GENE EXPRESSION AND GENETIC NETWORKS

BARBARA BRYANT
Millenium Pharmaceuticals
(barbara_bryant@msmail.mpi.com)

ALEKSANDAR MILOSAVLJEVIC
CuraGen Corporation
(amilosav@curagen.com; http://www.curagen.com)

ROLAND SOMOGYI
National Institutes of Health
(rolands@helix.nih.gov; http://rsb.info.nih.gov/mol-physiol/homepage.html;)

Powerful new molecular biological and computational technologies are beginning to give us access to the sequences of full genomes, the complete expression patterns of their genes, and the computational tools to store and interpret these data. Biologists are now in a position to retrieve genes that exhibit specific expression patterns in model systems. Such information can then be used to discover targets for medical intervention and genes of commercial importance. This session will address the computational methods for processing and interpreting gene expression data, and for functional gene discovery based on such information.

We believe that gene sequence and expression data contain much of the information required to determine the higher ordered structures of an organism and important aspects of its behavior. This knowledge is expected to significantly improve the detection and treatment of complex diseases (e.g. cancer), and eventually enable us to manipulate and engineer small organisms. Underlying these expectations is the general concept of information flow in biological systems, in particular from genotype to phenotype. But how exactly should we conceptualize information flow in biological systems?

The Central Dogma of biomolecular sequence information flow can be viewed as a theorem of coding theory in addition to being a biological phenomenon. A code is a mapping that determines how information in one domain (e.g. biopolymer) is transformed to information in another domain. Moreover, it is also generally agreed that the genetically encoded amino acid sequence of a protein determines its structure and biochemical properties. This relationship can be considered as a "structural code", essentially based on the laws of physics and chemistry. However, due to its complexity, this code is difficult to emulate using present computational modeling tools; we cannot yet expect to model the higher molecular structures of an organism directly from sequence. Furthermore, an organism receives inputs from the environment that cannot be predicted from its genome.

This session will deal with the first step of dynamic biomolecular information flow from genotype to phenotype: the expression of genes. Essentially, the product of an activated gene will interact with a variety of other biomolecules, all primary or secondary gene products, which in turn either directly or indirectly regulate the expression of genes through complex signaling cascades. We refer to this signaling loop as a genetic network (Fig. 1). Through careful study, we hope to identify the connections and codes (rules or functions) within such genetic networks. Through the classification of connectivity and rules, we hope to discover principles which they all have in common. Such principles could significantly accelerate functional inference from large scale gene expression data.

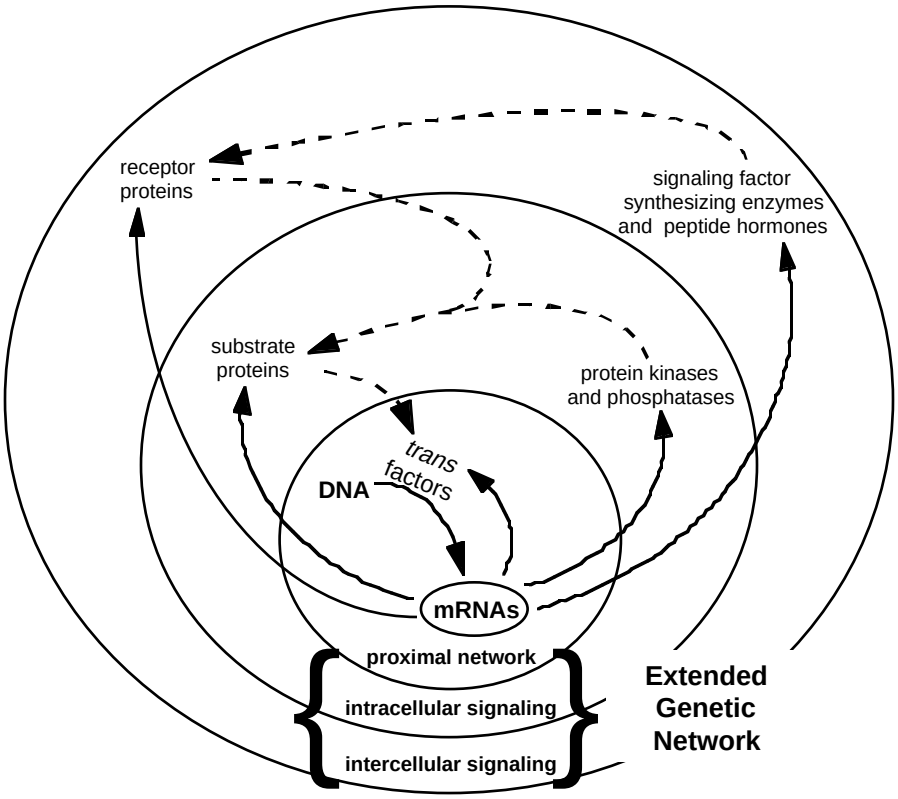

Fig. 1 Information flow in Genetic networks. Genes regulate the expression of genes through a hierarchy of signaling functions. Gene expression patterns represent the variables, while the signaling functions are determined by the gene structure. The solid lines refer to information flow from primary sources (DNA, mRNA). The broken lines correspond to information flow from secondary sources back to the primary source. Excerpted from *Somogyi R. and Sniegoski C.A. (1996) Complexity 1(6):45-63*

We anticipate progress to be made at several, complementary levels:

1. Large scale gene expression data acquisition, cataloging, inference of functional relationships ("top down"), and discovery of possible targets for intervention.
2. Reconstruction and modeling of moderately sized genetic networks in simpler model organisms, incorporating experimental knowledge of molecular interactions ("bottom up").
3. General continuous and discrete network modeling and inference strategies.

In this session we will be exposed to new strategies and recent advances, all posing new questions for the future. But most of all, we hope that discussions will take place that bring together the best of the diverse research strategies and model frameworks. By finding these commonalities, we hope to come closer to understanding the principles that underlie complex, dynamic biomolecular networks. We expect this to find important practical applications in extracting meaning from gene expression data and for optimal experimental design.