

Predicting differences in gene regulatory systems by state space models

R. Yamaguchi¹, S. Imoto¹, M. Yamauchi¹, M. Nagasaki¹,
R. Yoshida², T. Shimamura¹, Y. Hatanaka¹, K. Ueno¹,
T. Higuchi², N. Gotoh¹, and S. Miyano¹

1. Institute of Medical Science, University of Tokyo

2. Institute of Statistical Mathematics



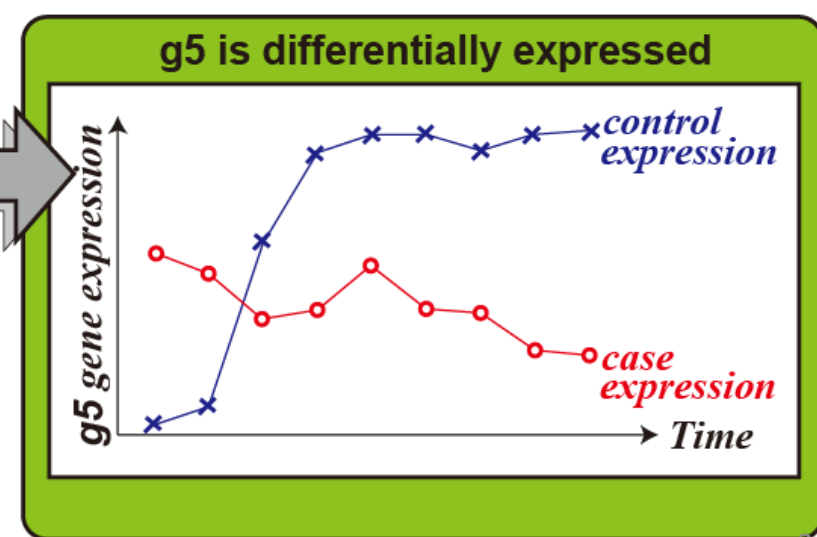
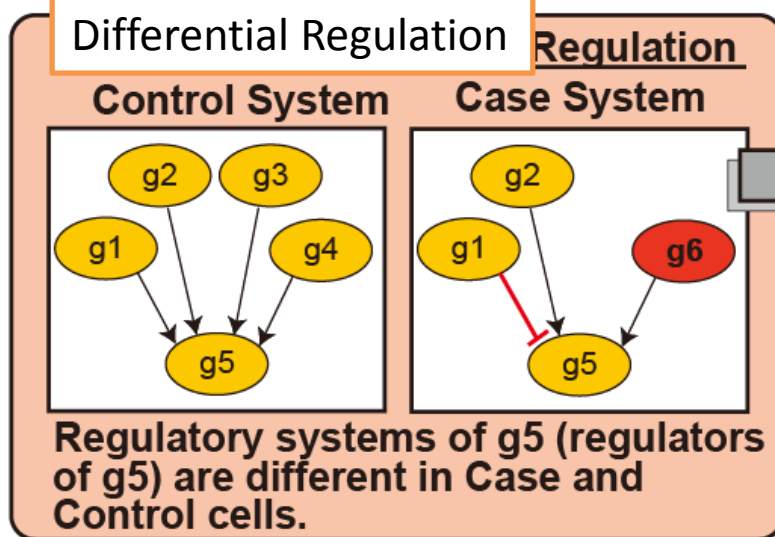
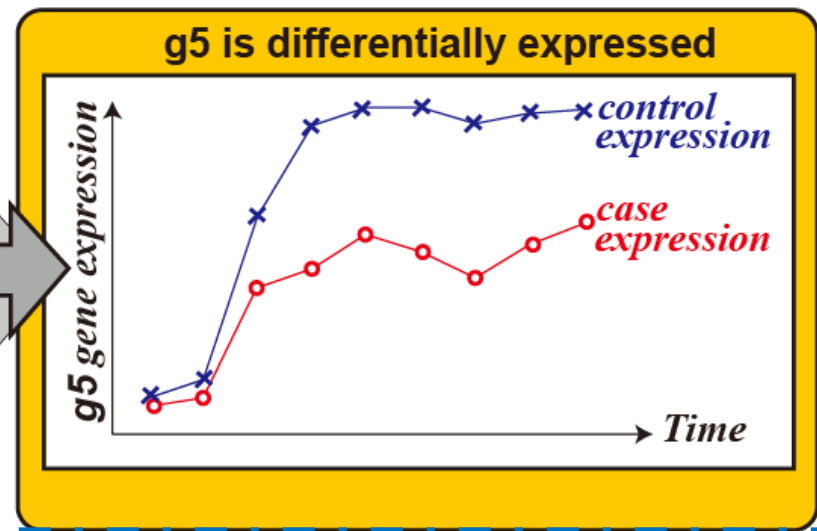
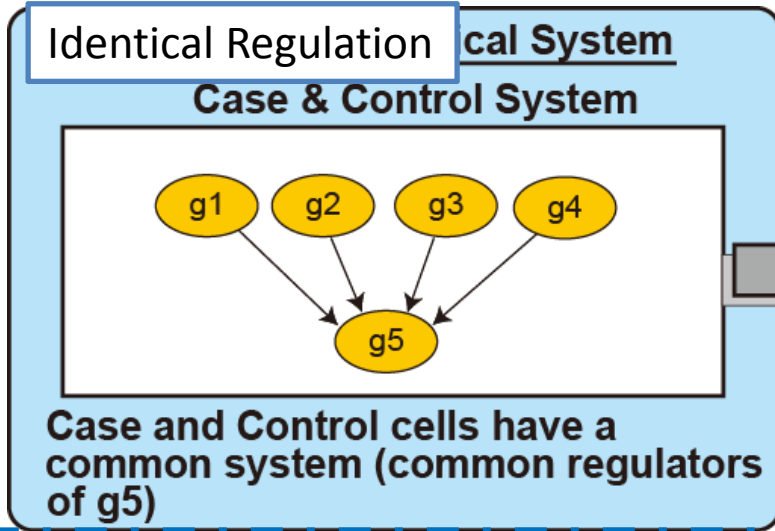
Presentation at GIW2008, December 2, 2008.

