Complex systems theory of Cancer Metastasis

Hiroshi Tanaka
Tokyo Medical and Dental University, professor emeritus
Tohoku University, Tohoku Medical Megabank Organization
special advisor to executive director
Today’s Topics

Systems approach to diseases in general
  – Systems (molecular) medicine, Precision medicine
  – Complex systems theory to disease

Complex systems approach to cancer
  – Complex systems approach of Cancer metastasis (EMT)
    • Essential process: EMT epithelial-mesenchymal transition
    • Tanaka H, Ogishima S., Network biology approach to epithelial-mesenchymal transition in cancer metastasis: three stage theory.
      
      *Journal of Molecular Cell Biology*, 7(3):253-66. 2015
Introduction of myself

• Activity in JAMI (Medical Informatics)☆
  – Board of Director (21yr: 1991-2013)
  – President (2003-2007)
  – Research: Japanese version of nation-wide EHR
• Activity in JSOM (Omics-based Medicine)☆
  – President (2008-present)
  – Research: Cancer metastasis,
    Integration of phenotype –genotype information in cancer (iCOD database)
• Activity in CBI association (Chemo-Bioinformatics)
  – President (2011-present)
  – Computational drug discovery, drug repositioning
• Working for realization (clinical implementation) of Japan-version of Precision Medicine and Genomic Cohort
Systems Molecular Medicine and Complex systems theory of disease in general
System-level understanding of disease to realize personalized/predictive/preemptive medicine

Based on application of the concept of “systems biology” to “disease”

Understanding a disease as an “unified system” (interconnecting molecular network)


Most of diseases, except for rare genetic diseases, is not caused by aberration of one/two genes but caused by "distortion (dysregulation) of molecular network".

- **Pathway-centric View of Diseases**, due to rapid advances in knowledge of molecular pathway/network and its alternation by disease.
- Inspired by success of systems biology in life science.
- Other names: p4-medicine, translational systems biology.

Employs disease omics profile and pathway knowledge to identify disease mechanism as caused by distorted pathways.
Three generations in molecular medicine

**1st generation** “Genomic Medicine” (1990～)
- Uses “inborn (germline) personalized differences of genome”
- Aiming at “Personalized medicine”
  - Estimation of “constitutional risk” of contracting disease
    - disease causative gene and disease susceptibility gene
    - Personalized medication (drug) PGx

**2nd generation** “Omics-based Medicine” (2000～)
- Uses “acquired somatic cell omics profile”
- Aiming at “Predictive/Preemptive medicine”
  - Using omics profile of disease (transcriptome, etc)
    - It changes depending on disease state and tissue
  - Estimate degree of on-going state of disease progression
    - Discover of disease subtype based on “omics profile”, ex. breast cancer
    - Directly related to prognosis or early detection of disease (Biomarker)

**3rd generation** Systems molecular Medicine (2010～)
- Disease is viewed as system distortion of molecular network
- Identify patient-specific dysregulated (distorted) pathway branches
Systems Molecular Medicine Disease Model

Individual-level clinical phenome

Organ-level Pathophysiology

Cell/tissue Pathomorphology

Disease Omics profile

Cell-cell communication Extracellular environment

Patient –specific Cellular molecular network

Mutated protein Mutated gene

Mutated protein Mutated gene

Mutated protein Mutated gene

rewiring

rewiring

Filtered
Basic strategy of systems molecular medicine

- Use “omics profile data” to estimate “patient-specific pathway dysregulation”
- Additionally use “pathway-activity-direct biomarker”, “Whole genome sequencing”
- Based on the “identified patient specific dysregulated pathway” make optimal plan of cancer treatment
- I would say precision medicine is systems molecular medicine
Complex systems approach to disease
Complex Systems Theory of Disease

- **Basis of Disease Occurrence**
  \[ \rightarrow \text{distortion of molecular network} \]

- **Progression of disease**:
  \[ \rightarrow \text{structural change of molecular network} \]
  - Trajectory in the molecular network space between quasi-potential basins in molecular network space
Gene regulatory network (GRN)

**Gene coded in Genome**

1. Genes are located *linearly in genome*

   ![Genes Linearly in Genome]

**Innate Intrinsic Structure of GRN**

1. All the possible sets of gene expression patterns in cellular molecular network
2. Wiring structure of cellular molecular network is innate but the set of its active nodes (proteins) and links (reactions) vary among the cell types and states
3. Equal to the space of all possible gene expression profiles

**Active Subnetwork of GRN**

1. Through transcription factor, gene expression regulations are *wired innately* in constitutional (germline) genome.
2. Same basic network structure is coded in all the cells.
3. But, actually, depending on the cell-type and subcellular/extracellular environment, a part of gene regulatory network is active. Subnetwork of gene regulation network is active and expressed.

