Systems Biology in Supercomputing Environment

Satoru Miyano
Human Genome Center, Institute of Medical Science, The University of Tokyo
BioPPN 2011
Thistle County Hotel, Newcastle upon Tyne, UK
June 20, 2011
My interest on systems biology is “Cancer”

Because …
Japanese

Cancer
No. 1

Cause-Specific Death Rate
Japanese

\[
\begin{array}{c}
1 \\
2
\end{array}
\]

Cancer
Japanese

\[ \frac{1}{3} \]

Die of cancer
Hack Cancer System with Supercomputer!

Caution: If you hack supercomputer, you will be arrested.
Cancer scientists had only slit-views on cancer

However, we have some insights on cancer ... 

And, molecular targets are discovered!

Lung cancer: Irressa
Breast cancer: Herceptin
Leukemia: Gleevec

But limited use ...
Farrah's Story: On A Wing & A Prayer

1970

2001 CML (Leukemia)

2006 Anal Cancer

2009 Died

2011 Active

Friday, May 15
9-11PM Pacific Time
NHKアーカイブス ビデオ・コンサート

天に響く歌声
～本田美奈子．情熱のステージ～

1985年（デビュー当時から）2004年（亡くなる前年）までの美奈子さんが心から愛したステージとNHKの番組からセレクトした交響曲です。アイドルからロック・ウィークに至るミス・サイゴンやミュージカル・スター、クラシックの歌姫へと成長していく彼女を描いた映像（50分）。

＜収録曲＞
「冷暖はビートに乘れない」「1986年のラモリ」「愛の誘惑」「新世界」「星に願う」（アメリカン・グラディス）他

日時
- 第1回 8月15日（金）13:00〜
- 第2回 8月15日（金）15:30〜
- 第3回 8月16日（土）13:00〜
- 第4回 8月16日（土）15:30〜

会場
NHKアーカイブス 2F
出会いの広場（シアター）にて

あなたのドナー登録をお待ちしています。

このポスターの制作にあたっては、「オートレース会員センター」の協力を頂きました。
Why not effective to me?

Can we hope new drugs and therapies in the future?
Cancer is Similar to Japanese Bureaucracy System

Angiogenesis
Known mechanisms involved in Angiogenesis is very complex

Model made by using Cell Illustrator
Hallmark of Cancer

D. Hanahan and R. A. Weinberg.
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
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.
But, Very Naïve Understanding and Representation as Systems

- Draw **pictures** and add English **narrations** for biological facts
- “Systems Knowledge” which is unambiguously represented is at most causal relations among molecules
- Huge gap between “what is represented” and “what is to be represented”
Cell Illustrator Online 5.0
Java Web Start Application

http://cionline.hgc.jp/
Hybrid Functional Petri Net with extension (HFPNe)

Extension
- Real (continuous)
- Integer (discrete)
  +
- Object (universal)
  Real, Integer, Boolean, String, vector, etc.


User’s Abilities Required for Cell Illustrator

- Biology
- Mobile phone
- Math at the level of junior high school
Place, Transition, Arc in Petri Net

<table>
<thead>
<tr>
<th>Type</th>
<th>Original symbols of HFPNe</th>
<th>Examples of biological images</th>
</tr>
</thead>
<tbody>
<tr>
<td>Entity</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Process</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Connector</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th></th>
<th>Discrete</th>
<th>Continuous</th>
<th>Generic</th>
<th>Discrete, Continuous, and Generic</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Cell Illustrator (software)

- **Process**: Arrow with label 'Process'
- **Association**: Dashed line with label 'Association'
- **Inhibitory**: Solid line with label 'Inhibitory'

GNI Ltd.
Foundations of Systems Biology
Using Cell Illustrator and Pathway Databases
Series: Computational Biology, Vol. 13
Nagasaki, M., Saito, A., Doi, A., Matsuno, H., Miyano, S.

2009, Approx. 170 p. 145 illus. in color. With CD-ROM., Hardcover
ISBN: 978-1-84882-022-7 26,95 €
2. Knowledge Representation of Dynamic Biopathways

Cell System Ontology (CSO 3.0)

Cell System Markup Language (CSML 3.0)

http://www.csml.org/
CSML 3.0 & CSO 3.0

- Native XML format for Cell Illustrator
- XML and Ontology for Biopathway
  - Modeling
  - Simulation
  - Visualization
Features of CSO 3.0

- System-dynamics centered ontology.
- The ontology is implemented with Web Ontology Language (OWL), which enables semantic validation and provide complete and consistent biological pathway models.


Features of CSO 3.0

- CSO equips mature core vocabularies (more than 350)
- Each core vocabulary has at least one standard icon
Features of CSML 3.0

1. Hybrid Functional Petri net with extension (HFPNe)
2. Logic based descriptions, e.g., temporal logic, can be defined with the format.
3. User can create sub-models for a model with a filtering concept.
4. User can define more than one views for a model, e.g., gene network view, simulation view.
5. All terms in CSML 3.0 has the background of Ontology: Cell System Ontology (CSO) 3.0.

These parts were missing in SBML and CellML. SMBM and CellML are subsets of CSML3.0.
3. Biopathway Layout

- Biologically sophisticated pathway layout algorithms are required
- Cell Illustrator Online 5.0 has more than 20 layout algorithms
Fast Grid Layout Algorithms for Biological Pathways


4. Transforming Pathway Databases for Cell Illustrator

XML Format
- TRANSPATH2CSML
- SBML2CSML
- CellML2CSML

Ontology Format
- BioPAX2CSO
TRANSPATH to CSML

• 16 modeling rules based on Hybrid Functional Petri Net with extension (HFPNe).
• TRANSPATH (Biobase): More than 115,000 cellular events in humans, mice, and rats, collected from over 31,500 publications.
• Petri net element is incorporated with Cell System Ontology (CSO) to enable semantic interoperability of models.
• 97% of the reactions in TRANSPATH are converted into simulation-based models in CSML.

TRANSPATH Pathway Library Module

- More than 1,000 TRANSPATH pathways (Signal Transduction Pathway and Gene Regulatory Network) are supplied. All pathways can be loaded, edited, saved and simulated on CIO.
  - 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.
Pathway Parameter Search Module

Data Assimilation Module

→ Li, Kuroyanagi et al.

