The DBRF Method for Inferring a Gene Network from Large-scale Steady-state Gene Expression Data

Koji M. Kyoda [1,2] Shuichi Onami [1,3,5] Hiroaki Kitano [1,3,4,5]

[1] Kitano Symbiotic Systems Project, ERATO, JST. M31 6A, 6-31-15 Jingumae, Shibuya-ku, Tokyo 150-0001, Japan

[2] Dept. of Fundamental Science and Technology, Keio University. 3-14-1 Hiyoshi, Kohoku-ku, Yokohama 223-8522, Japan

[3] Dept. of Control and Dynamical Systems 107-81, California Institute of Technology, 1200 East California Blvd. Pasadena, CA 91125, USA

[4] Sony Computer Science Laboratories, Inc. 3-14-13 Higashi-Gotanda, Shinagawa-ku, Tokyo 141-0022, Japan.

[5] The Systems Biology Institute M31 6A 6-31-15 Jingumae Shibuya-ku, Tokyo 150-0001 Japan

{kyoda,sonami,kitano}@symbio.jst.go.jp

ABSTRACT

Complete genome sequence has enabled whole-genome expression profiling and genome deletion projects, which are generating large-scale gene expression profiles corresponding to hundreds of deletion mutants. To extract novel biological information from those profiles is an important challenge in current biology.

We reported the Difference-Based Regulation Finding (DBRF) method, which infers the underlying gene network from those profiles [1]. The method 1) infers direct and indirect gene regulations by interpreting the difference of gene expression level between wild-type and mutant, and 2) eliminates the indirect regulations. A major characteristic of the method is its applicability to continuous-value expression data. Most of the other reported methods assume a gene network as a Boolean network, where a gene expression level is represented as a binary value. Binarization of original continuous-value gene expression data may cause the data to lack the information needed to infer regulatory relationships.

The performance of the method was evaluated using artificial gene networks by varying the network size, indegree of each gene, and the data characteristics (continuous- or binary- value). More than 70% of inferred gene regulatory relationships were consistent with the original gene network. The result showed that the performance of the DBRF method is superior to that of the predictor method [2].

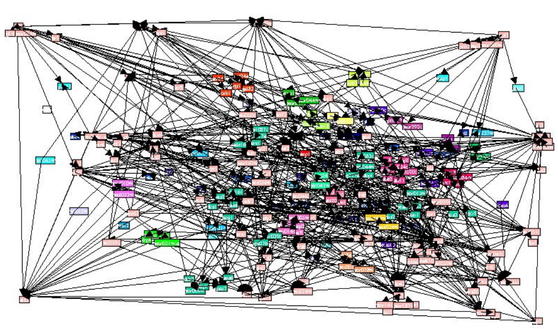

Figure 1: The inferred gene network from Yeast 249 gene expression data with the DRBF method.

We applied the DBRF method to yeast gene expression data [3]. Figure 1 shows the gene network inferred from yeast gene expression data with the method. Six hundred twenty-four gene regulatory relationships among 249 genes were inferred. Pentium II 450MHz Processor took 5 seconds for inferring this gene network, showing that the DBRF method has scalability for gene network size. In MAP kinase signal transduction cascade, the method inferred 9 consistent, 0 inconsistent, and 4 new gene regulatory relationships, compared with the known gene network. The result suggests that the DBRF method effectively infers a gene network from microarray gene expression data. The DBRF method will be a powerful tool for genome-wide gene network analysis.

REFERENCES

[1] Koji, K.M., Morohashi, M., Onami, S., and Kitano, H. A gene network inference method from continuous-value gene expression data of wild-type and mutants. Proceedings of the Eleventh Genome Informatics 2000 (GIW2000) Tokyo: Universal Academic Press Inc. pp 196-204, 2000.

[2] Ideker, T.E., Thorsson, V., and Karp, R.M. Discovery of regulatory interactions through perturbation: inference and experimental design. Proceedings of Pacific Symposium on Biocomputing 2000 (PSB2000) World Scientific. pp 305-316, 2000.

[3] Hughes, T.R., Marton, M.J., Jones, A.R., Roberts, C.J., Stoughton, R., Armour, C.D., Bennett, H.A., Coffey, E., Dai, H., He, Y.D., Kidd, M.J., King, A.M., Meyer, M.R., Slade, D., Lum, P.Y., Stepaniants, S.B., Shoemaker, D.D., Gachotte, D., Chakraburtty, K., Simon, J., Bard, M., and Friend, S.H. Functional discovery via a compendium of expression profiles. Cell 102:109-126, 2000