Computational Systems Biology

Satoru Miyano

Human Genome Center

Institute of Medical Science, The University of Tokyo

miyano@ims.u-tokyo.ac.jp

October 12, 2009

Outline (1/3)

1. Background

2. Supercomputer for Networks to Drugs and Therapy

3. Inferring and Analyzing Large-Scale Gene Networks

Outline (2/3)

4. Gene Networks in Human Umbilical Vein Endothelial Cell

   4.1 Gene Networks of HUVEC: Searching Drug Target Pathways Using Fenofibrate

   4.2 HUVEC Gene Networks: TNF-α and new hub genes regulating Inflammation and apoptosis

   4.3 Gene Networks of HUVEC and Fenofibrate: Dynamic activities of autocrine pathways that control drug-response transcriptome networks

Outline (3/3)

5. Gene Networks of Small Airway Epithelial Cell (SAEC) and Gefitinib: Growth Factor Signaling Systems Identify Critical Genes for Survival Prediction in Early Stage Lung Adenocarcinoma

6. Data Assimilation for Biological Systems

7. Cell Illustrator: A Computational Platform for Systems Biology

8. Predictive Personalized Medicine

1. Background

Biology is changing fast into a science of information management.

NIH Roadmap FOR MEDICAL RESEARCH

NIH Roadmap FOR MEDICAL RESEARCH

Biology has always been a haven for **2003**
microscopes, test tubes, and Petri dishes, but this
conventional picture of the field is expanding rapidly.
Sophisticated techniques adapted from physics,
chemistry, and engineering enable scientists to use
computers and robots to separate molecules in
solution, read genetic information, reveal the three-
dimensional shapes of natural molecules like proteins,
and take pictures of the brain in action. All of these
techniques generate large amounts of data, and
biology **is changing fast into a science of
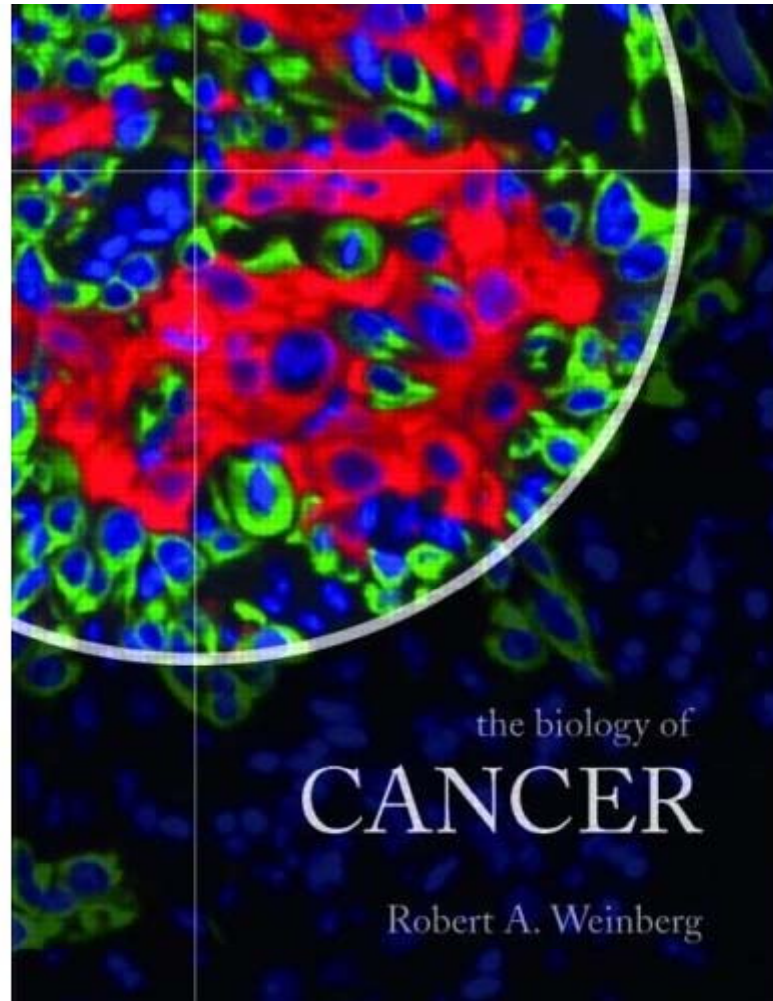information management.**

