From Bayesian Networks to Gene Networks: understanding expression data

Nir Friedman Iftach Nachman Dana Pe'er
Hebrew University of Jerusalem
ISRAEL
{ nir,iftach,danab }@cs.huji.ac.il
http://www.cs.huji.ac.il/labs/pmai2/expression/

DNA hybridization arrays can simultaneously measure the expression level of thousands of genes, giving a snapshot of the cell at varying time points and conditions. A major challenge for bioinformatics is to develop methods that extract new biological understanding from this wealth of data. Our ongoing research is to develop a framework that uncovers relationships between genes based on expression measurements.

We base this framework on the well developed statistical tool of *Bayesian networks* A Bayesian network [2] is a graphical model of joint probability distributions of many variables. This representation includes a directed acyclic graph that represents (in)dependencies among the variables. In recent years there has been significant progress on learning Bayesian networks (structure & parameters) from data [1]. We examine Bayesian networks as a model for gene interactions, mapping expression levels to Bayesian network variables. Some advantages of this learning model are:

- Ability to capture causal relations from observations.
- Flexibility in determining local distribution models.
- Efficient handling of noisy data.
- Effective inference for experiment planning.

To apply this framework on real expression data, we developed new methods for efficient learning of networks from a massive number of attributes [3] and for assessing the confidence in the resulting relations [4]. We applied our methods to 800 yeast cell-cycle regulated genes from [5]. Most of the relations found with high confidence are biologically plausible. An interactive graphic display allows the biologist to view relations at varying levels of confidence.

References

[1] D. Heckerman. A tutorial on learning with Bayesian networks. In *Learning in Graphical Models*, 1998.

[2] J. Pearl. *Probabilistic Reasoning in Intelligent Systems.* Morgan Kaufmann, San Francisco, Calif., 1988.

[3] N. Friedman, I. Nachman and D. Peer. Learning of Bayesian Network Structure for massive datasets: The "sparse candidate algorithm". In Uncertainty in Artificial Intelligence, 1999.

[4] N. Friedman, M. Goldszmidt and A. Wyner. Data anaylisis with Bayesian Networks: A bootstrap approach. IN Uncertainty in Artificial Intelligence, 1999.

[5] T. Spellman., G. Sherlock, M. Zhang, V. Iyer, K. Anders, M. Eisen, P. Brown, D. Botstein and B. Futcher. Comprehensive identification of cell-cycle regulated genes of the yeast Sacch. Cervisiae by microarray hybridization. *Mol. Bio. of the Cell* December 1998.