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.
Back to Cancer
Cancer Patient

Genetic Variations & Cancer Heterogeneity

Environmental Factors

30-800 Mutations on Genomes
Personal Genomes
Integration of Omics Data

10^3~10^4
Many patient samples

High-Throughput Technology
Microarray, ChIP-Chip, CGH array, SNP array, DNA-Seq, Exome-Seq, RNA-Seq, ChIP-Seq, ...

Can’t reach by run!

Limitations of Molecular Biology

Moon: Innovative cancer therapy, diagnosis, therapy

New Area of Cancer Research
Integrative Analysis
by Top Cancer Scientists

Systems Cancer

Computational Systems Biology
by Supercomputer
Systems Cancer Research Project by MEXT (2010-2014)
Integrative Systems Understanding for Advanced Diagnosis, Therapy and Prevention

http://systemscancer.hgc.jp

Oncology & Cancer Biology
Scientific research on innovative areas
Next-Generation Supercomputer Project
1,150,000,000,000 JPY for 7 years (2006-2012)

- Design, build, and set up the Next-Generation Supercomputer with a speed of 10 PETA FLOPS.
- Over 80,000 nodes & 640,000 cores.

- Kobe (神戸)
Kyoto

Hokkaido

Fukushima 3.11

Tokyo

250Km
Cancer System Hacker

Supercomputer
K computer No.1
International Supercomputing 2011
Hamburg, June 19, 2011

80% K computer (68,544 Nodes)
(8.774PFLOPS at Peak)

LINPACK 8.162PFLOPS (93.0%)
Big Contributor
Supercomputer System for 2009-2014

- **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
How to Hack Cancer Systems

Integrative Systems Understanding of Cancer for Advanced Diagnosis, Therapy and Prevention
How to Hack Cancer Systems

• Digitalizing Heterogeneity/Characteristics of Cancer Systems
• Bayesian Gene Networks of Cancer Systems
• Modeling Dynamics in Cancer Systems with State Space Model
• Comparing Networks in Cancer Systems under Different Conditions
• Extracting Functional Modules in Cancer Systems
• Software Platform – eXtensible integrative Pipeline & Cell Illustrator
1. Digitalizing Cancer Systems Towards Personal Gene Networks

Cancer Heterogeneity and Individual Variation

High-Throughput Technology

Microarray, ChIP-Chip, CGH array, SNP array, DNA-Seq, Exome-Seq, RNA-Seq, ChIP-Seq, …
Patient-Specific Gene Network

Gene Expression Data for Patient A

Gene 1 → Gene 2 → Gene 3 → Gene 4 → Gene 5

Gene 6 → Gene 8 → Gene 9 → Gene 11

Cellular System for Patient A

Drug X

Gene Expression Data for Patient B

Gene 1 → Gene 2 → Gene 3 → Gene 4 → Gene 5

Gene 6 → Gene 8 → Gene 9 → Gene 11

Cellular System for Patient B

Drug Y
Traditional Gene Network Estimation Problem

Knock-down Microarray Data for Sample A

<table>
<thead>
<tr>
<th></th>
<th>sample 1</th>
<th>sample 2</th>
<th>sample 3</th>
<th>....</th>
</tr>
</thead>
<tbody>
<tr>
<td>gene 1</td>
<td>1.5</td>
<td>5.2</td>
<td>1.4</td>
<td>....</td>
</tr>
<tr>
<td>gene 2</td>
<td>5.2</td>
<td>6.3</td>
<td>0.4</td>
<td>....</td>
</tr>
<tr>
<td>gene 3</td>
<td>3.4</td>
<td>9.3</td>
<td>0.3</td>
<td>....</td>
</tr>
<tr>
<td>gene 4</td>
<td>2.9</td>
<td>0.3</td>
<td>6.4</td>
<td>....</td>
</tr>
<tr>
<td>....</td>
<td>....</td>
<td>....</td>
<td>....</td>
<td>....</td>
</tr>
</tbody>
</table>

Gene Network for Sample A

Time-series Microarray Data for Sample A

<table>
<thead>
<tr>
<th></th>
<th>time 1</th>
<th>time 2</th>
<th>time 3</th>
<th>....</th>
</tr>
</thead>
<tbody>
<tr>
<td>gene 1</td>
<td>1.5</td>
<td>5.2</td>
<td>1.4</td>
<td>....</td>
</tr>
<tr>
<td>gene 2</td>
<td>5.2</td>
<td>6.3</td>
<td>0.4</td>
<td>....</td>
</tr>
<tr>
<td>gene 3</td>
<td>3.4</td>
<td>9.3</td>
<td>0.3</td>
<td>....</td>
</tr>
<tr>
<td>gene 4</td>
<td>2.9</td>
<td>0.3</td>
<td>6.4</td>
<td>....</td>
</tr>
<tr>
<td>....</td>
<td>....</td>
<td>....</td>
<td>....</td>
<td>....</td>
</tr>
</tbody>
</table>
Patient-Specific Gene Network Estimation

Microarray Data for Cancer Patients

<table>
<thead>
<tr>
<th></th>
<th>patient 1</th>
<th>patient 2</th>
<th>patient 3</th>
<th>...</th>
</tr>
</thead>
<tbody>
<tr>
<td>gene 1</td>
<td>1.5</td>
<td>5.2</td>
<td>1.4</td>
<td>...</td>
</tr>
<tr>
<td>gene 2</td>
<td>5.2</td>
<td>6.3</td>
<td>0.4</td>
<td>...</td>
</tr>
<tr>
<td>gene 3</td>
<td>3.4</td>
<td>9.3</td>
<td>0.3</td>
<td>...</td>
</tr>
<tr>
<td>gene 4</td>
<td>2.9</td>
<td>0.3</td>
<td>6.4</td>
<td>...</td>
</tr>
<tr>
<td>...</td>
<td>...</td>
<td>...</td>
<td>...</td>
<td>...</td>
</tr>
</tbody>
</table>

