grt.kyushu-u.ac.jp
Osamu Maruyama [1]
maruyama@ims.u-tokyo.ac.jp
Satoru Miyano [1]
miyano@ims.u-tokyo.ac.jp

[1] Human Genome Center, Institute of Medical Science, University of Tokyo
4-6-1 Shirokanedai, Minato-ku, Tokyo 108, Japan

[2] Graduate School of Genetic Resources Technology, Kyushu University
Higashi-ku, Hakozaki, Fukuoka 812, Japan

1 Introduction

The whole DNA sequences of many organisms are being determined by international collaborations, and some of them have already been made public at various ftp sites. Now the main topic of researches on such organisms has shifted to the systematic functional analysis of the genes.

Our research group has installed a systematic experimental method which allows both multiple gene disruptions and multiple gene overexpressions. By using this method, we have launched a project whose purpose is to reveal the gene regulatory networks between the 6,200 genes of *Saccharomyces cerevisiae* while many laboratories have also started similar projects. This project consists of two different tasks in addition to the task of experiments. One is to develop algorithms (strategies) for constructing gene regulatory networks from experimental data and testing the consistency of data with already known gene regulatory networks. The second is to construct a knowledge base of gene regulatory networks which have been already known for various organisms. By combining the results of these two tasks, our project will determine a rough figure of the gene regulatory network.

For the first task, we have developed several strategies. In this short abstract, we briefly describe the theoretical results implied by these strategies. Although the developed strategies are not yet practical, we believe that they give a foundation for developing practical strategies. Readers interested in details may refer Ref. [1].

2 Results

Fig. 1 shows an example of a gene regulatory network with 16 genes. Genes `A`, `C`, `I`, `K`, `N`, `X1`, `X2` express under no condition. Arrows with $\oplus$ and $\ominus$ mean *activation* and *inactivation*, respectively. Genes `B`, `E`, `H`, `J`, `M` express if their direct predecessors do not express. For gene `D`, it expresses if its all predecessors `C`, `F`, `X1`, `X2` express. The same holds for genes `L` and `G`. Gene `F` is *activated* by gene `A` and is also *inactivated* by gene `L`. `F` expresses if `A` expresses and `L` does not express. Three cases of gene expressions (normal, disruption of `A`, overexpression of gene `B`) are shown in Table 1, where `1` (`0`) means the gene expresses (does not express).

However, if the disruption of gene `K` yields the activation of `L` which inactivates `F` while gene `A` expresses and activates `F`. Such conflict may occur in the network. In such case the gene expression may be ambiguous or may change from time to time.

We investigate problems and algorithms for identifying gene regulatory networks from such experimental data in various situations. It might be better if analog values and dynamic behaviors were

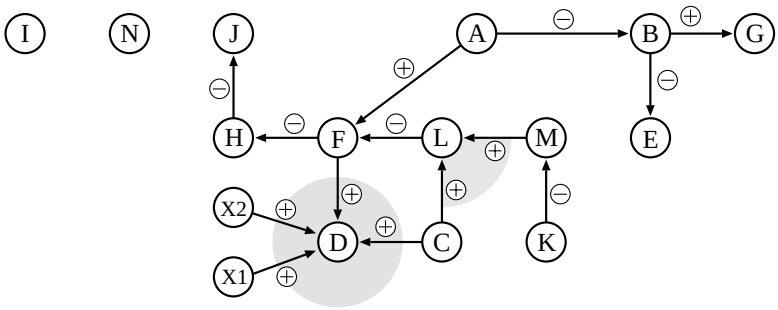

Figure 1: Example of a gene regulatory network.

taken into account since real biological regulatory networks are not discrete. However, we can observe only whether a gene expresses or not by our experiments. By this reason, we define a gene regulatory network as a boolean network where it deals with the cases "gene expresses (1)" and "gene does not express (0)" as in Table 1.

Table 1: Gene expressions by disruption and overexpression.

	Gene Expression															
Gene Name	A	B	C	D	E	F	G	H	I	J	K	L	M	N	X1	X2
Normal Condition	1	0	1	1	1	1	0	0	1	1	1	0	0	1	1	1
Disruption of A	0	1	1	0	0	0	1	1	1	0	1	0	0	1	1	1
Overexpression of B	1	1	1	1	0	1	1	0	1	1	1	0	0	1	1	1

We define the indegree of a gene by the number of genes directly affecting it in a gene regulatory network. The indegree of most genes may be only one or two except some special genes. Therefore, it has an important sense in practice to cope with gene regulatory networks with a small indegree constraint when the genes are restricted to a specific region.

As mentioned above, we consider a gene regulatory network model such that the expression of a gene is determined by a boolean function of the expressions of the genes directly affecting it. We prove upper and lower bounds of experiments required for identifying a gene regulatory network with n genes in regard to the indegree constraint and acyclicity. The results are summarized in Table 2.

Table 2: Summary of theoretical results on identification, where n denotes the number of genes.

	Number of Experiments	
Constraints	Lower Bounds	Upper Bounds
No constraint	$\Omega(2^{(n-1)/2})$	$O(n\,2^{n-1})$
Indegree $\leq D$	$\Omega(n^D)$	$O(n^{2D})$
Indegree $\leq D$ & AND-nodes (OR-nodes) only	$\Omega(n^D)$	$O(n^{D+1})$
Indegree $\leq D$ & Acyclic	$\Omega(n^D)$	$O(n^D)$
Indegree ≤ 2 & no inactivation edges	$\Omega(n^2)$	$O(n^2)$

References

[1] Akutsu, T., Kuhara, S., Maruyama, O. and Miyano, S., "Identification of Gene Regulatory Networks by Strategic Gene Disruptions and Gene Overexpressions," *Proc. Ninth ACM-SIAM Symp. Discrete Algorithms* (SODA'98, in press).