Detect the **time point** when organisms **toggle** gene regulatory networks from their expression profiles

竹中 要一
Y. Takenaka
Osaka Univ. Japan

---

**Self-Introduction**

- Yoichi Takenaka
  - Dept. Bioinformatic Engineering, Grad. Info. Sci. & Tech @ OU
  - Background : Computer Science
- **Motivation of research**
  - Cell Differentiation Analyses
    - MS cell to fat cell & bone cell
  - Dynamics of the networks interest me
Overview

**Biological System**  Gene Regulatory Network

**Data**  Time-course gene expression profile

**Goal**  Detect when the regulation changes

**Proposal**  Formulate the **strength** of the regulation at each **single time point**.

**Material**  Diauxic shift of E.coli.

---

Biological Networks

- **Signaling Pathway**
- **Metabolic Pathway**
- **Gene regulatory network**
Transcriptional Regulations

Known & unknown regulations

- All revealed
  - E.coli
- Most(?)
  - Yeast
- Partially
  - Others
    - Transfac

All will be known in near future.
Dynamics of regulations

• All the regulations will be known
  – Not the end of researches on regulations
• Next target: when the regulations work
  – Only a part of regulations work at once

Stable state
  • Environment (Nutrient Source, Temperature)
  • Cell states, cell types

Dynamics
  – Bridges stable states
  – When & How networks change

This is what I want to analyze.

Real Data

A Network in Adipo Cell

The reality is not simple

Mouse, Cell Differentiation from MS Cell to Adipo Cell
By Affymetrix GeneChip
Measurement of the regulations

- Correlation Coefficient
- Statistics
  - Null hypothesis
    - No regulation
    - p-value < level of significance
      -> the regulation exists
- Information Criterion
  - Measure of the relative quality
    \[ \text{AIC} = 2k - 2\ln(L) \]
    \( k \): #parameters
    \( L \): max value of likelihood function
  - Compare the value of networks w/wo regulations
  - The smaller is more plausible

Ex. use of Information Criterion

- Network Inference by Bayesian Network

<table>
<thead>
<tr>
<th>Expression</th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
<th>6</th>
<th>7</th>
<th>8</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>ON</td>
<td>ON</td>
<td>OFF</td>
<td>OFF</td>
<td>ON</td>
<td>ON</td>
<td>OFF</td>
<td>ON</td>
</tr>
<tr>
<td></td>
<td>ON</td>
<td>ON</td>
<td>ON</td>
<td>OFF</td>
<td>OFF</td>
<td>ON</td>
<td>OFF</td>
<td>OFF</td>
</tr>
</tbody>
</table>

Networks
**Ex. use of Information Criterion**

- **Network Inference by Bayesian Network**

<table>
<thead>
<tr>
<th>Expression</th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
<th>6</th>
<th>7</th>
<th>8</th>
</tr>
</thead>
<tbody>
<tr>
<td>ON</td>
<td>ON</td>
<td>OFF</td>
<td>OFF</td>
<td>ON</td>
<td>ON</td>
<td>OFF</td>
<td>ON</td>
<td>ON</td>
</tr>
<tr>
<td>ON</td>
<td>ON</td>
<td>ON</td>
<td>ON</td>
<td>OFF</td>
<td>OFF</td>
<td>ON</td>
<td>OFF</td>
<td>OFF</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Network</th>
<th>Freq.</th>
<th>prob.</th>
<th>-ln (p)</th>
<th>-2*Freq.*ln (p)</th>
<th>#param</th>
<th>AIC</th>
</tr>
</thead>
<tbody>
<tr>
<td>ON</td>
<td>5</td>
<td>5/8</td>
<td>0.47</td>
<td>4.70</td>
<td>1</td>
<td>4.7+5.88+2</td>
</tr>
<tr>
<td>OFF</td>
<td>3</td>
<td>3/8</td>
<td>0.98</td>
<td>5.88</td>
<td></td>
<td>=12.58</td>
</tr>
</tbody>
</table>