Reference: Yamaguchi *et al.*, *Genome Informatics*, 21: 101-113, (2008).

Outline

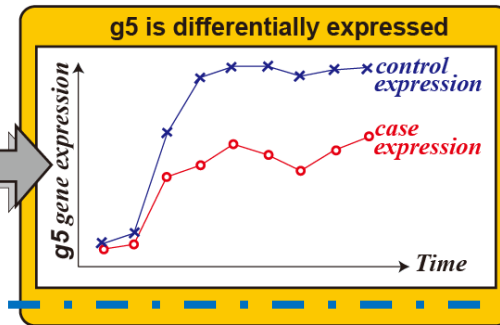
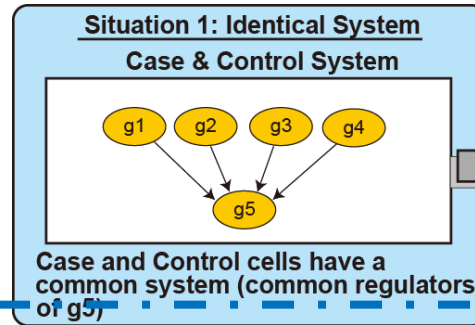
- Background
 - Finding Differentially Regulated Genes
- Method
 - State Space Models
- Analysis
 - Time Course Gene Expression Data with Different Conditions
- Summary and Future Works

Differentially-Expressed and Differentially-Regulated Genes

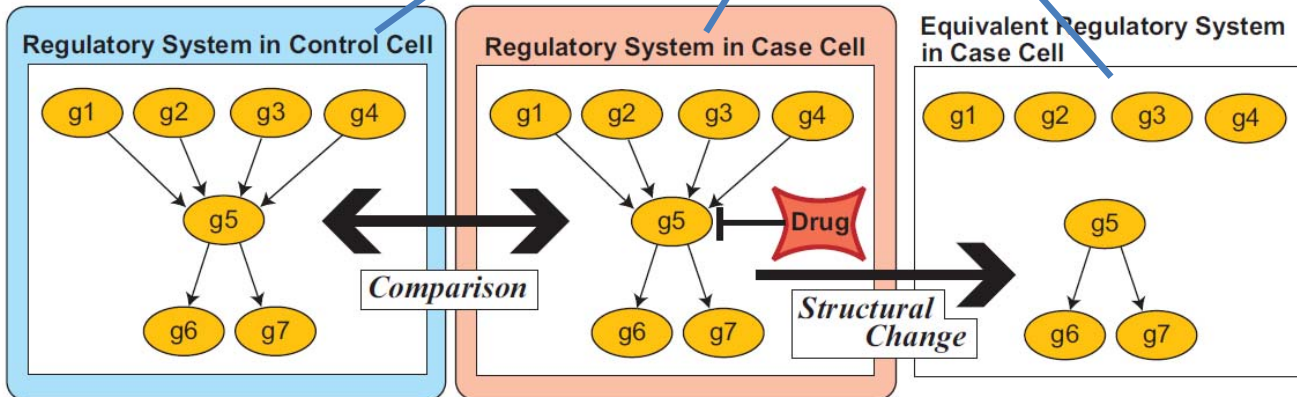
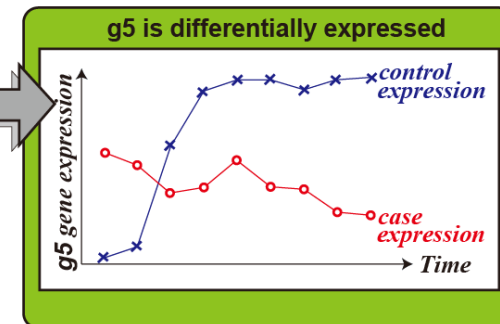
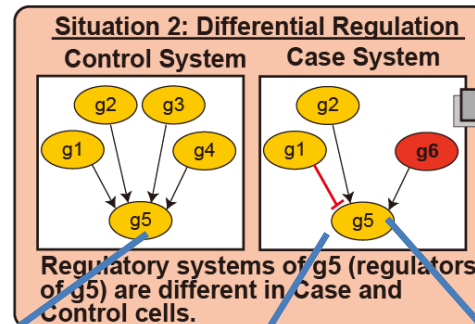


