

# A Method for Gene Regulations Inference from Gene Expression Data of Multiple Deletions Mutants

**Kotaro Baba**<sup>1,2</sup>

kotaro@sbi.jp

**Hiroaki Kitano**<sup>1,3,4,5</sup>

kitano@symbio.jst.go.jp

**Koji Kyoda**<sup>3,4</sup>

kyoda@symbio.jst.go.jp

**Shuichi Onami**<sup>1,3</sup>

sonami@sbi.jp

- <sup>1</sup> System Biology Institute, M31 6A, 6-31-15 Jingumae, Shibuya-ku, Tokyo 150-0001, Japan
- <sup>2</sup> National Institute of Agrobiological Sciences, 2-1-2 Kannondai, Tsukuba-shi, Ibaraki 305-8602, Japan
- <sup>3</sup> Graduate School of Science and Technology, Keio University, 3-14-1 Hiyoshi, Kohoku-ku, Yokohama 223-8522, Japan
- <sup>4</sup> Kitano Symbiotic Systems Project, ERATO, JST., M31 6A, 6-31-15 Jingumae, Shibuya-ku, Tokyo 150-0001, Japan
- <sup>5</sup> Sony Computer Science Laboratories, Inc., 3-14-13 Higashi-Gotanda, Shinagawa-ku, Tokyo 141-0022, Japan

**Keywords:** gene network, microarray, gene disruption, functional genomics, algorithm

## 1 Introduction

There are several significant studies to infer a gene regulatory network from gene expression data of large-scale mutant collections. Pe'er *et al.* presented a method using Bayesian network framework, which is applicable to various kinds of gene expression data including multiple deletion (deletion mutant) collection [2]. Their method outputs a gene network made up of “features”, many of which are undirected gene relations instead of directed gene regulations. Directed gene regulation displays regulatory flow in gene network and provides deeper insight into how the network is functioning than undirected gene relation. We presented DBRF method that infers gene network made up of signed directed gene regulations (SDGRs), that is, activations and repressions [1]. The method, however, is applicable only to single deletion collection.

To generate large-scale mutant collections, transposons are commonly used as a mutagen. A transposon often mutates at two or more genes at a time, and laborious efforts are required to obtain single mutants. For example, isolation of single mutation takes 3-5 years in *Oryza sativa* and this makes it difficult to generate large-scale single mutant collection. However, most genome-wide mutant collections have been dedicated to single mutants, partly due to the absence of effective methods to analyze multiple mutants.

Here, we present a method for inferring a gene network composed of SDGRs from a series of gene expression data of multiple deletants.

## 2 Method

Gene expression profiles of multiple deletants are sorted according to the number of deletions. In case that single deletants are included, SDGRs are deduced based on the difference of gene expression level between wild-type and single deletants as is done in the step 1 of DBRF method [1]. The deduced regulations are stored as “fixed regulations”.

Next, SDGRs are deduced from multiple deletants’ profiles by two separate processes. First, all the profiles of multiple deletants in which the same gene is disrupted are collected, and gene expression changes shared among all those multiple deletants, compared to that of wild-type, are sought in order

to deduce SDGRs.  $e_{ij}$  is deduced as activation (repression), unless it is a fixed regulation and if the expression level of  $v_j$  is reduced (elevated) in all deletants lacking  $v_i$  but having  $v_j$ , where  $v_i$  and  $e_{ij}$  represent  $i$ -th gene and regulation from  $v_i$  to  $v_j$ , respectively. The deduced SDGRs are stored as fixed regulations.

Second process is quite similar to the previous one, except that the possible competitive influences of co-existing deletions are intended to be eliminated. Namely,  $e_{ij}$  is deduced when the expression level of  $v_j$  changes in the same way in all deletants that lack  $v_i$  but have intact  $v_j$  and all genes that regulate  $v_j$  by any of fixed regulations. All possible SDGRs are tested in this way except the fixed regulations, and then, the deduced regulations are stored as fixed regulations. This process is repeated until no SDGR is deduced in a cycle. The set of all deduced regulations are called as the set of total inferred regulations.

Table 1: The efficiencies of inferences. Each measurement is an average of 100 simulated target networks, with standard error given in parentheses.

Num. of Genes	Num. of Deletions	Num. of Profiles	Sensitivity	Specificity
20	1	20	52.4%(20.1%)	94.5%(7.35%)
20	2	190	60.3%(17.4%)	94.5%(7.23%)
20	2	137	58.0%(17.2%)	93.3%(8.09%)
20	2	95	55.8%(19.7%)	94.3%(8.46%)
20	2	43	50.9%(19.3%)	83.6%(10.8%)

### 3 Results and Discussion

Artificial gene expression data generated as described in [1] was used to evaluate the performance of this method. Expression data of double deletants was applied to this method, obtaining total inferred regulations. Total inferred regulations contain indirect SDGRs, whose effects can be produced by combination of other two SDGRs. Such indirect SDGRs were removed from total inferred regulations so as to extract essential regulations, as is done in the step 2 of DBRF method [1]. Sensitivity and specificity of the obtained gene network [1] are summarized in Table 1. DBRF method was used to infer gene networks from the single deletants' data. The inference from all combinations of double deletant profiles (190 profiles) exhibits similar or even higher efficiency compared with the inference from all single deletant profiles. This method was also applied to randomly chosen three quarters (137 profiles), half (95 profiles) and a quarter (43 profiles) of all possible combinations of double deletant profiles. The performances became lower as the number of profiles decreased, though, this method exhibited considerable performance even with a quarter of all possible profiles.

This method is the first method for inferring SDGRs from multiple deletant profiles, and it would provide a possible way to analyze multiple deletants without isolation of each single deletant. Since this method exhibited considerable performance even with a quarter of all possible profiles, profiles of far less complete set of multiple deletants would be sufficient to provide considerable inference. A trade-off between the experimental cost of isolating single deletants from multiple deletants, and that of generating and profiling different combination of multiple deletants should be a subject to be discussed.

### References

- [1] Kyoda, K., Morohashi, M., Onami, S., and Kitano, H., A gene network inference method from continuous-value gene expression data of wild-type and mutants, *Genome Informatics*, 11:196–204, 2000.
- [2] Pe'er, D., Regev, A., Elidan, G., and Friedman, N., Inferring subnetworks from perturbed expression profiles, *Bioinformatics*, 17:S215–S224, 2001.