AIC = 2k − 2ln(L)

k: #parameters
L: max value of likelihood function

**Ex. use of Information Criterion**

- **Network Inference by Bayesian Network**

<table>
<thead>
<tr>
<th>Expression</th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
<th>6</th>
<th>7</th>
<th>8</th>
</tr>
</thead>
<tbody>
<tr>
<td>ON</td>
<td>ON</td>
<td>OFF</td>
<td>OFF</td>
<td>ON</td>
<td>ON</td>
<td>OFF</td>
<td>ON</td>
<td>ON</td>
</tr>
<tr>
<td>ON</td>
<td>ON</td>
<td>ON</td>
<td>ON</td>
<td>OFF</td>
<td>OFF</td>
<td>ON</td>
<td>OFF</td>
<td>OFF</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Network</th>
<th>Freq.</th>
<th>prob.</th>
<th>-ln (p)</th>
<th>-2*Freq.*ln (p)</th>
<th>#param</th>
<th>AIC</th>
</tr>
</thead>
<tbody>
<tr>
<td>ON</td>
<td>4</td>
<td>4/8</td>
<td>0.69</td>
<td>5.55</td>
<td></td>
<td></td>
</tr>
<tr>
<td>OFF</td>
<td>4</td>
<td>4/8</td>
<td>0.69</td>
<td>5.55</td>
<td>1</td>
<td>5.55+5.55+2</td>
</tr>
</tbody>
</table>

12.58

25.68

13.10

=13.10
### Ex. use of Information Criterion

- **Network estimation by Bayesian Network**

<table>
<thead>
<tr>
<th>Expression</th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
<th>6</th>
<th>7</th>
<th>8</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>ON</td>
<td>ON</td>
<td>OFF</td>
<td>OFF</td>
<td>ON</td>
<td>ON</td>
<td>OFF</td>
<td>ON</td>
</tr>
<tr>
<td></td>
<td>ON</td>
<td>ON</td>
<td>ON</td>
<td>OFF</td>
<td>OFF</td>
<td>ON</td>
<td>OFF</td>
<td>OFF</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Network</th>
<th>Freq.</th>
<th>prob.</th>
<th>-ln (p)</th>
<th>-2*Freq.*ln (p)</th>
<th>#param</th>
<th>AIC</th>
</tr>
</thead>
<tbody>
<tr>
<td>ON</td>
<td>5</td>
<td>5/8</td>
<td>0.47</td>
<td>4.70</td>
<td>1</td>
<td>12.58</td>
</tr>
<tr>
<td>OFF</td>
<td>3</td>
<td>3/8</td>
<td>0.98</td>
<td>5.88</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

\[ \text{AIC} = 2k - 2 \ln(L) \]

\[ k: \# \text{parameters} \]

\[ L: \text{max value of likelihood function} \]
### Ex. use of Information Criterion

- **Network estimation by Bayesian Network**

<table>
<thead>
<tr>
<th>Expression</th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
<th>6</th>
<th>7</th>
<th>8</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>ON</td>
<td>ON</td>
<td>OFF</td>
<td>OFF</td>
<td>ON</td>
<td>ON</td>
<td>OFF</td>
<td>ON</td>
</tr>
<tr>
<td></td>
<td>ON</td>
<td>ON</td>
<td>ON</td>
<td>OFF</td>
<td>OFF</td>
<td>ON</td>
<td>OFF</td>
<td>OFF</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Network</th>
<th>Freq.</th>
<th>prob.</th>
<th>-ln (p)</th>
<th>-2*Freq.*ln (p)</th>
<th>#param</th>
<th>AIC</th>
</tr>
</thead>
<tbody>
<tr>
<td>ON</td>
<td>ON</td>
<td>3</td>
<td>3/5</td>
<td>0.51</td>
<td>3.06</td>
<td></td>
</tr>
<tr>
<td>OFF</td>
<td>ON</td>
<td>2</td>
<td>2/5</td>
<td>0.92</td>
<td>3.67</td>
<td></td>
</tr>
<tr>
<td>OFF</td>
<td>OFF</td>
<td>1</td>
<td>1/3</td>
<td>1.10</td>
<td>2.20</td>
<td></td>
</tr>
<tr>
<td>OFF</td>
<td>OFF</td>
<td>2</td>
<td>2/3</td>
<td>0.41</td>
<td>1.62</td>
<td></td>
</tr>
</tbody>
</table>