Differential Regulations by Drug Dosing

Identical Regulation



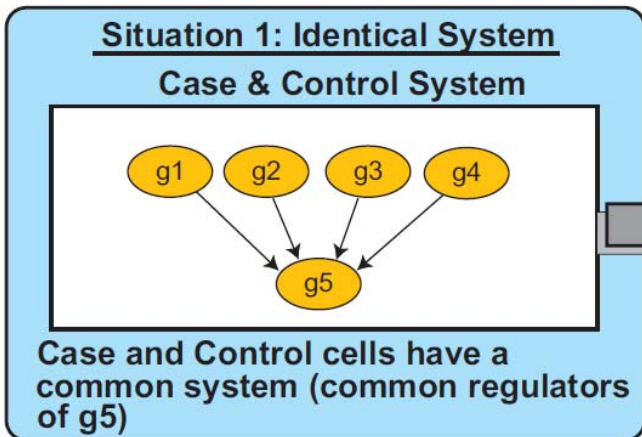
Differential Regulation



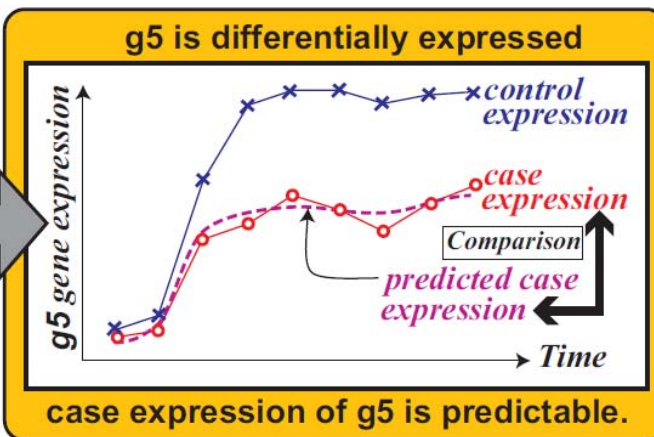
Biomarkers
Drug Targets
etc.

Predicting Case Data by Control's System to Discriminate the two Situations

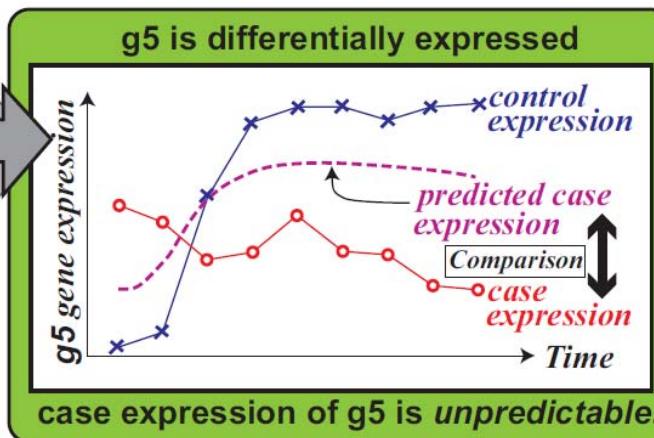
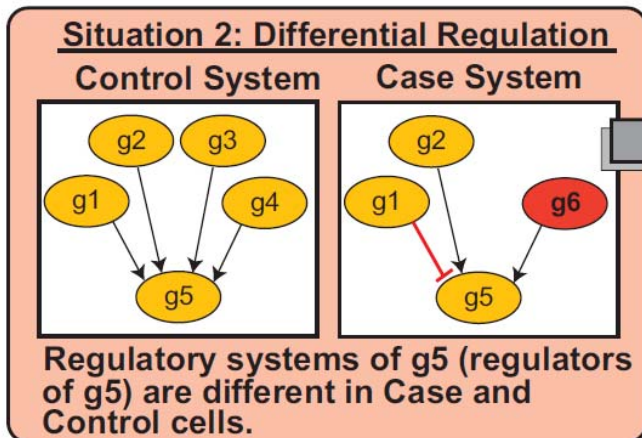
A Thought Experiment: If we know the Control's System, we use it to predict the Case data.



Prediction of case expression by control system



Case is **predictable** from Control's System



Case is **unpredictable** From Control's System

We use a statistical model for inferring gene regulatory systems.

State Space Model for Modeling Gene Regulatory Systems

SSM

$$x_n = Fx_{n-1} + v_n, n \in \mathcal{N},$$

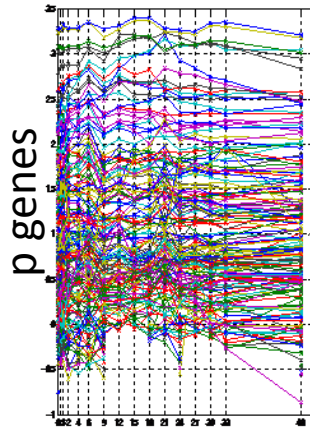
$$y_n = Hx_n + w_n, n \in \mathcal{N}_{obs},$$

SSM(θ) $\theta = \{H, F, R, x_0\}$

Time-Series Data

Short- and High-dimensional Data

$$N_{obs} \ll p$$



N_{obs} Time points

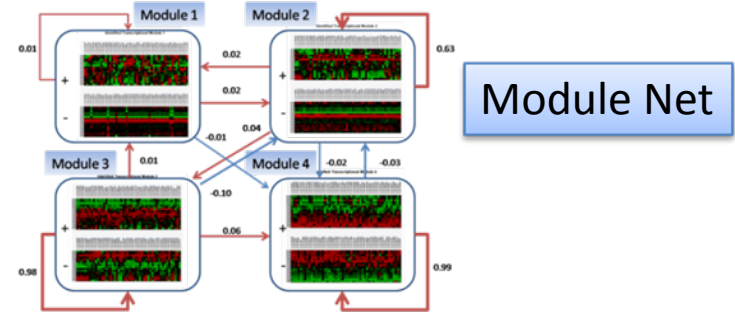
$$Y_{N_{obs}} = \{y_n\}, n \in \mathcal{N}_{obs}$$