Gene Network for Patient A

Gene Network for Patient B
Corelation between 2 genes

No relation?
Change can be found if we look with a modulator
Finding Gene Relation by Data Stratification

\[ X_2 = -0.92 \times X_1 + \epsilon_2 \]

Modulator A is low

\[ X_2 = -0.07 \times X_1 + \epsilon_2 \]

Modulator A is high

Stratification of Data
\[ X_A \perp X_B \mid X_{\Gamma \backslash \{A, B\}} \]
Concept of NetworkProfiler

Usual graphical model

\[ X_A \perp X_B \mid X_{\Gamma \setminus \{A,B\}} \]

Modulator-dependent graphical model
(NetworkProfiler)

Modulator: Cofactor modulating relationship between genes A and B
Examples of Modulator

- Tumor progression (Stage I, Stage II, …)
- Drug sensitivity (IC50, GI50, …)
- Disease-free survival
- Molecular characteristics (Metastasis, EMT…)
- Pathway activity
- …
What is System?

Node: gene transcript
Edge: conditional dependence (not equal to correlation)

Co-expression Network

Structural Equations

\[
\begin{align*}
    x_1 &= 4.6 \times x_5 + 4.2 \times x_6 + 0.7 \times x_7 + \epsilon_1 \\
    x_2 &= -2.7 \times x_1 + 0.95 \times x_8 + \epsilon_2 \\
    x_3 &= -1.4 \times x_1 + \epsilon_3 \\
    x_4 &= -0.5 \times x_1 + \epsilon_4
\end{align*}
\]

Gene Network

Samples
NetworkProfiler

\( p \text{ (genes)} \times n \text{ (patients)} \) gene expression data matrix

\( x_{aj} \); expression value of \( j \)-th gene for \( \alpha \)-th patient

\( m_\alpha \); modulator value for \( \alpha \)-th patient

**Structural equation model of \( k \)-th gene for \( \alpha \)-th patient**

\[
x_{\alpha k} = \sum_{j=0, j \neq k}^{p} \beta_{jk\alpha} x_{aj} + \varepsilon_{\alpha k}, \quad \beta_{jk\alpha} = \beta_{jk}(m_\alpha)
\]

**\( \alpha \)-th Patient**

\[
X_1 = \varepsilon_1 \\
X_2 = \varepsilon_2 \\
X_3 = 0.3X_1 + 1.1X_2 + \varepsilon_3 \\
X_4 = -0.9X_3 + 2.4X_6 + \varepsilon_4 \\
X_5 = \varepsilon_5 \\
X_6 = -0.9X_5 + \varepsilon_6
\]
NetworkProfiler

$p \times n$ gene expression data matrix

$x_{bj} \beta$; expression value of $j$-th gene for $\beta$-th patient

$m_{\beta}$; modulator value for $\beta$-th patient

**Structural equation model of $k$-th gene for $\beta$-th patient**

$$x_{\beta k} = \sum_{j=0, j \neq k}^{p} \beta_{jk\beta} x_{bj} + \epsilon_{\beta k}, \quad \beta_{jk\beta} = \beta_{jk}(m_{\beta})$$

**6-th Patient**

$$X_1 = \epsilon_1$$

$$X_2 = -0.1X_3 + \epsilon_2$$

$$X_3 = 0.3X_1 + \epsilon_3$$

$$X_4 = -0.5X_3 + \epsilon_4$$

$$X_5 = \epsilon_5$$

$$X_6 = -0.3X_5 + 1.1X_2 + \epsilon_6$$
Concept of NetworkProfiler

**Purpose:** estimation of *coefficients* for $\alpha$-th sample $\beta_{jk\alpha}$

**Problem:** microarray for $\alpha$-th sample is only one!!

**Idea:** data integration by *sample weighting*

Network Estimation
Concept of NetworkProfiler

**Purpose:** estimation of *coefficients* for $\beta$-th sample $\beta_{jk\beta}$

**Problem:** microarray for $\beta$-th sample is only one!!

**Idea:** data integration by *sample weighting*
NetworkProfiler

Structural equation model of $k$-th gene for $\alpha$-th patient

$$x_{ck} = \sum_{j=0, j \neq k}^{p} \beta_{jk\alpha} x_{cj} + \varepsilon_{ck}, \quad \beta_{jk\alpha} = \beta_{jk}(m_{\alpha})$$

Elastic net-type regularized **weighted** loss function

$$S(\beta_{1k\alpha}, \ldots, \beta_{pk\alpha} | m_{\alpha}) = \sum_{i=1}^{n} \left[ K_{h}(m_{i} - m_{\alpha}) \left\{ x_{ik} - \sum_{j \neq k}^{p} \beta_{jk\alpha} x_{ij} \right\} \right]^{2} + \lambda_{1} \sum_{j \neq k}^{p} w_{jk\alpha} | \beta_{jk\alpha} | + \frac{\lambda_{2}}{2} \sum_{j \neq k}^{p} \beta_{jk\alpha}^{2}$$

$K_{h}(m_{i} - m_{\alpha})$: Gaussian kernel function ($m_{\alpha}$: center, $h$: width)

$\lambda_{1}, \lambda_{2}$: regularization parameters

$$\hat{\beta}_{k\alpha} = \left( \hat{\beta}_{1k\alpha}, \ldots, \hat{\beta}_{pk\alpha} \right)^{T} = \arg\min_{\beta_{k\alpha}} S(\beta_{1k\alpha}, \ldots, \beta_{pk\alpha} | m_{\alpha})$$

The neighborhood samples in terms of modulator have **similar network structure**
Model Selection

Performance of NetworkProfiler $\rightarrow$ Selection of $\lambda_1$, $\lambda_2$, and $h$

