系统生物学

Introduction to Computational Systems Biology

生命科学技术学院
王卓

zhuowang@sjtu.edu.cn
Topics to be discussed in this course

• What is systems biology?
• Methodologies and techniques to understand systems biology
• Structure and function of biological network
• Application of systems biology - precision medicine and synthetic biology

• Skills: Cytoscape; MATLAB; COBRA toolbox
Can a biologist fix a radio?

Lazebnik, Cancer Cell, 2002
Building models from parts lists

- Sequencing
- Gene knock-out
- Microarrays
- etc.

- Genetic interactions
- Protein-Protein interactions
- Protein-DNA interactions
- Subcellular Localization
- Microarrays
- Proteomics
- Metabolomics

Model Generation
Systems biology and emerging properties
Systematic viewpoint

http://www.newvisions.ucsb.edu/background/images/elephant.gif
Outline

- Ch1  Introduction to systems biology
- Ch2  Omics technology
- Ch3  Biological network
- Ch4  Network topology and visualization
- Ch5  Gene regulatory network
- Ch6  Metabolic network and metabolic flux analysis
- Ch7  Integration of regulatory and metabolic network
- Ch8  Systems biology for precision medicine
- Ch9  Systems biology for metabolic engineering
- Ch10  Network comparison and alignment
Lab exercises

- Database and Resource  1
- Network visualization: Cytoscape  2
- Guide to MATLAB:  1
- Metabolic flux simulation: COBRA  2

Case study: E.coli, yeast, rice, human, cancer...
Team project

- Students will form a 5-person-team and each team will select one case-study;
- Implement the case-study using modeling software and write an article-like report (5-10 pages);
- Present your final results at the end of term.
Evaluation

- Attendance  5%
- Lab work    20%
- Team project 25%
- Final exam  50%

Course website:  http://cbb.sjtu.edu.cn/~zwang/sysbio/
林标扬  系统生物学  浙江大学出版社  2012


Klipp E.著. 贺福初等译. 系统生物学的理论、方法和应用. 复旦大学出版社, 2007


邹权, 陈启安, 曾湘祥, 刘向荣 著  系统生物学中的网络分析方法  西安电子科技出版社 2015

[美] D.詹姆森，M.魏玛，H.V.维斯特霍夫 著  系统生物学方法（导读版）  科学出版社 2013
Ch1 Introduction to Systems Biology
The Genomic Era

2001-2-16 Science

2001-2-15 Nature
Genomics, Post-genome & Systems Biology


Genomics

Post-genomic projects

Systems Biology
PubMed abstracts indicate interest in Systems Biology

Human genome completed
One framework for Systems Biology

1. The components. Discover all of the genes in the genome and the subset of genes, proteins, and other small molecules constituting the pathway of interest. If possible, define an initial model of the molecular interactions governing pathway function (how?).

2. Pathway perturbation. Perturb each pathway component through a series of genetic or environmental manipulations. Detect and quantify the corresponding global cellular response to each perturbation.
3. Model Reconciliation. *Integrate* the observed mRNA and protein responses with the current, pathway-specific model and with the global network of protein-protein, protein-DNA, and other known physical interactions.

4. Model verification/expansion. Formulate new *hypotheses* to explain observations not predicted by the model. Design additional perturbation experiments to test these and iteratively repeat steps (2), (3), and (4).
Continuum of modeling approaches

**Top-down**
- **Abstracted**
  - High-level models (L1)
- Statistical mining
- Bayesian networks
- Boolean models
- Components and connections
- Influences and information flow

**Bottom-up**
- **Specified**
  - Low-level models (L2)
- Markov chains
- Mechanisms (Including structure)
- Differential equations

*TRENDS in Biotechnology*
Data integration and statistical mining

Need computational tools able to distill pathways of interest from large molecular interaction databases (top-down)
Types of information to integrate

- **Data that determine the network (nodes and edges)**
  - protein-protein
  - protein-DNA, etc…

