Human Genome Center
Institute of Medical Science,University of Tokyo

Computational Strategy for Systems Biology and Drug Target Pathway Discovery

Satoru Miyano

Human Genome Center
Institute of Medical Science, University of Tokyo

Hotel Zürichberg, Zürich
September 15, 2008

10 PETA FLOPS COMPUTER will operate in 2011

RIKEN Next-Generation Supercomputer (Kobe, Japan)

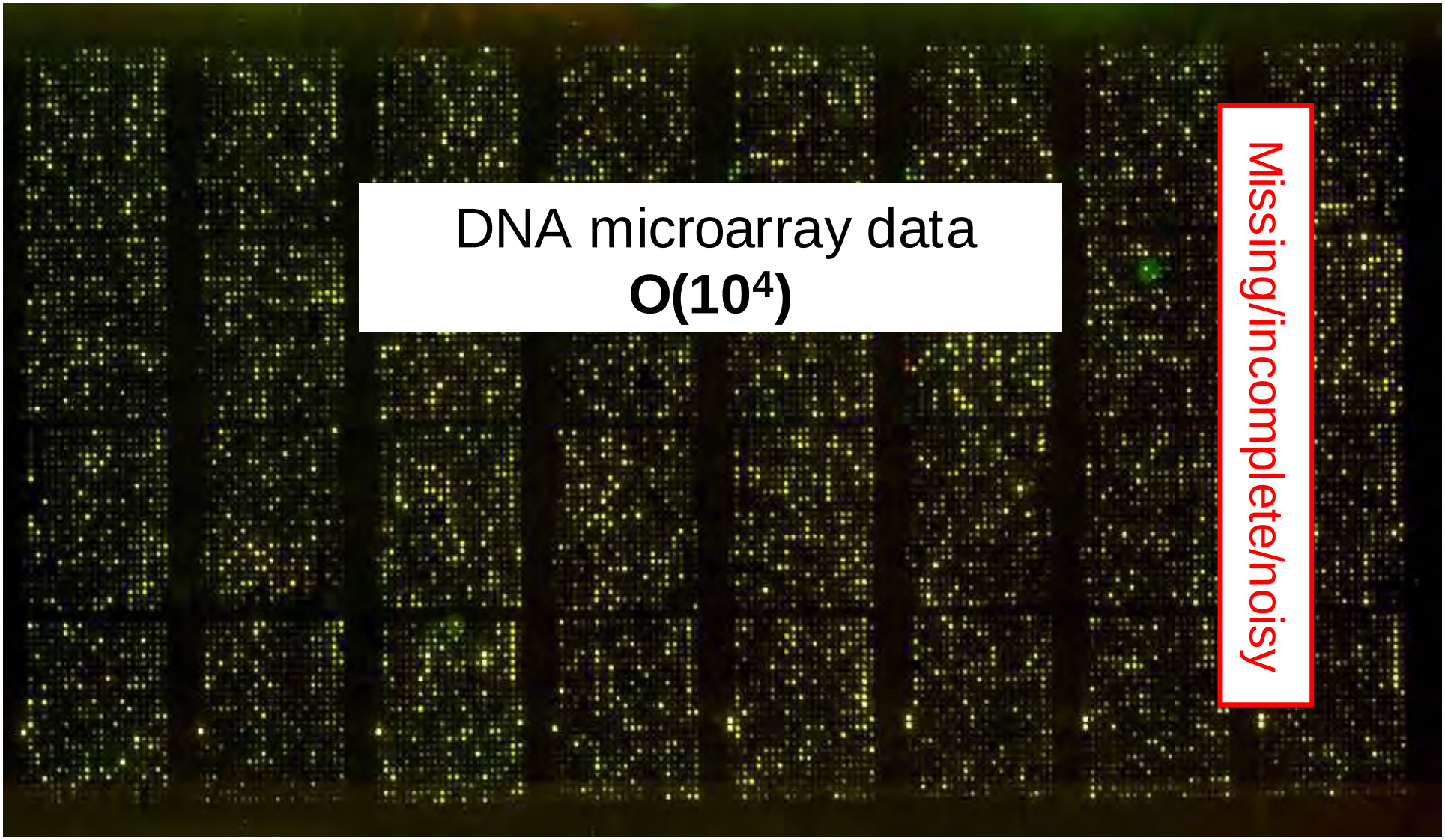
We are facing with

high dimensional, heterogeneous,
huge data related to genes and
their products.

Computational resources

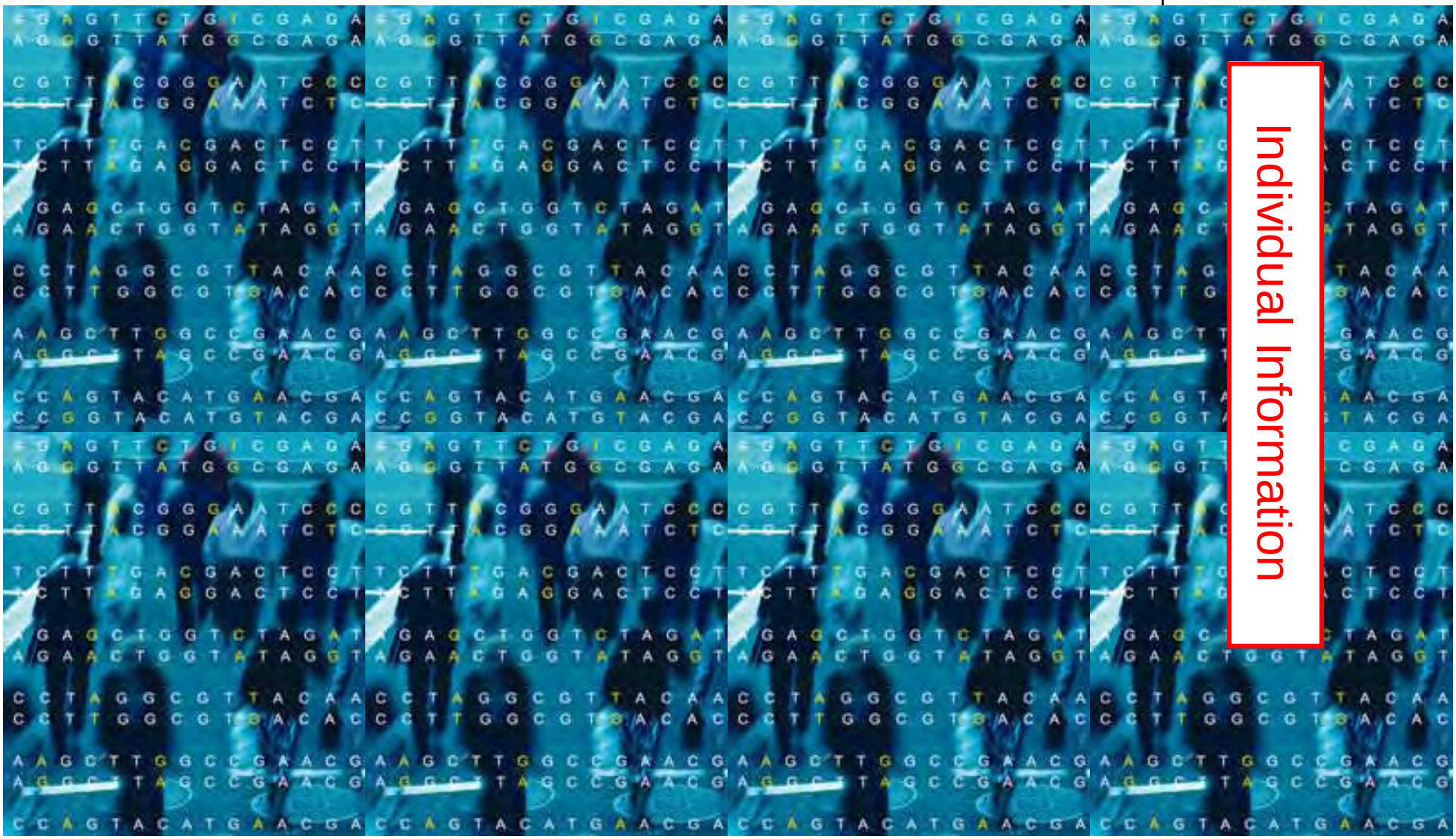
are enormously required.

Large-Scale High Dimensional Data



SNPs (Single Nucleotide Polymorphisms)

$O(10^5)\sim$



Association Analysis of Haplotypes and Phenotypes

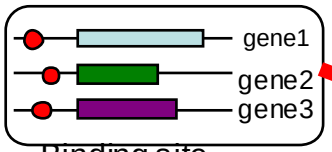
Dr. Kamatani
(RIKEN Center for Genomic Medicine) said:

- Within 20,000 haplotype blocks, there are 500 haplotype blocks with more than 20 loci. But it requires 1,200 days for computation on 10 TPLOPS computer
- It just requires only 12 days on 10 PFLOPS computer.

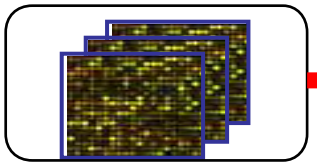
Computational Strategy for Understanding Biological Systems

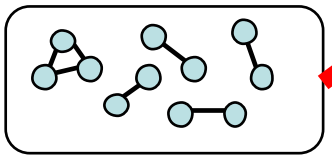
Gene Network Computation from Data

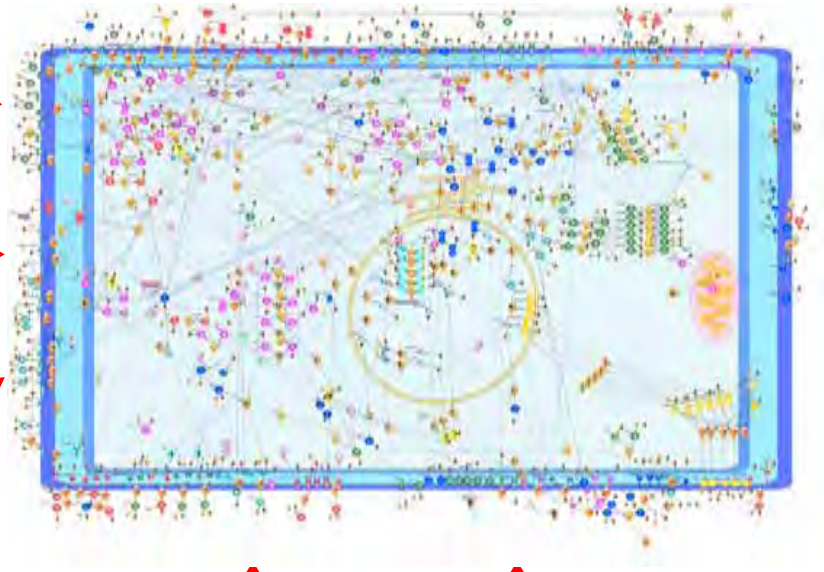
Database Management System for Dynamic Biological Pathways

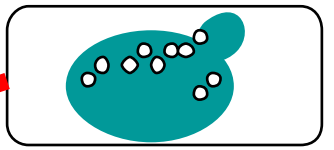


Binding site

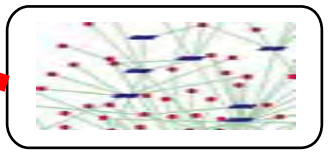

Expression data

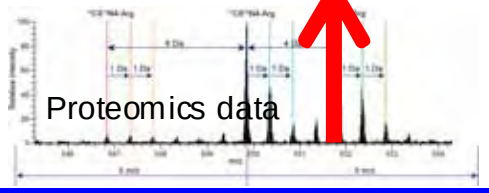

P-P interaction




Protein subcellular localization

Literature


microRNA network


Proteomics data


SNPs

Data Assimilation for Fusing Simulation Models and Personal Data with Supercomputer

Software Platform for Systems Biology

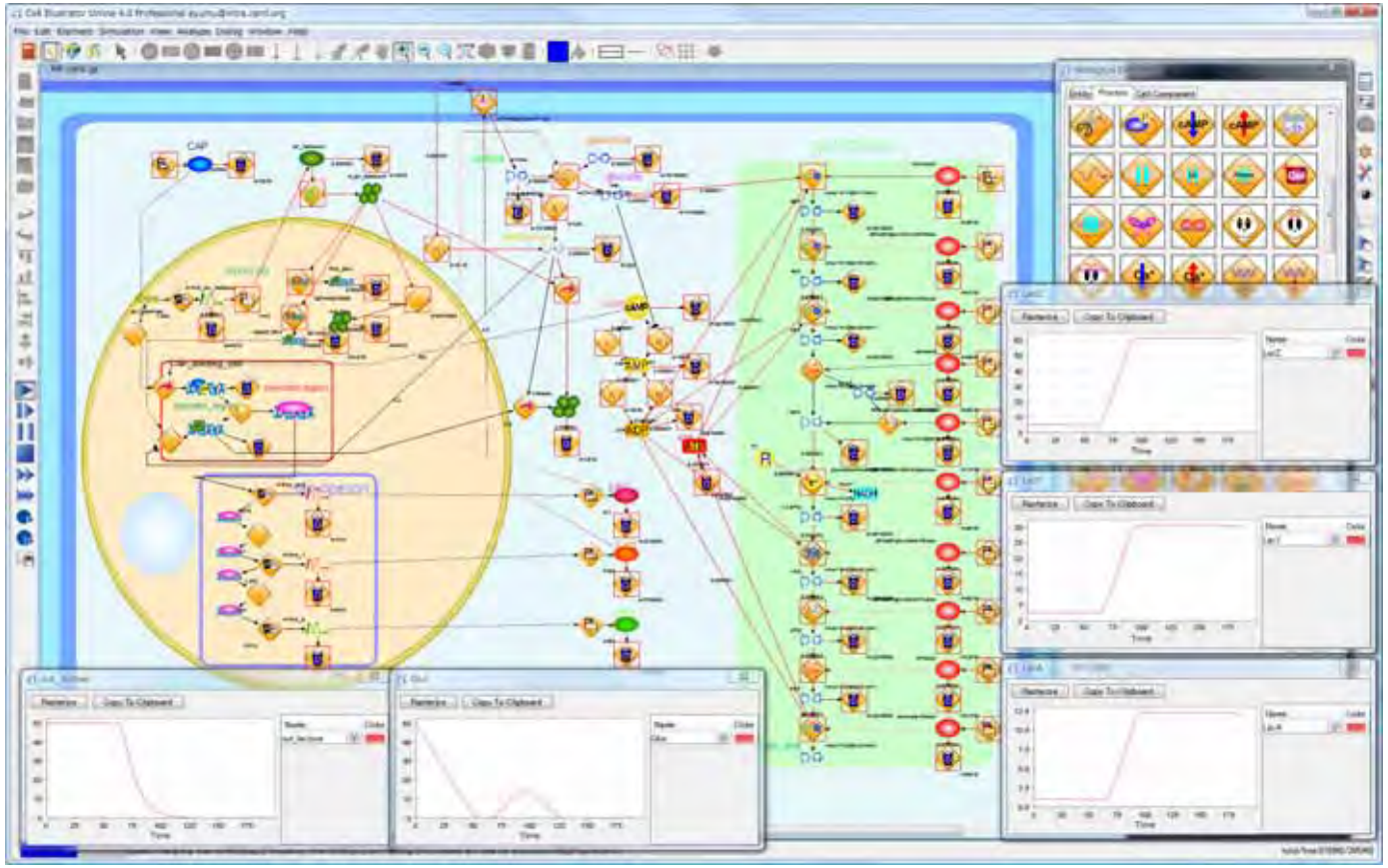
Cell Illustrator Online

https://cionline.hgc.jp

Commercially available from BIOBASE

Software Tool for Modeling and Simulation

XML format Cell System Markup Language CSML and Cell System Ontology CSO for describing biological systems with dynamics and ontology

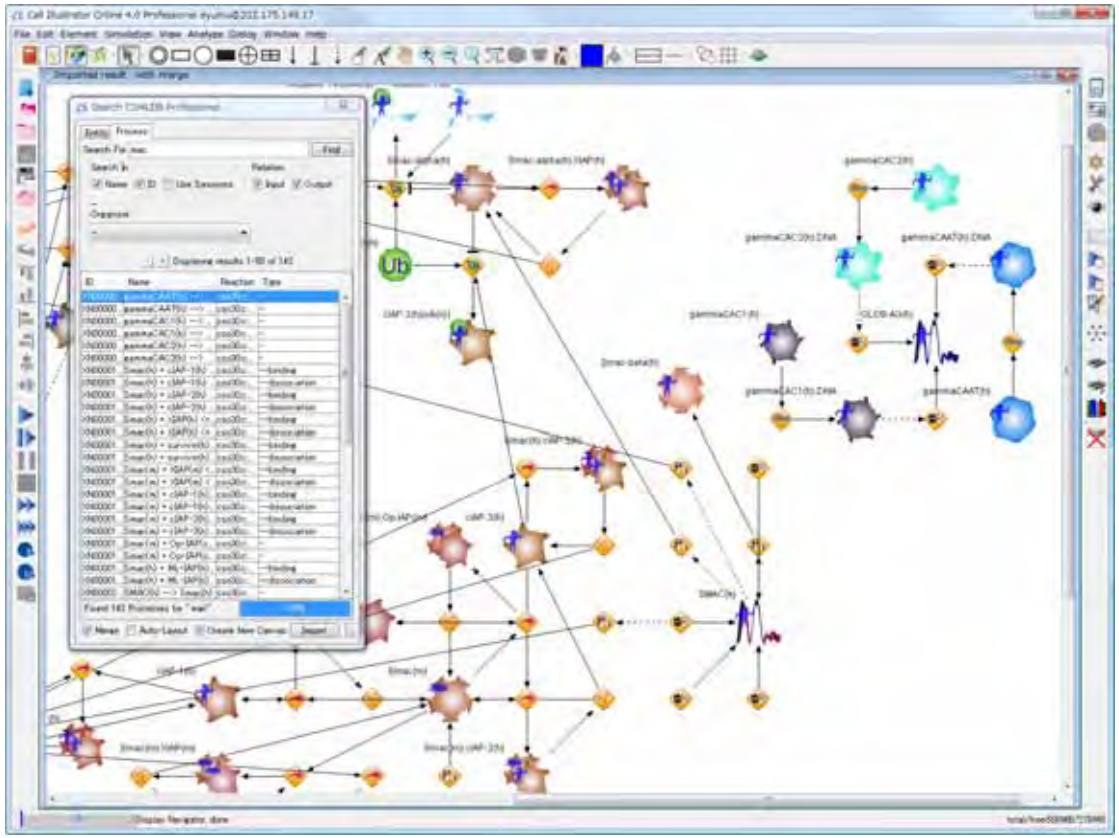


Nagasaki M, Doi A, Matsuno H, Miyano S. Genomic Object Net: I. A platform for modeling and simulating biopathways. Applied Bioinformatics. 2003; 2: 181-4.

Pathway Database Search Module

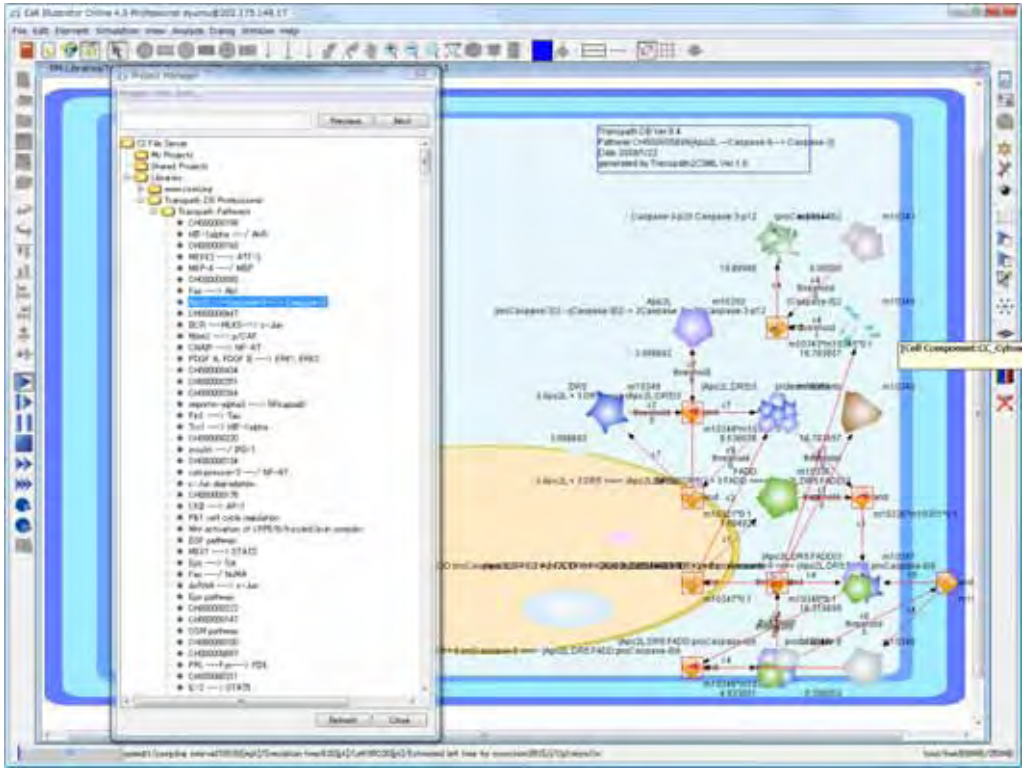
- Pathway models in CSML format are stored into one uniform database and it is possible to search the database with various search options via GUI interface.