12.58 + 14.55 = **27.13**

---

### Ex. use of Information Criterion

- **Network estimation by Bayesian Network**

<table>
<thead>
<tr>
<th>Expression</th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
<th>6</th>
<th>7</th>
<th>8</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>ON</td>
<td>ON</td>
<td>OFF</td>
<td>OFF</td>
<td>ON</td>
<td>ON</td>
<td>OFF</td>
<td>ON</td>
</tr>
<tr>
<td></td>
<td>ON</td>
<td>ON</td>
<td>ON</td>
<td>OFF</td>
<td>OFF</td>
<td>ON</td>
<td>OFF</td>
<td>OFF</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Network</th>
<th>Freq.</th>
<th>prob.</th>
<th>-ln (p)</th>
<th>-2*Freq.*ln (p)</th>
<th>#param</th>
<th>AIC</th>
</tr>
</thead>
<tbody>
<tr>
<td>ON</td>
<td>ON</td>
<td>3</td>
<td>3/4</td>
<td>0.29</td>
<td>1.73</td>
<td></td>
</tr>
<tr>
<td>OFF</td>
<td>ON</td>
<td>1</td>
<td>1/4</td>
<td>1.39</td>
<td>2.77</td>
<td></td>
</tr>
<tr>
<td>ON</td>
<td>OFF</td>
<td>2</td>
<td>2/4</td>
<td>0.69</td>
<td>2.78</td>
<td></td>
</tr>
<tr>
<td>OFF</td>
<td>OFF</td>
<td>2</td>
<td>2/4</td>
<td>0.69</td>
<td>2.78</td>
<td></td>
</tr>
</tbody>
</table>

14.04 + 13.10 = **27.14**
Ex. use of Information Criterion

- Network estimation by Bayesian Network

<table>
<thead>
<tr>
<th>Expression</th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
<th>6</th>
<th>7</th>
<th>8</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>ON</td>
<td>ON</td>
<td>OFF</td>
<td>OFF</td>
<td>ON</td>
<td>ON</td>
<td>OFF</td>
<td>ON</td>
</tr>
<tr>
<td></td>
<td>ON</td>
<td>ON</td>
<td>ON</td>
<td>OFF</td>
<td>OFF</td>
<td>ON</td>
<td>OFF</td>
<td>OFF</td>
</tr>
</tbody>
</table>

Network

25.68

27.13

27.14

Dynamics of organisms

- Organisms are not static, but dynamic
  - Cell cycles
  - Stimulation/stress from outside
  - Environmental Change
  - Cell Differentiation
Dynamics of organisms

- The example was static, no time-line
- Dynamics
  - Cell cycles
  - Stimulation/stress from outside
  - Environmental Change
  - Cell Differentiation

Many Time-course Data in DBs

Dynamic Analysis methods are Proposed

Dynamic Bayesian Network

- Infer the dynamic network
- Assumption
  - Regulation of Gene expression holds Markov Property
    - Nearest future depends only on the present
Network inference by Dynamic Bayesian Network

Markov Property on Discrete-time Model

Time $t$ controls Time $t+1$

$$AIC = 2k - 2 \ln(L)$$

$k$: #parameters
$L$: max value of likelihood function

Network inference by Dynamic Bayesian Network

<table>
<thead>
<tr>
<th>Time</th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
<th>6</th>
<th>7</th>
<th>8</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>ON</td>
<td>ON</td>
<td>OFF</td>
<td>OFF</td>
<td>ON</td>
<td>ON</td>
<td>OFF</td>
<td>ON</td>
</tr>
<tr>
<td></td>
<td>ON</td>
<td>ON</td>
<td>ON</td>
<td>OFF</td>
<td>OFF</td>
<td>ON</td>
<td>OFF</td>
<td>OFF</td>
</tr>
<tr>
<td></td>
<td>1</td>
<td>2</td>
<td>3</td>
<td>4</td>
<td>5</td>
<td>6</td>
<td>7</td>
<td>8</td>
</tr>
</tbody>
</table>