- **Data that determine the state of the system**
  - mRNA expression data
  - Protein modifications
  - Protein levels
  - Growth phenotype
  - Dynamics over time
Mapping the phenotypic data to the network

- Systematic phenotyping of 1615 gene knockout strains in yeast
- Evaluation of growth of each strain in the presence of MMS (and other DNA damaging agents)
- Screening against a network of 12,232 protein interactions

Mapping the phenotypic data to the network

Mapping the phenotypic data to the network

Network models can be predictive

Green nodes represent proteins identified as being required for MMS resistance; gray nodes were not tested as part of the 1615 strains used in this study; blue lines represent protein-protein interactions.

The untested gene deletion strains (ylr423c, hda1, and hpr5) were subsequently tested for MMS sensitivity; all were found to be sensitive (bottom).

• In *Escherichia coli*, for instance, there are 225,000 proteins, 15,000 ribosomes, 170,000 tRNA-molecules, 15,000,000 small organic molecules and 25,000,000 ions inside the a few µm cell.

• There are estimated $10^{14}$-$10^{16}$ biochemical reactions in a cell.
A complex problem

- 35,000 genes either on or off (huge simplification!) would have $2^{35,000}$ solutions.

- Things can be simplified by grouping and finding key genes which regulate many other genes and genes which may only interact with one other gene.

- In reality there are lots of subtle interactions and non-binary states.
Computing Biological Complexity

- Protein machine interactions
- Molecular machine classical simulation
- Community metabolic regulatory, signaling simulations
- Constraint-based flexible docking
- Genome-scale protein threading
- Comparative genomics
- Constrained rigid docking
- Cell, pathway, and network simulation
- Molecule-based cell simulation

Current U.S. Computing

*Teraflops

1 TF

10 TF

100 TF

1000 TF
Shanghai Jiao Tong University

Protein-DNA interactions
- ▲ Chromatin IP
- ▼ DNA microarray

Gene levels (up/down)

Protein-protein interactions
- ▲ Protein coIP
- ▼ Mass spectrometry

Protein levels (present/absent)

Biochemical reactions
- ▲ none
- ▼ Metabolic flux measurements

Biochemical levels
Systems Level Perspective

• Reductionism
  – Focus on smallest system components
  – Molecular biology – genes & proteins
  – Primarily an experimental effort
  – Attempt to fully understand all components
  – Hope that systems level understanding will emerge

• Integration
  – Seek to understand system behavior from complex interactions of components
  – Cell – genes, proteins, enzymes, networks, etc.
  – Requires both experimental & modeling work
  – Interactions between components most important
  – Accept limited understanding of components
Two ways of looking a biological problem

Systems biology: Top-down approach

Molecular Biology: Bottom-up approach

Life’s Complexity Pyramid (Oltvai-Barabasi, Science 10/25/02)
Top-down approach

• Biology went top-down for the last 50 years
  – From cell to protein to gene ...
  – Huge amounts of data produced
• Top-down approach tries to make use of high-throughput data using DNA microarray and other new measurement technologies.
Bottom-up approach

• Bottom up is traditional approach
  – You would study a pathway in detail not worrying about how that pathway might interact with other elements in the cell.
  – You would strive to understand a gene or pathway in great detail, eventually you might extend this knowledge to other organisms and compare and contrast.

• This approach is particularly suitable for the end-game scenario where most of the pieces are known and one is trying to find the last few pieces

• Tries to construct a gene regulatory network based on the compilation of independent experimental data
Top-down or bottom-up

• Either look at the whole organism and abstract large portions of it
• Or try to understand each small piece and then after understanding every small piece assemble into the whole
• Both are used, valid and complement each other
Historical Perspective

- **Systems Theory**
  - Cybernetics - Norbert Wiener (1948)
  - Biochemical systems theory (1960s)
  - Metabolic control analysis (1970s)
  - Early work suffered from inadequate data