※TRANSPATH 8.4 (BIOBASE) is supported. Mar/2008.
※It is possible to support other pathway models if converted into the CSML format.



BIOBASE TRANSPATH Pathway Library Module

- More than 1,000 TRANSPATH pathways (Signal Transduction Pathway and Gene Regulatory Network) are supplied. All pathways can load, edit, save and simulate on CIO4.0.
  - Support pathways supplied in TRANSPATH 8.4 (BIOBASE).
  - Academic user can register and use the academic version of TRANSPATH.
  - Curated 100,000 reactions and 100,000 molecules in Human and Mouse.

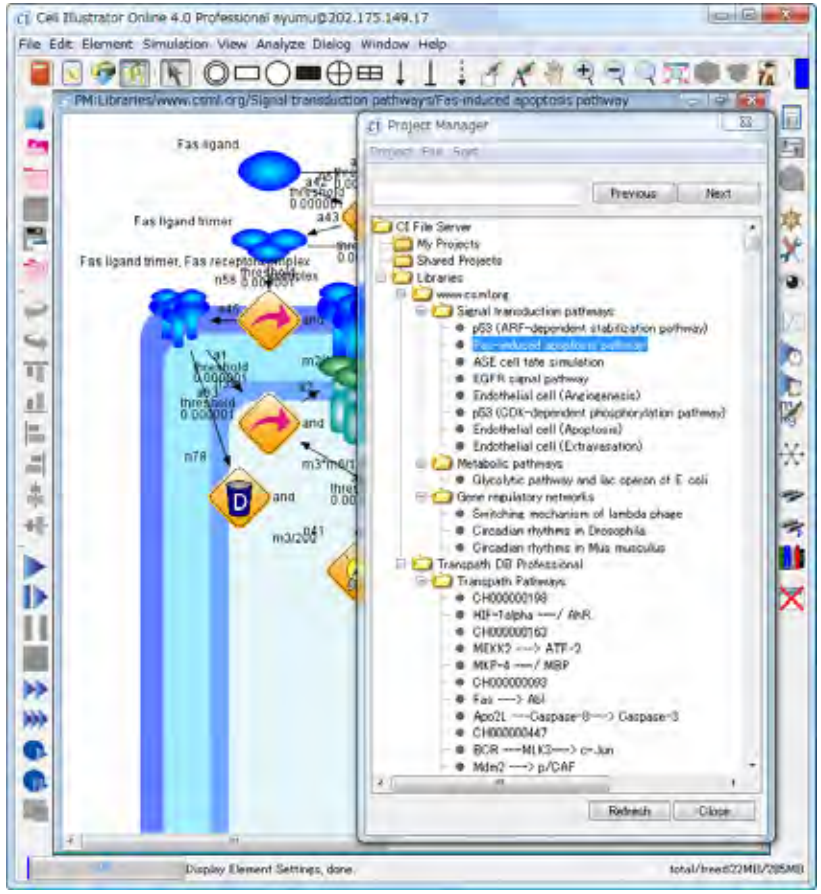


GNI Ltd. and the University of Tokyo

Project Management Module

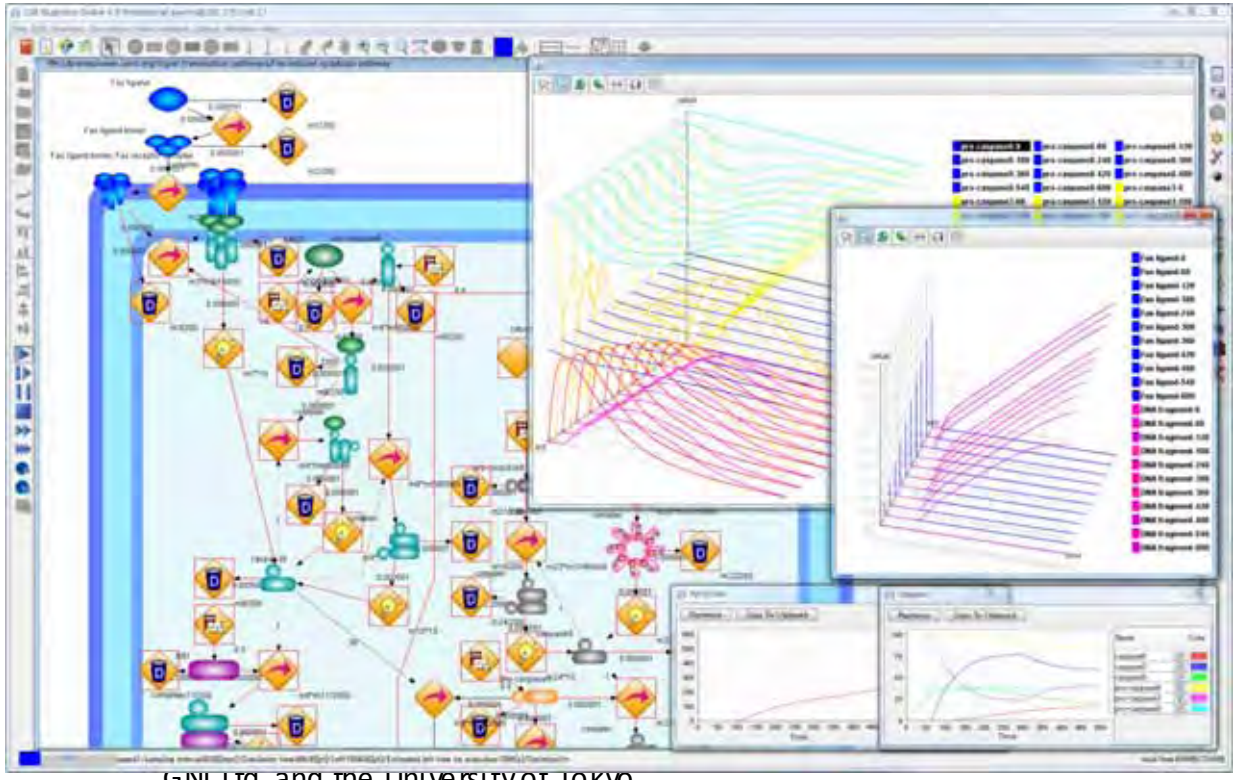
- User can store the pathway model, related experimental data and report to the server side.
- The each stored project on server can be shared with other permitted users (read, write or both permission.)

- Public pathway models – latest signal transduction pathway, metabolic pathway and gene regulatory network – (same models in http://www.csml.org/) can access from the GUI interface of the module.



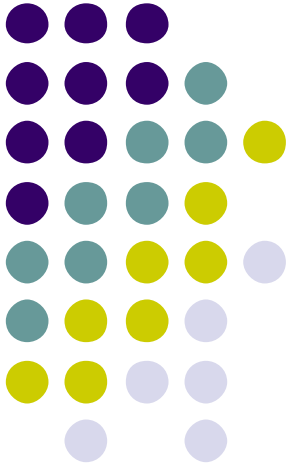
Pathway Parameter Search Module

- For a CIO pathway model, the module executes the user specified multiple initial conditions at once and displays the result with 2D or 3D plots.
  (※The module needs to activate other two simulation related modules in advance.)

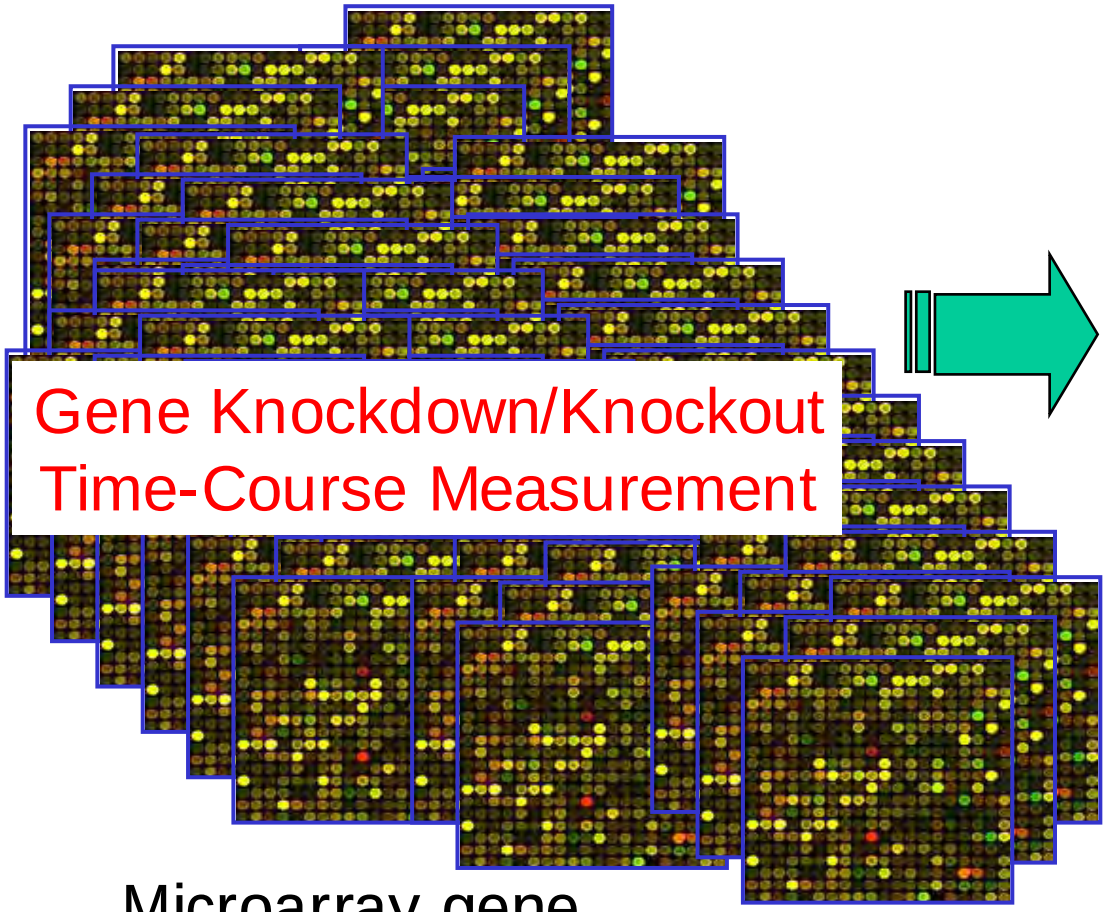


Mining Large-Scale Gene Network Structures from Gene Expression Data

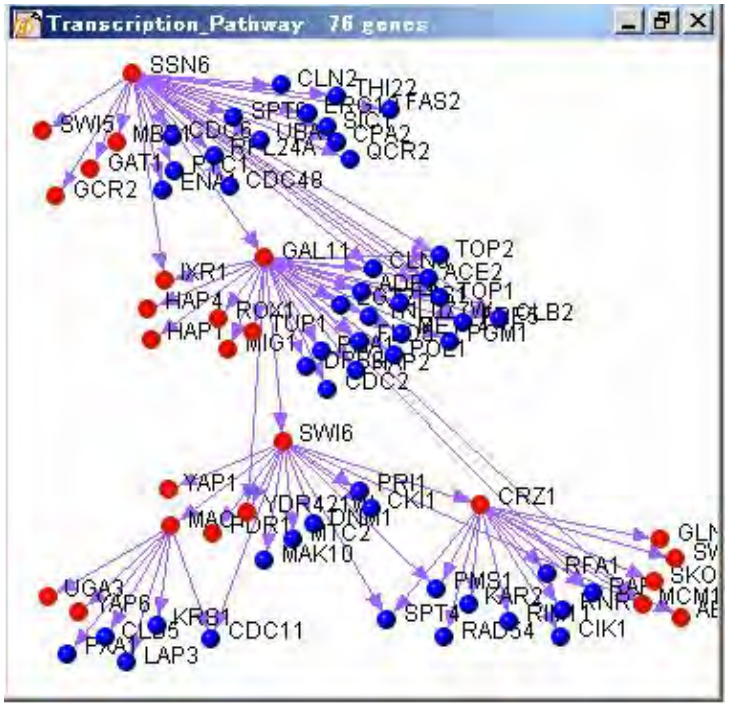
- Large-scale (>300) siRNA gene knock-down
- Drug responses in time-course
- Microarray measurements



Bayesian Network and Nonparametric Regression + α

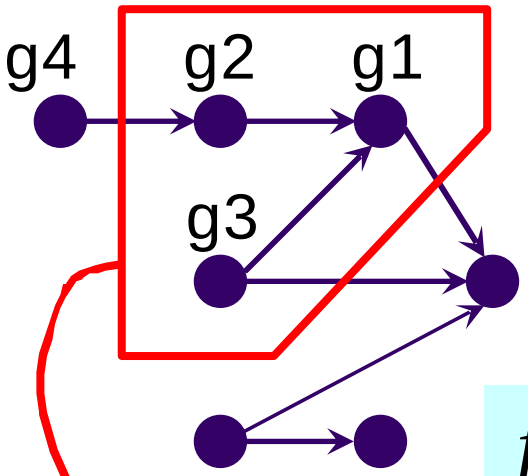


Microarray gene expression data



Gene network

Bayesian networks



DAG encoding the Markov assumption.

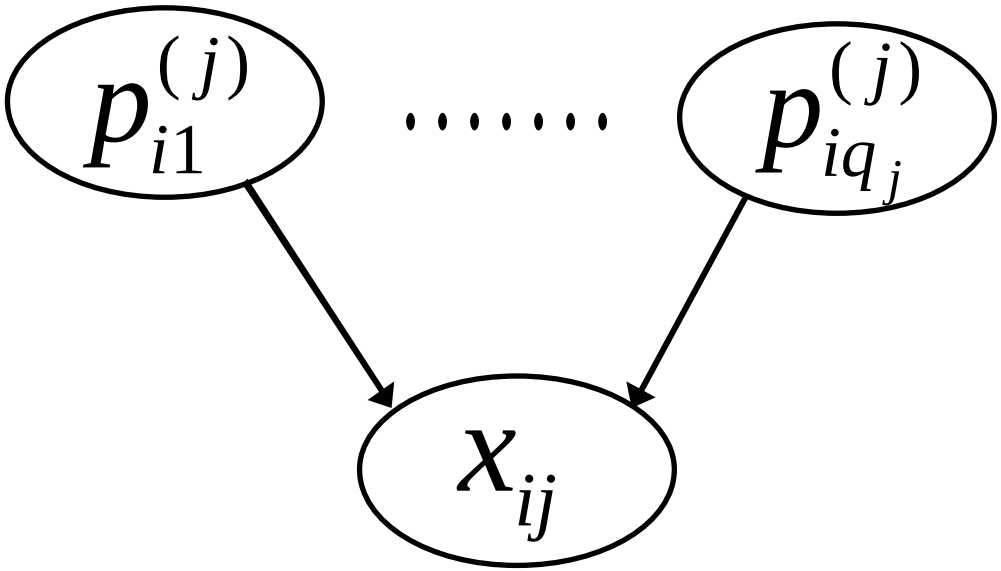
The joint density can be computed by the product of the conditional densities.

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

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

- Imoto, S., Goto, T., Miyano, S. Estimation of genetic networks and functional structures between genes by using Bayesian network and nonparametric regression. Pacific Symposium on Biocomputing. 7:175-186, 2002.
- Imoto, Kim, Goto, Aburatani, Tashiro, Kuhara, Miyano (2003). Bayesian network and nonparametric heteroscedastic regression for nonlinear modeling of genetic networkJ. *Bioinformatics and Comp. Biol.*, 1(2), 231-252

Nonparametric regression



We consider the additive regression model:

$$x_{ij} = m_1(p_{i1}^{(j)}) + \cdots + m_{q_j}(p_{iq_j}^{(j)}) + \varepsilon_j ,$$

$$where \quad \varepsilon_j \sim N(0, \sigma_j^2) \; and \; \mathbf{p}_{ij} = (p_{i1}^{(j)}, ..., p_{iq_j}^{(j)}).$$

Here $m_k(\cdot)$ is a smooth function from R to R.

Nonlinear Bayesian network model

$$f(x_{i1},\ldots,x_{ip};\boldsymbol{\theta}_G)=\prod_{j=1}^{p} f_j(x_{ij}\,|\,\mathbf{p}_{ij};\boldsymbol{\theta}_j),$$

$$f_j(x_{ij}\,|\,\mathbf{p}_{ij};\boldsymbol{\theta}_j)=\frac{1}{\sqrt{2\pi\sigma_j^2}}\exp\left\{-\frac{(x_{ij}-\mu_{ij})^2}{2\sigma_j^2}\right\}$$

$$\mu_{ij}=m_1(p_{i1}^{(j)})+\Lambda+m_{q_j}(p_{iq_j}^{(j)})$$

$$=\sum_{k=1}^{q_j}\sum_{m=1}^{M_{jk}}\gamma_{mk}b_{mk}^{(j)}(p_{ik}^{(j)})$$

Criterion for selecting good networks

BNRC Score
Bayesian Network and Nonparametric Regression Criterion

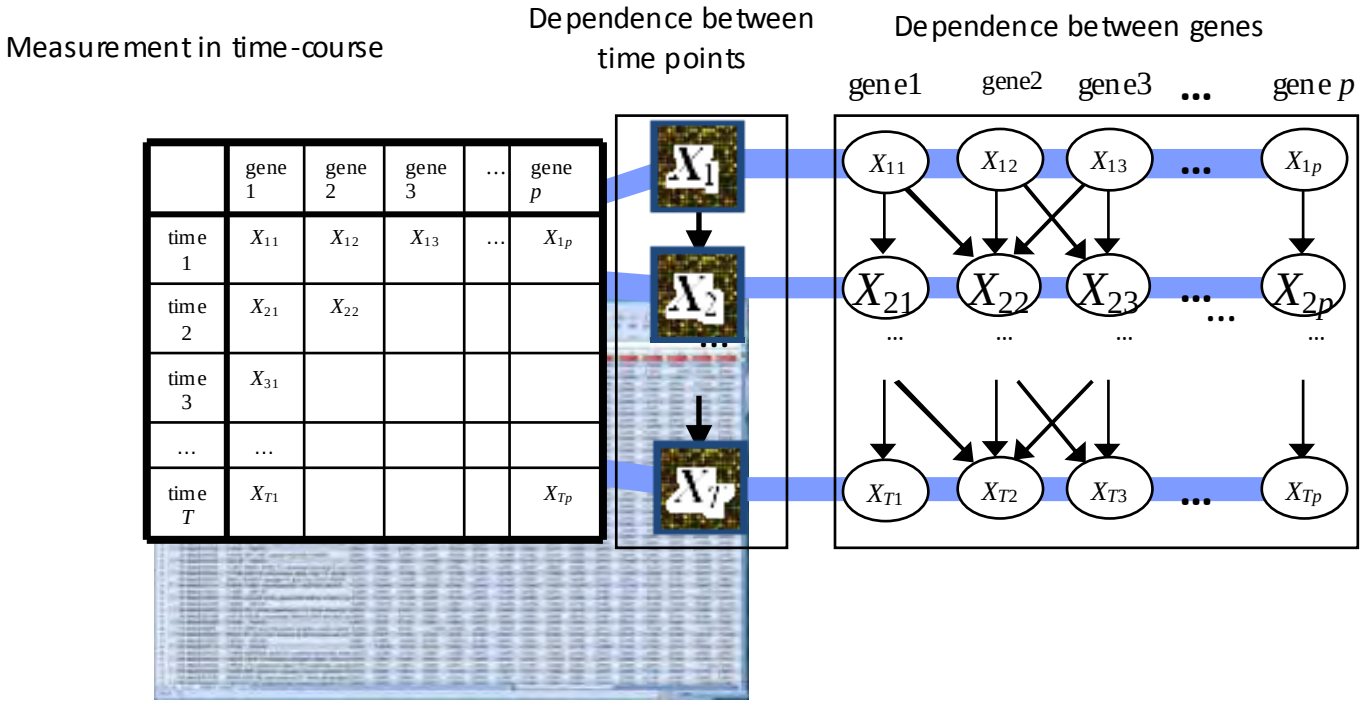
$$\mathrm{BNRC}(G) = -2\log\pi_G \int \prod_{i=1}^{n} f(\mathbf{x}_i;\boldsymbol{\theta}_G)\pi(\boldsymbol{\theta}_G \mid \boldsymbol{\lambda})d\boldsymbol{\theta}_G$$

$$= -2\log\pi_G - r\log(2\pi n^{-1})$$

$$+\log\left|J_\lambda(\hat{\boldsymbol{\theta}}_G)\right| - 2nl_\lambda(\hat{\boldsymbol{\theta}}_G \mid \mathbf{X}_n)$$

We choose the graph that minimizes the value of the BNRC score.

Dynamic Bayesian Network Model for Time-course Gene Expression Data



1. Imoto, S., Higuchi, T., Goto, T., Tashiro, K., Kuhara, S., Miyano, S. Combining microarrays and biological knowledge for estimating gene networks via Bayesian networks. *J. Bioinfomatics and Computational Biology*. 2(1):77-98, 2004.
2. Kim, S., Imoto, S., Miyano, S. Dynamic Bayesian network and nonparametric regression for nonlinear modeling of gene networks from time series gene expression data. Biosystems, 75(1-3), 57-65, 2004.