Huang Bioessay 34(2011)
State Space for activation of gene regulation network

“GRN space”

1. All the possible sets of gene expression patterns in cellular molecular network
2. Wiring structure of cellular molecular network is innate but the set of its active genes and links (regulations) vary among the cell types and states
3. Equal to the space of all possible gene expression profiles

GRN space comprise all possible activation patterns of gene network, as a basic framework to deal with its temporally varying activity.
Quasi-potential in GRN space

- **Quasi-potential distribution**
  - the relative stability of each state in the GRN space
  - Stable state: bottom state in potential basin
  - elevation of the potential surface is inversely related to the likelihood of occurrence of the corresponding cell state

- **Mathematical calculation of quasi-potential**
  - Hung et.al. and Wang et. Al
  - Master eq \( \Rightarrow \) Stochastic Model

\[
\begin{align*}
\frac{dx_1}{dt} &= \frac{a_1 x_1^n}{S^n + x_1^n} + \frac{b_1 S^n}{S^n + x_1^n} - k_1 x_1 = F_1(x_1, x_2) \\
\frac{dx_2}{dt} &= \frac{a_2 x_2^n}{S^n + x_2^n} + \frac{b_2 S^n}{S^n + x_2^n} - k_2 x_2 = F_2(x_1, x_2) \\
\frac{dx}{dt} &= F(x) = [F_1(x_1, x_2), F_2(x_1, x_2)] \\
\frac{da}{dt} &= -\lambda a
\end{align*}
\]

Wang J, Zhanga K, Xua L, Wang E, PNAS 108(20) 2011
Waddington Epigenetic Landscape

- Waddington proposed the “metaphor” of cell fate
- Cell fate: cell differentiation process
  - Ball rolling down in the landscape from multipotent stem cell through progenitor cell to matured cell type
  - Ball is rolling down along valleys separated from ridge line
- Cell choosing the one of the branched path at the foot of the ridge
- Dividing ridge: epigenetic gene regulation mechanism → "Epigenetic Barrier"
- Matured stable cell type forms the basin in the Waddington landscape
- **GRN space equipped with quasi-potential** is thought to be “quantitative Waddington Epigenetic Landscape”
Complex Systems Approach of Cancer Metastasis
Epithelial-mesenchymal transition (EMT)

- Primary function of EMT is a necessary step for basic developmental processes such as gastrulation or neural crest formation.
- EMT is acknowledged as an essential process of cancer metastasis.
  - Non-motile, polarized epithelial cells, embedded via cell-cell junctions, dissolve their cell-cell junctions and convert into individual, non-polarized, motile and invasive mesenchymal cells.
- EMT is a cell type conversion, hence, a structural change of gene regulation network during metastasis.

Trajectory of EMT process in GRN space

- EMT is structural change of GRN
- Hence, can be depicted as a trajectory traversing in GRN space
- Matured cell types stay in the bottom of basins of GRN space
- EMT is considered as transition from epithelial cell basin to mesenchymal cell basin
Our Study on Complex systems approach to cancer metastasis: EMT

In real problem, to create the quasi-potential mathematically is vertically impossible. We developed “empirical” approach and execute “Attractor Analysis”

Article

Network biology approach to epithelial–mesenchymal transition in cancer metastasis: three stage theory

Hiroshi Tanaka.2,† and Soichi Ogishima2,†

1 Department of Bioinformatics, Medical Research Institute, Tokyo Medical and Dental University, Tokyo, Japan
2 Department of Bioclinical Informatics, Tohoku Medical Megabank Organization, Tohoku University, Sendai, Japan
† These authors contributed equally to this work.
* Correspondence to: Hiroshi Tanaka, E-mail: tanaka@bioinfo.tmd.ac.jp

Journal of Molecular Cell Biology, 7(3):253-66. 2015
Our empirical method to create GRN space

(1) Create frequency distribution ($\varphi$) of cell states in GRN space
   - Collect a number of samples of gene expression profiles, a particular GRN activation pattern, from
   - Public DBs of gene expression profiles: GEO (Gene Expression Omnibus), ArrayExpress
   - Created an empirical frequency distribution of GRN states.

(2) Create quasi-potential distribution
   - Based on the Boltzmann’s principle
   - The quasi-potential ($\psi$) is given by

\[(q\text{-potential}) \ \psi = - \log \varphi \ (\text{frequency})\]
Creation of GRN subspace to depict EMT

Gene Expression Profile Data

(1) Samples of gene expression profiles of Epithelial and Mesenchymal cell types in general
   - Obtained 14 datasets
   - 13 datasets from GEO and 1 dataset from ArrayExpress

(2) Time course data of gene expression data on EMT process
   - Human retinal pigment epithelium (RPE) cell line (ARPE-19)
   - induces cancer EMT in response to the combined treatment of TGF-β and TNF-α.
Gene expression profiles during EMT retinal pigment epithelial cell, using GeneChip (Affy HGU133 Plus 2.0).

