

Analysis of sequence and expression data in the context of biological interactions

Jeffrey L. Blanchard
National Center for Genome Resources
email: jlb@ncgr.org
http://www.ncgr.org

Tutor's backgrounds and experience

Jeff is a research scientist and program manager at the National Center for Genome Resources. His group uses informational networks as an interpretive layer for understanding the relationship between traits and the underlying genes, proteins and metabolites. To facilitate their research they have developed PathDB, a cellular interaction database with tools for the visualization and analysis of cellular networks. Jeff has developed and continues to lead summer workshops ("Techniques in Bioinformatics and Comparative Genomics" and "Database Design and Development for Genomics Research") at a non-profit educational center, the BioPharmaceutical Technology Center Institute in Madison, Wisconsin.

Goals for the tutorial

The goals for the tutorial are: (1) Understand the role of biological databases in data annotation and analysis. (2) Understand the ways in which biological interactions and associations can be used to analyze sequence and expression data. (3) Understand how to plug your experimental data into existing systems. (4) Help identify requirements for new algorithms and software tools.

Target audience

Research scientists and software developers with an interest in genomic data. A basic understanding of molecular biology as it applies to genomic research will be assumed.

Abstract

What is the difference between two cells? Intuitively, the answer will be dependent on the genotype, development, location, and current environment of the cells. The answer will also be dependent on integrating knowledge derived from many different chemical and biological experimental data sets. Several new databases and tools have recently been developed to visualize and analyze the connectivity between cellular components. In essence they use informational networks as an interpretive layer between a trait and the underlying genes, proteins and metabolites. In the first hour we will review available data sources and approaches to creating and visualization connections between data.

In the second hour, we will demonstrate how a cellular interaction database, PathDB, can be used to interpret large-scale measurements by using NCGR's integration platform (ISYS) synchronized with other publicly available database and tools. PathDB has been designed to allow the users to analyze their sequence, expression and phenotypic data in the context of known cellular interaction data and pathway models and is integrated with the Gene Ontology (GO) consortium model and tools. Several tools have been implemented that allow the de novo construction, extension, visualization and manipulation of cellular networks. Using information in the PathDB relational database the user can distinguish between known routes and novel (newly) discovered routes and ascertain the statistical likelihood of the routes in the context of a set of expression measurements. By providing a database on cellular processes, basic tools and a GUI interface we allow mathematicians, engineers, and scientists to focus on developing software and algorithms (that can be plugged into the system). ISYS integration also allows synchronization with tools developed at Stanford, Berkeley, Washington University, University of Manchester and other sites as they are added on to the platform. Here is a subset of questions we intend to allow users to address.

- * What is the difference between two cell types (i.e. between a normal and cancer cell)?
- * What if the effect of a knockout mutation on the cellular network.
- * What "classical" pathways are up or down regulated in my expression data.
- * How well does a set of expression data support my cellular network model?
- * How does a drug perturb a cellular network as judged through expression data?

Outline

Introduction
Curated databases
Creating the component inventory
Interaction sources

Using Natural Language Processing techniques to create associations.
Connectivity diagrams
Genotype and Environmental Specific Connectivity diagrams
Overlaying sequence, expression and phenotype information
Graph (network) analysis tools
Using network analysis tools to select drug targets
Use as predictive systems for evaluating gene knockouts and drug response