Yoshida *et al.*, CSB, (2005)

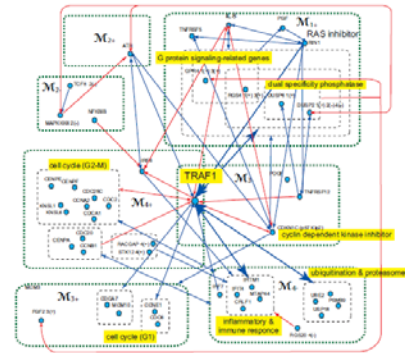
Yamaguchi *et al.*, IEEE SPM, (2007)

Hirose *et al.*, Bioinformatics, (2008)

Yamaguchi *et al.*, GIW2008.

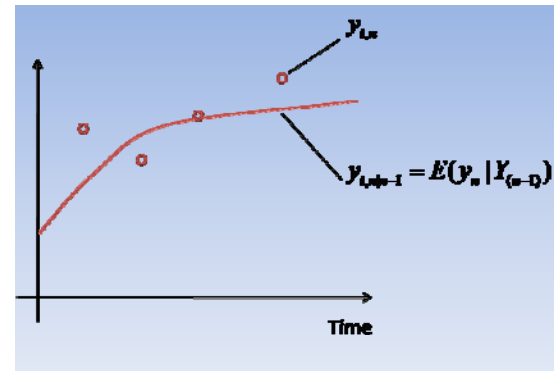


Module Net



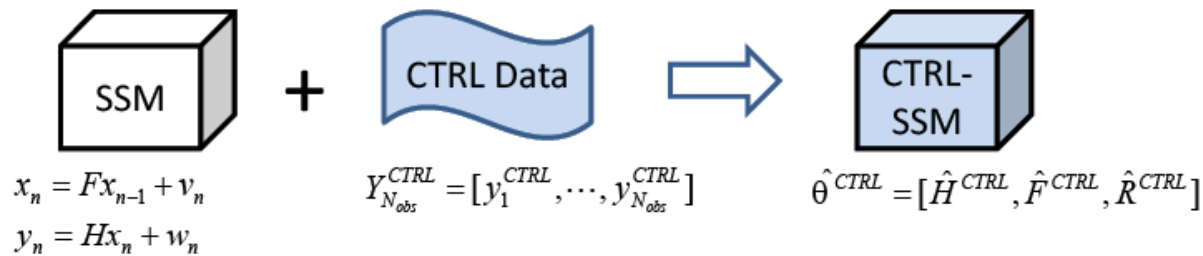
Gene Net

Prediction

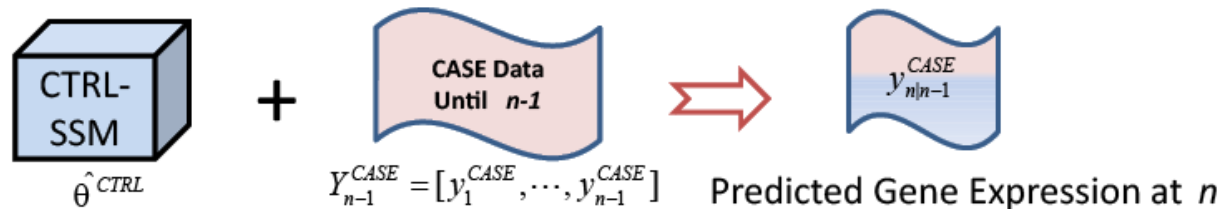


Strategy to Predict Differentially Regulated Genes

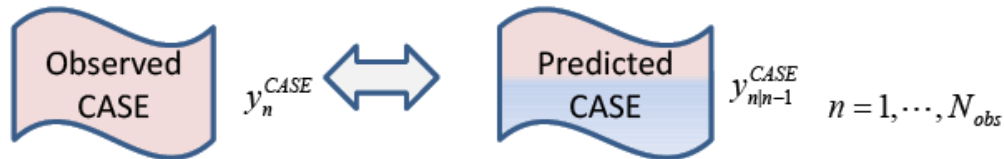
1. Train SSM by CTRL time-course data and Estimate an CTRL-SSM System