Computational Complexity of Searching Good Networks is Very High!

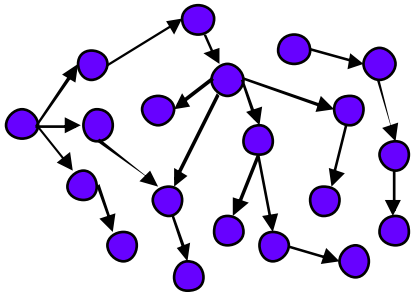
- Determining the optimal Bayesian network is computationally intractable (NP-hard)

- 2.34×10^{72} possible networks for 20 genes
- 2.71×10^{158} possible networks for 30 genes
- 1.21×10^{15} possible networks for 9 genes

A brute force approach would take years of computation time even on a supercomputer.

Optimal Gene Networks are Hard to Find

- Optimal networks can be found for 30 genes with SUN Fire 15K (100CPU) supercomputer in a day.



•Finding Optimal Models for Small Gene Networks. Ott, S., Imoto, S., Miyano, S. Pacific Symposium on Biocomputing, 9: 557-567, 2004.
•Ott, S., Miyano, S. Finding optimal gene networks using biological constraints. Genome Informatics. 14:124-133, 2003.
•Ott, S., Hansen, A., Kim, S.-Y., and Miyano, S. Superiority of network motifs over optimal networks and an application to the revelation of gene network evolution. Bioinformatics. 21(2):227-238, 2005.

Supercomputer System (2003-2008)

The Computational Center for Genome Research

- **Renewed in January 2003**

  HITACHI HA8000, 8xSunFire 15K, 2xSunFire 6800, SGI Origin3900T

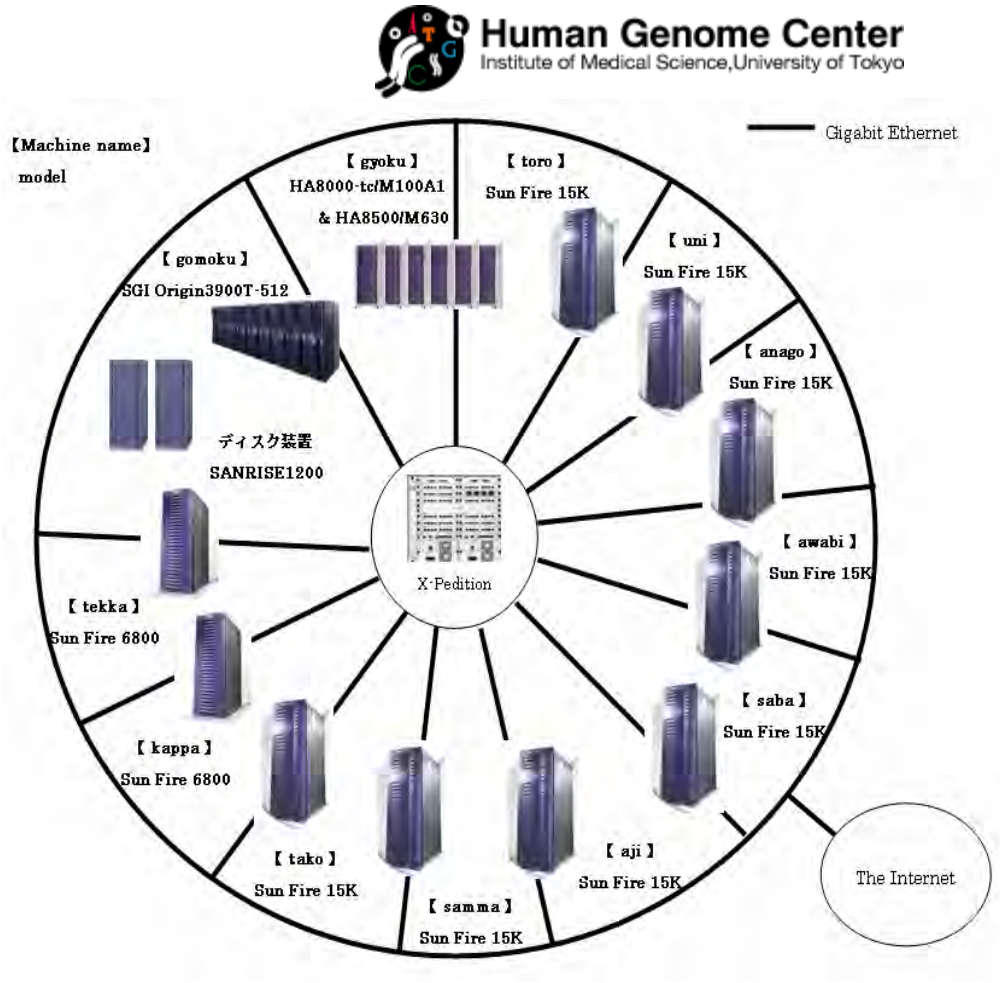
  1,428 CPUs, 145 TB
- **Budget:**

  100,000,000JPY/year for 6 Year Lease, 80,000,000JPY for electricity/year
- **All Japan Users: 500**

  75% from U. Tokyo,

  25% from Others

  50 very intensive users



Strategic Computational Initiative

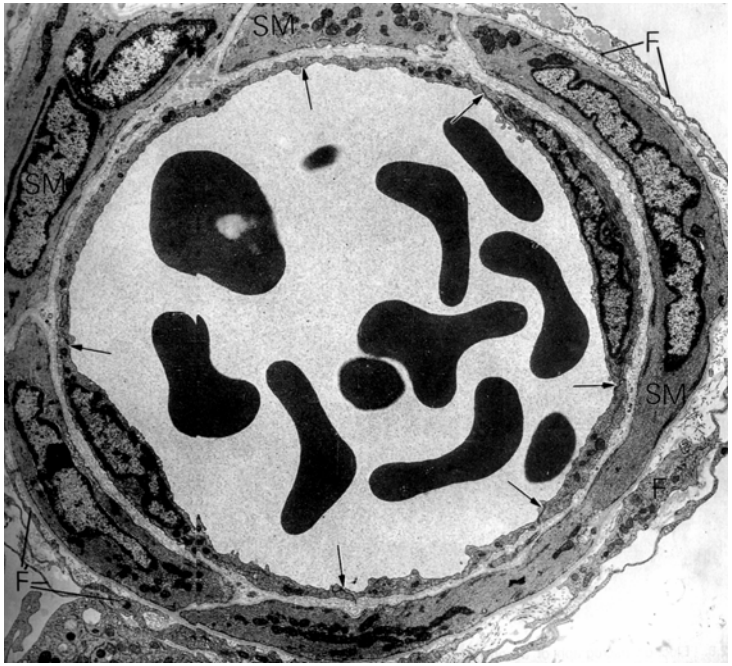
Next Supercomputer System for 2009-2014

Human Genome Center
Institute of Medical Science, University of Tokyo

Renewed in January 2009

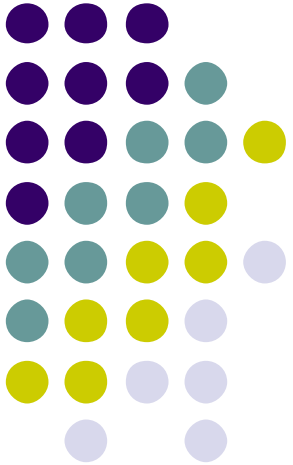
- January 2009: 75 TFLOPS at peak & 1 PB Disk Space
  - PC Cluster (Sun Microsystems)
  - Large Shared Memory Machine (SGI Altix)
  - Lustre File System (Sun Microsystems)
- January 2011: 225 TFLOPS at peak & 4PB Disk Space

Mining Gene Networks in Human Umbilical Vein Endothelial Cell (HUVEC)

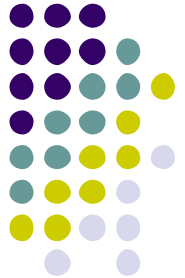


Courtery by Cristin Print, University of Auckland

Search for Drug Target Pathways



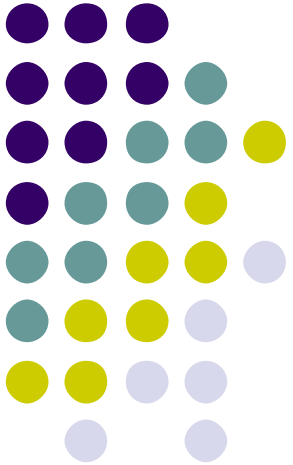
Endothelial Cells (EC) play key roles in disease



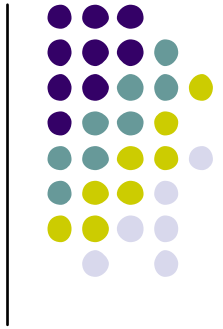
- Vessel growth (angiogenesis)
- Vessel regression (apoptosis)
  - Cancer
  - Cardiovascular disease
  - etc.
- Inflammation
  - Atherosclerosis
  - Vasculitis
  - etc.

First Case
HUVEC Gene Networks

Searching Drug Target Pathways
Using Fenofibrate



HUVEC treated with Fenofibrate

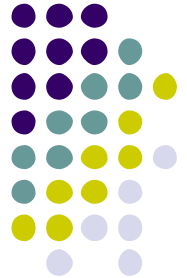


- Fenofibrate is:
  - Agonist of **PPARα**
  - Drug for ***disorder of lipid metabolism (hyperlipidaemia)***

- Our aim is to:

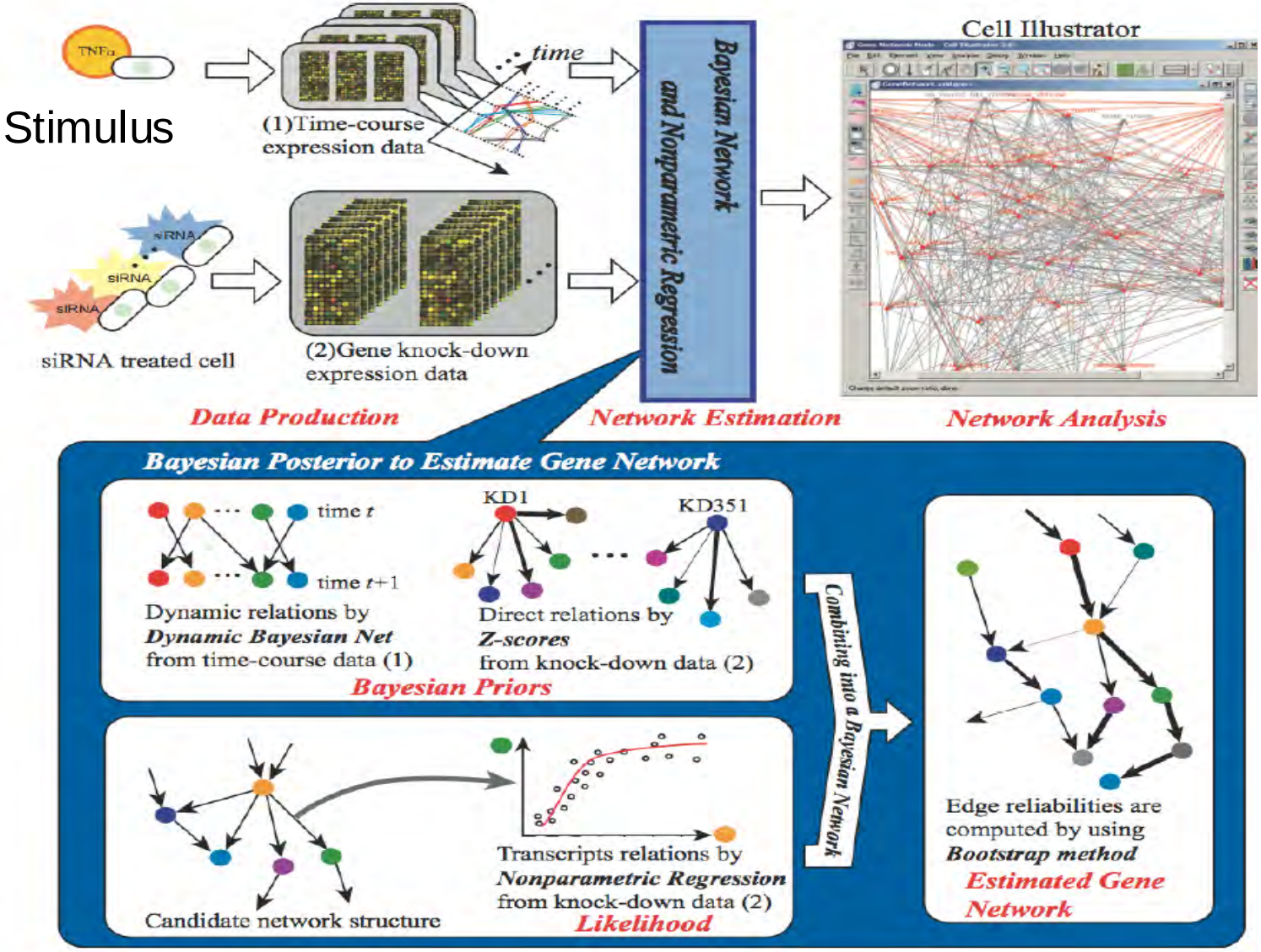
  Elucidate fenofibrate-related gene network based on
  - 25 μM fenofibrate dosed
  - Time-course response arrays against fenofibrate (six time points (0, 2, 4, 6, 8 and 18 hours) in duplicate)
  - 270 gene knock-down arrays by siRNA

Selection of Genes for Knock-Down



- 351 transcription factors, signaling molecules, receptors and ligands were selected based on knowledge of their relevance to transcriptional regulation in EC.

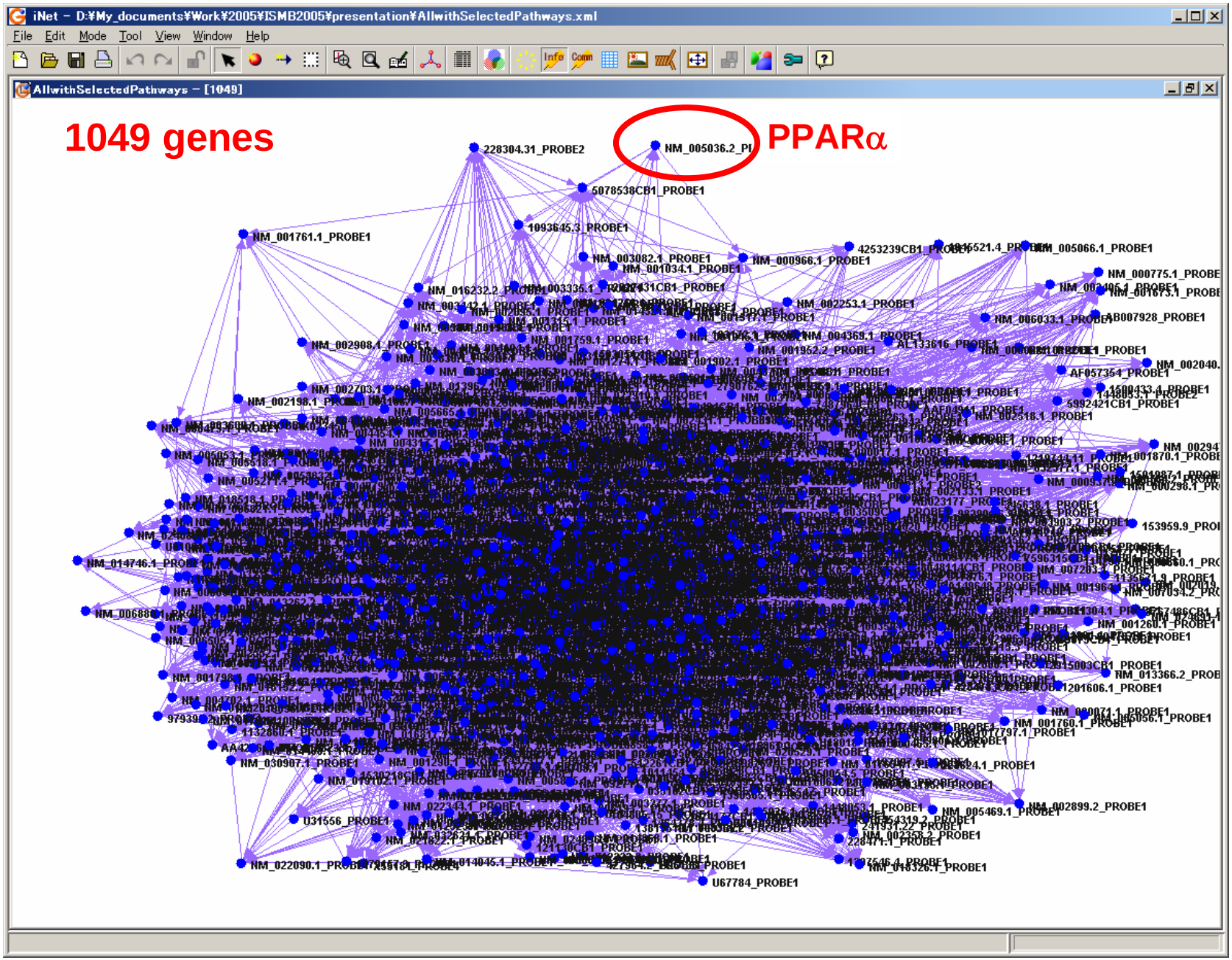
Computational Strategy



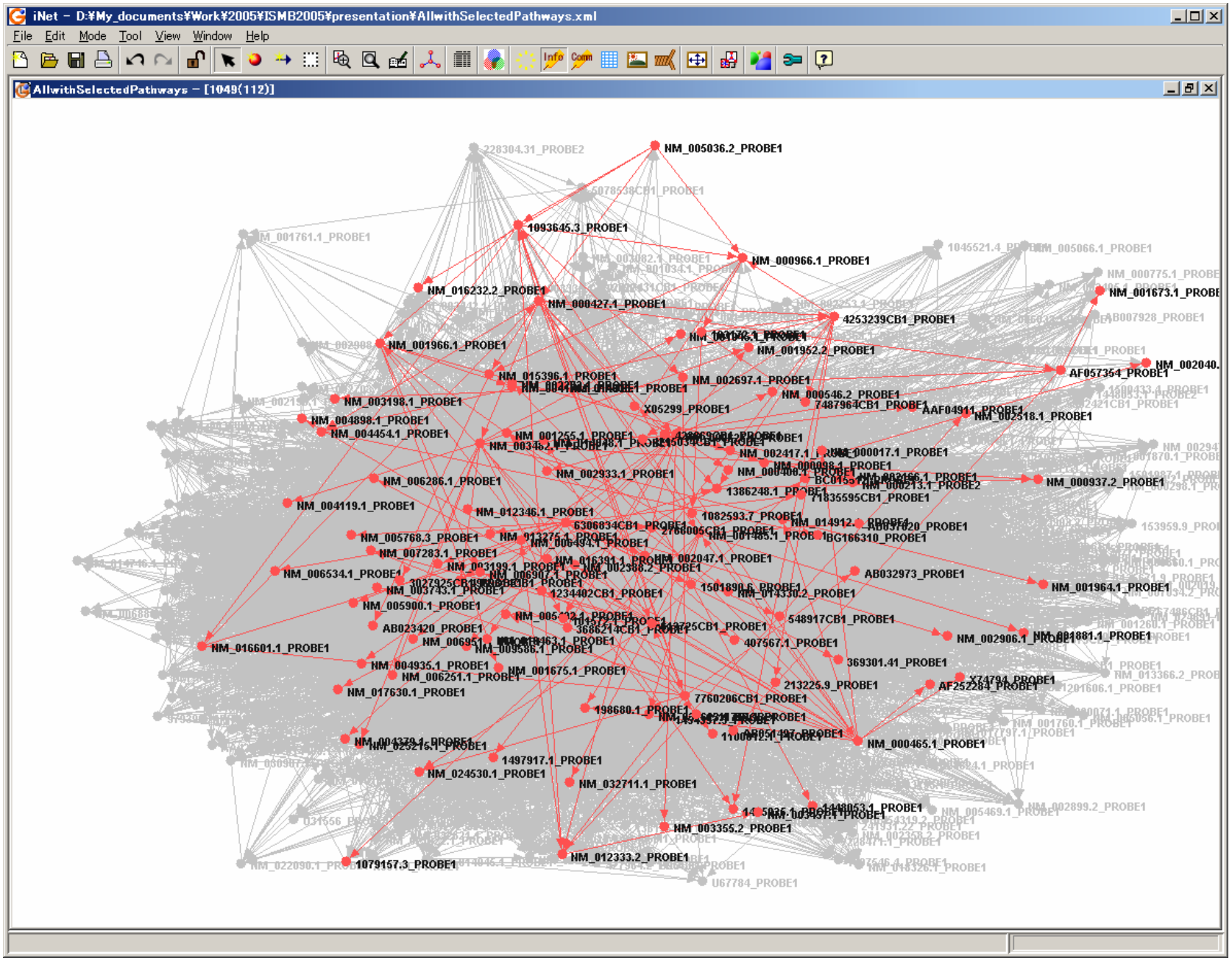
Computational Strategy

- Imoto S, Tamada Y, Araki H, Yasuda K, Print CG, Charnock-Jones SD, Sanders D, Savoie CJ, Tashiro K, Kuhara S, Miyano S. Computational strategy for discovering druggable gene networks from genome-wide RNA expression profiles. Pacific Symposium on Biocomputing, 11, 559-571, 2006.