- Select $\lambda_1$ and $\lambda_2$ based on WAICc (Shimamura et al., 2010b)
- Select $h$ based on cross-validation

\[
S^{(-i)}(\beta_{1k\alpha}, ..., \beta_{1k\alpha} \mid m_\alpha) = \sum_{i \neq \alpha} K_h (m_i - m_\alpha) \left\{ x_{ik} - \sum_{j \neq k} \beta_{jk\alpha} x_{ij} \right\}^2 + \lambda_1 \sum_{j \neq k} w_{jk\alpha} |\beta_{jk\alpha}| + \frac{\lambda_2}{2} \sum_{j \neq k} \beta^2_{jk\alpha}
\]

\[
\hat{\beta}^{(-i)}_{k\alpha} = \arg\min \left\{ S^{(-i)}(\beta_{1k\alpha}, ..., \beta_{1k\alpha} \mid m_\alpha) \right\}
\]

\[
CV_k(h) = \sum_{\alpha=1}^n \left\{ x_{ck} - \sum_{j \neq k} \hat{\beta}^{(-\alpha)}_{jk\alpha} x_{cj} \right\}^2
\]

Select $h$ minimizing $CV_k(h)$
Epithelial-Mesenchymal Transition (EMT)

- Key developmental remodeling program, where cells alternate between Epithelial-like and Mesenchymal-like phenotypes
- Relate to tumor grade and metastasis
- Contribute to increasing in drug resistance

System related to EMT is a “black box”
We selected coherent 50 genes from 121 EMT signature genes to define the modulator for EMT (EEM, Niida et al., 2009)
We selected coherent 50 genes from 122 EMT signature genes to define the modulator for EMT (EEM, Niida et al., 2009)

Signature-based hidden modulator extraction algorithm
2. Then, narrowed the set to 50 functionally coherent genes with $p < 10^{-5}$ by using the extraction of expression module (EEM).
3. Computed the first principal component of these 50 genes as hidden values of the EMT-related modulator.
Elucidating Systems Responsible for EMT

Input
- Transcriptome data of 762 cancer cells
  (22,283 transcripts = 22,277 mRNA + 581 miRNA)
- EMT Modulator

Output
762 structural varying systems with 22,283 genes related to EMT

Computation with 1,024 cores (12.3 TFLOPS)
3 months

Compute how networks change from low to high
miRNAs highly expressed in Epithelial

ZEB1, SIP1 (ZEB2) (E-cadherin Inhibitor) Gregory et al., Nature Cell Biology, 2008
System Changes Related to EMT

- Functional loss of E-cadherin = a hallmark of EMT
- Focus on regulators of E-cadherin

Epithelial-Like Cell → Mesenchymal-Like Cell

E-cadherin → ZEB1

Expression profiles of ZEB1

Regulation profiles of ZEB1 → CDH1
Upstream Regulatory Changes of E-cadherin

Coefficient profiles of E-cadherin regulators through EMT

Low (Epithelial)  |  EMT-related modulator  |  High (Mesenchymal)

- miR-141
- GRHL2
- miR-135b
- LSR
- NAT8
- miR-211
- IRF6
- OVOL2
- TCF3
- ZEB2
- ZEB1
- DAXX

Coefficient from miR-141 to E-cadherin

miR-141  →  E-cad  →  EMTness
Table 1. 25 top-ranked regulators of E-cadherin for the change in the regulatory effect change among the EMT with published evidence

<table>
<thead>
<tr>
<th>regulator</th>
<th>type</th>
<th>regulatory effect change</th>
<th>Evidence</th>
</tr>
</thead>
<tbody>
<tr>
<td>IRF6</td>
<td>A</td>
<td>101.04</td>
<td>[8]</td>
</tr>
<tr>
<td>miR-141</td>
<td>A</td>
<td>87.58</td>
<td>[8]</td>
</tr>
<tr>
<td>GRHL2</td>
<td>A</td>
<td>64.13</td>
<td></td>
</tr>
<tr>
<td>ZEB1 (SIP1)</td>
<td>I</td>
<td>50.72</td>
<td>[9]</td>
</tr>
<tr>
<td>LSR</td>
<td>I</td>
<td>46.89</td>
<td></td>
</tr>
<tr>
<td>miR-200b</td>
<td>A</td>
<td>31.55</td>
<td>[8]</td>
</tr>
<tr>
<td>KLF4</td>
<td>A</td>
<td>26.28</td>
<td>[10]</td>
</tr>
<tr>
<td>OVOL2</td>
<td>A</td>
<td>22.08</td>
<td></td>
</tr>
<tr>
<td>miR-200a</td>
<td>A</td>
<td>17.70</td>
<td>[8]</td>
</tr>
<tr>
<td>FOXA2</td>
<td>A</td>
<td>17.26</td>
<td>[11]</td>
</tr>
<tr>
<td>TCF4 (E2.2)</td>
<td>I</td>
<td>14.15</td>
<td>[12]</td>
</tr>
<tr>
<td>ELF3</td>
<td>A</td>
<td>13.58</td>
<td></td>
</tr>
<tr>
<td>ZNF217</td>
<td>A</td>
<td>13.53</td>
<td></td>
</tr>
<tr>
<td>MYB</td>
<td>A</td>
<td>12.50</td>
<td></td>
</tr>
<tr>
<td>KLF5</td>
<td>A</td>
<td>12.42</td>
<td></td>
</tr>
<tr>
<td>miR-192</td>
<td>A</td>
<td>12.30</td>
<td>[13, 14]</td>
</tr>
<tr>
<td>FOXA1</td>
<td>A</td>
<td>11.69</td>
<td>[11]</td>
</tr>
<tr>
<td>ZNF165</td>
<td>A</td>
<td>11.39</td>
<td></td>
</tr>
<tr>
<td>NKK2-1</td>
<td>A</td>
<td>11.21</td>
<td></td>
</tr>
<tr>
<td>HNF1B</td>
<td>A</td>
<td>11.08</td>
<td></td>
</tr>
<tr>
<td>TFE3</td>
<td>A</td>
<td>11.01</td>
<td></td>
</tr>
<tr>
<td>ZEB2 (δEF)</td>
<td>I</td>
<td>10.66</td>
<td>[15]</td>
</tr>
<tr>
<td>TRIM29</td>
<td>I</td>
<td>9.87</td>
<td></td>
</tr>
<tr>
<td>SNAI2</td>
<td>I</td>
<td>9.74</td>
<td>[16]</td>
</tr>
</tbody>
</table>

