

Oral presentation

Open Access

Inferring gene regulatory networks from classified microarray data: Initial results

Stuart Aitken*, Thanyaluk Jirapech-Umpai and Ronan Daly

Address: Department of Informatics, Edinburgh University, Edinburgh, UK.

Email: Stuart Aitken* - stuart@inf.ed.ac.uk

* Corresponding author

from BioSysBio: Bioinformatics and Systems Biology Conference
Edinburgh, UK, 14–15 July 2005

Published: 21 September 2005

BMC Bioinformatics 2005, **6**(Suppl 3):S4

Using a method of selecting genes on the basis of their utility for classification [2], we apply optimal gene network inference to the 24 most highly-ranked genes in a leukemia data set [1]. In order to have confidence in the resulting Bayesian gene networks, we first validate the network inference methodology on synthetic data and establish that the methodology has very high specificity, i.e. if an edge is inferred then it is highly likely to be correct. However, we are unable to confidently predict directed edges in the network.

Microarray data analysis poses a number of challenges arising from the high dimensionality of the data, the small number of samples, and sample noise. Consequently, significant methodological questions arise. Statistical techniques can identify correlations between the expression levels of genes, while evolutionary computational techniques can be used to learn classifiers that accurately distinguish categories such as AML and ALL (tumour types) in leukaemia data. The genes of most use in classifying samples can be identified in this way, but the relationships between them are not uncovered. To find these relationships, we apply Bayesian network inference.

The network inference methodology we present is based on the optimal network search algorithm proposed by Ott [3] which is applied in a resampling framework. ROC analysis of networks recovered from synthetic data provides a measure of the performance of this approach. Having selected a small number of genes from the 7070 assayed in the microarray experiment, we are able to perform network inference having solved the feature selection problem. The class labels inform our analysis of the resulting networks. We show that distinct sub-networks

associated with AML and with T-cell responses emerge. Evaluation of the biological plausibility of the results is on-going.

References

1. Golub TR, Slonim DK, Tamayo P, Huard C, Gaasenbeek M, Mesirov JP, Coller H, Loh ML, Downing JR, Caligiuri MA, Bloomfield CD, Lander ES: **Molecular Classification of Cancer: Class Discovery and Class Prediction by Gene Expression Monitoring.** *Science* 1999, **286**:531-537.
2. Jirapech-Umpai S, Aitken S: **Feature selection and classification for microarray data analysis: Evolutionary methods for identifying predictive genes.** *BMC Bioinformatics* 2005, **6**:148.
3. Ott S, Imoto S, Miyano S: **Finding Optimal Models for Small Gene Networks.** *Pacific Symposium on Biocomputing* 2004, **9**:557-567.