AIC of Green

<table>
<thead>
<tr>
<th>Freq.</th>
<th>prob.</th>
<th>$-\ln(p)$</th>
<th>$-2*\text{Freq.}*\ln(p)$</th>
<th>#param</th>
<th>AIC</th>
</tr>
</thead>
<tbody>
<tr>
<td>ON</td>
<td>ON</td>
<td>3</td>
<td>3/4</td>
<td>0.29</td>
<td>1.73</td>
</tr>
<tr>
<td>OFF</td>
<td>ON</td>
<td>1</td>
<td>1/4</td>
<td>1.39</td>
<td>2.77</td>
</tr>
<tr>
<td>ON</td>
<td>OFF</td>
<td>0</td>
<td>0</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>OFF</td>
<td>OFF</td>
<td>3</td>
<td>1</td>
<td>0</td>
<td>0</td>
</tr>
</tbody>
</table>

AIC = 6.50
**Network inference by Dynamic Bayesian Network**

<table>
<thead>
<tr>
<th>Time</th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
<th>6</th>
<th>7</th>
<th>8</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>ON</td>
<td>ON</td>
<td>OFF</td>
<td>OFF</td>
<td>ON</td>
<td>ON</td>
<td>OFF</td>
<td>ON</td>
</tr>
</tbody>
</table>

- **AIC of Yellow:** 13.10

**Markov Property on Discrete-time Model**

Time $t$ controls Time $t+1$

---

**Network inference by Dynamic Bayesian Network**

<table>
<thead>
<tr>
<th>Time</th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
<th>6</th>
<th>7</th>
<th>8</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>ON</td>
<td>ON</td>
<td>OFF</td>
<td>OFF</td>
<td>ON</td>
<td>ON</td>
<td>OFF</td>
<td>ON</td>
</tr>
</tbody>
</table>

|      | ON| ON| ON| OFF| OFF| ON| OFF| OF |

- **AIC of Yellow:** 6.50

<table>
<thead>
<tr>
<th>Freq.</th>
<th>prob.</th>
<th>-ln (p)</th>
<th>-2*Freq.*ln (p)</th>
<th>#param</th>
<th>AIC</th>
</tr>
</thead>
<tbody>
<tr>
<td>ON</td>
<td>ON 1</td>
<td>1/4</td>
<td>1.39</td>
<td>2.77</td>
<td>2.73 +1.73+</td>
</tr>
<tr>
<td>OFF</td>
<td>ON 3</td>
<td>3/4</td>
<td>0.29</td>
<td>1.73</td>
<td>0 + 2*2 =6.50</td>
</tr>
<tr>
<td>ON</td>
<td>OFF 3</td>
<td>2/4</td>
<td>0</td>
<td>0</td>
<td>2</td>
</tr>
<tr>
<td>OFF</td>
<td>OFF 0</td>
<td>-</td>
<td>-</td>
<td>0</td>
<td>6.50+13.10 = 19.60</td>
</tr>
</tbody>
</table>
Network Inference

AIC values

25.68

19.08

19.60

DBN infers this is the most plausible.