A: Activator
I: Inhibitor
Upstream Regulatory Changes of E-cadherin

(a). epithelial-like cell

(b). mesenchymal-like cell

EMT

miR-141  ZEB1  E-cadherin

miR-141  ZEB1  E-cadherin

(c). The green and red colors indicate epithelial- and mesenchymal-like cells, respectively.
miR-141, ZEB1, and E-cadherin

- NetworkProfiler revealed regulatory changes in miR-141, ZEB1, and E-cadherin.
- Specifically, it suggested that decreased expression of miR-141 in mesenchymal cells disrupts the negative feedback loop between miR-141 and ZEB1, which would allow ZEB1 to decrease the expression of E-cadherin during the EMT.
• We predicted 45 EMT-dependent putative master regulators that control sets of genes involved in cell adhesion, migration, invasion and metastasis, namely,

• 17 of which are in the downstream of TGFB1, a master switch of the EMT, in our prediction.
Table 1. 25 top-ranked regulators of E-cadherin for the change in the regulatory effect change among the EMT with published evidence

<table>
<thead>
<tr>
<th>regulator</th>
<th>type</th>
<th>regulatory effect change</th>
<th>Evidence</th>
</tr>
</thead>
<tbody>
<tr>
<td>IRF6</td>
<td>A</td>
<td>101.04</td>
<td></td>
</tr>
<tr>
<td>miR-141</td>
<td>A</td>
<td>87.58</td>
<td>[8]</td>
</tr>
<tr>
<td>GRHL2</td>
<td>A</td>
<td>64.13</td>
<td></td>
</tr>
<tr>
<td>ZEB1 (SIP1)</td>
<td>I</td>
<td>50.72</td>
<td>[9]</td>
</tr>
<tr>
<td>LSR</td>
<td>I</td>
<td>46.89</td>
<td></td>
</tr>
<tr>
<td>miR-200b</td>
<td>A</td>
<td>31.55</td>
<td>[8]</td>
</tr>
<tr>
<td>KLF4</td>
<td>A</td>
<td>26.28</td>
<td>[10]</td>
</tr>
<tr>
<td>OVOL2</td>
<td>A</td>
<td>22.08</td>
<td></td>
</tr>
<tr>
<td>miR-200a</td>
<td>A</td>
<td>17.70</td>
<td>[8]</td>
</tr>
<tr>
<td>FOXA2</td>
<td>A</td>
<td>17.26</td>
<td>[11]</td>
</tr>
<tr>
<td>TCF4 (E2.2)</td>
<td>I</td>
<td>14.15</td>
<td>[12]</td>
</tr>
<tr>
<td>ELF3</td>
<td>A</td>
<td>13.58</td>
<td></td>
</tr>
<tr>
<td>ZNF217</td>
<td>A</td>
<td>13.53</td>
<td></td>
</tr>
<tr>
<td>MYB</td>
<td>A</td>
<td>12.50</td>
<td></td>
</tr>
<tr>
<td>KLF5</td>
<td>A</td>
<td>12.42</td>
<td></td>
</tr>
<tr>
<td>miR-192</td>
<td>A</td>
<td>12.30</td>
<td>[13, 14]</td>
</tr>
<tr>
<td>FOXA1</td>
<td>A</td>
<td>11.69</td>
<td>[11]</td>
</tr>
<tr>
<td>ZNF165</td>
<td>A</td>
<td>11.39</td>
<td></td>
</tr>
<tr>
<td>NKK2-1</td>
<td>A</td>
<td>11.21</td>
<td></td>
</tr>
<tr>
<td>HNF1B</td>
<td>A</td>
<td>11.08</td>
<td></td>
</tr>
<tr>
<td>TFE3</td>
<td>A</td>
<td>11.01</td>
<td></td>
</tr>
<tr>
<td>ZEB2 (δEF)</td>
<td>I</td>
<td>10.66</td>
<td>[15]</td>
</tr>
<tr>
<td>TRIM29</td>
<td>I</td>
<td>9.87</td>
<td></td>
</tr>
<tr>
<td>SNAI2</td>
<td>I</td>
<td>9.74</td>
<td>[16]</td>
</tr>
</tbody>
</table>

A: Activator
I: Inhibitor
A novel regulator KLF5 of EMT

- Krueppel-like factor 5 (KLF5) from a list of the remaining candidate regulators and conducted *in vitro* validation experiments.

- As a result, we found that knockdown of KLF5 by siRNA significantly decreased E-cadherin expression and induced morphological changes characteristic of EMT.
We define “relapse risk score” by using relapse makers
Patient Samples (Microarray Data)
Each patient network is computed

Network of the lowest risk patient

Network of the highest risk patient
Differences of Hubness Suggest Key Genes
System-Oriented Personalized Medicine

EMT/Prognosis/Metastasis/etc.
Thorough Analysis Requires 500 Times

25 years with 1000 cores
10 days with 1,000,000 cores
Prologue

Mapmakers in Systems Biology
伊能忠敬 Tadataka Inoh
A man who walked 40,000,000 steps

・伊能忠敬は、江戸幕府の事業として、1800年から1816年にかけて全国を歩いて測量をし、1821年に「大日本沿海輿地全図」が幕府に納められたといいます。伊能忠敬はその完成を見ずに1818年に死去しましたが、その後、仕上げの編纂作業が行われ、全部で21年の歳月をかけてこの地図は完成しています。
2001年のNHKの正月時代劇「四千万歩の男・伊能忠敬」（原作：井上ひさし「四千万歩の男」（講談社））として放映
The Mapmakers

The story of the great pioneers in cartography – from antiquity to the space age

• John Noble Wilford
2. Mining Gene Networks from Gene Expression Profiles

Mapmaking of Molecular Networks

- Bayesian Network Gene Networks
  - Gene knock-down/knock-out
  - Various shocks
  - Time-course data

Yoshinori Tamada; Seiya Imoto; Masao Nagasaki; Satoru Miyano
2. Bayesian Networks + Nonparametric Regression

- We want to estimate gene networks from high-throughput biological data e.g. gene expression data.

| Gene   | KD1   | KD2   | KD3   | ...
|--------|-------|-------|-------|-------
| Gene 1 | 1.45  | -1.54 | 1.23  | ...
| Gene 2 | 3.21  | -2.1  | 1.44  | ...
| ...    | ...   | ...   | ...   | ...