Prologue

The Tale of the Heike (13ᵗʰ Century, Author Unknown)

It tells the story of the rise to glory and eventual downfall of the Heike clan in the late twelfth century, a theme based on the Buddhist concept that the proud will surely be destroyed.
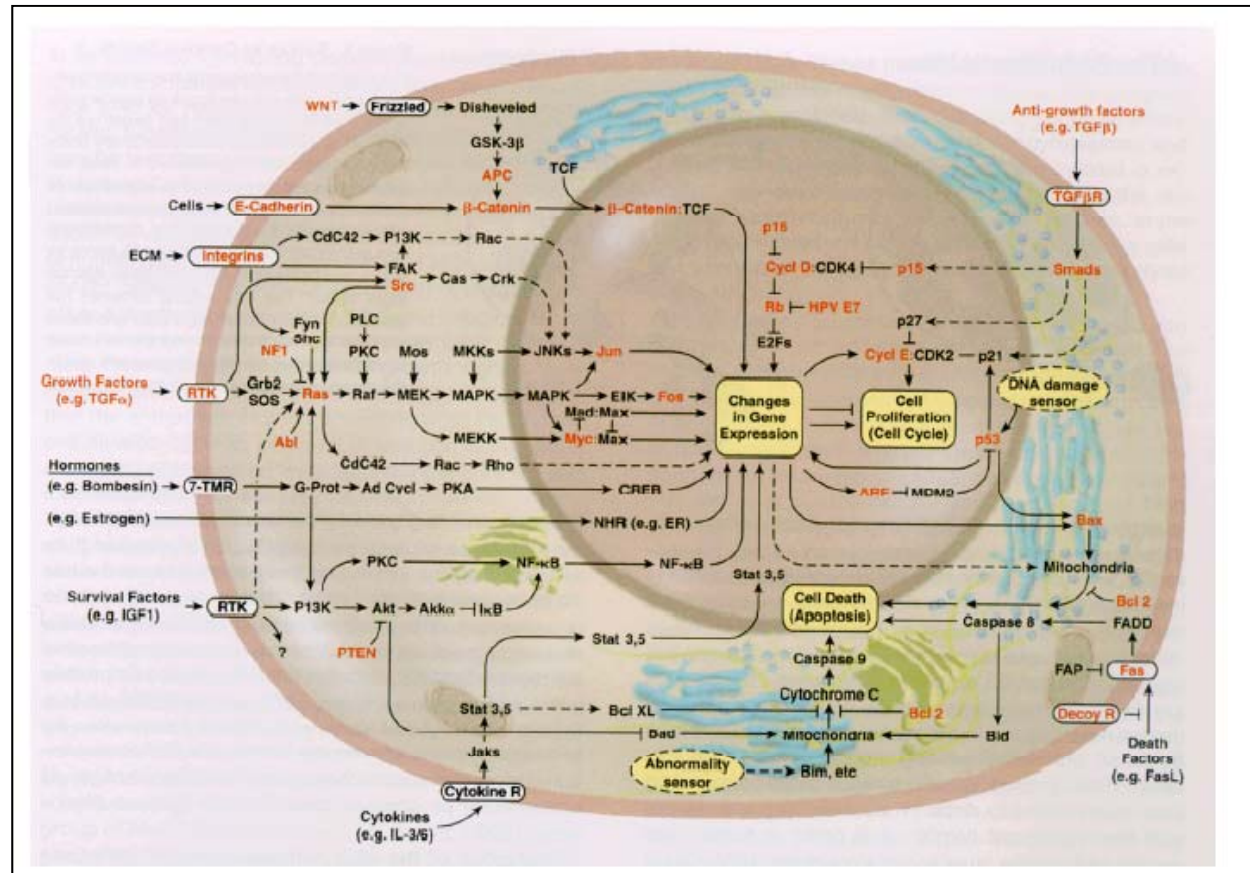
The Biology of Cancer