2. Predict CASE time-course data by the CTRL-SSM System



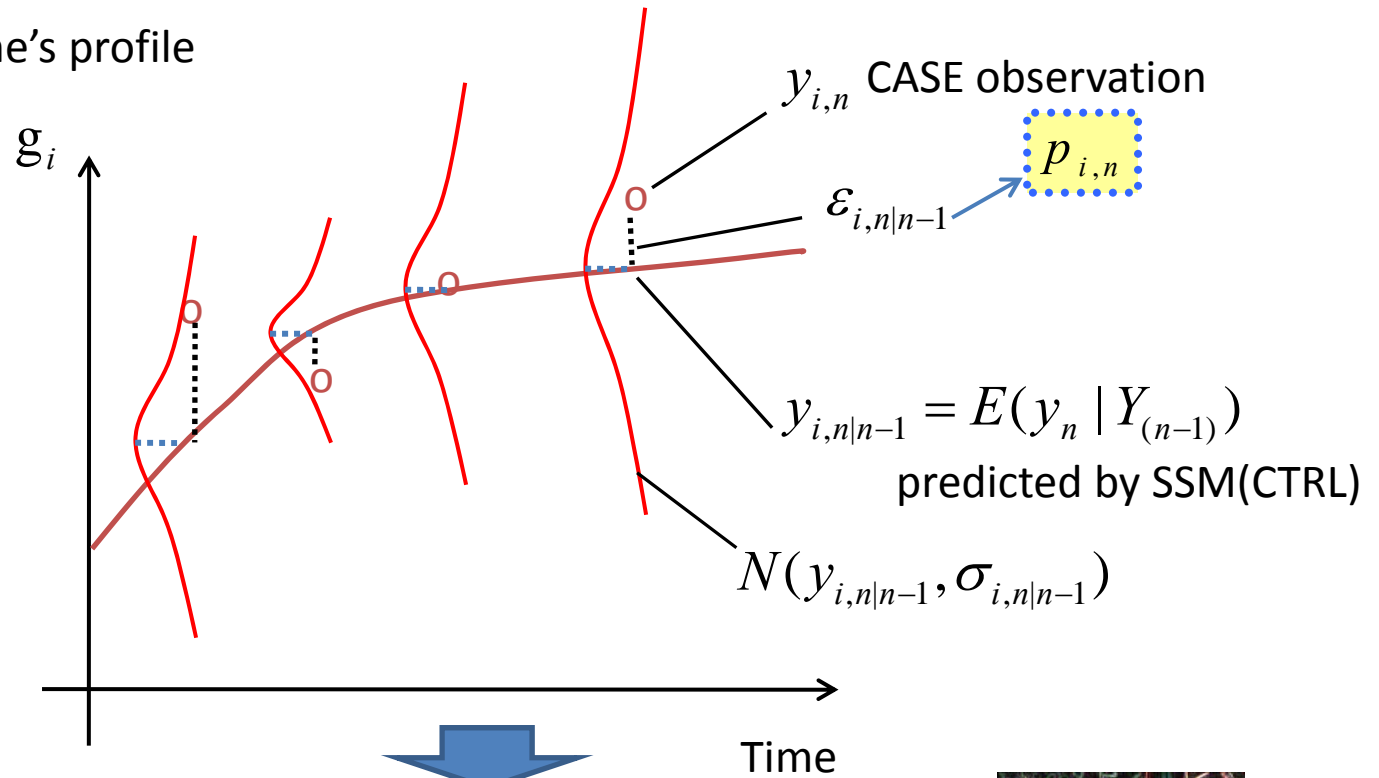
3. Compare Predicted and Observed Time Courses



Candidates of Differentially Regulated Genes:
Unpredictable genes in the Case data by the Control's SSM.

Significance Test for Prediction Errors

The i^{th} gene's profile



Integration $\{p_{i,n}\}, n \in N_{obs}$ by MetaGP



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

Finding Significantly
Differentially Regulated
Genes

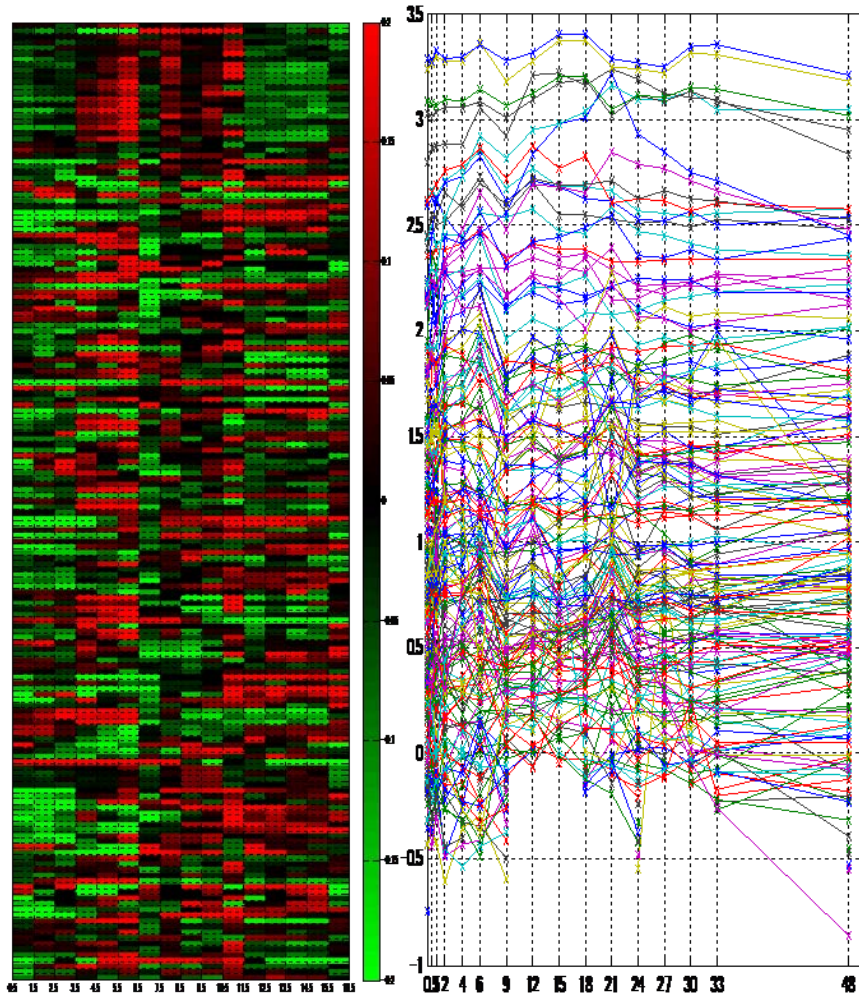
Integrated p -value p_i for the i^{th} gene ($i = 1, \dots, p$)