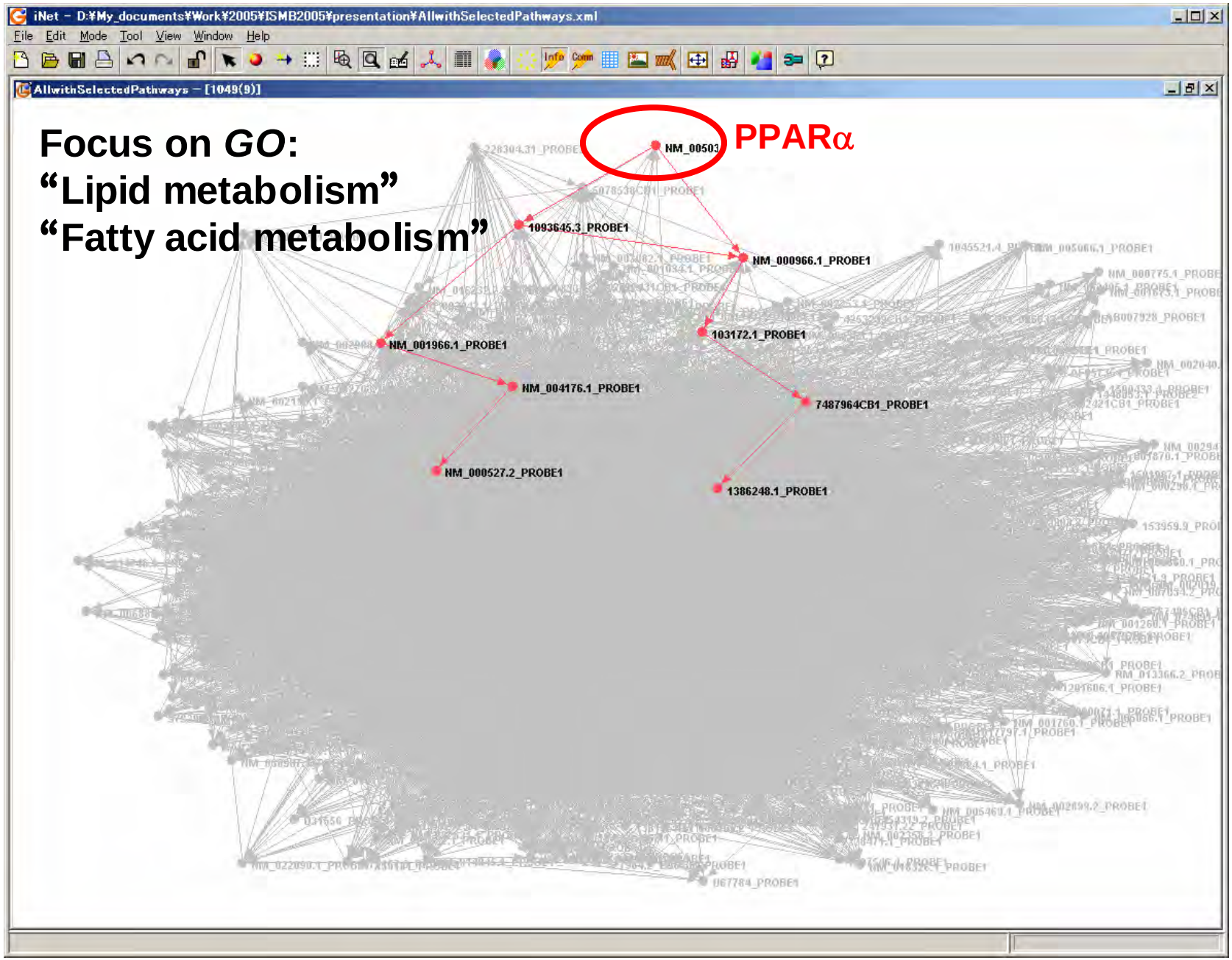
Estimated Feno-related Network

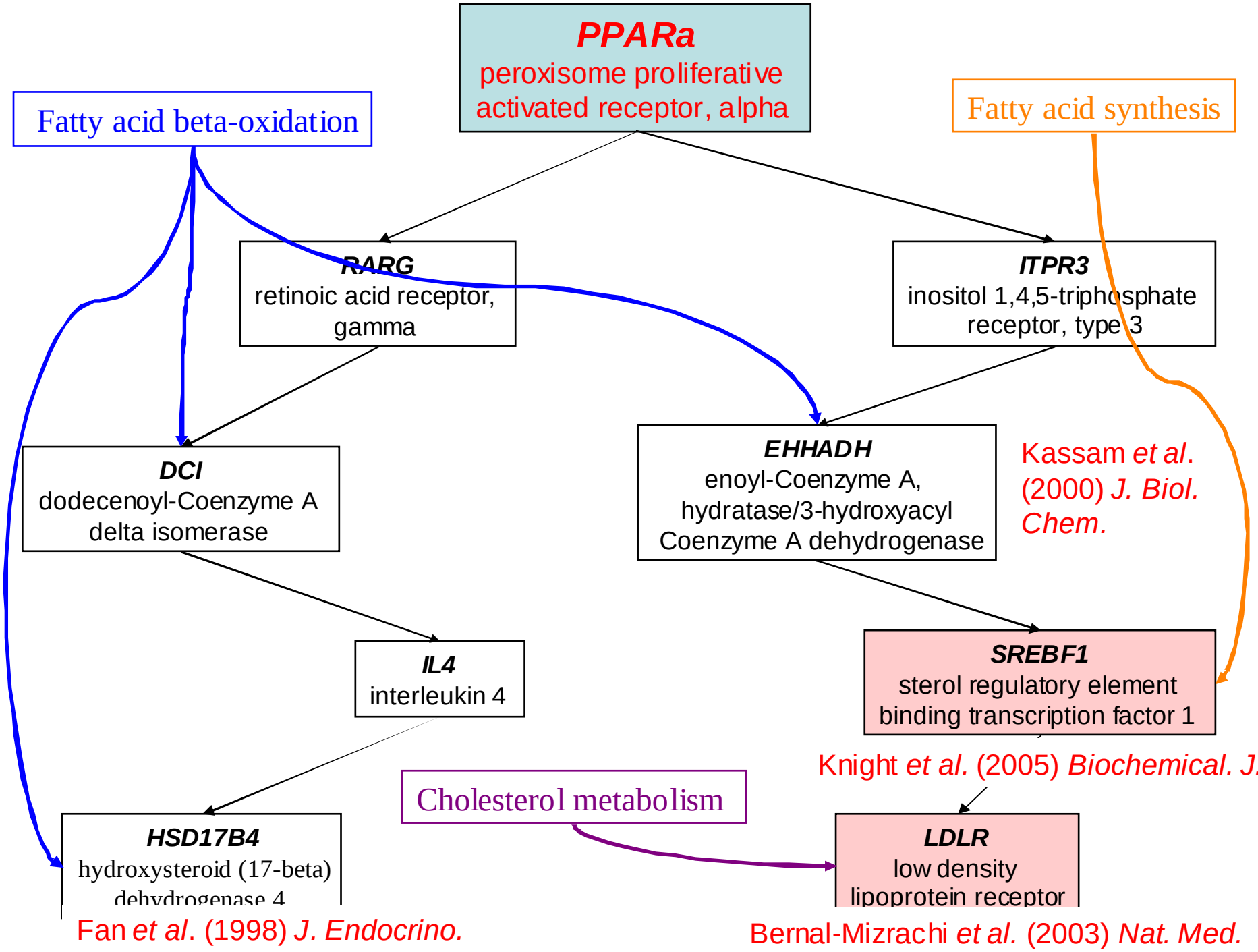


Downstream of PPARα



Evaluation (An Example)




PPARa
peroxisome proliferative activated receptor, alpha
Fatty acid beta-oxidation
Fatty acid synthesis
RARG
retinoic acid receptor, gamma
ITPR3
inositol 1,4,5-triphosphate receptor, type 3
DCI
dodecenoyl-Coenzyme A delta isomerase
EHHADH
enoyl-Coenzyme A, hydratase/3-hydroxyacyl Coenzyme A dehydrogenase
Kassam et al. (2000) J. Biol. Chem.
IL4
interleukin 4
SREBF1
sterol regulatory element binding transcription factor 1
Knight et al. (2005) Biochemical. J.
Cholesterol metabolism
HSD17B4
hydroxysteroid (17-beta) dehydrogenase 4
LDLR
low density lipoprotein receptor
Fan et al. (1998) J. Endocrino.
Bernal-Mizrachi et al. (2003) Nat. Med.

Druggable Gene Network?

- **17 (out of 42) lipid metabolism genes have more children than PPARα (listed in the Table below).**
- **Some of listed genes in the Table have**

Children#	GeneTitle	GeneName	Druggable	Description
30	Gene1	Gene1		
22	Gene2	Gene2		
20	Gene3	Gene3		
16	LIPG	lipase, endothelial	Druggable	NatGenet 21:424-8
15	Gene4	Gene4		
13	Gene5	Gene5		
13	HMGCR	3-hydroxy-3-methylglutaryl-Coenzyme A reductase	Druggable	Many Company's Targets
11	Gene6	Gene6		
10	Gene7	Gene7	Druggable	
9	Gene8	Gene8		
9	Gene9	Gene9		
9	LSS	lanosterol synthase (2,3-oxidosqualene-lanosterol cyclase)	Druggable	Roche's Target
8	AKR1C3	aldo-keto reductase family 1, member C3	Druggable	CancerRes 63:505-512
7	PTGS2(COX2)	prostaglandin-endoperoxide synthase 2	Druggable	Many Company's Targets
7	Gene10	Gene10		
7	PPARA	peroxisome proliferative activated receptor, alpha	Druggable	Fournier Pharma's Target

***Druggable*: Nat. Rev. Drug Discov. 1:727-30, 2002**

In Silico Search of Drug Target Pathways with Gene Network Computation

HUVEC

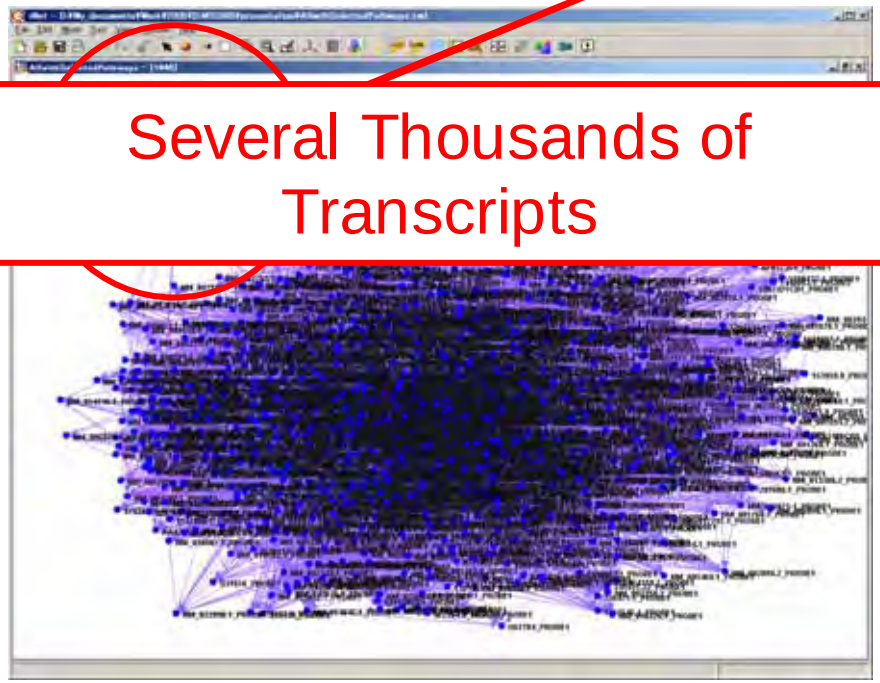
With PETA FLOPS

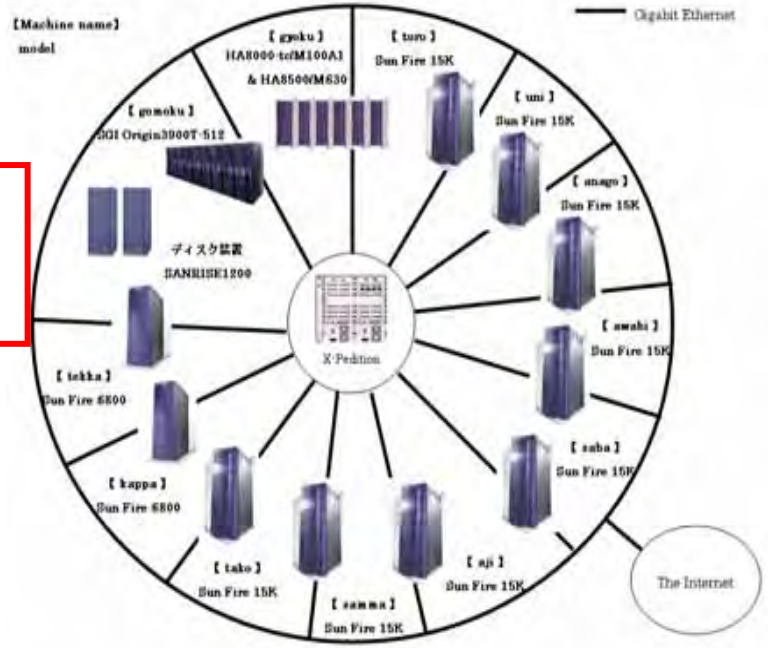
Compution of gene network of 1000 genes affected by Fenofibrate

Current Supercomputer (HGC)

Only 3% of Human Genes

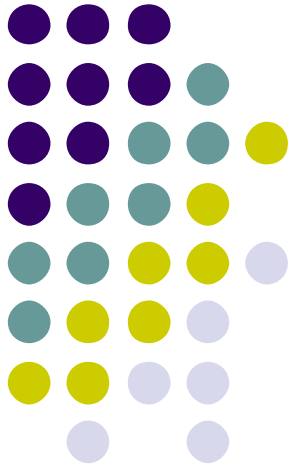
Several Thousands of Transcripts



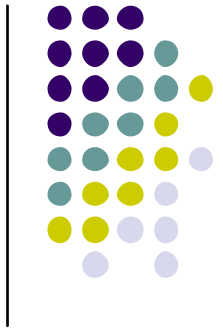


Second Case
HUVEC Gene Networks

TNF-α and New Hub Genes Regulating Inflammation and Apoptosis



HUVEC treated with TNF-α



Tumor Necrosis Factor (TNF)-α

EC regulates

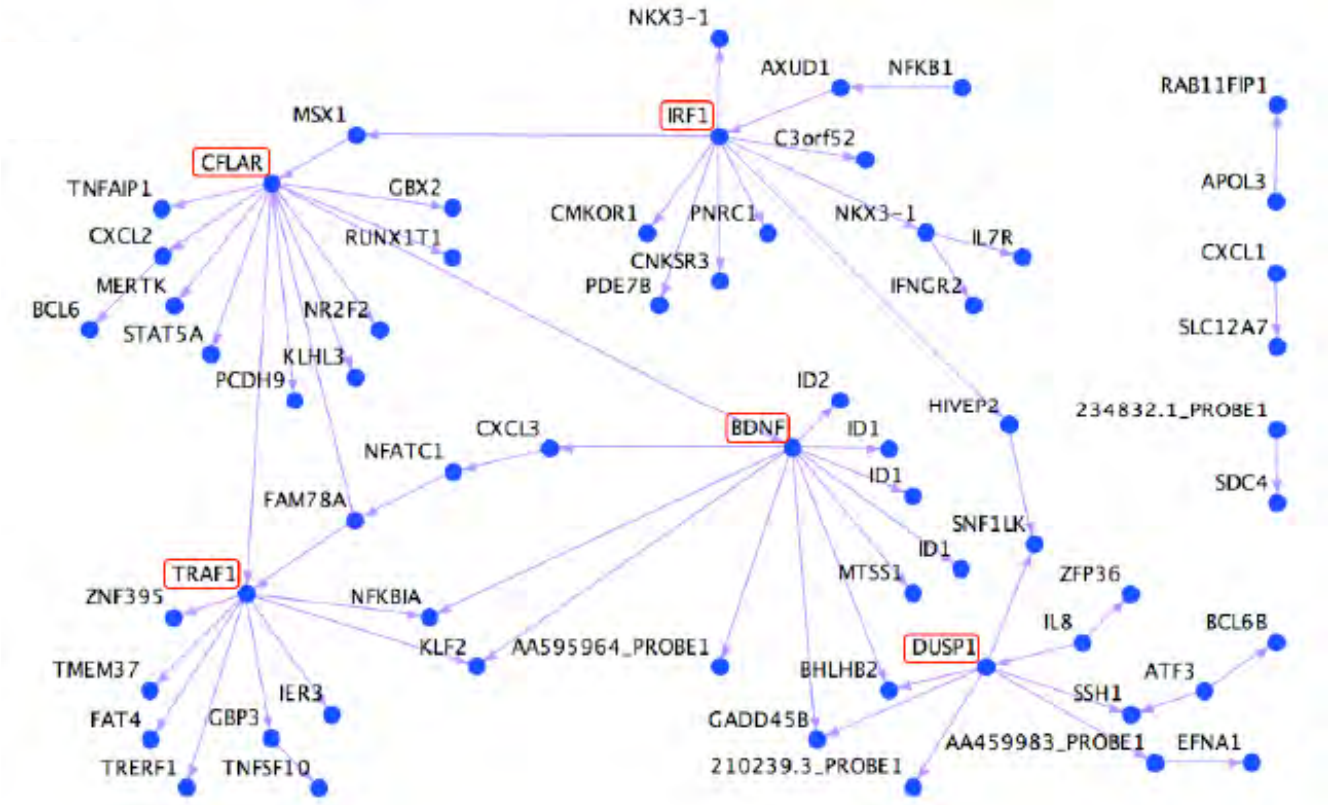
- the entry of leucocytes into damaged tissues and their activation
- blood vessel structure by their coordinated morphogenesis into vessels
- Vessel regression (appoptosis)

EC functions are influenced by TNF-α

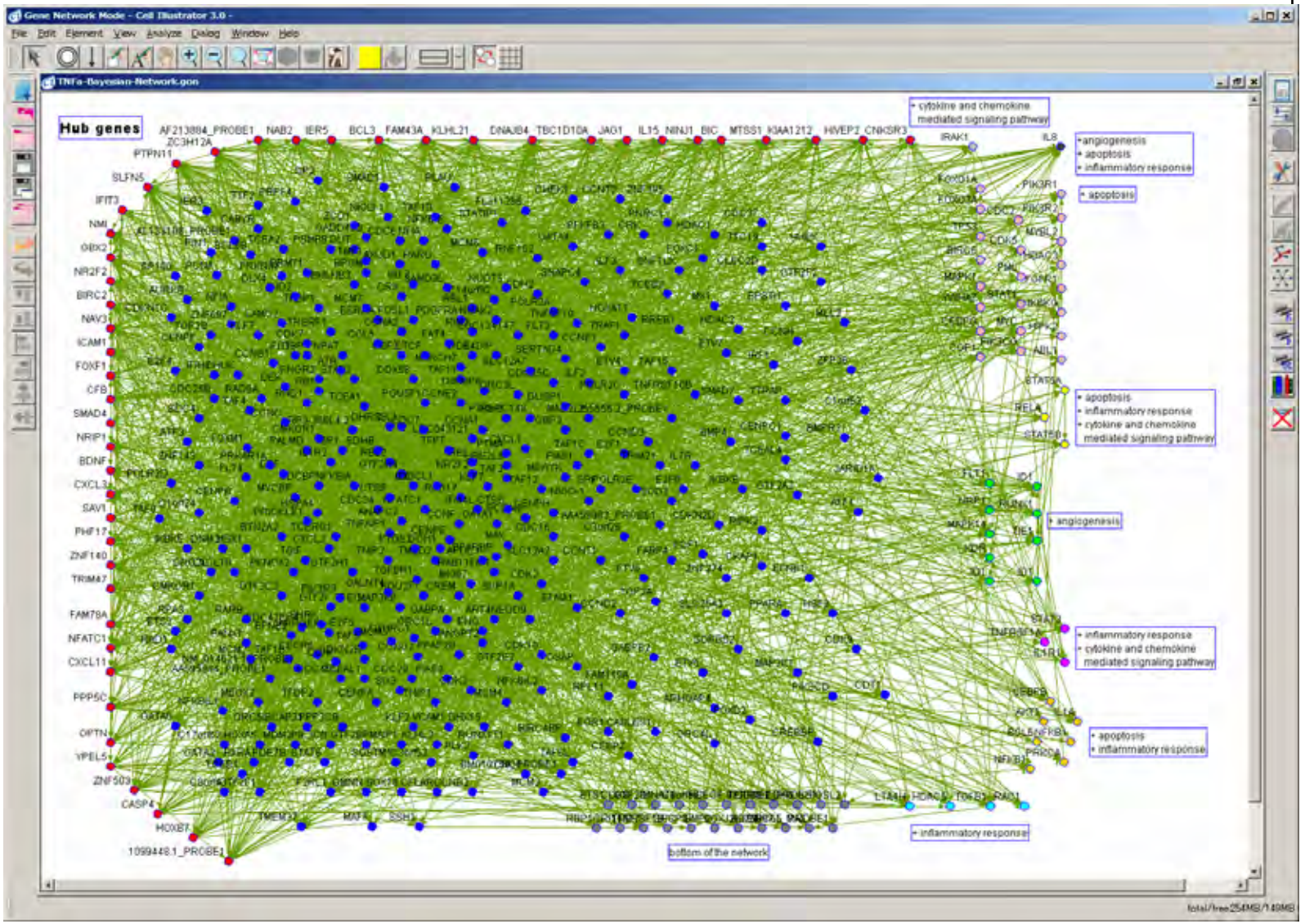
Elucidate TNF-α stimutaed gene network

- Stimulation with TNF-a (10ng/mL)
- Time-course response arrays
- 8 time points (0, 1, 1.5, 2, 3, 4, 5, 6) in triplicate
- 351 gene knock-down arrays by siRNA

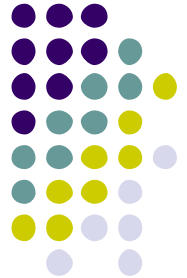
Dynamic Bayesian Network with Nonparametric Regression found five hubs all of which are known to play key roles in TNF-α related EC processes.