Robert A. Weinberg

Golden Days of Molecular Biology

- **Phenotypes and Genes**
  - Paradigm of Molecular Biology

- **Super Stars in Biology**
  - Novel Prize Laureates

- **Dramatic History**
  - "The Biology of Cancer", R.A. Weinberg

Hallmark of Cancer

D. Hanahan and R. A. Weinberg.
Cell., 100(1):57–70 Review, 2000.

But, Very Naïve Understanding and Representation as Systems

- Draw pictures and add English narrations for biological facts

- "Systems Knowledge" or "Model" which is unambiguously represented is at most causal relations between molecules

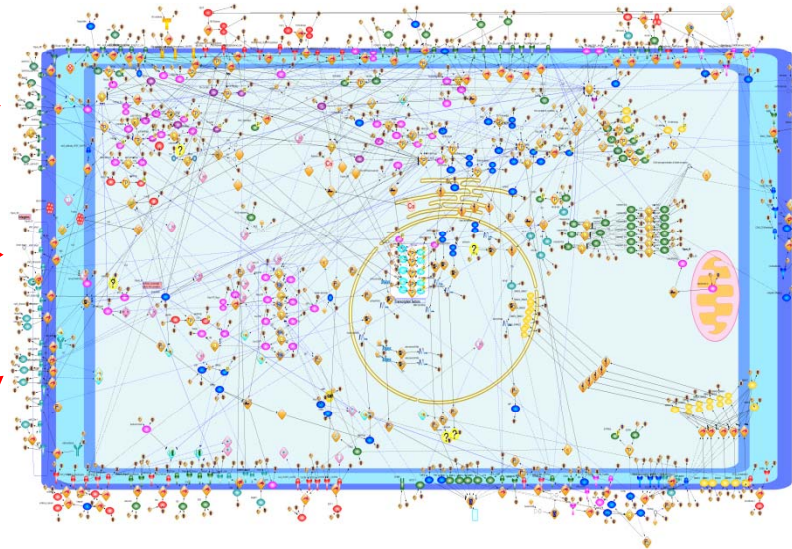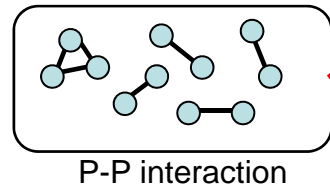- Huge gap between "what is represented" and "what is to be represented"
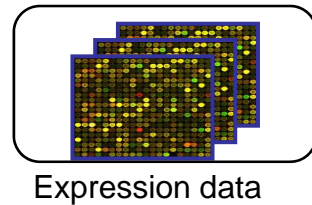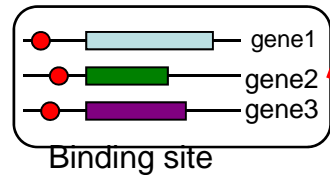
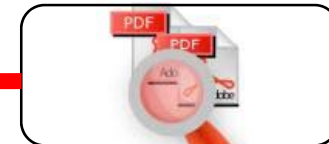# 2. Supercomputer for Networks to Drugs and Therapy

My view of

Understanding Biological Systems through Computation

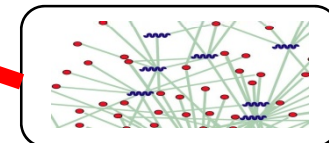**Mining Molecular Networks from Experimental Data**
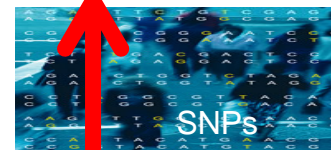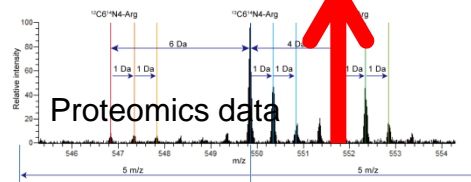
Modeling & Simulation of Biological Networks

Binding site

Expression data

P-P interaction
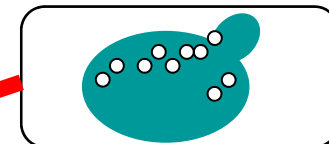
Protein subcellular localization

