Discovery of Gene-Regulation Pathways in Mouse Asbestos Using Background Knowledge

Changwon Yoo[1], Mark Pershouse[2], Elizabeth Putnam[2]

Keywords: Bayesian networks;systems biology; causal discovery;asbestos;gene networks

1 Introduction.

A gene expression study using DNA microarrays usually involves two major steps. The first step typically consists of performing initial experiments to narrow the set of genes to study further in more detail. The experimenter can avoid this first step if he or she already knows the set of genes of interest. For example, the genes involved in galactose metabolism in yeast are relatively well known, so an experimenter could skip the experiments in the first step in the study. After choosing those genes, the experimenter has to produce an experimental design for further study of how those genes regulate each other.

Asbestos fibers small enough to be inhaled, and numerous enough to overcome the normal host defenses can lodge in the lungs, leading to chronic inflammation, pulmonary fibrosis (asbestosis), pleural thickening as well as cancers of the lung and pleura. Asbestos fibers are clearly associated with induction of mesothelioma, a neoplasm derived from the mesothelial lining of the pleural cavity, but the mechanism is unclear [1]. We describe our initial effort in a gene expression study that is designed to learn causal relationships among genes that play an important role in mouse asbestos. In this paper, we concentrate on assessing expert biologist's knowledge of pairwise relationships.

2 Analysis.

Causal networks represent causal relationships using a graphical model. Graphical models hold great promise as representations of molecular biological processes, because they are both expressive and intuitive. In a recent issue of *Science*, the authors of four separate review articles on bioinformatics and related topics described graphical models as one of the most promising methods for representing cellular pathways [2-5]. Three of the articles specifically mention causal Bayesian networks as a promising type of graphical model. Particularly, Kitano [5] refers to one of our causal analysis papers [6] and emphasize the importance of causal discovery in systems biology research. We use Bayesian networks to model interactions of genes in mouse asbestos.

A causal Bayesian network is a directed acyclic graph in which each arc is interpreted as a direct causal influence between a parent node and a child node, relative to the other nodes in the network [7]. One of the challenges in applying Bayesian methods for causal discovery is the assessment of informative priors on possible causal structures and on the parameters of those structures. On the one hand, the ability to represent such prior information is a great strength of the Bayesian approach. With it, we can potentially express prior causal knowledge that comes from many sources other than the observational data. While good progress has been made in facilitating the expression of priors on Bayesian network structures and parameters [8], assessing such prior probabilities (particularly when

[1] Dept. of Computer Science, Univ. of Montana, Missoula, MT. E-mail: cwyoo@cs.umt.edu
[2] School of Pharmacy, Univ. of Montana, Missoula, MT. E-mail: {markp,lizp}@selway.umt.edu

there is a large set of variables) can still be difficult. Note that it will be impractical to assess priors of all of the possible pairwise relationships especially if you are dealing with more than 5,000 genes. We asked expert biologists to assess priors of pairwise relationships that they think are relatively well known, and we assume default prior probabilities for the remaining relationships. We currently use the following assumptions (unless the expert biologist specifies his or her knowledge): $p(E_i^{XY}) = 1/3$ for i=1,2,3. For genes X and Y, E_i^{XY} represents $X \leftarrow Y$, $X \rightarrow Y$, and $X \; Y$ for i=1,2,3 respectively. We list relatively well known pairwise relationships assessed by expert biologists in Table 1.

Relationship	Prior	Relationship	Prior	Relationship	Prior	Relationship	Prior
SRA→TNFα	0.9	*SRA → IL6*	0.6	*SRA → IL1*	0.6	*IFNγ →TNFα*	0.5
IL1 → IL6	0.8	*IFNγ → IL6*	0.5	*IFNγ → IL1*	0.5		

Table 1: Priors assessed by expert biologists about pairwise relationships among genes that play an important role in mouse asbestos.

3 Future Research.

We will combine the background knowledge shown in Table 1 with microarray experiments with 108 mice that were exposed in different asbestos agents [9]. We are planning to use a pairwise causal algorithm, Implicit Latent Variable Scoring (ILVS) method [6] and its extension Local ILVS Method (LIM) that analyzes local structures with more than pairwise variables [10]. There are many challenges in the planned analysis, e.g., (1) causal discovery with no direct manipulation of genes; (2) global network inference with local networks.

Once we develop a model that combines expert's background knowledge and the results of microarray experiments, we are planning to develop and evaluate a system that recommends experimental design of a gene expression study, e.g., what gene to knock out; how many experimental repetitions to make.

References.

[1] Fung, H., et al., *Patterns of 8-hydroxydeoxyguanosine (8)HdG) formation in DNA and indications of oxidative stress in rat and human pleural mesothelial cells after exposure to crocidolite asbestos.* Carcinogenesis, 1997. **18**: p. 101-108.

[2] Karp, P.D., *Pathway databases: A case study in computational symbolic theories.* Science, 2001. **293**: p. 2040-2044.

[3] Gifford, D.K., *Blazing pathways through genetic mountains.* Science, 2001. **293**(2049-2051).

[4] Mjolsness, E. and D. DeCoste, *Machine learning for Science: State of art and future prospects.* Science, 2001. **293**: p. 2051-2055.

[5] Kitano, H., *Systems Biology: A Brief Overview, , March 1, 2002.* Science, 2002. **295**: p. 1662-1664.

[6] Yoo, C., V. Thorsson, and G.F. Cooper. *Discovery of a gene-regulation pathway from a mixture of experimental and observational DNA microarray data*. in *Pacific Symposium on Biocomputing*. 2002. Maui, Hawaii: World Scientific.

[7] Pearl, J., *Probabilistic Reasoning in Intelligent Systems*. Representation and Reasoning, ed. R.J. Brachman. 1988, San Mateo, CA: Morgan Kaufmann.

[8] Heckerman, D., D. Geiger, and D. Chickering, *Learning Bayesian networks: The combination of knowledge and statistical data.* Machine Learning, 1995. **20**: p. 197-243.

[9] Driscol, K., et al., *Intratracheal instillation as an exposure technique for the evaluation of respiratory tract toxicity: uses and limitations.* Toxicol. Sci., 2000. **55**: p. 24-35.

[10] Yoo, C. and G. Cooper. *Discovery of gene-regulation pathways using local causal search*. in *AMIA*. 2002. San Antonio, Texas.