Gene Expression Data
What we wanted to do

Microarray gene expression data

Gene Knockdown/Knockout
Time-Course Measurement

Gene network
Bayesian Network and Nonparametric Regression

Network of 521 genes constructed from 120 yeast microarrays obtained by disrupting 120 genes, where 78 of them are transcription factors.

Nonlinear Bayesian network model

\[
f(x_{i1}, \ldots, x_{ip}; \theta_G) = \prod_{j=1}^{p} f_j(x_{ij} | p_{ij}; \theta_j),
\]

\[
f_j(x_{ij} | p_{ij}; \theta_j) = \frac{1}{\sqrt{2\pi\sigma_j^2}} \exp \left\{ -\frac{(x_{ij} - \mu_{ij})^2}{2\sigma_j^2} \right\}
\]

\[
\mu_{ij} = m_1(p_{i1}^{(j)}) + \cdots + m_{q_j}(p_{iq_j}^{(j)})
\]

\[
= \sum_{k=1}^{q_j} \sum_{m=1}^{M_{jk}} \gamma_{mk} b_{mk}^{(j)}(p_{ik}^{(j)})
\]
Criterion for Selecting Good Networks

**BNRC Score**

*Bayesian Network and Nonparametric Regression Criterion*

\[
\text{BNRC}(G) = -2 \log \pi_G \int \prod_{i=1}^{n} f(x_i; \theta_G) \pi(\theta_G | \lambda) d\theta_G
\]

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

\[
+ \log |J_\lambda(\hat{\theta}_G)| - 2nl_\lambda(\hat{\theta}_G | X_n)
\]

We choose the graph that minimizes the value of the BNRC score.
A Series of Programs on Supercomputer for mining gene networks of size from 30 to 20,000 (genome-wide)

Optimal to Locally Optimal

WR: Optimal Bayesian Networks of 32 Nodes
April 2010
Parallel Computing with 8192 Cores

Genome-Wide Bayesian Network Computation
SiGN
Extracting Dynamic Changes in Cancer Gene Network by Dynamic Bayesian Network

Joint Probability by a DAG

\[ \pi(G | X) \propto \pi(G) \prod_{i=1}^{n} f(x_{i1}, ..., x_{ip} | \theta_G) \pi(\theta_G) \]

= BNRC(G): Network Score

\[ G : \text{Network (DAG)} \]

Regression data

Nonparametric regression by B-spline

\[ y_{ij} = \sum_{m} m_{ij}(p_{ij}^{(m)}) + \varepsilon_{ij}, \varepsilon_{ij} \sim N(0, \sigma_j^2) \]

Network Estimation

\[ G = \arg \min_G \text{BNRC}(G) = \arg \min_G \sum_j \text{BNRC}_j(G) \]

\[ \text{BNRC}_j(G) = 2(q_j + 1) - \left( \sum_{\lambda_j} M_{\lambda j} + 1 \right) \log \left( \frac{2\pi}{n} \right) + n \log(2\pi \hat{\sigma}_j^2) + n \]

+ \[ \sum_{\lambda_j} \log \left| \lambda_j \log(\hat{\sigma}_j^2) \right| - \log(2\hat{\sigma}_j^2) \]

+ \[ \sum_{\lambda_j} \left( M_{\lambda j} - 2 \right) \log \left( \frac{2\pi \hat{\sigma}_j^2}{n\hat{\beta}_{\lambda j}} \right) + \log | \mathbb{K}_{\lambda j} | + \frac{n\hat{\beta}_{\lambda j}}{\hat{\sigma}_j^2} \gamma_{\lambda j} \]

Mathematics is harmful to your health.
Anti-cancer Drug Response Gene Network of Melanoma

Dynamic Bayesian + Nonparametric Regression

- Melanoma A-375 + Taxol (Paclitaxel)
- Inferring and chasing the changes of gene networks of 2,000 genes for 24 hrs at 14 time points (triplicate at each time)
- 1 hour computation using 1024 cores
The size of a node corresponds to the hub size

RBM23

A gene making a complex with Tubulin alpha-4A chain, a target of Taxol.
Known as a Taxol resistance gene in breast cancer
EGR1

TXNIP
Known as a Taxol resistant gene in breast cancer.
3. Gene Networks of Small Airway Epithelial Cell and Gefitinib - State Space Model -

Growth Factor Signaling Systems Identify Critical Genes for Survival Prediction in Early Stage Lung Adenocarcinoma

Yamaguchi, R., Imoto, S., Yamauchi, M., Nagasaki, M., Yoshida, R., Shimamura, T., Hatanaka, Y., Ueno, K., Higuchi, T., Gotoh, N., Miyano, S.
Gene Expression Profiles

State Space Model:
\[
\begin{align*}
    x_n &= Fx_{n-1} + v_n, & x_n \in \mathbb{R}^k \\
    y_n &= Hx_n + w_n, & y_n \in \mathbb{R}^p
\end{align*}
\]

High-Dimensional Short Time Series Data:
\[
Y = \{y_1, \cdots, y_{N_{\text{obs}}} \} \quad N_{\text{obs}} << p \approx 10^3
\]

System Estimation with Dimension Reduction:
\[
\dim(x_n) = k < \dim(y_n) = p
\]
\[
\begin{align*}
    L(\theta) &= \int p(Y, X | \theta) dX \\
    H^T R^{-1} H &= \Lambda = \text{diag}(\lambda_1, \cdots, \lambda_k)
\end{align*}
\]

Gene Expression Prediction:
\[
y_{n_{\text{pred}}} = H \int x_n p(x_n | y_1, \cdots, y_{n-1}, \theta) dx_n
\]

Module-Based Gene Network Estimation:
\[
R^{-1/2}(y_n - w_n) = \Psi R^{-1/2}(y_{n-1} - w_{n-1}) + R^{-1/2} H v_{121}
\]
Epidermal Growth Factor Receptor Pathway
**State Space Model** for Inferring Transcriptional Module Networks from Time-Course Gene Expression Data