Literature

microRNA network

Proteomics data

SNPs

Laboratory Experiments

Strategic Computational Initiative
Supercomputer System for 2009-2014

Human Genome Center
Institute of Medical Science, University of Tokyo

Renewed in January 2009

- **January 2009**: 75 TFLOPS at peak & 1 PB Disk Space
  PC Cluster (Sun Microsystems)
  Large Shared Memory Machine (SGI Altix)
  Lustre File System (Sun Microsystems)
- **January 2012**: 225 TFLOPS at peak & 4PB Disk Space

RIKEN Next-Generation Supercomputer (Kobe, Japan)
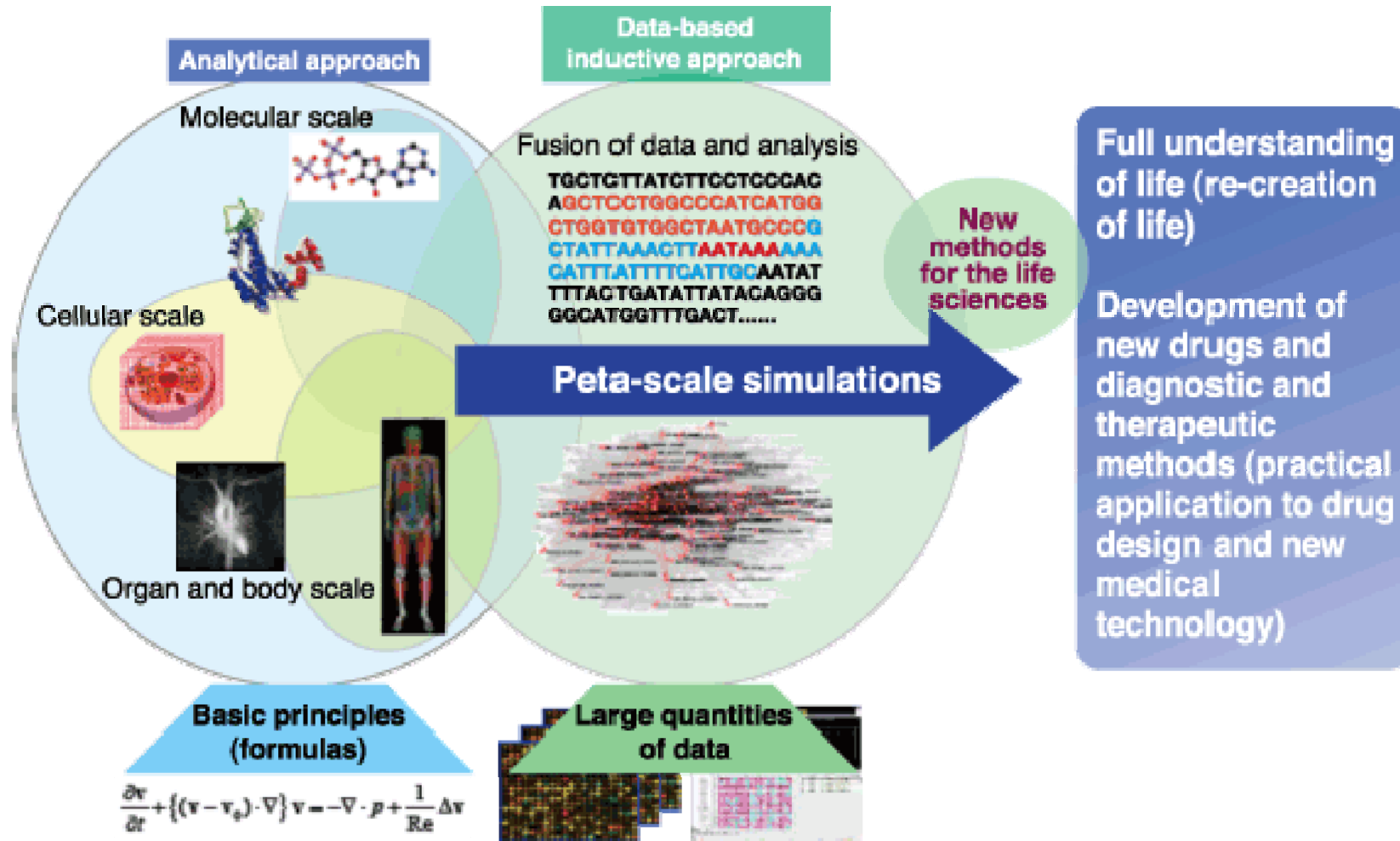10 PETA FLOPS COMPUTER