Application to Real Data

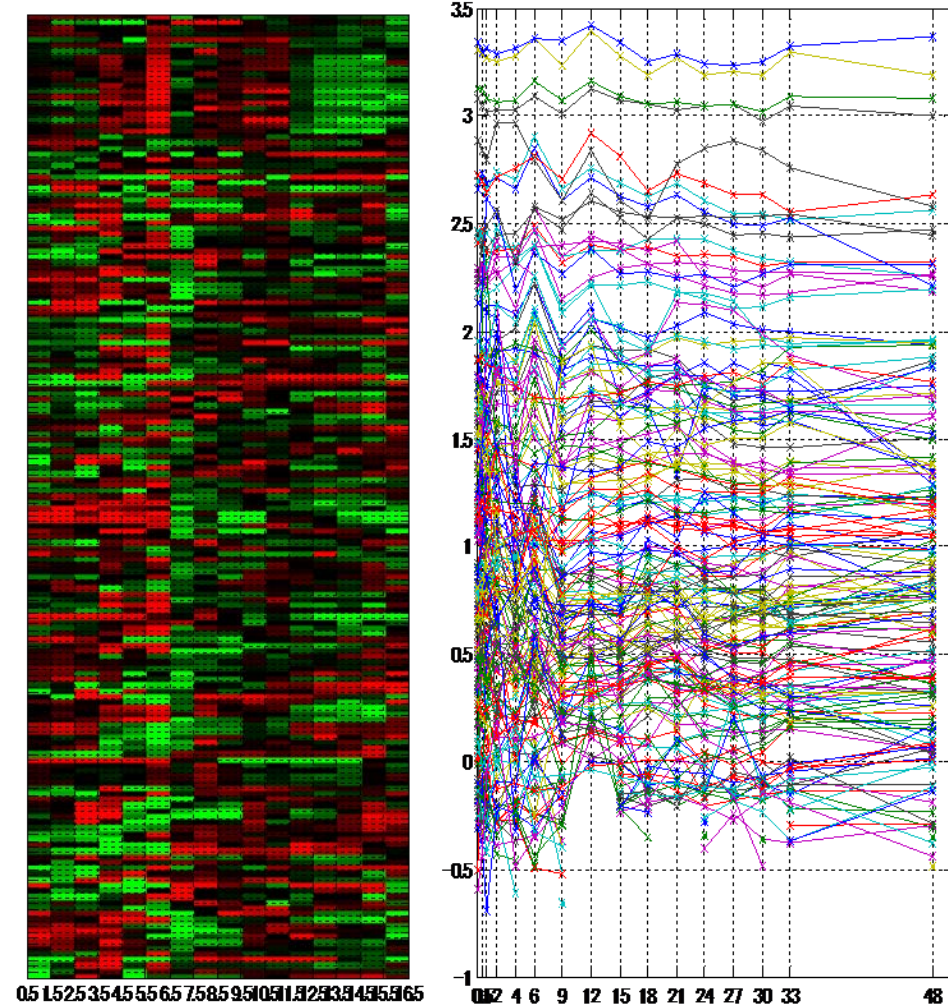
- EGFR pathway
- Data
 - Time Course Gene Expression Microarray Data
 - 19 time points during 48 hours
 - Human Small Airway Epithelial Cell (SAEC)
 - Two Conditions (Different Drugs)
 - EGF Dosed (Control Data)
 - EGF + Gefitinib Dosed (Case Data)
 - Gene Set
 - 500 genes extracted based on Coefficient of Variation
- SSM Analysis
 - Predict Case (EGF+GFT) with SSM for Control (EGF) to find differentially regulated genes

Time course of Genes (probes) on the EGFR pathway

EGF (Control)



EGF+GFT (Case)