TNF-α Network computed from microarray data of 351 siRNA knock-downs



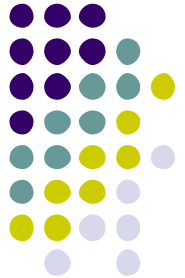
Evaluation of TNF-α Networks and Discoveries



(1) **Many of the topological hubs in the network are already known to occupy key positions in signaling cascades that ultimately control transcription.**

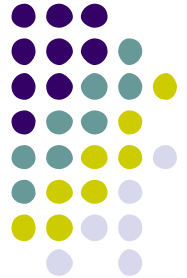
(2) **Literature analysis of ten networks topological hubs (with more children)**

Discovery: Gene X is a key role in inflammation regulated in EC



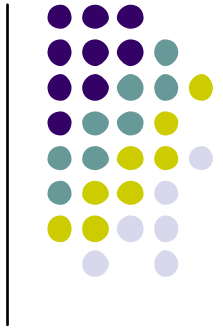
1. X has 38 children in our network
2. Knocking down X and analyzed EC secretion of five chemokines using cytometric bead arrays.
3. It was proved that gene X plays a key role in inflammation EC.

Discovery: Gene Y regulates TNF-induced appoptosis



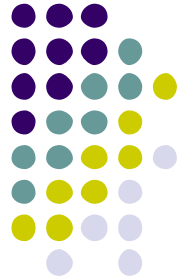
1. **Y has 20 children in our network**
2. **Knocking down Y and analyzed EC with/without TNF-α.**
3. **Y modulates the effect of TNF-α on EC apoptosis pathway.**

Summary of Discovery



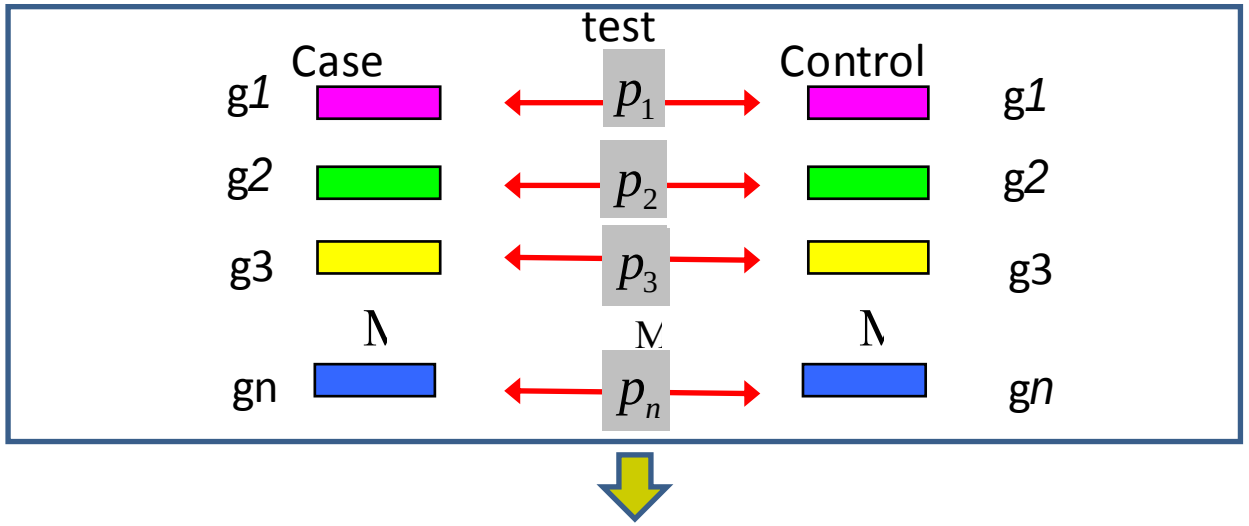
- The network model predicted known regulatory hubs and previously uncharacterized hubs, which our experiments confirmed were regulators of EC apoptosis and inflammation.

Literature (IPA) and Our Network



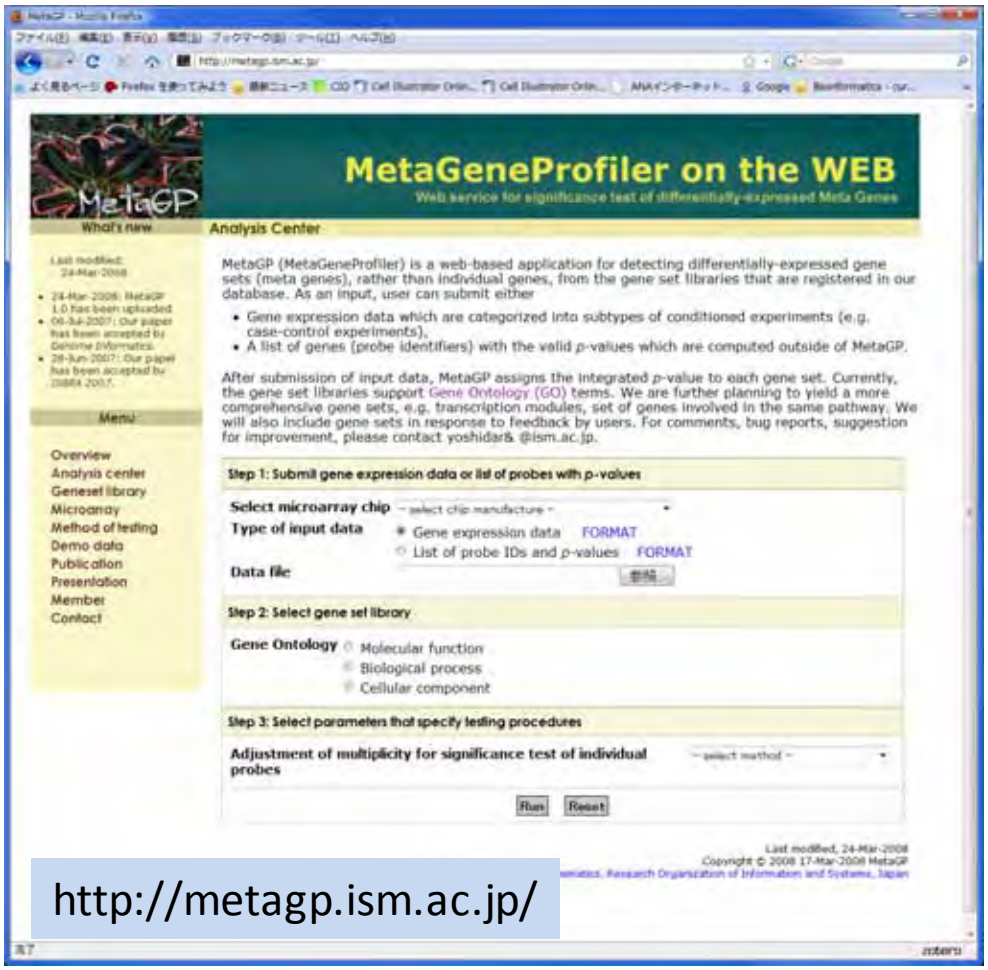
- We found that transcript-to-transcript relationships predicted by the published literature (IPA) did not correlate well with those found within our data.
- It suggests that lineage-specific data sets may be very important for systems biology.

Mining Pathways from Case-Control Analysis Microarray Data



Can we see the difference of the systems?

Meta Gene Profiler (MetaGP)



http://metagp.ism.ac.jp/

MetaGP is a statistical test for detecting differentially-expressed gene sets (meta genes), rather than individual genes, from the gene set libraries (e.g., **pathways**, **GO** terms, etc.).

$$p_{Integrated} = f_{MetaGP}(p_1, \Lambda\ , p_n)$$

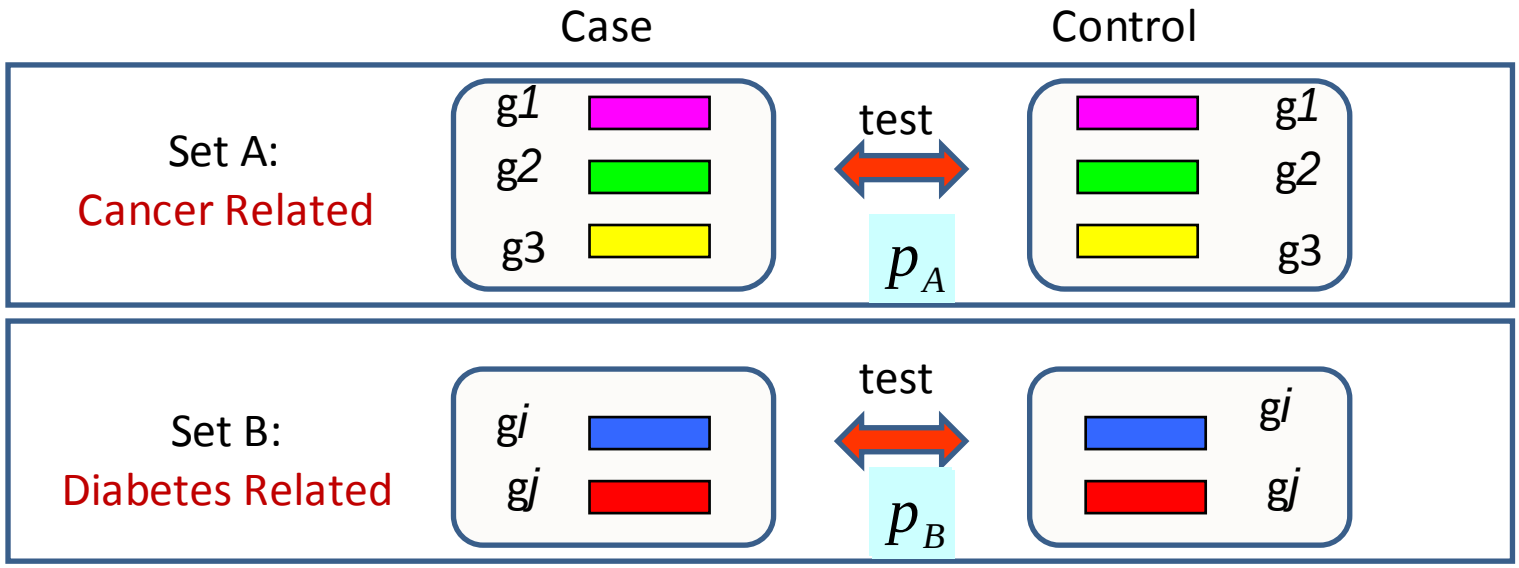
p_i : the p-value of the *i*th gene in the gene set

Gupta, P.K., R. Yoshida, S. Imoto, R. Yamaguchi, and S. Miyano, Statistical absolute evaluation of gene ontology terms with gene expression data, LNBI, 4464: 146-157, 2007.

Test for a Set of Genes

Secondly analysis:
test for a **set of genes** with the same functional annotation

Functional Annotations: **Pathways**, **Gene Ontology**, etc.



Obtain *p*-values for the sets of genes with **Meta Gene Profiler**

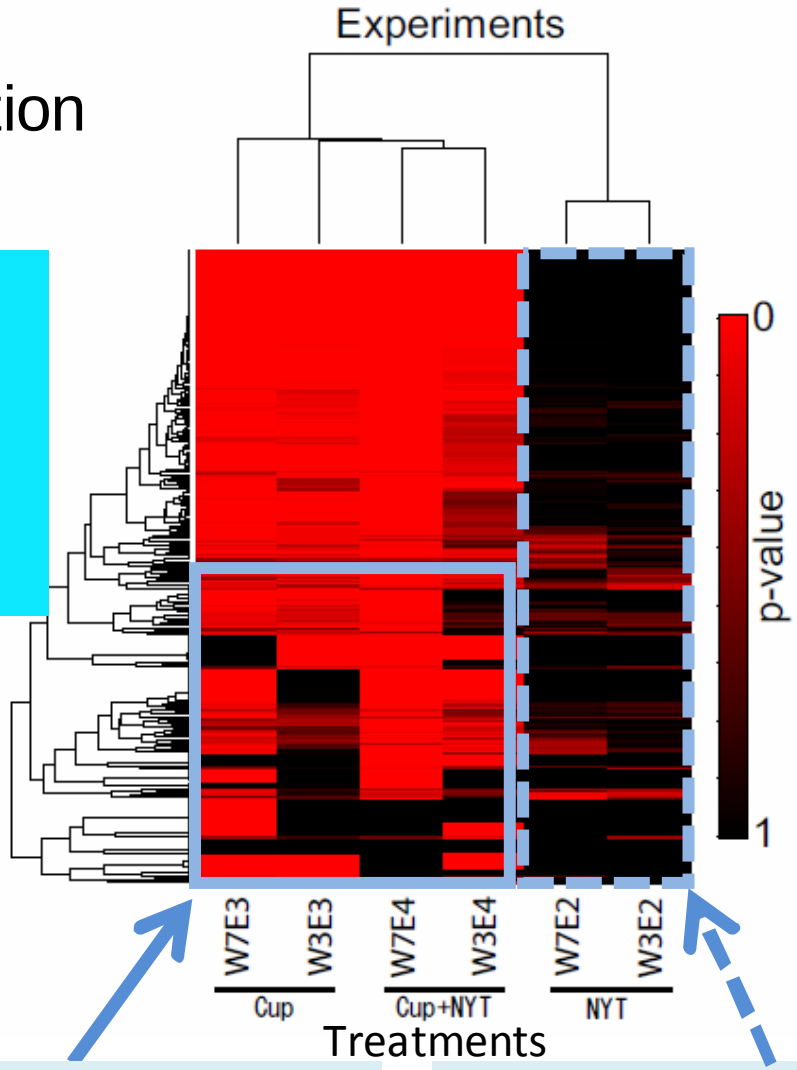
Higher interpretability

Ninjin'yoeito (NYT) for remedying degraded myelin sheath of nerves

人参養栄湯: NYT
Cuprizone (CUP) for demyelination

投与条件の異なる二群比較に基づく各転写因子(306)の活性度変化の
MetaGP test
(多発性硬化症マウス)

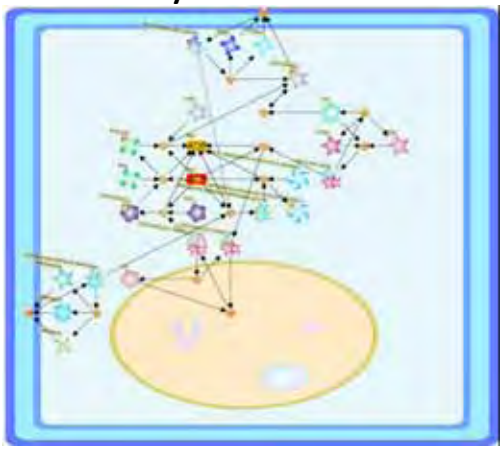
MetaGP analysis of gene expression data from CUP/NYT treated-mice using 306 TFs and their binding gene sets



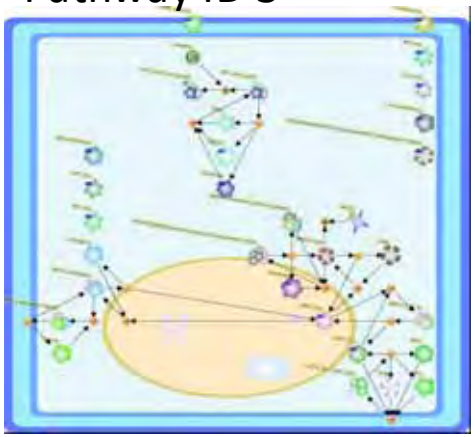
Yamaguchi, R., Yamamoto, M., Imoto, S., Nagasaki, M., Yoshida, R., Tsujii, K., Ishiga, A., Asou, H., Watanabe, K., Miyano, S. Identification of activated transcription factors from microarray gene expression data of Kampo-medicine treat mice. Genome Informatics. 18, 119-129, 2007.