次世代スーパーコンピュータ施設 完成イメージ図

正面

北西から研究棟側を見る

H21.04.09

Total Project Budget for 7 Years:
115,000,000,000JPY

Grand Challenge
"The Next-Integrated Life Simulations"

Team: Data Analysis Fusion

Satoru Miyano (U. Tokyo)

Network Inference and Applications

Tomoyuki Higuchi (Institute of Mathematical Statistics)

Data Assimilation Technology for Biological Systems

Large-scale network inference technology	Bayesian statistical methods	Data assimilation technology
Linkage & Linkage Disequilibrium	Haplotype Analysis	PPI prediction technolgy

Naoyuki Kamatani (RIKEN SNP Research Center)

Large-scale SNP Data Analysis

Yutaka Akiyama (Tokyo Institute of Technology)
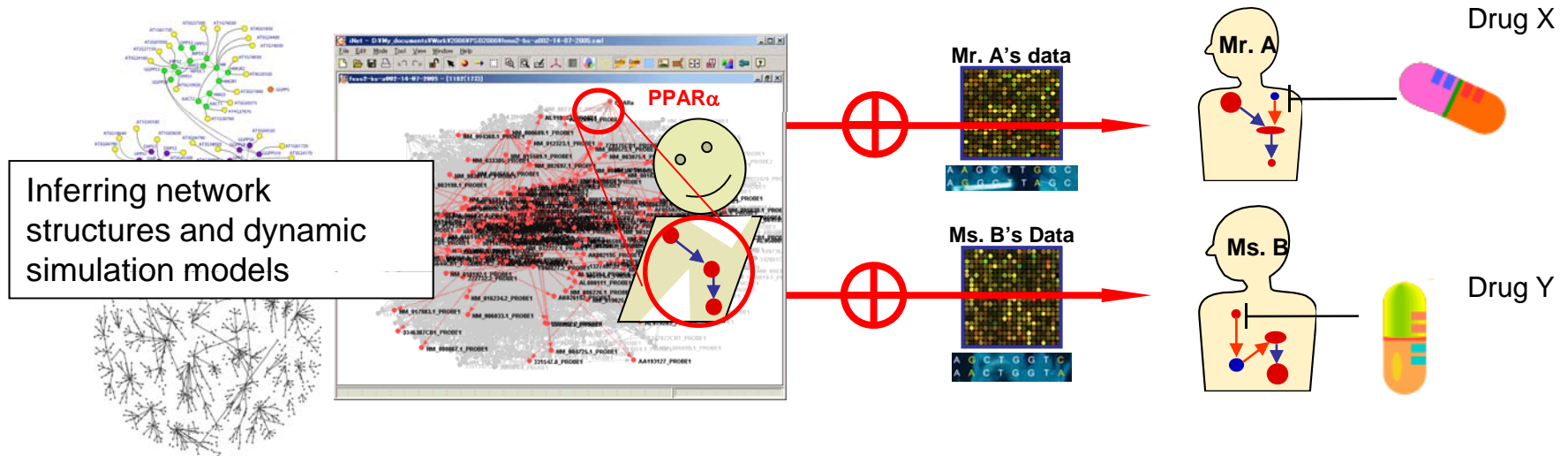
1000x1000 PPI Challenge

Networks to Drugs and Therapy

Drug target discovery

Diagnosis/Therapy

"General" Data/Model

"Personal" Data/Model

Inferring network structures and dynamic simulation models

PPARα

Drug X

Mr. A's data

Mr. A

Ms. B's Data

Ms. B

Drug Y

Predict large molecular networks, and use them as "maps" for searching drug/disease related genes and pathways.