**SSM**

\[
x_n = F x_{n-1} + v_n, \quad n \in \mathcal{N}, \\
y_n = H x_n + w_n, \quad n \in \mathcal{N}_{\text{obs}},
\]

**SSM(\theta)** \[ \theta = \{H, F, R, x_0\} \]

Time-Series Data

Short- and High-dimensional Data

\[ Y_{\text{obs}} = \{y_n\}, \quad n \in \mathcal{N}_{\text{obs}} \]

Module Network

Gene Network

Prediction by simulation
SSM Application
Predicting Differences in Gene Regulatory Systems

• Focus: EGFR pathway
• Data
  – Time Course Gene Expression Microarray Data (20K)
    • 19 time points during 48 hours
  – Human Small Airway Epithelial Cell (SAEC)
  – Two Conditions (Different Drugs)
    • EGF Stimulation (Control Data)
    • EGF + Gefitinib Dosed (Case Data)
SAEC #52 mRNA

Starvation Start
Gefitinib (0.5 µM) Treatment
EGF (100 ng/ml) Stimulation
RNA Sampling

(-) : 0 hr ~ 48 hr
EGF : 0 hr ~ 48 hr
EGF+Gefitinib : 0 hr ~ 48 hr
Gefitinib : 0 hr ~ 9 hr

Cell Morphology
(-) 48 hr  EGF 48 hr  EGF+Gefitinib 48 hr
Selection 1500 Genes for Network Analysis

**a**

- Starvation Start
- EGF (100 ng/ml) Stimulation
- Geﬁtinib (0.5 µM) Treatment
- RNA Sampling

**b**

- 19,633 genes
  - Signiﬁcantly induced-genes by EGF stimulation: 43 genes
  - Up (or down) regulated-genes (more than 1.5 fold) at a certain time point within 9 hr after EGF stimulation: 470 genes
  - Literature based knowledge: 987 genes

# Remove genes which have a large margin of error, same expression pattern between EGF+geﬁtinib and geﬁtinib-treated samples

- 1,500 genes
Time-Course of Drug-Stimulated Human Normal Lung Cell

Cell: SAEC (Human Small Airway Epithelial Cell)
Conditions: Control  EGF (mimic of Cancer Cells)
          Case    EGF + Gefitinib (mimic of Cancer Cells + Drug)
Time-Course: 19 Time Points in 48 hours

Gene Selection for SSM Analysis
Selection of System Dimension: k
Parameter Estimation: θ
Meta Analysis of Transcriptional Modules
Construction of Gene Network

1500 Genes: Literature DB (Ingenuity) + Variation Filter (variance)

Learning SSM with EGF data (Control’s System)
k = 8 minimizes prediction errors for hold-out sample.

Prediction of EGF-Gefitinib data by EGF-Learned SSM
Obtaining p-value for each Gene

Genes with small p-values are considered to be induced differential regulations by Gefitinib
Differentially-Expressed and Differentially-Regulated Genes

Time-Course Gene Expression Profiles from a Case-Control Experiment.

What kind of systems are behind differential expressions?
Differential Regulations by Drug Dosing

**Situation 1: Identical System**
Case & Control System

Case and Control cells have a common system (common regulators of g5).

![Gene expression graph](Time)

**Situation 2: Differential Regulation**
Control System & Case System

Regulatory systems of g5 (regulators of g5) are different in Case and Control cells.

![Gene expression graph](Time)

**Comparison**

**Structural Change**

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.

We use a statistical model for inferring gene regulatory systems.
Strategy to Predict Differentially Regulated Genes

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

\[ x_n = Fx_{n-1} + v_n \]
\[ y_n = Hx_n + w_n \]
\[ Y_{N_{obs}}^{CTRL} = [y_1^{CTRL}, \ldots, y_{N_{obs}}^{CTRL}] \]
\[ \hat{\theta}^{CTRL} = [\hat{\Theta}^{CTRL}, \hat{F}^{CTRL}, \hat{R}^{CTRL}] \]

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

\[ \hat{\Theta}^{CTRL} \]
\[ y_{n-1}^{CASE} = [y_1^{CASE}, \ldots, y_{n-1}^{CASE}] \]
\[ y_{n-1}^{CASE} \]

Predicted Gene Expression at n

3. Compare Predicted and Observed Time Courses

Candidates of Differentially Regulated Genes:
Unpredictable genes in the Case data by the Control’s SSM.
Prediction of EGF Data and EGF-GFT Data by SSM of EGF System

EGF Data Prediction: Check for Model Accuracy
EGF+GFT Data Prediction: Exploration for Differences between Systems

Good Prediction

1. Small prediction error (Integrated p value high)
   - Observed EGF data
   - Predicted by EGF SSM
   - Prediction error: \( \epsilon_{i,n}^{\text{EGF}} \)
   - Successful modeling

Bad Prediction

2. Large prediction error (Integrated p value low)
   - Observed EGF data
   - Predicted by EGF SSM
   - Prediction error: \( \epsilon_{i,n}^{\text{EGF}} \)
   - Unsuccessful modeling

Diff. Reg. Gene

3. Small prediction error (Integrated p value high)
   - Observed EGF-GFT data
   - Predicted by EGF SSM
   - Prediction error: \( \epsilon_{i,n}^{\text{EGF-GFT}} \)
   - Gefitinib insensitive

4. Large prediction error (Integrated p value low)
   - Observed EGF-GFT data
   - Predicted by EGF SSM
   - Prediction error: \( \epsilon_{i,n}^{\text{EGF-GFT}} \)
   - Gefitinib sensitive
Examples: Selected Diff Reg Genes

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

Differentially Regulated Genes

Similarly Regulated Genes
Estimated Gene Networks with SSMs

\[ R^{-1/2}(y_n - w_n) = \Psi R^{-1/2}(y_{n-1} - w_{n-1}) + R^{-1/2}Hv_n \]

Diff gene (gene 162)

Control (EGF)

Case (EGF+GFT)

Predicted disappearance of parent nodes by drug dosing

Diff gene (gene 192)
State Space Model computed from Melanoma + Taxol

The same gene is identified

Observation

Difference

Simulation result

24 h
Gene expression–based survival prediction in lung adenocarcinoma: a multi-site, blinded validation study

Although prognostic gene expression signatures for survival in early-stage lung cancer have been proposed, for clinical application, it is critical to establish their performance across different subject populations and in different laboratories. Here we report a large, training–testing, multi-site, blinded validation study to characterize the performance of several prognostic models based on gene expression for 452 lung adenocarcinomas. The hypothesis–based approach examined whether microarray measurements of gene expression alone or combined with basic clinical covariates (stage, age, sex) could be used to predict overall survival in lung cancer subjects. Several models examined produced risk scores that substantially correlated with actual subject outcome. Most methods performed better with clinical data, supporting the combined use of clinical and molecular information when building prognostic models for early-stage lung cancer. This study also provides the largest available set of microarray data with extended pathological and clinical annotation for lung adenocarcinomas.