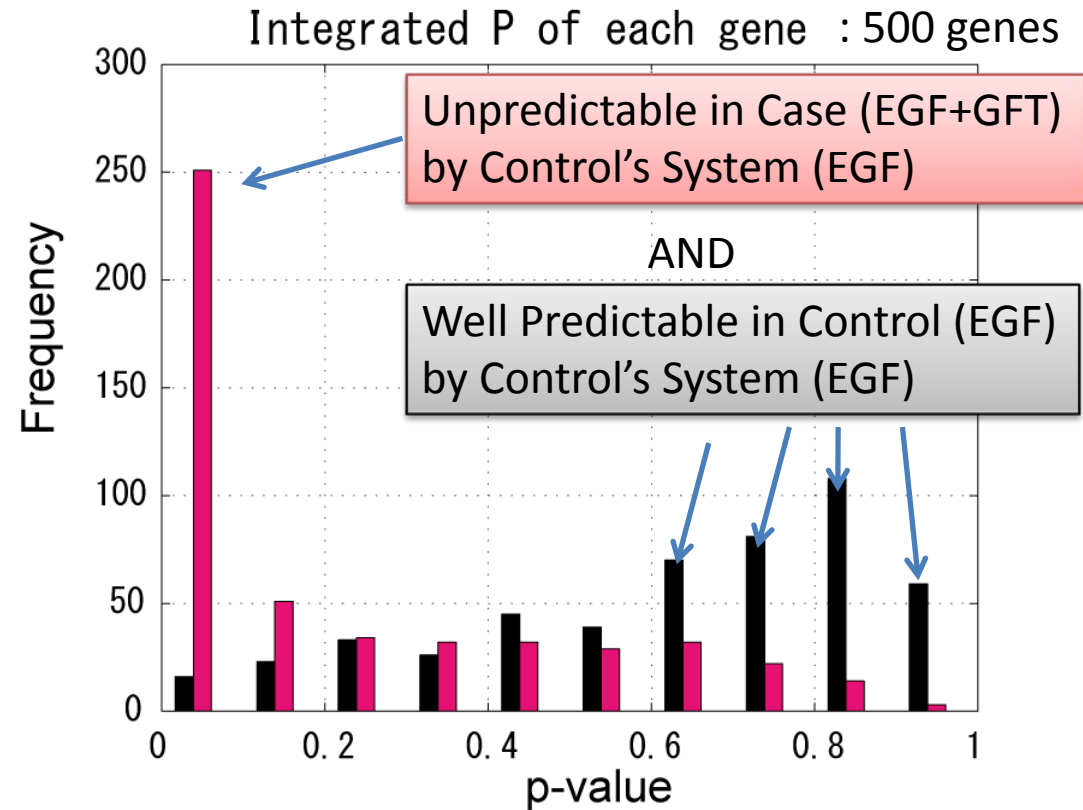


Finding Differentially Regulated Genes

Predicted Data	Prediction System	Bar Color
Control (EGF)	Control (EGF)	Black
Case (EGF+GFT)	Control (EGF)	Red

Each gene is characterized by the two p-values (in black and red).

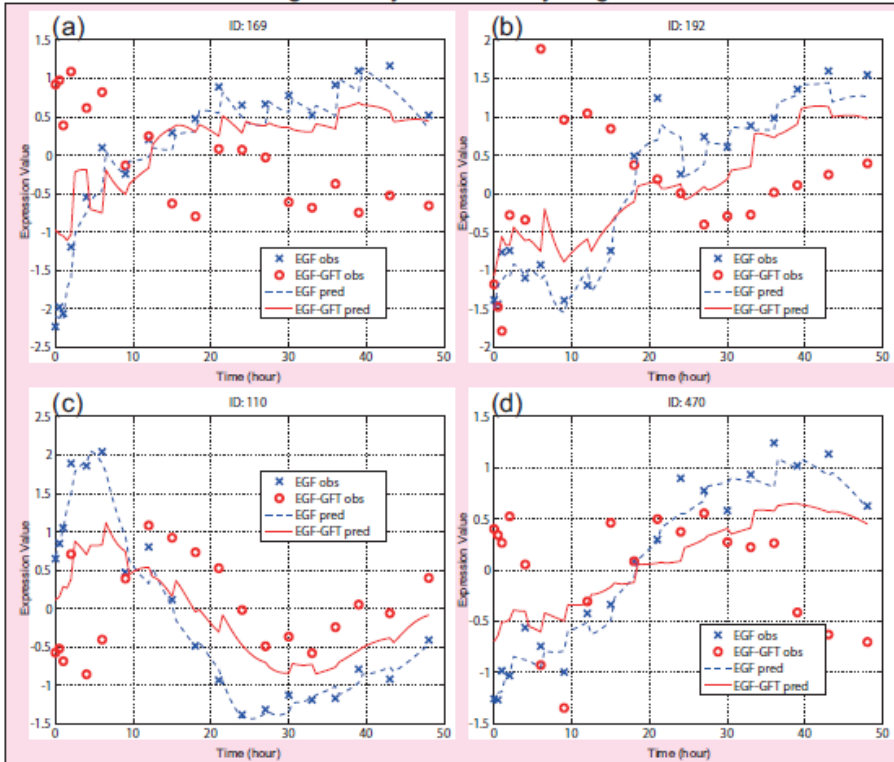
We selected genes with small p (<0.01) in red and with large p (>0.5) in black