- **Molecular biology**
  - Isolation of DNA (1869)
  - Double-helix structure of DNA (1953)
  - Transgenic & knockout mice (1980s)
  - Human genome sequence (2000)
  - Further advances require data integration & analysis

- **Systems biology**
  - Represents integration of the systems & molecular approaches
  - Motivated by need to relate genotype - phenotype
  - Enabled by high throughput measurement technologies & advances in computer hardware & algorithms
Challenge

• Put the pieces back together again
• Attempts to create predictive models of cells, organs, biochemical processes and complete organisms
  – Data combined with computational, mathematical and engineering disciplines
  – Model <-> simulations <-> experiment
What is Systems Biology?

The study of the mechanisms underlying complex biological processes as integrated systems of many interacting components. Systems biology involves:

1. collection of large sets of experimental data
2. proposal of mathematical models that might account for at least some significant aspects of this data set
3. accurate computer solution of the mathematical equations to obtain numerical predictions, and
4. assessment of the quality of the model by comparing numerical simulations with the experimental data.

Leroy Hood, 1999
Definitions

• Leroy Hood
  – As global a view as possible
  – Fundamentally quantitative
  – Different scales integrated

• H. Kitano

Aims at systems-level understanding, which requires a set of principles and methodologies that links the behaviors of molecules to systems characteristics and functions.
Defining Systems Biology

• “Integrative approaches in which scientists study and model pathways and networks, with an interplay between experiment and theory.” (Henry, 2003)

• “Systems biology has two distinct branches: Knowledge discovery and data mining, which extract the hidden pattern from huge quantities of experimental data, forming hypothesis as result and simulation-based analysis, providing predictions to be tested by in vitro and in vivo studies.” (Kitano, 2003)

• “There are many different definitions of Systems Biology; like an elephant it is easy to recognize and hard to define.” (Kell, 2004)

• “An interdisciplinary approach for integrating experimental data with mathematical modeling tools to analyze & predict the behavior of biological systems.” (Henson, 2005)
什么是系统生物学

是研究一个生物系统中所有组成成分（基因、mRNA、蛋白质等）的构成，以及在特定条件下这些组分间的相互关系的学科。系统生物学不同于以往的实验生物学——仅关心个别的基因和蛋白质，它要研究所有的基因、所有的蛋白质、组分间的的所有相互关系。

系统生物学是以整体性研究为特征的一种大科学
Systematic viewpoint

http://www.newvisions.ucsb.edu/background/images/elephant.gif
Hypothesis testing and knowledge discovery

B. Biological and physiological knowledge and data

\[\text{Models of gene regulation, biochemical networks, cells and organs (including physiome and virtual human)}\]

\[\text{Computational 'dry' experiments and analysis to screen hypotheses}\]

E. Experimental design to test hypotheses

D. Development of experimental techniques

W. 'Wet' experiments to verify or reject hypotheses
Developing models helps us improve our understanding on biological networks.

Observe behaviours of the model and compare to experimental data from real networks.

Make predictions about real metabolic networks based on the properties of the model.

Test the predictions by checking our understanding, as comparison of a model with reality is the way to prove whether the model works.
1. System structure identification:
   ✓ genes, proteins, small molecules, cells involved

2. System behavior
   ✓ dynamics, stability

3. System control
   ✓ to change system variables in order to control other system variables

4. System design
   ✓ to design and modify systems for desired properties
The year 2000 was significant:

• Completion of the Human Genome Project

• Occurrence of the First International Conference on Systems Biology in Tokyo

• Founding of the Institute for Systems Biology in Seattle (headed by Leroy Hood)

• Initiation of activities for SBML (Systems Biology Mark-up Language) mainly led by John Doyle at Caltech
Human Genome Project

3 billion dollars → 3 billion base pairs

The milestone of modern biology

2001-2-15 Nature

2001-2-16 Science
Institute for Systems Biology

ISB was co-founded in 2000 in Seattle, Washington by Dr. Leroy Hood, an immunologist and technologist; Dr. Alan Aderem, an immunologist and Dr. Ruedi Aebersold, a protein chemist. It has since grown to more than 300 staff members, including 13 faculty members and laboratory groups.