Apply "data assimilation" to revise "general models" with "personal data".

Discover genes related to phenotypes (diseases/drug responses), and predict them from SNPs and environmental data.

3. Inferring and Analyzing Large-Scale Gene Networks

Two Models for Gene Networks

1. Bayesian Networks with Nonparametric Regression

   - Nonlinear Quantitiative Causal relation

2. State Space Models (SSM)

   - Dynamics & Simulation

Data for Modeling

DNA Sequences
Amino Acid sequences
Protein Structures
Gene Expression Data
    cDNA Microarrays
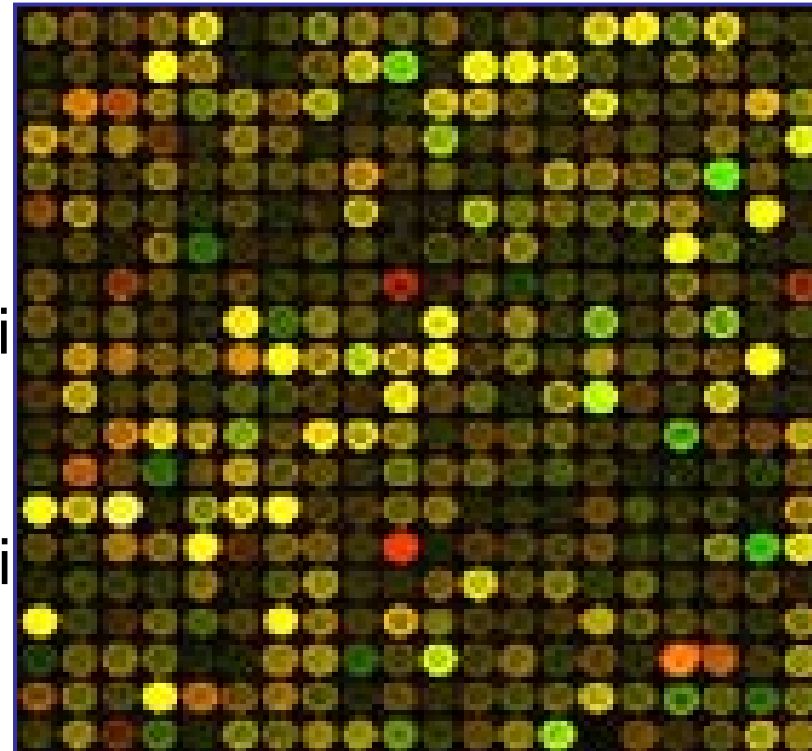   DNA Chips
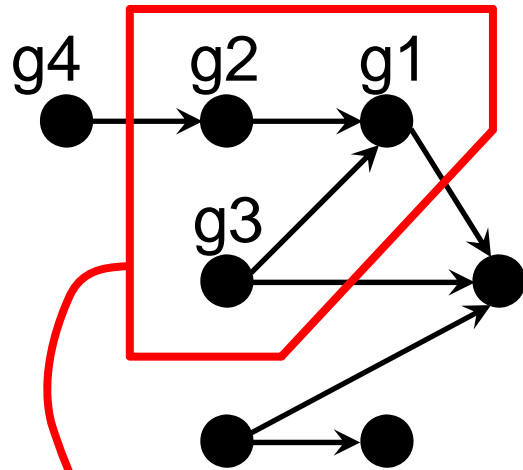Protein-Protein/DNA Interacti
    Protein Chips
    MALDI/TOF MS
Protein Subcellular Localizati
Literature
    MEDLINE database

1. Bayesian network

g4   g2   g1

g3

DAG encoding the Markov assumption.

The joint density can be computed by the product of the conditional densities.

$$x_{i1} \iff \mathbf{p}_{i1} = (x_{i2}, x_{i3})^T$$

$$f(x_{i1}, ..., x_{ip} \mid \boldsymbol{\theta}_G) = \Pi_{j=1}^{p} f_j(x_{ij} \mid \mathbf{p}_{ij}, \boldsymbol{\theta}_j)$$