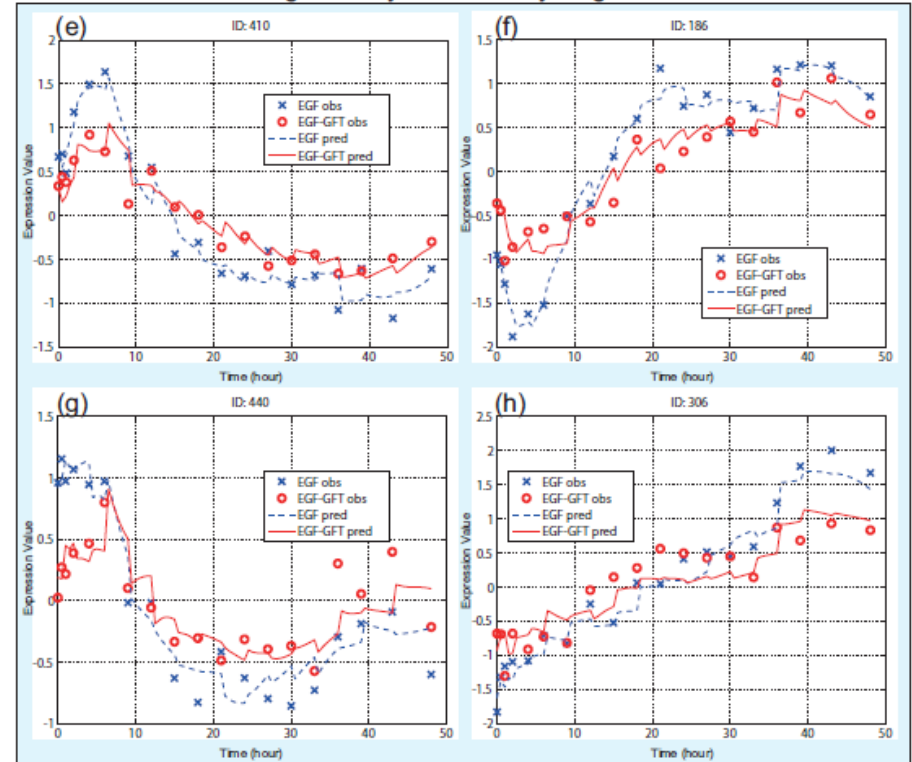
Selected Genes

- X : EGF obs (Control)
- O : EGF+GFT obs (Case)
- : EGF pred by SSM(EGF)
- : EGF+GFT pred by SSM(EGF)

Significantly Differentially Regulated



Insignificantly Differentially Regulated



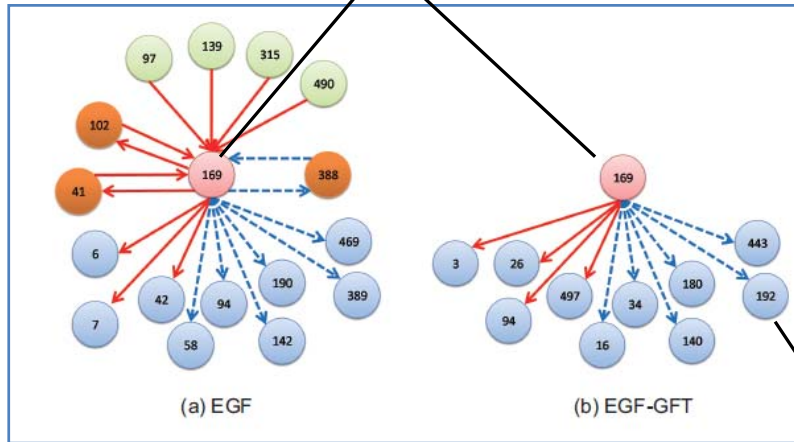
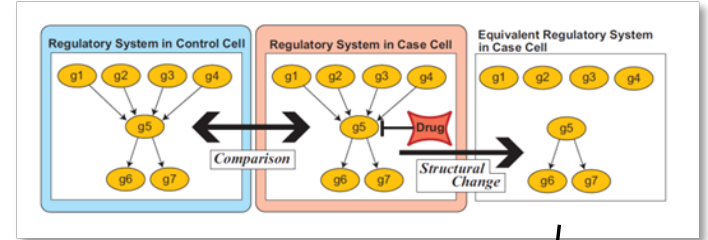
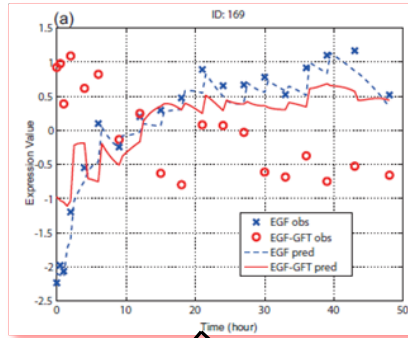
Prediction of EGF-GFT by SSM(EGF) (red solid line) deviate from the observations (circles)

Prediction of EGF-GFT by SSM(EGF) (red solid line) trace the observations (circles)

Estimated Gene Networks with SSMs

$$R^{-1/2}(y_n - w_n) = \Psi R^{-1/2}(y_{n-1} - w_{n-1}) + R^{-1/2} H v_n$$

Diff gene
(gene 162)

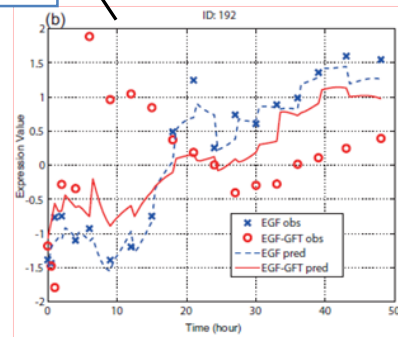


Control
(EGF)

Case
(EGF+GFT)

Diff gene
(gene 192)

Predicted
disappearance
of parent nodes
by drug dosing



Summary and Future Works

- Method to Find Differentially Regulated Genes
 - State Space Model
 - Predicting Case Data with Control's System
 - Unpredictable Genes: Differentially Regulated Genes
- Real Data Analysis
 - SAEC dosed with EGF and EGF-GFT
 - Consistent View of Predicted disappearance of parent nodes by drug dosing
- Future Works
 - Finding Drug Targets
 - Finding Biomarkers