www.systemsbiology.org
The Seventh International Conference on Systems Biology

October 9-13, 2006 in Yokohama, Japan
Pacifico Yokohama

The 7th International Conference on Systems Biology continues an annual series of conferences initiated by Hiroaki Kitano in Tokyo in 2000. Subsequent conferences were held at Caltech in Pasadena, at the Karolinska Institute in Stockholm, and at Washington University in St. Louis, Missouri. In 2004 in Heidelberg, Germany. and in 2005, ICSB was held at at Harvard Medical School in Boston.

Venue: Pacifico Yokohama

Schedule:
October 8 Tutorials
October 9-11 Conference
October 12&13 Workshops

Announcements on ICSB 2006

(2006:05:01) Notice of Acceptance for The Round One submissions are being sent out today. For those who submitted abstract by April 1st, and have not received notification yet, please check your email address at the abstract submission site and update the data. If you have any question about your abstract acceptance, please contact the ICSB 2006 Secretariat at secretariat@icsb-2006.org.
SBML is a machine-readable format for representing models. It's oriented towards describing systems where biological entities are involved in, and modified by, processes that occur over time. An example of this is a network of biochemical reactions. SBML's framework is suitable for representing models commonly found in research on a number of topics, including cell signaling pathways, metabolic pathways, biochemical reactions, gene regulation, and many others.
Complex systems of simple elements have functions that emerge from the properties of the networks they form. Biological systems have functions that rely on a combination of the network and the specific elements involved.
The Omics-Cascade

What can happen

GENOME

What appears to be happening

TRANSCRIPTOME

Bioinformatics

What makes it happen

PROTEOME

What actually happens

Systems Biology

METABOLOME

PHENOTYPE

WHY WE CARE!
Scientific Approach:
The “4 M’s” of Systems Biology at MIT

Mathematical Analysis:
- elucidate hypotheses (*mining*)
- facilitate predictions (*modeling*)

Systematic experimentation:
- pathway focused - multi-variate
  - resolved in time and space
Welcome to E-Cell Project

E-Cell Project is an international research project aiming to model and reconstruct biological phenomena in silico, and developing necessary theoretical support, technologies and software platforms to allow precise whole cell simulation.

Highlights

- **Qiu D. paper accepted in Development, Growth and Differentiation** - 2004/07/13
  "Sustained MAPK activation depends on continual stimulation of NGF receptors: an alternative mechanistic model based on computer simulation" by Qiu D., Mao L., Kikuchi and Tomita M. has been accepted for publication in *Development, Growth and Differentiation*.

- **E-Cell Simulation Environment Ver.3.1.102 Released** - 2004/06/02
  E-Cell Simulation Environment Ver.3.1.102 has been released. The latest version of the software can be downloaded from the developer site. See [here](#) for a full changelog.

- **Ishii N. paper accepted in Journal of Biotechnology** - 2004/04/27
  "Toward large-scale modeling of the microbial cell for computer simulation" by Ishii N., Robert M., Nakayama Y., Kanai A. and Tomita M. has been accepted for publication in *Journal of Biotechnology*.
Virtual Cell

Developed by Loew and Schaff of the University of Connecticut Health Center.
http://www.nrcam.uchc.edu/

The National Resource for Cell Analysis and Modeling (NRCAM), developer of the Virtual Cell Model Simulation Framework, is a national resource supported by the National Center for Research Resources (NCRR), at the National Institutes of Health (NIH).

NRCAM is located at the University of Connecticut Health Center and is part of the Center for Cell Analysis and Modeling, CCAM.