<table>
<thead>
<tr>
<th>Database</th>
<th>Dataset ID</th>
<th>Sample ID</th>
</tr>
</thead>
<tbody>
<tr>
<td>NCBI GEO</td>
<td>GSE32474</td>
<td>GSM803648, GSM803707, GSM803765</td>
</tr>
<tr>
<td>NCBI GEO</td>
<td>GSE58605</td>
<td>GSM1414981, GSM1414982, GSM1414983</td>
</tr>
<tr>
<td>NCBI GEO</td>
<td>GSE50832</td>
<td>GSM1230054, GSM1230055, GSM1230056</td>
</tr>
<tr>
<td>NCBI GEO</td>
<td>GSE57083</td>
<td>GSM1374389, GSM1374390</td>
</tr>
<tr>
<td>NCBI GEO</td>
<td>GSE25146</td>
<td>GSM618009, GSM618010</td>
</tr>
<tr>
<td>NCBI GEO</td>
<td>GSE50832</td>
<td>GSM1230782, GSM1230783, GSM1230784, GSM1230728, GSM1230729, GSM1230730, GSM1230701, GSM1230702, GSM1230703, GSM1230045, GSM1230046, GSM1230047</td>
</tr>
<tr>
<td>NCBI GEO</td>
<td>GSE12548</td>
<td>GSM315430, GSM315431, GSM315432 (Takahashi et al. 0h hour data for epithelium)</td>
</tr>
<tr>
<td>NCBI GEO</td>
<td>GSE10315</td>
<td>GSM260657, GSM260658, GSM260659, GSM260660, GSM260661, GSM260662</td>
</tr>
<tr>
<td>EBI ArrayExpress</td>
<td>E-MEXP-858</td>
<td>31mOhC, CW31M0hA, CW31M0hB</td>
</tr>
<tr>
<td>NCBI GEO</td>
<td>GSE24612</td>
<td>GSM606843, GSM606844</td>
</tr>
<tr>
<td>NCBI GEO</td>
<td>GSE9440</td>
<td>GSM239829, GSM239830, GSM239831</td>
</tr>
<tr>
<td>NCBI GEO</td>
<td>GSE9832</td>
<td>GSM248201, GSM248202, GSM248204, GSM248209, GSM248210, GSM248213, GSM248214</td>
</tr>
<tr>
<td>NCBI GEO</td>
<td>GSE9865</td>
<td>GSM249026, GSM249027</td>
</tr>
<tr>
<td>NCBI GEO</td>
<td>GSE12583</td>
<td>GSM315617, GSM315618</td>
</tr>
<tr>
<td>NCBI GEO</td>
<td>GSE14897</td>
<td>GSM372142, GSM372144, GSM372146</td>
</tr>
<tr>
<td>NCBI GEO</td>
<td>GSE18226</td>
<td>GSM452726, GSM452730, GSM452734, GSM452735, GSM452736, GSM452737</td>
</tr>
<tr>
<td>NCBI GEO</td>
<td>GSE23583</td>
<td>GSM579884, GSM579885, GSM579886, GSM579887, GSM579888, GSM579889, GSM579916, GSM579917, GSM579918</td>
</tr>
<tr>
<td>NCBI GEO</td>
<td>GSE12548</td>
<td>GSM315446, GSM315447, GSM315448 (Takahashi et al. 60h hour data for mesenchyme)</td>
</tr>
</tbody>
</table>
Creation of GRN subspace to depict EMT

Creation of two axes for EMT depiction

- Selection of the genes for the subspace coordinates
  - 61 genes were selected
  - differ significantly between E and M cell states (significance level $p=0.1 \times 10^{-5}$).

- Principal components used for depiction of E & M cell states and EMT process
  - applied PCA to GRN subspace of 61 genes
  - PC1 represents the general component between E and M cell states
  - PC2 is related to cancer metastasis,
Frequency distribution of E, M states and EMT process

Experimental data (Takahashi et al. 2010)

Epithelial cells

Mesenchymal cells
Quasi-potential distribution of E,M states and EMT process
EMT trajectory in q-potential in GRN-space

Two-dimensional depiction of the trajectory of EMT

EMT trajectory in PC1-PC2 coordinates

Ward’s cluster analysis

Three stages of EMT process
EMT trajectory in q-potential in GRN-space

Three-dimensional depiction of the trajectory of EMT
Inference of gene regulatory network (GRN)