In the United States and in many Western countries, lung cancer represents the leading cause of cancer-related death. The 5-year overall survival rate is 15% and has not improved over many decades. This is mainly because approximately 20% of lung cancers are discovered at an advanced stage, for which cure by surgical resection is no longer an option. Furthermore, even among early-stage patients who are treated primarily by surgery with curative intent, 30–50% will develop and die of metastatic recurrence. Recent multinational clinical trials (IACT, IPASS, ANITA UPT, ACE) conducted in several continents have demonstrated that adjuvant chemotherapy significantly improves the survival of patients with early-stage (I-II) disease.

Nevertheless, it is clear that a proportion of patients with stage I disease have poorer prognosis and may benefit significantly from adjuvant chemotherapy, whereas some with stage II disease with relatively good prognosis may not benefit significantly from adjuvant chemotherapy. It remains possible, however, that the latter patients could derive additional benefit from advanced targeted therapies. Therefore, there is an urgent need to establish new diagnostic paradigms and subclasses to clinical trials methods for improving the selection of stage I–II patients who are most likely to benefit from adjuvant chemotherapy.

Global gene-expression profiling using microarray technologies has helped to improve our understanding of the molecular heterogeneity of non-small cell lung cancer (NSCLC) and has identified potential biomarkers and gene signatures for classifying patients with significantly different survival outcomes. However, the performance and general applicability of published classifiers has not been easy to establish because of small numbers of subjects examined and inclusion of heterogeneous tumor types. Furthermore, there have not been uniform criteria for sample inclusion, annotation, sample processing and data analysis. To address these concerns and to generate a large microarray database of NSCLC samples that have been collected and studied using a common protocol, we conducted a large retrospective, multi-site, blinded study. The study included a blinded validation step to characterize the performance of several newly developed prognostic models using a total of 452 lung adenocarcinomas, the specific type of lung cancer that is increasing in incidence.

To ensure scientific validity of the results, subject samples along with all relevant clinical, pathological and outcome data were collected by investigators at four institutions using data from six lung-cancer treatment sites with subject inclusion criteria defined a priori. Gene
Survival Predictions for Lung Cancer Patients with Gene Sets Identified by SSM analysis for SAEC data

1. DRGs Identified from SAEC data by SSM

2. Lung Cancer Patients Data
   (Shedden et al. 2008)

3. Classifier Construction with Expressions of the DRGs of Patients in Training Sample

4. Stratifications of Test Patients by Predicted Risk Scores from the Classifier

Q: Can Expressions of DRGs in Lung Cancer Patients Predict Prognosis? (Esp. for Stage I)

DRG: Differentially Regulated Gene
Classifier: Risk Score Function

• **Partial Cox Regression Model**

  – Hazard Function:  \( \lambda(t, X_i) = \lambda_0(t) \exp[f(X_i)] \)

  – Risk Score Function

  \[
  f\left(X_i^{\text{Valid}}\right) = \sum_{j=1}^{p} \beta_j^{*\text{Train}} \left( X_{ij}^{\text{Valid}} - \bar{X}_j^{\text{Train}} \right)
  \]

  where

  \[
  X_i^{\text{Valid}} = \begin{bmatrix} X_{i1}^{\text{Valid}} & \ldots & X_{ip}^{\text{Valid}} \end{bmatrix}^T
  \]

  The p gene expressions for the \( i \)th patient in a validation set

  The \( i \)th patient is classified into a high risk group when

  \[ f(X_i^{\text{Valid}}) > 0 \]

  or a low risk group when

  \[ f(X_i^{\text{Valid}}) \leq 0 \]

  Li and Gui, 2004
Survival Predictions for Lung Cancer Patients with Gene Sets Identified by SSM analysis for SAEC data

- **All Stages with 277 genes**
- **Stage 1 with 277 genes**
- **Stage 1 with 139 genes**
- **Stage 1 with 139 genes and Covariates**

![Graphs showing survival predictions for different stages with different gene sets and covariates.](image)
4. Gene Networks of Lung Cancer

Gene Networks of Gefitinib Sensitive PC9 and Gefitinib Resistant PC9GR2
PC9GR2
(Lung Cancer Cell Line with Gefitinib resistance)

Andre Fujita: Statistician
Yoshinori Tamada: Computer Scientist
Dynamic Bayesian Network + Nonparametric Regression

Gefitinib
5. Building Data Analysis and Simulation Pipeline at ONE STOP

XiP (eXtensible integrative Pipeline)
Systems Biology integrative Pipeline

NetComparator analysis pipeline
NetComparator


Regularized Weighted Recursive Elastic Net
XiP

State Space Model
SiGN
NetComparator
MetaGeneProfiler

EEM & CNV
Network Profiler
Cell Illustrator R & Bioconductor
Statistical Genetics

Supercomputer
Difference between Gefitinib Sensitive Lung Cancer and Resistant Lung Cancer

E2F1 is known as a TF regulating apoptosis and cell cycle.

Gefitinib Sensitive Lung Cancer

Gefitinib Resistant Lung Cancer

NCOA3: a nuclear receptor coactivator that interacts with nuclear hormone receptors to enhance their transcriptional activator functions.
A New Paradigm for Understanding Cancer

- Super-early biomarker
- Prediction of efficacy and relapse
- New molecular targets
- Mechanism of drug resistance

Systems Understanding of Cancer