3rd International Symposium on Computational Biology
March 19-23, 2005
Scope of Systems Biology

• Basic elements
  – Integration of experimental & theoretical approaches
  – Focus on complex systems that involve multiple scales
  – Strong emphasis on mathematical modeling & analysis
  – Highly interdisciplinary

• Related research areas
  – Theoretical biology – applied mathematicians
  – Bioinformatics – computer scientists
  – Computational biology – computational scientists
  – Systems biology is more general & more difficult to define

• Long-term potential
  – Genotype - phenotype
Network Biology

• Structure of the systems (Network)
• The dynamics of such systems
• Methods of control systems
• Methods to design and modify for desired properties
Gene regulatory network

1. Transcription

2. Translation

Protein synthesis
Protein Interaction Network

Finding Proteins That Interact

One technique, called the yeast two-hybrid system, relies on bringing into close proximity two halves (a and b) of a protein that activates a gene that causes a yeast cell to turn blue. It is used to determine which of a pool of unknown "prey" proteins binds to a known "bait" protein.

1. Insert DNA encoding a known "bait" protein linked to DNA for half (a) of the activator protein.

2. Insert DNA for the other half (b) of the activator protein linked to DNA encoding random "prey" proteins.

3. Look for color change, which indicates "prey" protein binding to "bait"
Metabolic Pathway/KEGG
Signal transduction pathway
TRANSPATH – p53 pathway
What is SBML?

• A machine-readable format for representing computational models in systems biology
  – Expressed in XML using an XML Schema
  – Intended for software tools—not for humans
    • (Although it is text-based and therefore readable)
    • Think HTML

• Intended to be a tool-neutral exchange language for software applications in systems biology
  – Simply an enabling technology
What kind of models can you express in SBML?

- **Focus:** systems of biochemical reactions

- **Models can also include:**
  - Compartments
  - Rules/constraints
  - Discrete events
Structure of models expressed in SBML

• Basic structure of SBML is straightforward:
• A model is a list of its components:
  – Beginning of SBML model definition
    • List of function definitions
    • List of unit definitions
    • List of compartments
    • List of molecular species
    • List of parameters
    • List of rules
    • List of reactions
    • List of events
  – End of SBML model definition
<?xml version="1.0" encoding="UTF-8"?>
<sbml xmlns="http://www.sbml.org/sbml/level1"
      level="1" version="2">
    <model name="gene_network_model">
      ...
    </model>
</sbml>
The Systems Biology Markup Language (SBML): a medium for representation and exchange of biochemical network models


SBML Level 2

• SBML is being defined in *levels*
  – Higher levels add more functionality & complexity
  – Defined collaboratively with many software developers and modelers (the *SBML Forum*)

• SBML Level 2 features finalized in June 2003
  – MathML for mathematical expressions
  – Support for user-defined functions
  – Support for “events”
  – Metadata annotations
  – Miscellaneous fixes
SBML Level 3

- Addition of "modules" providing facilities for representing more information, including:
  - Graphical diagrammatic visual layout of models
  - Model composition (submodels)
  - Multistate complex species
  - Arrays of elements
  - Representing 2-D & 3-D spatial geometry
  - And several others…
  - *If interested, please join discussions at [http://sbml.org](http://sbml.org)*
Application Support for SBML

- BASIS — UK
- Bio Sketch Pad — BBN
- CellDesigner — ERATO Kitano Symbiotic Systems Project
- Cellerator — NASA JPL & University of California Irvine
- Cytoscape — Institute for Systems Biology & MIT
- Gepasi — Virginia Tech
- Jarnac — Keck Graduate Institute
- JDesigner — Keck Graduate Institute
- JigCell — Virginia Tech
- NetBuilder — University of Hertfordshire
- SigPath — Mount Sinai
- StochSim — Cambridge University
- Virtual Cell — University of Connecticut Health Center
- WinSCAMP — Keck Graduate Institute
- COBRA — UCSD
Software from the SBML Team

• Embeddable software library for using SBML
  – LibSBML 2.3.0
    • Provides API for C, C++, Java, Python, others
    • Supports Linux, Mac, Windows
• Interfaces to popular general math environments
  – MathSBML 2.3.38 for Mathematica
  – SBMLToolbox 2.0 for MATLAB
• Conversion tools
  – KEGG2SBML 1.1.0
  – CellML2SBML 1.0
• Web-based facilities
  – Validation, visualization, example models
The Systems Biology Markup Language (SBML) is a computer-readable format for representing models of biochemical reaction networks. SBML is applicable to metabolic networks, cell-signaling pathways, regulatory networks, and many others.

