Towards the virtual organism: Networks and expression

Tutor 1: Jeff Blanchard

since 2000	Program Leader Biochemical Pathways NCGR, New Mexico, USA
1999-2000	Post-Doc Promega Corporation, Madison, Wisconsin, USA
1996-1998	Post-Doc University of Oregon, Eugene, Oregon, USA
1995	PhD in botany, University of Georgia, Athens, GA, USA

Relevant Publications

Waugh ME, Bulmore DL, Farmer AD, Gonzales M, Steadmen PA, Wlodek ST, Mendes P, Kueffner RM and Blanchard JL, "PathDB: A database and tools for modeling metabolic networks.", *Bioinformatics*, submitted

Blanchard JL and Lynch M., "Organelle genes: Why do organellar genes end up in the nucleus?" *Trends in Genetics* (2000) **16***:* 315-320

de Visser AJM, Zeyl CW, Gerrish PJ, Blanchard JL, and Lenski RE, "Diminishing returns from mutation supply rate in asexual populations.", *Science* (1999) **283**: 404-406

Teaching experience (excerpt)

- bioinformatics-related workshops at BioPharmaceutical Technology Center Institute
- introductory lectures on bioinformatics for students at New Mexico Highlands University and at the Singapore National University

Tutor 2: Robert Küffner

since 2000	Research Scientist Biochemical Pathways NCGR, New Mexico, USA
1998-2000	Post-Doc Bioinformatics; group Prof. Lengauer GMD, Germany
1997	PhD in molecular biology, University of Düsseldorf, Germany

Relevant Publications

Zien A, Küffner R, Zimmer R, Lengauer T, "Analysis of gene expression data with pathway scores.", ISMB 2000

Küffner, R, Zimmer, R, Lengauer, T, "Analysis of metabolic databases via Differential Metabolic Display (DMD).", Bioinformatics (2000) 16: 825-836

Prospective audience

Scientists with an interest in biological pathways or in the analysis of large scale measurements in biology (e.g. gene expression, Yeast2Hybrid, metabolic profiling).

Level: Advanced
Length: Half-day

Tutorial's presentation: motivation, goals, content

The availability of expression data is expected to have a significant impact on the understanding of basic cellular processes and the various mechanisms by which cells control the transcription of their genes. Conversely, the interpretation of gene expression measurements requires to utilize existent knowledge about biological networks.

We use our biochemical pathway database PathDB for compiling and exploiting existing knowledge to interpret large-scale measurements. Knowledge thereby generated can be fed back into the process iteratively. Our goal is to allow the modeling, visualization and analysis of complete networks of organisms via their underlying structure of molecular processes including all relevant hierarchies. We try to help the computational scientist to focus on developing algorithms (that can be plugged into our system) by providing a database on cellular processes, basic tools and a GUI interface. We try to enable the molecular biologist to construct and evaluate hypotheses on cellular processes or networks and to assess them.

Here is a subset of questions we hope to allow scientists to address.

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

In the first part of this tutorial we review publicly available pathways resources. We will then guide the participant through one of these, PathDB, and outline how it relates to, complements and utilizes other approaches. We will show how PathDB can be used to select a set of cellular processes and (from this set) to construct and visualize pathways de novo.

In the second part we summarize recently published approaches to analyze expression data or to relate cellular processes to expression data. We will demonstrate the visualization of gene expression data in the context of classic textbook pathways. From there, these pathways can be merged and manipulated to form customized pathways. The validity of constructed pathways can be assessed using statistical scores in the context of gene expression experiments.

The third part aims to demonstrate the value of the combination of gene expression and networks to solve our current research task of elucidating organelle targeted pathways.

Detailed outline

1) **Basics: Databases and Tools**
   - introduce biochemical pathways resources
   - introduce PathDB, a database and tools for storing and manipulating molecular and cellular processes
   - How is the PathDB system complementary to other pathway projects ?
   - describe PathDB data model and content development
   - construct custom, organism specific pathways using PathDBs Query- and Discovery tools
   - visualize constructed pathways in the PathwayViewer

2) **Relating Expression Data and Pathways**
   - summarize state-of-the-art methods for analyzing expression data
   - contrast purely data driven approaches to methods utilizing existing knowledge (e.g. Pathway scores: Zien at al., 2000)
   - demonstrate a connection between PathDB and GeneX, an open source software system for expression data
   - apply the pathway-score metric to evaluate how well pathways are supported by (sets of) expression measurements
   - complement analysis by visualization of gene expression in the context of pathways

3) **Application: our current research work**
   - discover and complement pathways related to the import of proteins into mitochondria and chloroplasts
   - screen literature for known facts about organelle-related pathways
   - utilize techniques for the prediction of the subcellular localization of proteins
   - incorporate this data into PathDB
   - generate hypotheses for the completion of known pathways using PathDB
   - construct and evaluate putative pathways using gene expression data and pathway scores
   - use established techniques like sequence comparison and phylogenic trees for further validation
   - compare resulting pathways with complementary large scale experiments like Yeast2Hybrid