Limitation of DBN

- Purpose of inferring Dynamic networks
  – Bridge the two stable networks

Stable Dynamic Stable

Cell Cycle:  G1/S Checkpoint  S/G2 Checkpoint

Mesenchymal stem cell  Adipo Cell

Cell Differentiation:
Limitation of DBN

- DBN infers only One network
  - Edges are plausible throughout the period

Higher time-resolution is required

Scenario A

Scenario B
Usual Methods

- Separation

- Sliding window

Expression

<table>
<thead>
<tr>
<th></th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
<th>6</th>
<th>7</th>
<th>8</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>ON</td>
<td>ON</td>
<td>OFF</td>
<td>OFF</td>
<td>ON</td>
<td>ON</td>
<td>OFF</td>
<td>ON</td>
</tr>
<tr>
<td></td>
<td>ON</td>
<td>ON</td>
<td>ON</td>
<td>OFF</td>
<td>OFF</td>
<td>ON</td>
<td>OFF</td>
<td>OFF</td>
</tr>
</tbody>
</table>

Disadvantages of Usual Methods

- Separation

- Sliding window

1. Time-resolutions are still rough
2. Less the data, lose the quality of network
   - #time points are small number in biological data.

My Objective: Single-time-point resolution, with minimum loss of data
**Proposed Method**  
**[Key Idea]**

- Use *t* subprofiles without a time point

<table>
<thead>
<tr>
<th>Time</th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>...</th>
</tr>
</thead>
<tbody>
<tr>
<td>subProf.1</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>subProf.2</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>subProf.3</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

**Expression**

<table>
<thead>
<tr>
<th>Time</th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
<th>6</th>
<th>7</th>
<th>8</th>
</tr>
</thead>
<tbody>
<tr>
<td>subProf.1</td>
<td>ON</td>
<td>ON</td>
<td>OFF</td>
<td>OFF</td>
<td>ON</td>
<td>ON</td>
<td>OFF</td>
<td>ON</td>
</tr>
<tr>
<td>subProf.2</td>
<td>ON</td>
<td>ON</td>
<td>ON</td>
<td>OFF</td>
<td>OFF</td>
<td>ON</td>
<td>OFF</td>
<td>OFF</td>
</tr>
</tbody>
</table>

**Relative Plausibility of a network**

<table>
<thead>
<tr>
<th>Time</th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>...</th>
</tr>
</thead>
<tbody>
<tr>
<td>subProf.1</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>subProf.2</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Relative Plausibility at each time

![Diagram](image)
• Smaller the value is, the regulation at the time is more plausible
• It can replace IC value in BN, DBN

Plausibility Value

\[
\text{value} = \text{average} + 2 \times \text{Ave}(\text{AIC}_{g,t})
\]

- \text{AIC}_{g,1} + 2 \times \text{Ave}(\text{AIC}_{g,t})
- \text{AIC}_{g,2} + 2 \times \text{Ave}(\text{AIC}_{g,t})
Experiment for Validation

Detect the time of regulation change
[Material] Diauxic Shift of E. Coli
  - Switch the nutrient source from Glucose to Lactose.
    • 17 Enzymes from KEGG
    • 14 Transcription Factors (RegulonDB)
  - Expression Data
    • GSE7265 by Traxler

Metablic Pathways
Expression Profile

Gene Regulatory Networks

11 Enzymes from Glucose Pathway
7 Enzymes from Lactose Pathway
14 TF from RegulonDB
50 gene regulations
Result: Glucose

Score = Subtraction of Plausibility values with/without regulations.

Result: Lactose &TF

Score = Subtraction of Plausibility values with/without regulations.
Detecting shifts in gene regulatory networks during time-course experiments at single-time-point temporal resolution

Yoichi Takenaka*, Shigeto Seno and Hideo Matsuda
Graduate School of Information Science and Technology
Osaka University
Yamadaoka 1-5, Suita, Osaka, Japan
*takenaka@ist.osaka-u.ac.jp

Accepted 13 September 2015
Published 27 October 2015

Comprehensively understanding the dynamics of biological systems is one of the greatest challenges in biology. Vastly improved biological technologies have provided vast amounts of information that must be understood by bioinformatics and systems biology researchers. Gene regulations have been frequently modeled by ordinary differential equations or graphical models based on time-course gene expression profiles. The state-of-the-art computational approaches for detecting gene regulatory networks that elucidate these time changes are effective at the present time.