Internationally Supported and Widely Used

SBML has been evolving since mid-2000 through the efforts of an international group of software developers and users. Today, SBML is supported by over 80 software systems, including the following (where "*" indicates SBML support in development):

- BALSA
- BASIS
- BIOCHAM
- BioCharon
- biocy2SBML
- BioGrid
- BioModels
- BioNetGen
- BioPathway Explorer
- Bio Sketch Pad
- BioSPICE Dashboard
- BioSpreadsheet
- BioTapestry
- BioUML
- BTLab
- CALIVE
- CellDesigner
- Cellerator
- CellML2SBML
- Cellware
- CL-SBML
- COPASI
- Cytoscape
- DBsolve*
- Dizzy
- E-CELL
- ecellJ
- ESS
- FluxAnalyzer
- Fluxor
- Gepasi
- INSILICO discovery
- Jarnac
- JDesigner
- JigCell
- JSim
- JWS Online
- Karyote*
- KEGG2SBML
- Kin solver*
- libSBML
- LION Target Engine
- MATLAB SBToolbox
- MathSBML
- MesoRD
- MetaboLogica
- MetaFluxNet
- MMCT2
- Modesto
- Moleculizer
- Monod
- NetBuilder
- PANTHER Pathway
- PathArt
- PathScout
- Pathway Tools
- PathwayBuilder
- PaVESy
- Reactome
- ProcessDB*
- PySCeS*
- runSBML
- SBML ODE Solver
- SBMLEditor
- SBMLR
- SBMLToolbox
- SBW
- SCiPath
- Sigmoid*
- SigPath
- SigTran
- SIMBA
- Simpaticha
- SimWiz
- SmartCell
- StochSim
- STOCKS
- TERANODE Suite
- Trellis
- Virtual Cell
- WinSCAMP

Nature supports SBML, BioModels


read more

BioletGen 1.1 Release

(May 4, 2005) BioletGen is a software package for generating computational models accounting for the full spectrum of molecular species implied by a reaction system.

read more

LibSBML 2.3.0 released!

(May 3, 2005) The latest version of libSBML, a portable library for working with SBML files and data streams, is now available. It has exciting new features such as an overhauled validation system.

read more

RunSBML Released

(April 14, 2005) Ariadne Genomics has released runSBML, a free, GPL-licensed simulation tool that uses SBML as its input format and produces time series of concentration values as output.
Impact and Potential of Systems Biology

- Predictive and Personalized Medicine
- Synthetic Biology
- Physics and Chemistry
- Computer Science
Impact and Potential of Systems Biology

• Toward predictive and Personalized Medicine
  – New **P4 Medicine** (Leroy Hood)
    • Predictive, preventive, personalized and participatory
    • A personalized medicine that will revolutionize health care
  – Drug companies will have the opportunity for more effective means of drug discovery
    • Guided by diagnostics
    • Smaller patient populations but higher therapeutic effectiveness
系统生物学为合成生物学提供理论基础
合成生物学为系统生物学提供验证手段

设计新的生物网络：合成生物学

- 合成生物学就是全新设计建立一个生命系统，使其能够按预设的方式运行，同时具有复杂的动力学和逻辑特征。包括：
  - 设计和构建新的生物元器件和系统；
  - 对于已有生物体系的重新组装与设计，使其完成特定的功能。
Impact and Potential of Systems Biology

● On Computer Science
  – Concurrency theory methods to biological systems
    • Encouraged the community to propose a distinct “algorithmic” or “executable” approach to Systems Biology
  – Evolutionary computing
    • Network inference and estimation of parameters (canonical ODE models)
  – Information mining approaches
    • data and text mining
  – Information systems supporting various forms of collaboratories
Impact and Potential of Systems Biology