MetaGP with BIOBASE TRANSPATH Database

Pathway ID 1

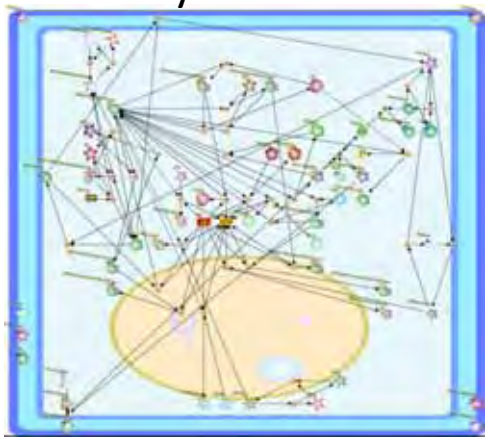


Pathway ID 3



Pathway ID 2

Pathway ID 4

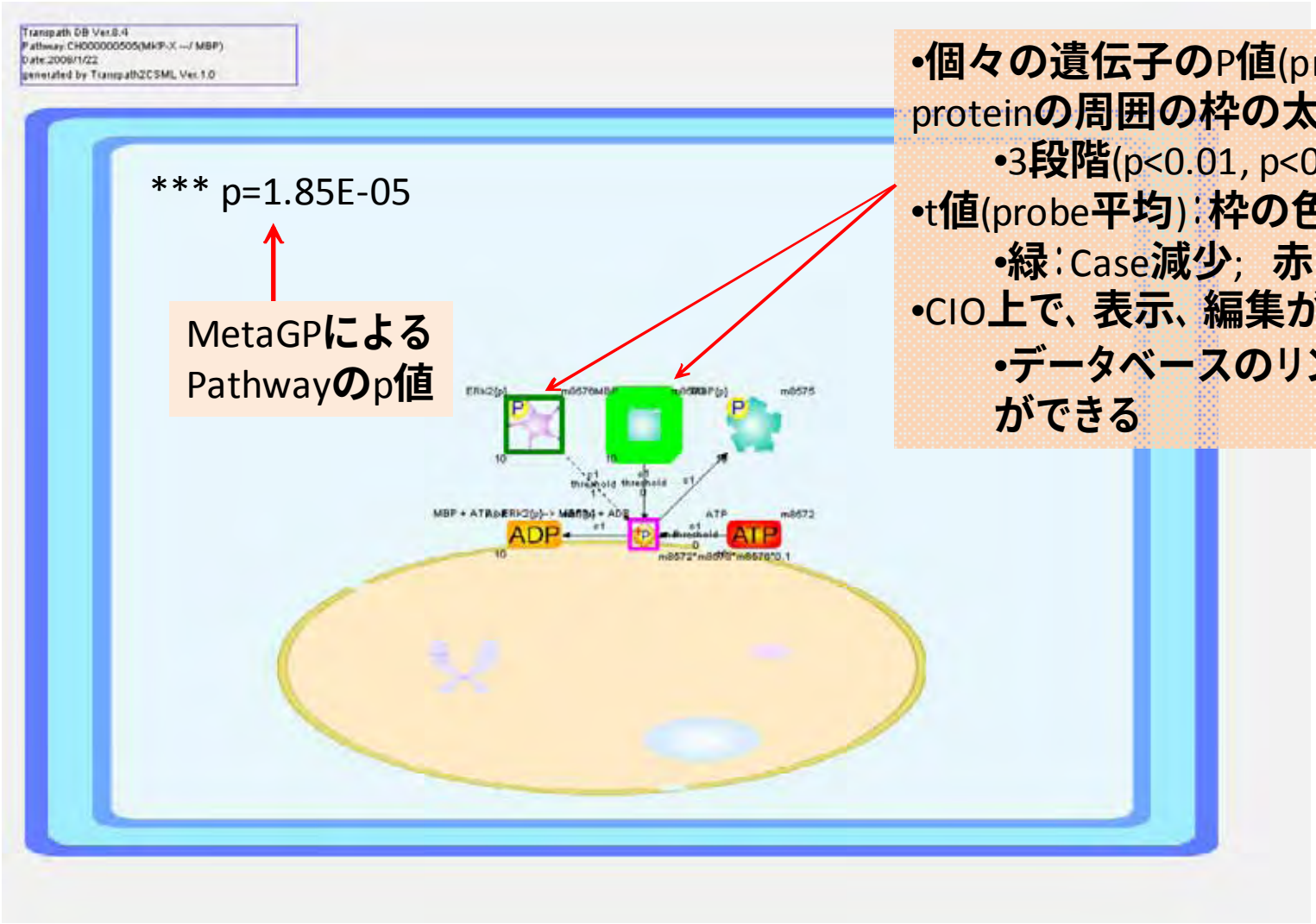


819 pathways are screened by Cell Illustrator Online +TRANSPATH

Cell Illustrator Online Analysis

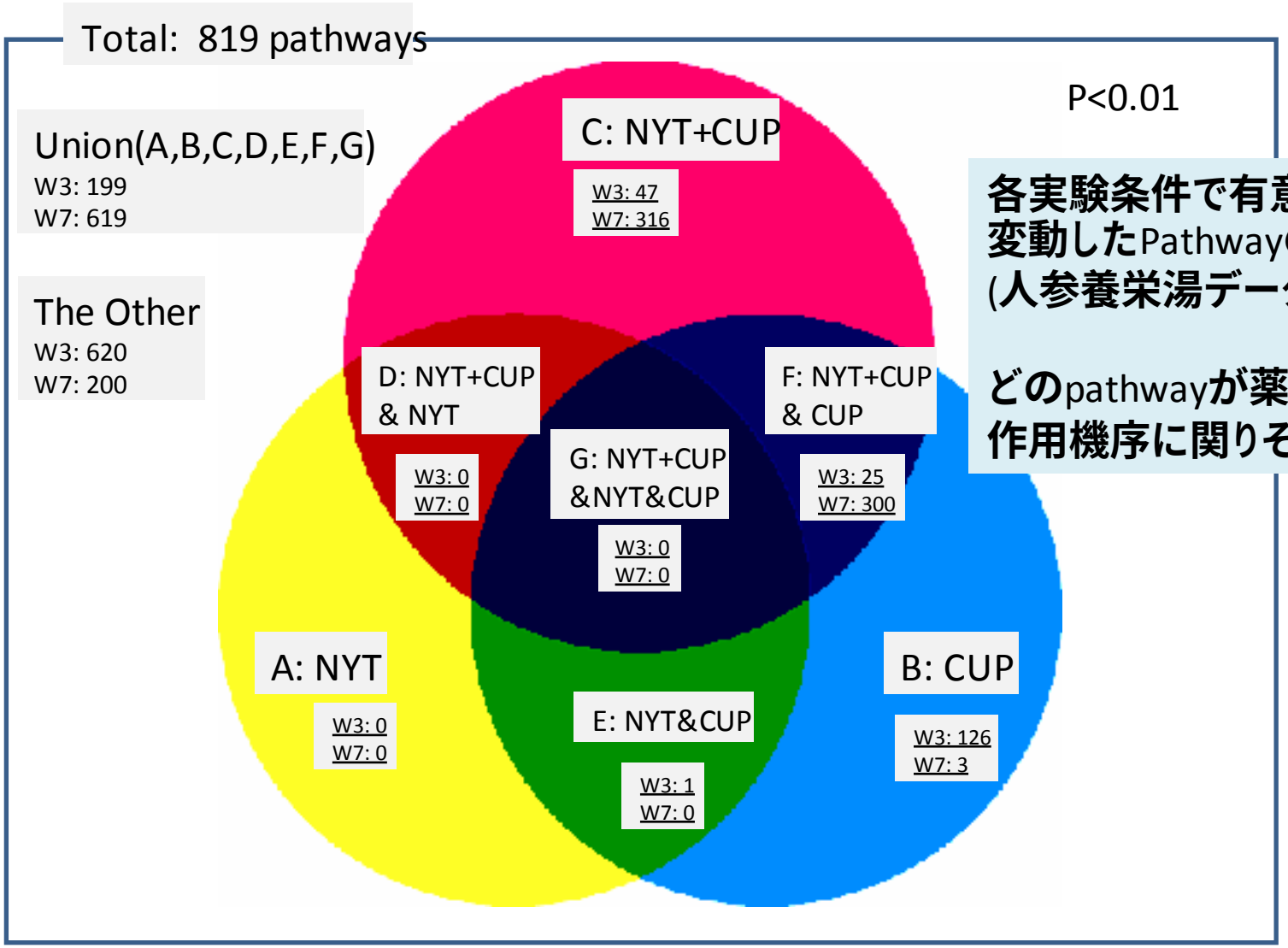
- Pathway ID: CH000000505
- Pathway Name: MKP-X ---/ MBP
- ERK2 とMBPのパスウェイ
- **下の図は7週目(EXP4)の結果**

Week	Ven Diag	P-Exp2	P-Exp3	P-Exp4
3rd	F	0.896887	1.09E-08	1.32E-09
7th	C	0.989038	0.120053	1.85E-05



- **個々の遺伝子のP値(probe平均)：proteinの周囲の枠の太さで表現**
  - **3段階(p<0.01, p<0.05, p>0.05)**
- **t値(probe平均)：枠の色で表現**
  - **緑：Case減少； 赤：Case増加**
- **CIO上で、表示、編集が可能**
  - **データベースのリンク情報も見ることができる**

Sets of Significant Pathways (p<0.01)



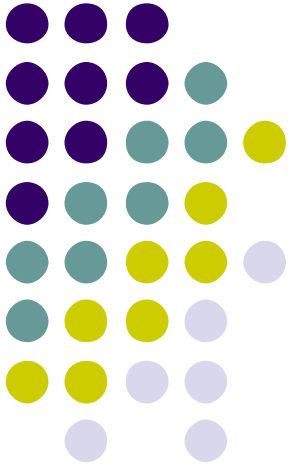
各実験条件で有意に変動したPathwayのベン図(人参養栄湯データ)

どのpathwayが薬の作用機序に関りそうかを示唆

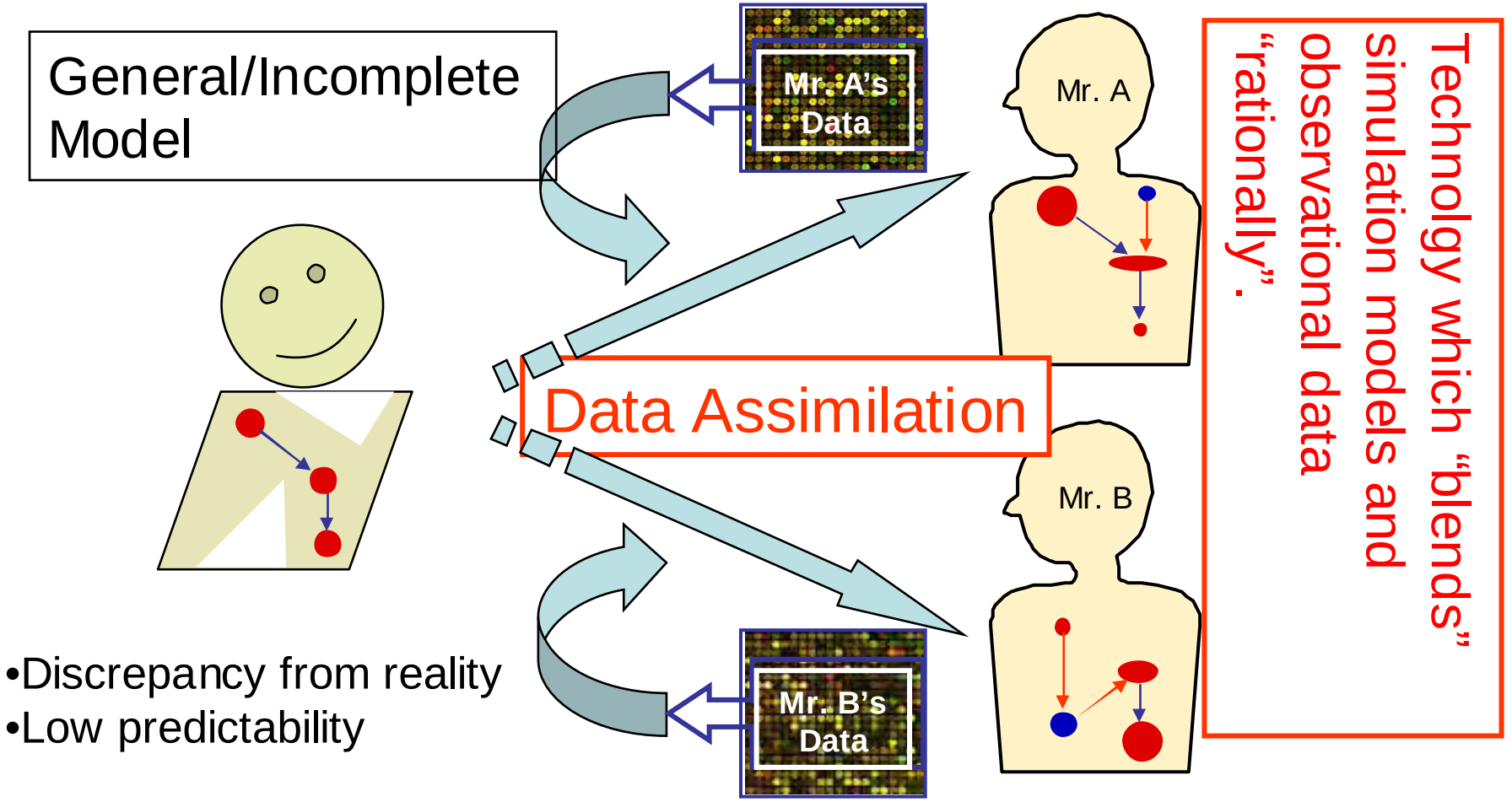
Data Assimilation for Biological Systems

Technology which “blends” simulation models and observational data “rationally”.

Peta FLOPS Computing for Biomedical Research Applications



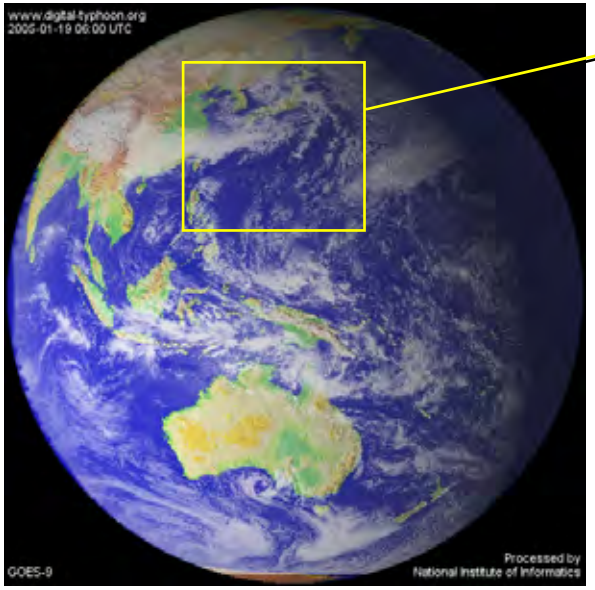
Application of Data Assimilation Technology



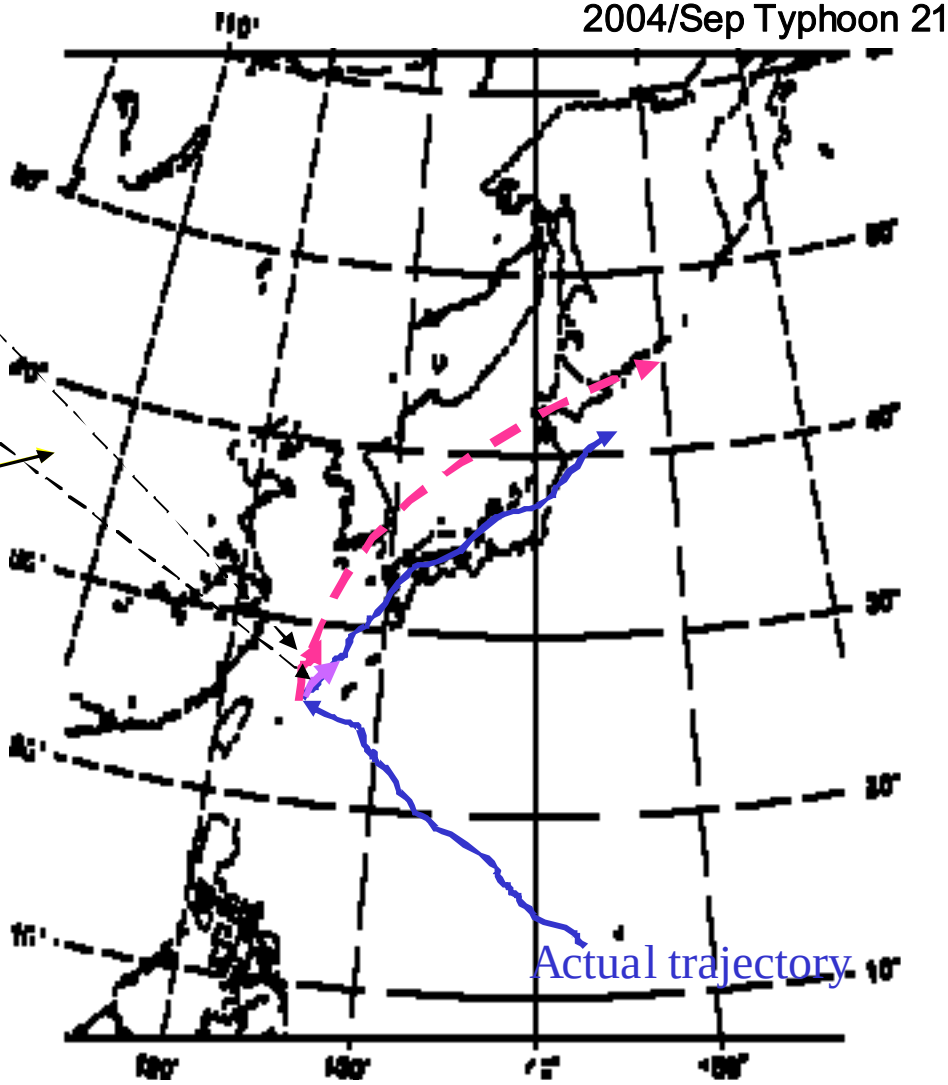
Prediction of Typhoon Trajectory

→ : by Simulation only.

→ : by Data Assimilation

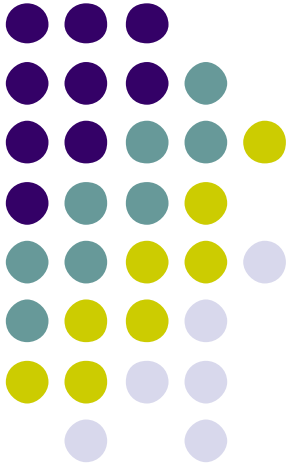


(Taken from NII, Japan)

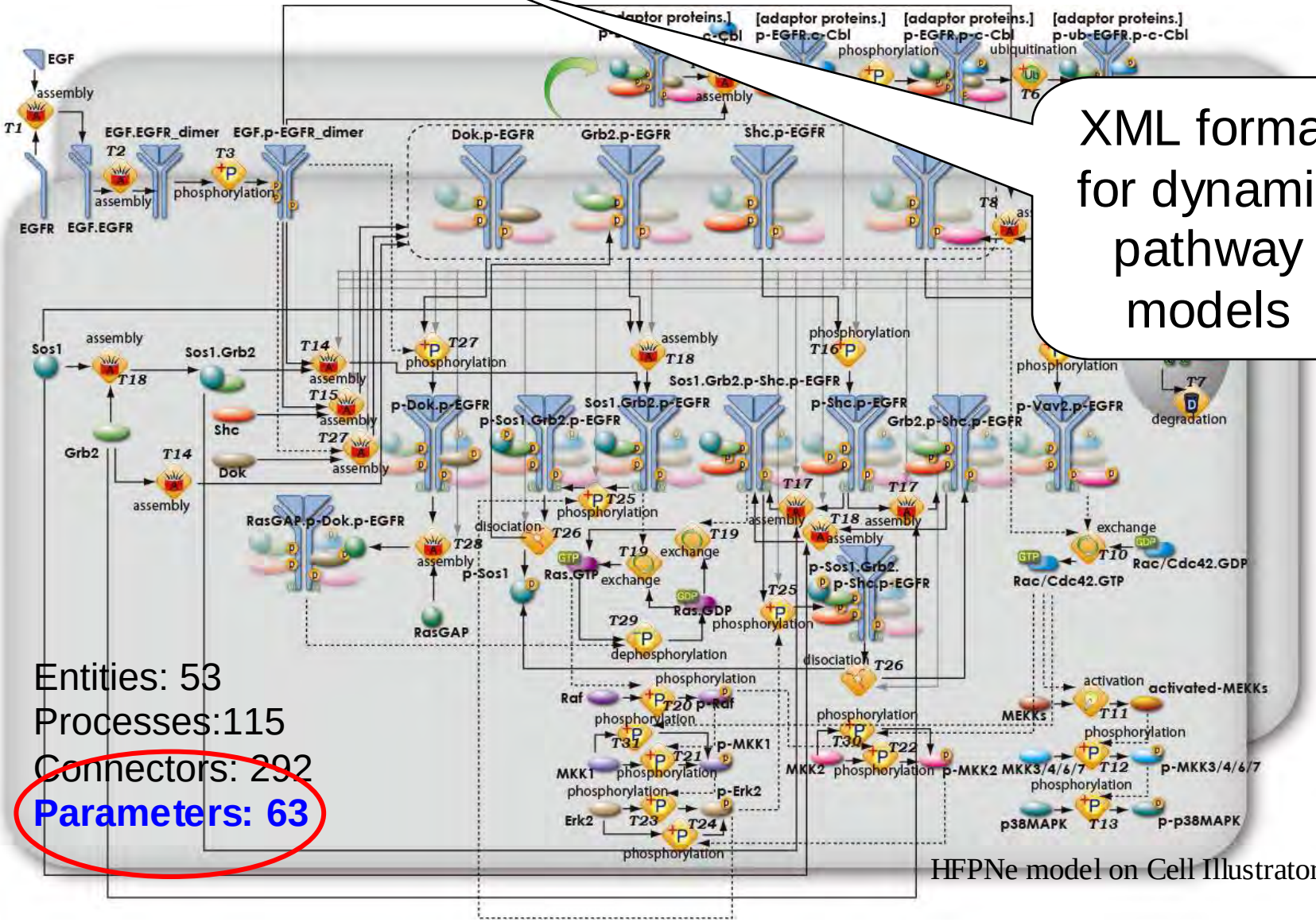


A First Step

Data Assimilation of EGF Receptor Pathway Dynamic Model and SILAC Proteome Time-Course Data



EGF Receptor Pathway Dynamic Model in CSML using Cell Illustrator



List of Main Processes

Table 1: Biological facts obtained from the literature and assigned to processes in the HFPNe model in Figure 2. #1: Corresponding processes in the HFPNe. #2: Corresponding sub-pathways in Figure 2.

