Pacific Symposium on Biocomputing 5:413-414 (2000)

Identification of Coordinated Gene Expression and Regulatory Sequences

G.D. Stormo
Dept of Genetics, Washington University
St. Louis, MO 63110, USA
(stormo@genetics.wustl.edu)

In the last few years genome sequencing projects have provided us with enormous amounts of DNA sequence, and the rate of new sequence accumulation continues to grow rapidly. Other sources of sequence data, such as EST projects, have further increased the information about what each genome contains. Computational tools have been essential for extracting important biological information from those sequences. Together these approaches have allowed us to establish a fairly extensive "parts list" for many species. Among the remaining challenges are understanding how the various parts interaccombine to form complexes, regulate each others synthesis and activity, and in general interact to perform all of the functions necessary for the organism. Methods developed in the last few years can help to decipher these interactions, and undoubtedly new methods will emerge that will provide even greater details. Computational approaches to organize and analyze those data are already playing essential roles, but there are many open problems to be addressed in the coming years.

The focus of this session is on the analysis of mRNA expression data. The data itself may be obtained in various ways, including ESTs and expression array methods. These techniques allow one to capture information about the level of gene expression for all of the genes simultaneously, and to monitor how the expression changes in a variety of experiments. For example, one can determine how expression changes over time from some fixed point of reference, how it changes in response to environmental stimuli, and how it differs depending on the genetic background of the cells being studied. The raw data from such experiments, which consists of the expression levels for each gene, can be enormous and also undecipherable without computational tools to extract and highlight the most significant results. Papers in this session fall into two categories. The first are those to identify sets of genes with related expression patterns, which can be used to infer functional relationships. Theoretical aspects of the problem are described and examples are presented from several different systems, including yeast, Xenopus and human genes. The second category can be considered “downstream” of the first, where sets of coordinated genes have been defined and the problem is to identify sequence patterns related to the clustering. Algorithms are presented for defining the significant patterns, and examples are given using yeast expression data. Yeast is a fairly simple system for this type of analysis, and one of the challenges ahead is to develop appraoches capable of similar results on much more complicated regulatory systems.

On behalf of the organizers I would to thank all of the authors who submitted papers. We regret that there were more good ones than we could include in the presentations. We would also like to thank the reviewers who provided many useful comments on the papers and helped us to pick the ones that seemed most relevant for the session.

Note added in remembrance:

As these proceedings were going to press I was informed that Dan Prestridge has died from the esophageal cancer that he was diagnosed with earlier this year. Dan was one of the organizers of this session and, in fact, helped bring together the rest of us to organize this program. Dan got his Ph.D. from Duke University in 1991, with much of his thesis research being done with Christian Burks at Los Alamos National Laboratory in New Mexico. He then did a one year post-doc with me before taking the position of Director of the Molecular Biology Computing Center at the University of Minnesota, St. Paul. He moved to Axys Pharmaceuticals in 1998, only to decide to move back into academia again this year. He had just moved to the Mayo Clinic as the Director of the Research Computing Facility when he was diagnosed with the cancer.

Those who didn’t know Dan personally remember him primarily for his work on methods to predict and analyze eukaryotic promoters, especially for human genes. Those of us who knew him also remember his enthusiasm for his work and for teaching others who were interested to learn from him. He was an excellent colleague who was always willing to help anyone who asked. He prodded us to move into new areas of research, and his influence on my group far outlasted his brief stay there. Those of us who knew him the best will miss him the most.