Approaching Causality: Discovering Time-Lag Correlations in Genetic Expression Data with Static and Dynamic Relevance Networks

Ben Y. Reis[1,2] ben.reis@alum.mit.edu
Atul J. Butte[1,3] butte_a@ai.tch.harvard.edu
Isaac S. Kohane[1,3] isaac_kohane@harvard.edu

[1] Children's Hospital Informatics Program, Harvard Medical School, 350 Longwood Avenue, Boston, MA 02115, USA
[2] Harvard-MIT Division of Health Sciences and Technology, 77 Massachusetts Avenue, Cambridge, MA 02139, USA
[3] Department of Endocrinology, The Children's Hospital, 300 Longwood Avenue, Boston, MA 02115, USA

Keywords: genetic expression, dynamics, causality, phase shift

1 Introduction

Recent advances in micro-array technology have allowed gene expression measurements to be made on a whole-genome scale. Previous research has focused on identifying related genes by studying related simultaneous patterns of gene expression [2, 4]. Other research has focused on studying the simultaneous dynamics, or rate of change, of gene expression [5]. In this study, we focus on identifying genes based on non-simultaneous correlated behavior patterns of expression.

2 Methods and Results

Expression data of yeast, measured by Eisen, et al. [3] at 79 data points under different conditions are analyzed. Both static and dynamic Relevance Networks [1] are constructed from the data, taken with time lags of one and two data points. The issues of appropriate time scale of measurements and time lag of analysis complicate the analysis and serve to increase the already large search-space. It is hypothesized that this methodology can identify causal relationships in which the expression behavior of one gene leads to a delayed pattern of expression of another.

References

[1] Butte, A.J. and Kohane, I.S., Unsupervised knowledge discovery in databases using relevance networks, *Proc. AMIA Symp.*, 1999.

[2] Butte, A.J. and Kohane, I.S., Mutual information relevance networks: Functional genomic clustering using pair-wise entropy measurements, *Proc. Pacific Symp. Biocomputing*, 2000.

[3] Eisen, M.B. *et al.*, Cluster analysis and display of genome-wide expression patterns, *Proc. Natl. Acad. Sci. USA*, 95(25):14863-14868, 1998.

[4] Michaels, G.S. *et al.*, Cluster analysis and data visualization of large-scale gene expression data, *Proc. Pacific Symp. Biocomputing*, 1998.

[5] Reis, B.Y., Butte, A.J. and Kohane, I.S., Dynamics-based functional genomics clustering using mutual information relevance networks, *Proc. Pacific Symp. Biocomputing*, 2000.