Biological phenomena from experimental data in the literature	#1	#2	Type of biological process	Reference
Binding of EGF to EGFR induces the dimerization of the receptors resulting in autophosphorylation of the receptors.	***T1***	***i***	Association	[29]
	T2	***i***	Association	
	T3	***i***	Phosphorylation	
c-Cbl binds to the tyrosine-phosphorylated EGFR and simultaneously c-Cbl is tyrosine phosphorylated.	***T4***	***i***	Association	[43]
	T5	***i***	Phosphorylation	
c-Cbl catalyses ubiquitination of EGFR.	***T6***	***i***	Ubiquitination	
The ubiquitinated EGFR is degraded by the proteasome/lysosome.	***T7***	***i***	Degradation	
Vav2 binds to tyrosine-phosphorylated EGFR via its SH2 domain and is tyrosine phosphorylated by EGFR.	***T8***	***ii***	Association	[24][33][42]
	T9	***ii***	Phosphorylation	
Vav2 activates Rac1 by promoting the exchange of bound GDP for GTP.	***T10***	***ii***	Exchange	[7]
Activated Rac/Cdc42 induces activation of MEKKs. We modeled one MEKK as representative of MEKKs that mediate p38MAPK phosphorylation.	***T11***	***ii***	Activation	[10][11][18][43]
Activated MEKK phosphorylates MKK3/4/6/7. Phosphorylated MKK3/4/6/7 phosphorylate p38MAPK. We modeled MKK3/4/6/7 as one protein for simplification.	***T12***	***ii***	Phosphorylation	
	T13	***ii***	Phosphorylation	
Grb2 associates with tyrosine-phosphorylated EGFR.	***T14***	***iii***	Association	[18][19]
Shc binds to the tyrosine-phosphorylated EGFR.	***T15***	***iii***	Association	[35][37]
Shc is tyrosine phosphorylated and interacts with Grb2	***T16***	***iii***	Phosphorylation	[40][41][46]
	T17	***iii***	Association	
Sos1 binds to Grb2.	***T18***	***iii***	Association	[4]
Complex of Sos1 associated with EGFR catalyzes Ras GTP/GDP exchange.	***T19***	***iii***	Exchange	[20]

Generalized State Space Model

$$m_t = f(m_{t-1}, w_t, \theta)$$ System model

$$y_t = Hm_t + \varepsilon_t$$ Observation model

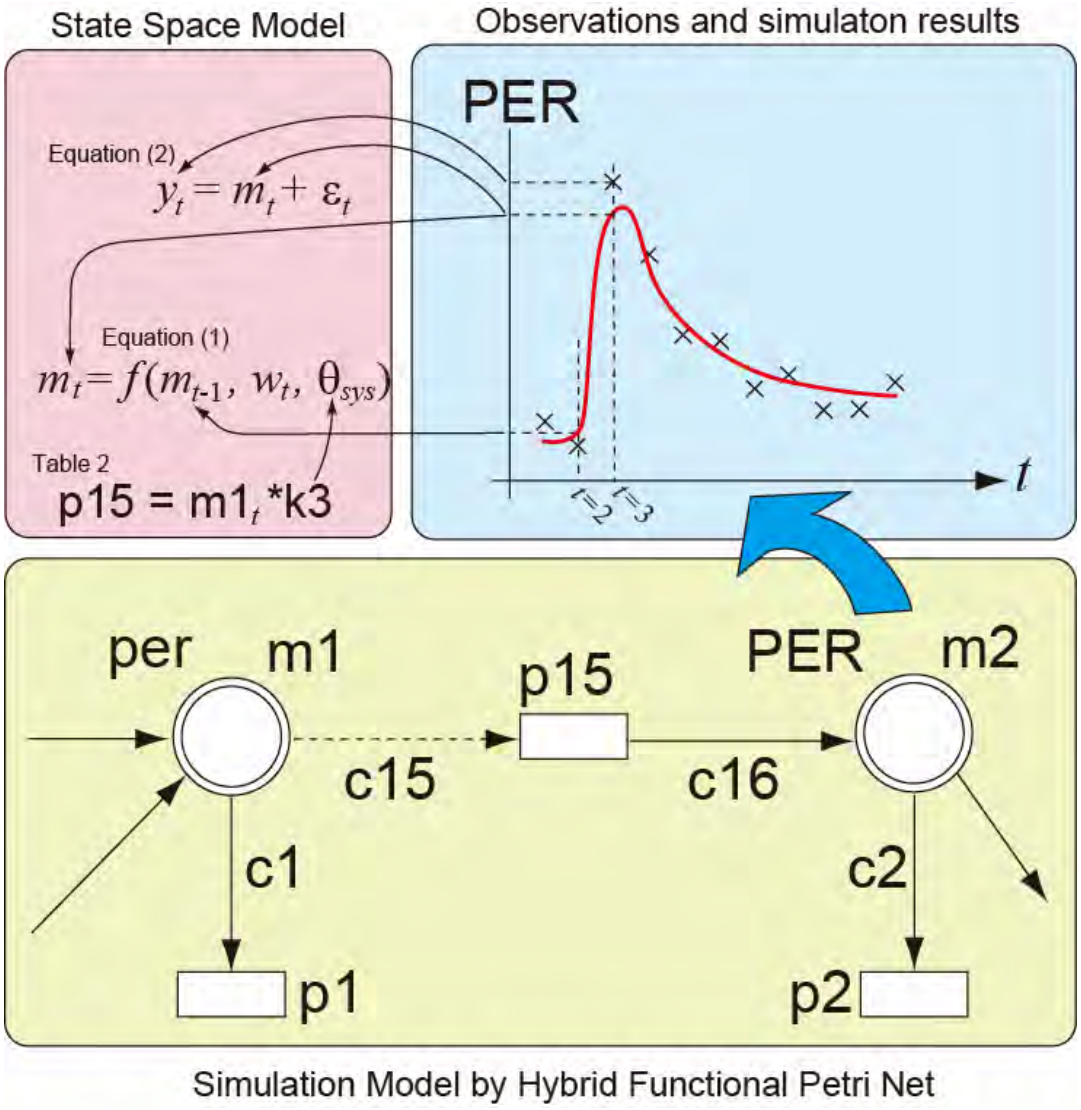
m_t : state vector at time t, f : simulation devise, $t = 1, \mathrm{K}, T$

w_t : system noise, θ : parameter vector,

y_t : observation vector at time t, H : observation matrix,

$\varepsilon_t \sim N(0, \sigma^2)$: observation noise

State Space Model and HFPN



Simulation Model by Hybrid Functional Petri Net

DA to obtain

$$P(M_T, \theta \mid Y_T)$$

$$M_T = \{m_1, \mathrm{K}, m_T\}$$

$$Y_T = \{y_1, \mathrm{K}, y_T\}$$

Using recursive estimation algorithm: **Particle Filter**

For parameter estimation, we used 10,000 particles in this study

Posterior distributions of the parameters $P(\theta \mid Y_N)$

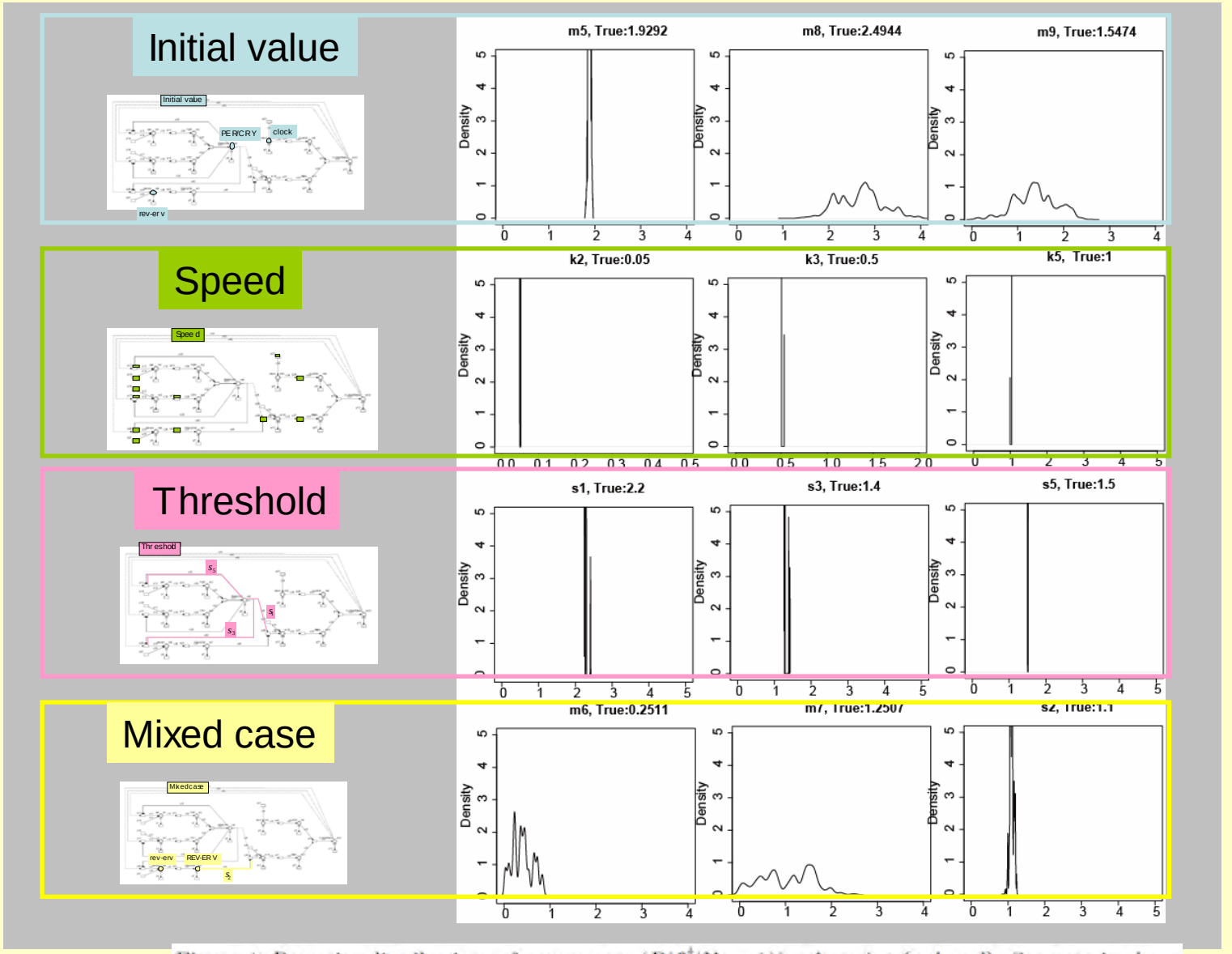
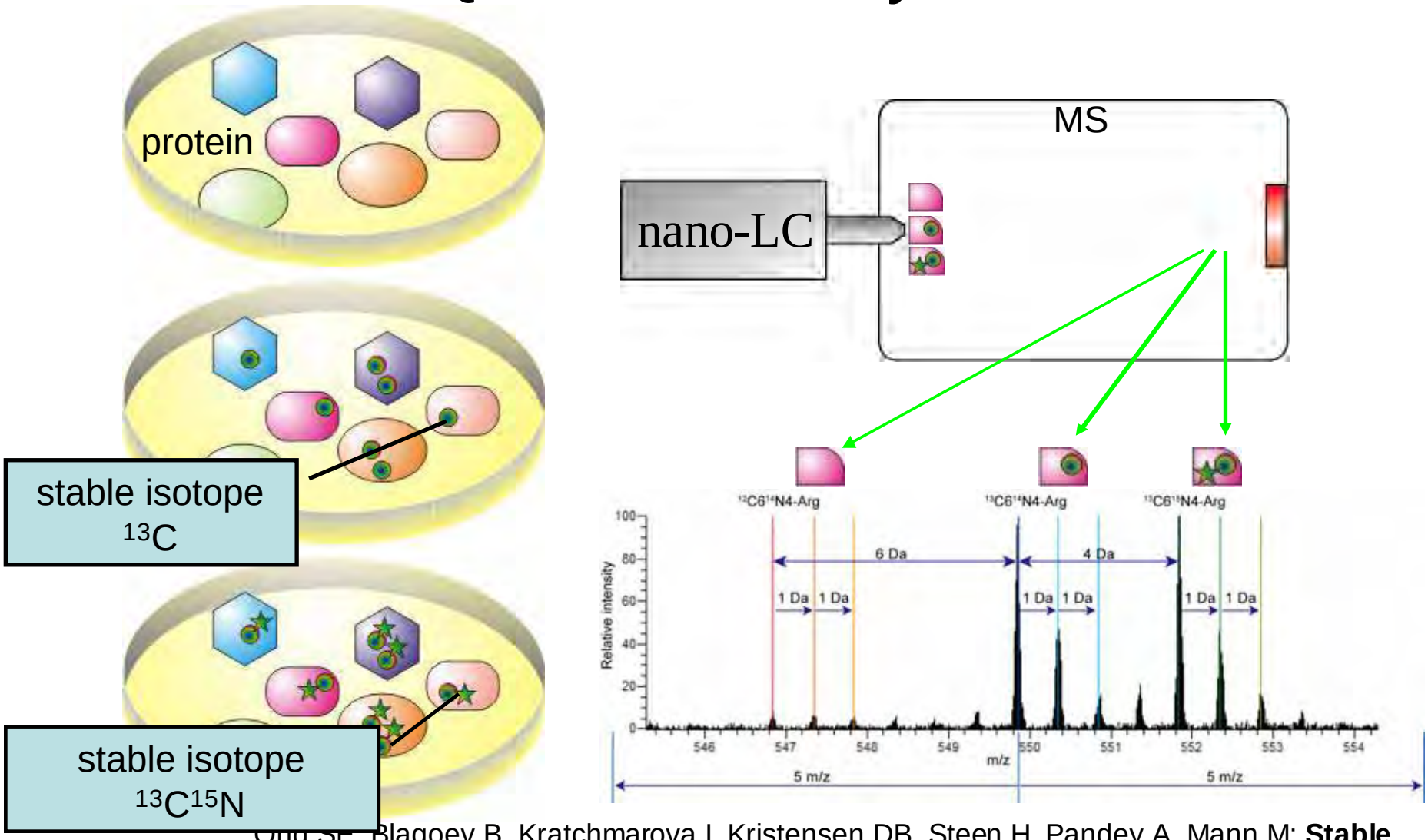


Figure 4: Posterior distributions of parameters ($P(\boldsymbol{\theta}^{\dagger}|\mathcal{Y}_{syn}, \hat{\sigma})$), where $\dagger \in \{a, b, c, d\}$. See note in the text. The panels on the first, the second, the third, and the fourth rows are for $\boldsymbol{\theta}^a$, $\boldsymbol{\theta}^b$, $\boldsymbol{\theta}^c$, and $\boldsymbol{\theta}^d$, respectively.

Observation Data:

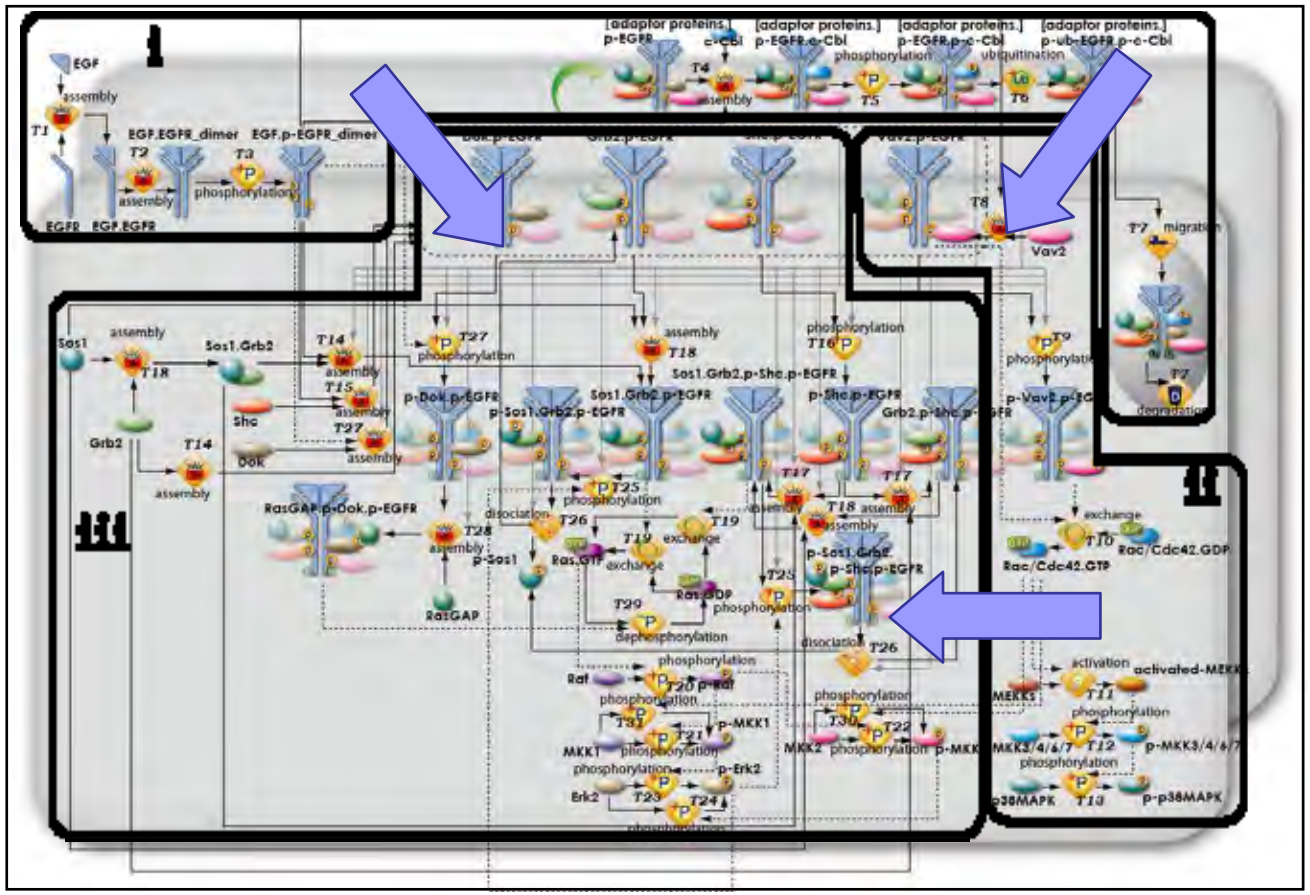
Protein Quantification by LC-MS/MS



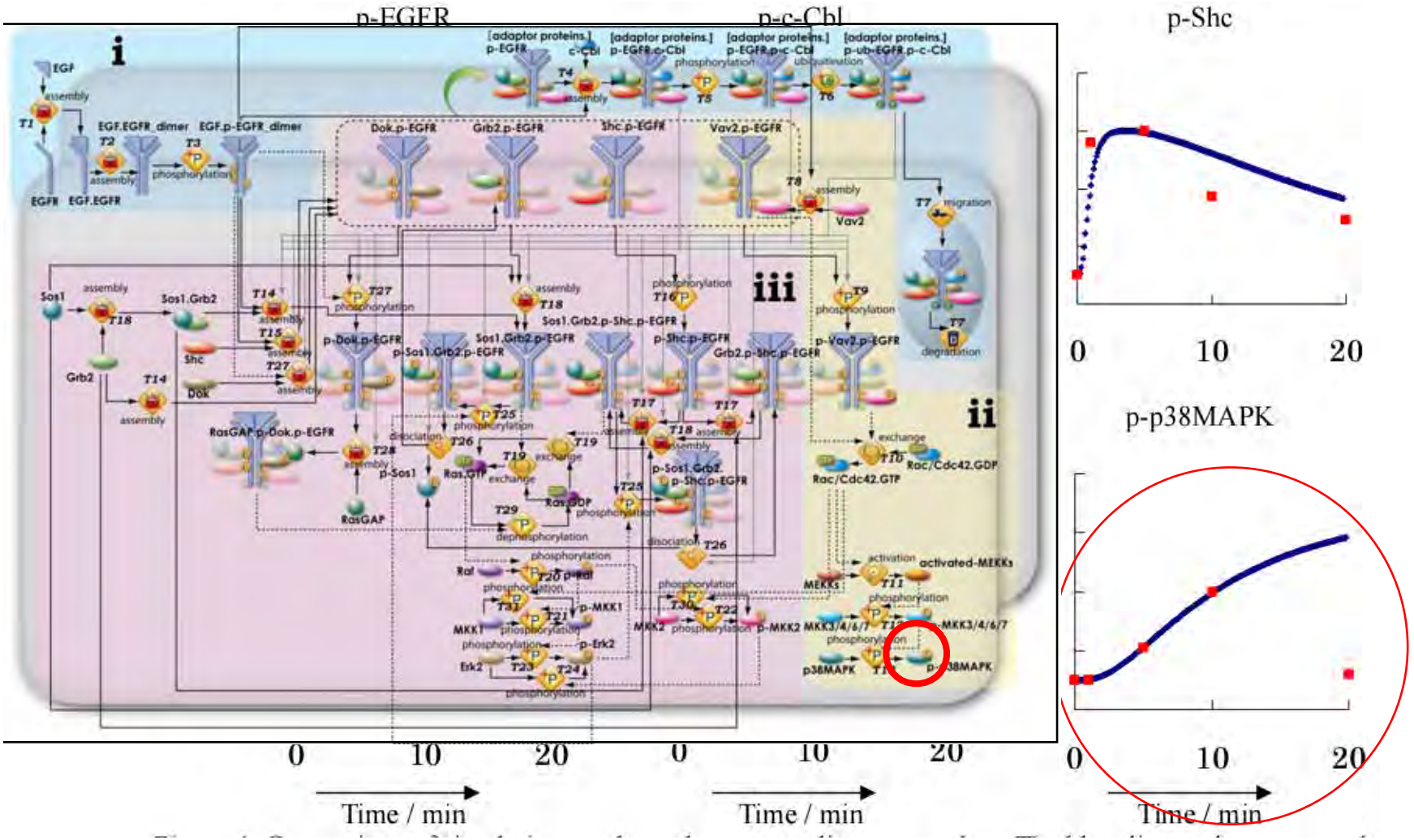
Ong SE, Blagoev B, Kratchmarova I, Kristensen DB, Steen H, Pandey A, Mann M: **Stable isotope labeling by amino acids in cell culture, SILAC, as a simple and accurate approach to expression proteomics.** *Mol Cell Proteomics* 2002, **1:**376–386.

Pathway Decomposition

Too many (63?) parameters!