• On Biology, Physics and Chemistry
  – Bionanotechnology (Biomimetics or Bionik)
    • Where bio-inspired methods are used in effecting nanotechnological advances
  – Nanobiotechnology
    • Uses advances in nanoscience and nanotechnology to study biological processes
  – Bioimaging (microscopy and spectroscopy)
    • Producing data on dynamics so essential for modelling in systems biology
Impact and Potential of Systems Biology

- On Biology, Physics and Chemistry
  - Structural Systems Biology
    - Biophysics (molecular motors)
    - Quantum Chemistry
    - Computational techniques
      - Molecular dynamics
      - Coarse-grained methods
        » Mechanistic details of complex motors such as ribosome or bacterial photosynthetic factories
Systems Biology Research

- Experimental data are essential for modeling and understanding biological processes and systems.
- Without models and hypotheses, accumulated experimental data are generally unstructured and uninformative.
- Systems biology research integrates experimental data of diverse types with coherent models, with the goal of understanding the biological processes and systems being investigated.
Technologies which support the research activities

• Data generation
  – Collect data on the organism under study. Ongoing technologies development aims to increase throughput and efficiency, improve accuracy, and decrease the cost of this work

• Data management
  – Provide us with the means to automate portions of collecting, processing, annotating, and integrating experimental data

• Data visualization and analysis
  – Bioinformatic tools and databases
  – Modeling software to simulate the dynamics of biological processes or systems
Data Generation

• **Probing genetic frameworks:** What is the genomic parts list of an organism? What genes interact in concert to regulate or create a molecular interaction network? How does genetic variation influence gene expression and protein function?
  – Representative technologies: DNA sequencing, genotyping, large-scale gene deletion constructs; RNAi knockouts

• **Probing gene expression patterns:** What genes are up-regulated or down-regulated in response to a genetic or environmental perturbation? What genes are expressed in what tissues under what conditions?
  – Representative technologies: microarrays and DNA tagging procedures
• **Probing DNA-protein interactions:** What genes does a particular transcription factor regulate under defined experimental conditions?
  – Representative technology: chromatin-immunoprecipitation and gene chips to localize binding sites (ChIP-chip)

• **Probing protein-protein interactions:** What proteins are present in enzyme complexes, nuclear pore complexes, the cytoskeleton? Which proteins modify other proteins in signaling cascades?
  – Representative technologies: two-hybrid-based interactions; affinity purification; mass spectrometry; quantitative proteomics

• **Probing subcellular protein localization:** When during development is a protein made and where in the cell does it go?
  – Representative technologies: cell sorting, molecular imaging based on reporter genes or antibody staining
Data management

• Bioinformatics pipelines (BioPerl – https://bioperl.org)
  – Collect, extract, store, and interpret data at several different levels of analysis

• Database frameworks (MySQL)
  – Store data, allow data access by query, and facilitate data curation

Example:
SBEAMS (Systems Biology Experiment Analysis Management System) Platform for managing data derived primarily from microarray and proteomics experiments

www.sbeams.org
Data visualization and analysis

• Sources
  – Literature and curated databases
    • Biochemical pathways, annotated genomes, known protein complexes, or gene ontology tables
  – Large-scale computational tools
    • Gene prediction, binding site prediction, location of genome-wide repetitive elements, protein structure predictions
  – Large-scale wet lab data collection
    – Tables based on microarrays, proteomics, genome sequencing, protein structures
Experimental (left) and computational (right) hierarchies will become increasingly codependent as the research community models greater biological complexity.
Conclusion

• System biology is a new and emerging field in biology
• A long ways to go before understanding biological systems
• We believe that systems biology will be the dominant paradigm in biology, and many medical applications as well as scientific discoveries are expected.

Key words: network; omics; integration; perturbation
Thanks!