Data Assimilation Result

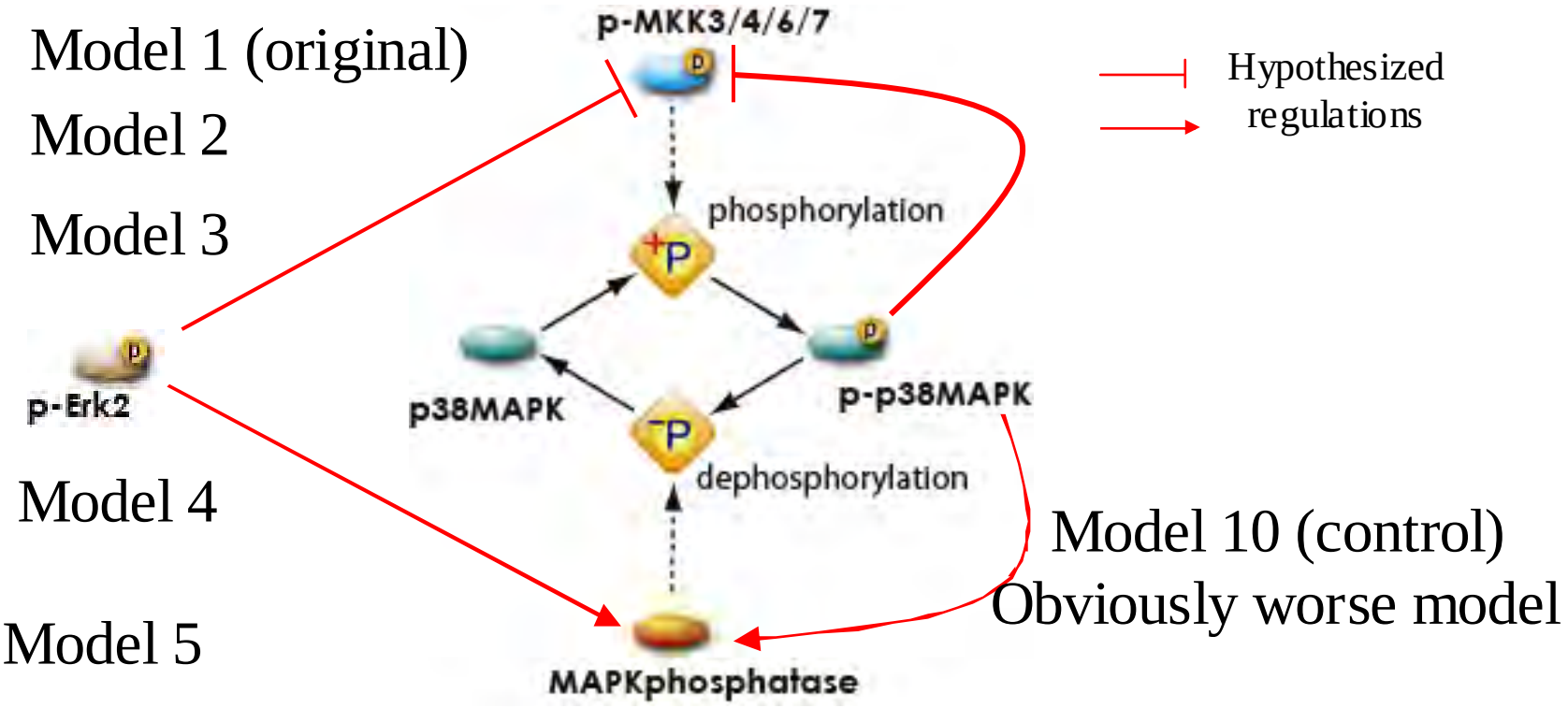


Preparing Hypothesis Models

Model 1 (original)

Model 2

Model 3



Model 4

Model 5

Model 6

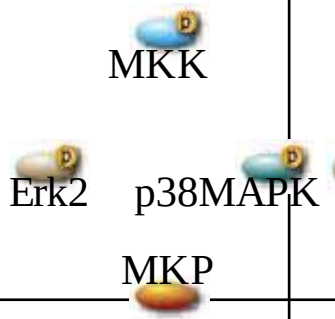
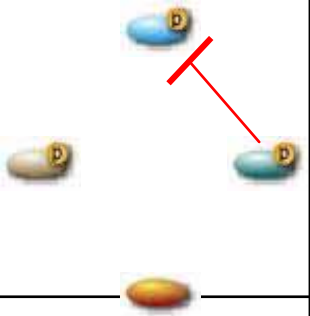
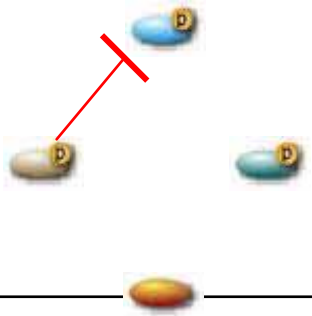
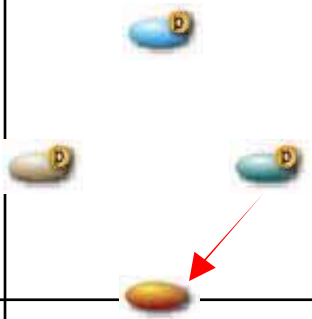
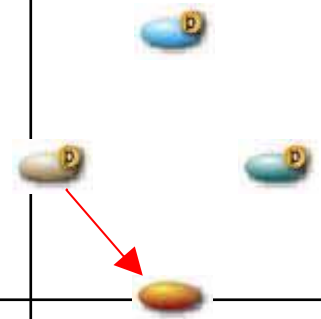
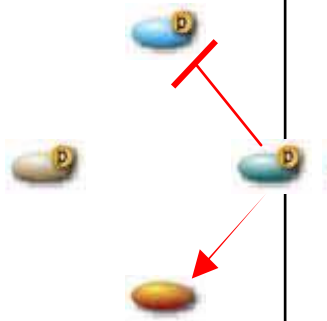
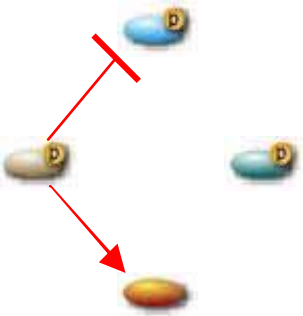
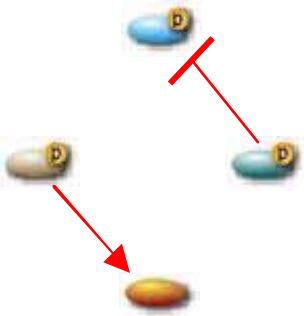
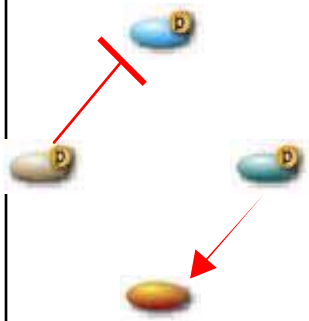
Model 7

Model 8

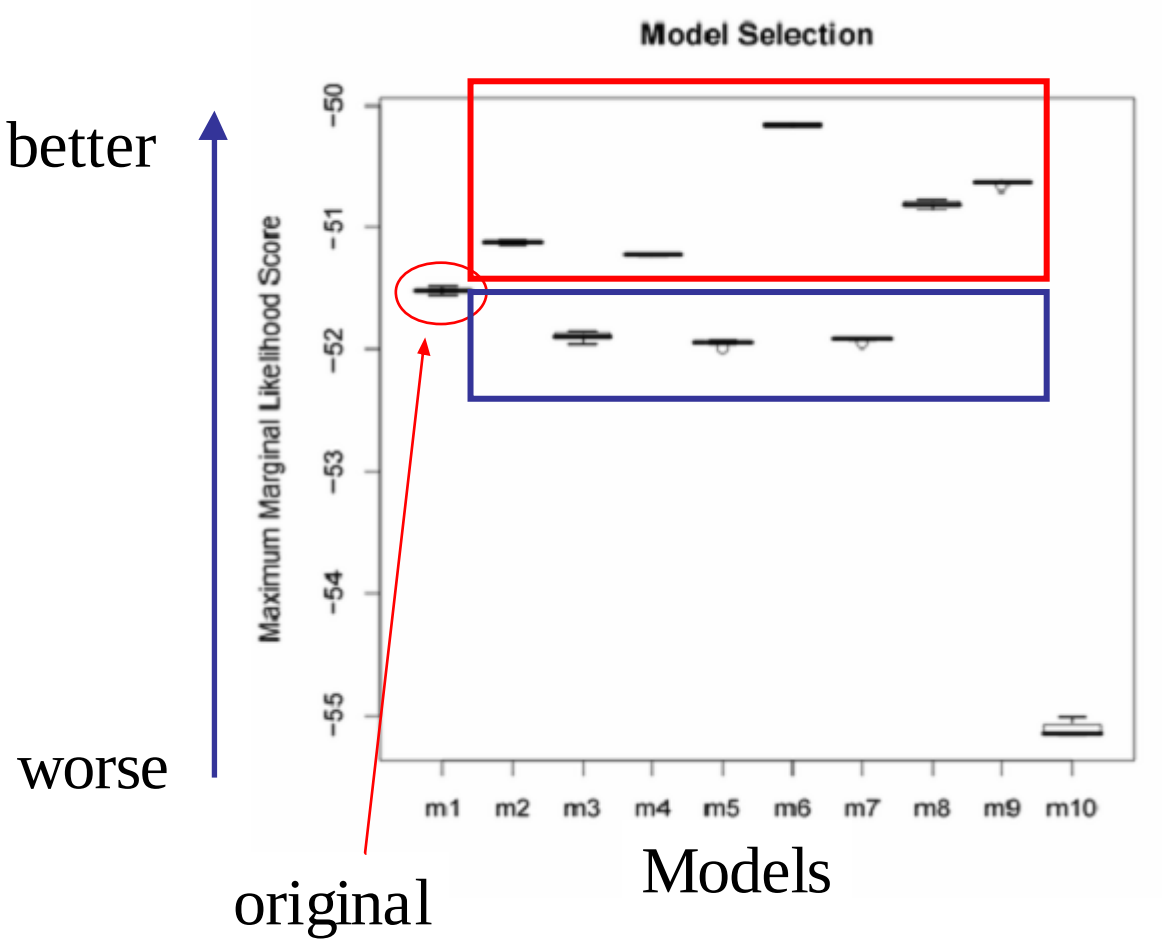
Model 9

Model 10 (control)
Obviously worse model

Preparing Hypothesis Models

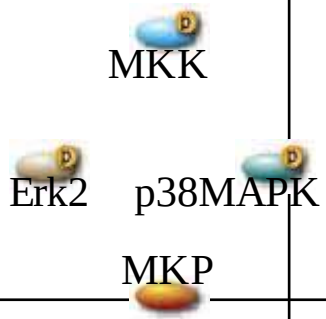
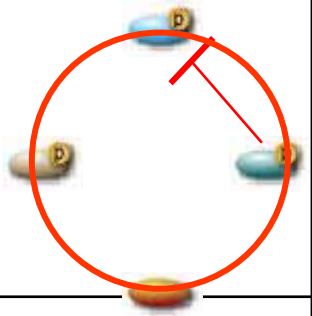
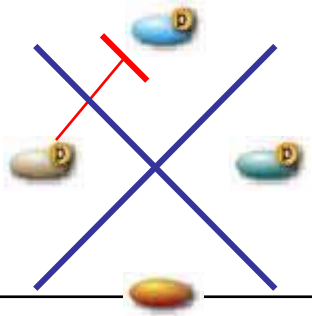
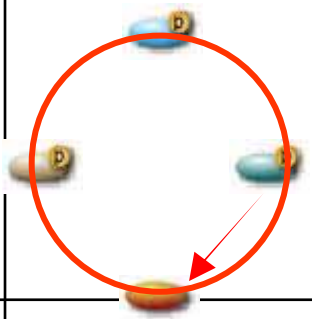
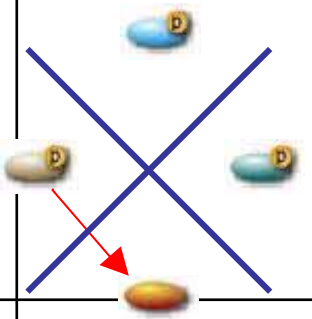
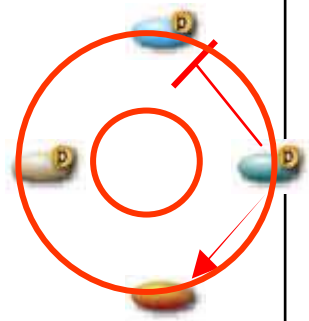
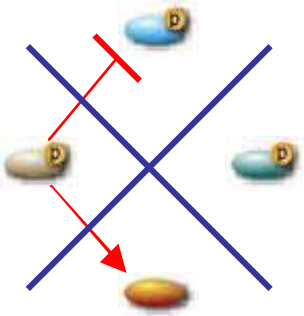
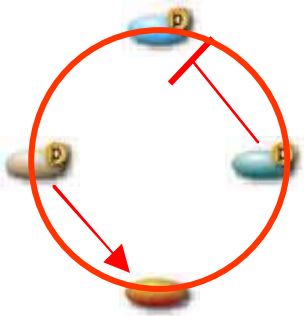
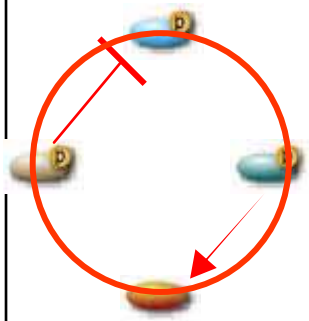
Model 1	Model 2	Model 3	Model 4	Model 5
				
Model 6	**Model 7**	**Model 8**	**Model 9**	**Model 10**
				control

Model Selection

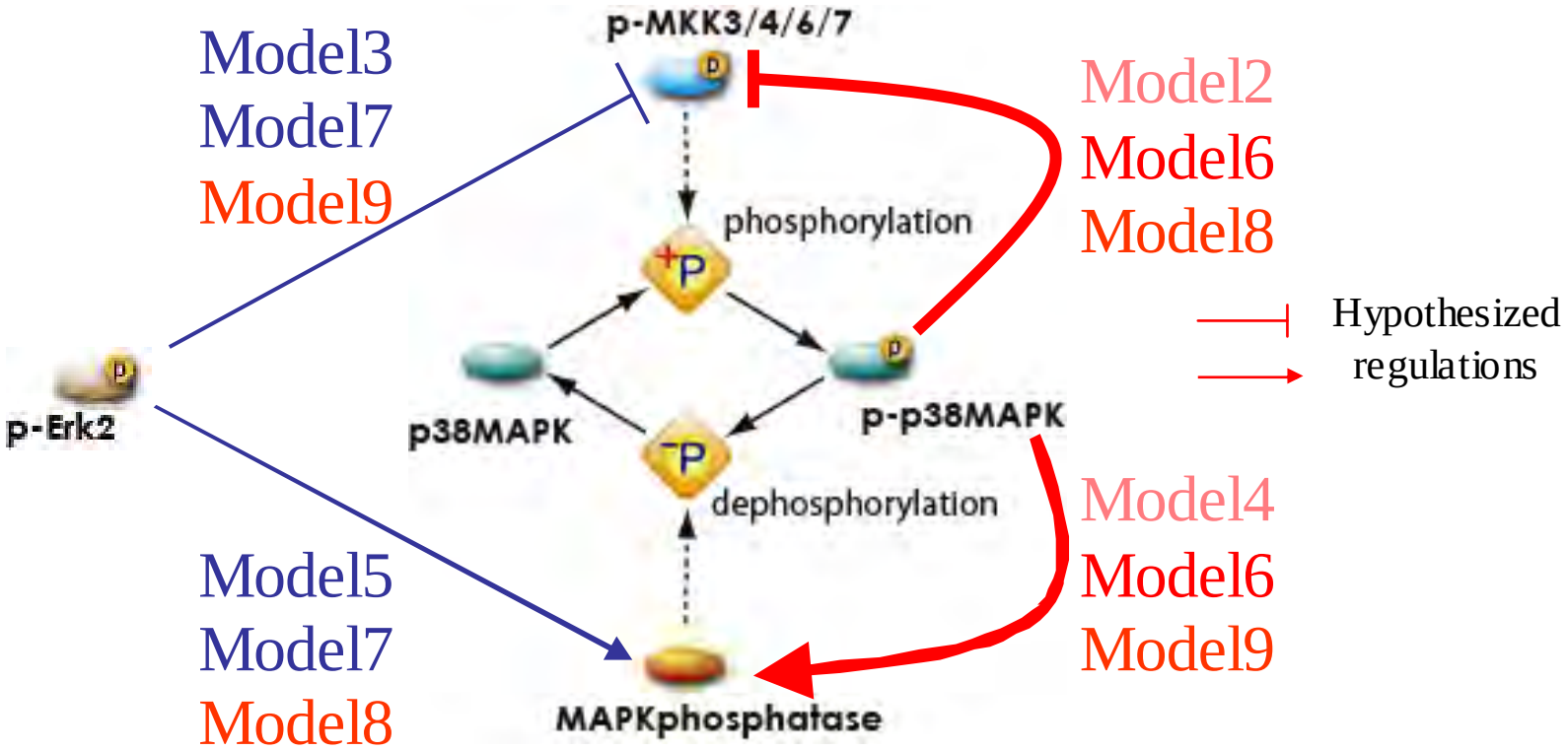


Nagasaki, M., Yamaguchi, R., Yoshida, R., Imoto, S., Doi, A., Tamada, Y., Matsuno, H., Miyano, S., Higuchi, T. Genomic data assimilation for estimating hybrid functional Petri net from time-course gene expression data. Genome Informatics. 17(1). 46-61, 2006.

Preparing Hypothesis Models

Model 1	Model 2	Model 3	Model 4	Model 5
				
Model 6	**Model 7**	**Model 8**	**Model 9**	**Model 10**
				control

Model Selection



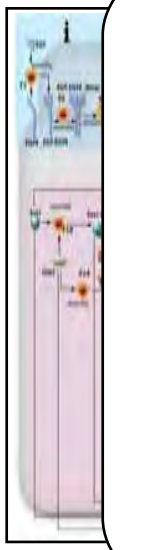
Currently …

Data assimilation technology is successfully applied to a small scale simulation model and course data.

Simulation model of EGF Pathway on C

Quantitative proteome time-course data by

New hypotheses



63 p

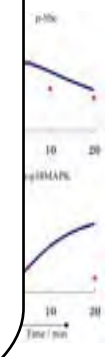
Very recently …

One billion particles are proven effective for parameter estimation from short time-course data.

Nakamura, K., Yoshida, R., Nagasaki, M., Miyano, S., Higuchi, T. Parameter estimation of *in silico* biological pathways with particle filtering towards a petascale computing. Pacific Symposium on Biocomputing. 14. In press.

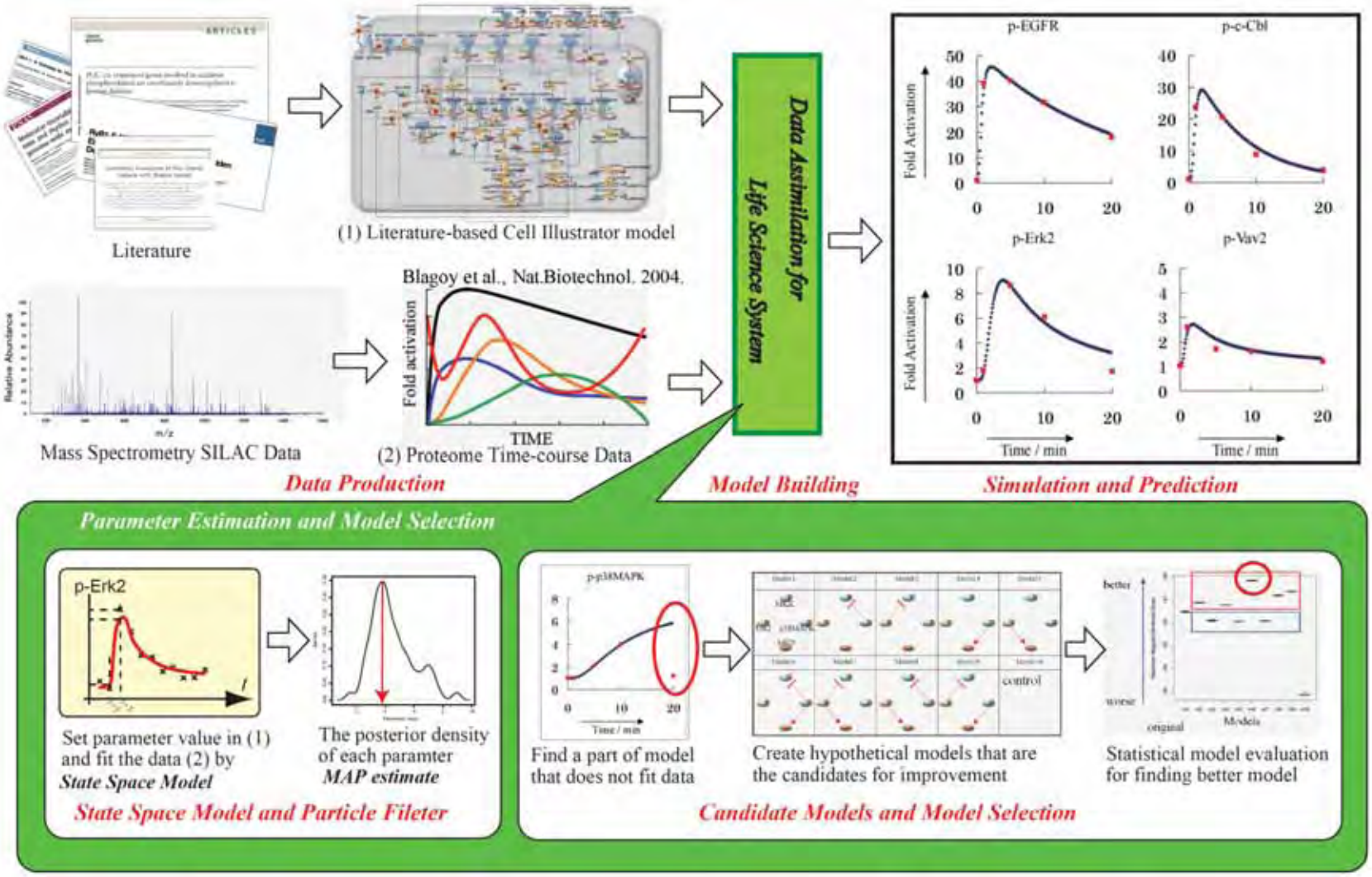
othesize

lations



Current Supercomputer is NOT Enough.

PETA FLOPS Computing!



Thank you for patience