Selection of genes used to infer gene regulatory network

5,183 probe sets; 2,988 genes

1) Genes that showed large variations between time points during EMT; p<0.001; FC>1.5
   3421 probe sets; 1776 genes

2) Genes monotonously increasing and decreasing during EMT (accumulated chi-squared test; p<1.0×10^{-10})
   1,689 probe sets; 1,203 genes

3) Collection of known epithelial/mesenchymal marker genes and EMT related genes
   34 genes
Cancer EMT gene regulatory network (GRN)

ARACNe inference of cancer EMT gene regulatory network (17,368 regulations) \((p<1.0 \times 10^{-10}; \text{bootstrap} >90\%)\).
Master regulator analysis on cancer EMT GRN

Inference of master regulators which regulate more differentially expressed genes than expected (Fisher exact test p<0.05).

Cancer EMT GRN inferred by ARACNe algorithm

### Transcription Factors Table

<table>
<thead>
<tr>
<th>Transcription Factor</th>
<th># of DEG (SAM; p&lt;0.0001, FC&gt;1.5)</th>
<th>P-value (Fisher exact test)</th>
</tr>
</thead>
<tbody>
<tr>
<td>TCF3</td>
<td>86/135</td>
<td>$2.20 \times 10^{-16}$</td>
</tr>
<tr>
<td>ZEB1</td>
<td>46/592</td>
<td>$8.58 \times 10^{-16}$</td>
</tr>
<tr>
<td>SMAD2</td>
<td>36/317</td>
<td>0.000209</td>
</tr>
<tr>
<td>TWIST1</td>
<td>16/154</td>
<td>0.005634</td>
</tr>
<tr>
<td>TP63</td>
<td>6/7</td>
<td>0.01302</td>
</tr>
<tr>
<td>FOSL2</td>
<td>5/6</td>
<td>0.02516</td>
</tr>
<tr>
<td>PPARA</td>
<td>5/7</td>
<td>0.03732</td>
</tr>
<tr>
<td>ARNTL2</td>
<td>15/38</td>
<td>0.03843</td>
</tr>
<tr>
<td>MXD1</td>
<td>4/5</td>
<td>0.04918</td>
</tr>
<tr>
<td>MITF</td>
<td>4/5</td>
<td>0.04918</td>
</tr>
<tr>
<td>NR2F2</td>
<td>4/5</td>
<td>0.04918</td>
</tr>
</tbody>
</table>

Among MRs regulating genes and their neighbor genes (within two edges in protein-protein interaction network BIOGRID), we determined EMT related genes (including epithelial and mesenchymal marker genes), and then constructed “gene-gene network”.
Cancer EMT network

Gene regulatory networks directly regulated by 11 master regulators.
Cancer EMT network

KRT18, TP63 and TCF3 expressed in the epithelial cells. A master regulator TP63 regulates KRT18, well-known epithelial marker.
Cancer EMT network

CTNNB1 was predicted to induce the expression of ZEB1 as reported before. The CTNNB1 was known to be key factor inducing the EMT process.

ZEB1 was also a master regulator which expressed from the beginning of EMT process, and was known to down-regulate CDH1 gene expressions.
Cancer EMT network

SMAD2 was a master regulator, and induce, together with ARNTL2, the mesenchymal marker genes; e.g., MMP9 and FN1, and SERPINE2 genes.

TWIST1 was also a master regulator, and in fact, was reported to be essential in induction of the EMT process.
Three Stage Theory of EMT: cancer metastasis

- During the process of cancer EMT, three groups of states form the separate basins
  - The first basin is epithelial and the final basin is mesenchymal
  - Trajectory shows not directly transit from epithelial to mesenchymal state
  - It makes a detour to stay at Intermediate state basin
  - Relatively stable state to form a potential basin

- Three stage theory for EMT/Cancer Metastasis
  - But what is the biological meaning of intermediate state
One possibility: Chen’s critical transition state (DNB)

- Dynamical network theory of disease proposed by Luonan Chen
- Chen divided the disease progression process into three stages: a normal state, a pre-disease state (critical transition state), and a disease state.
- In critical transition state, strongly and dynamically correlated gene subnetwork (“dominant group of molecules”) emerges
- Expression level of the members of this subnetwork increasingly fluctuates in cooperative way (DNB: Dynamical Network Biomarker)
- Other possibility: return to the undifferentiated state (stem cell state)
The standard deviations and absolute values of correlation coefficients among the expressions of TCF3, ZEB1 and CTNNB1

(a) Standard deviation

(b) Absolute value of correlation
Summary

• EMT is not just an incidental aberration of the cellular process but a biological deep structural change of the gene network, “phase transition” in physics.
• Succeeded to depict the trajectory of the time course gene expression profiles of cancer EMT process onto the quasi-potential distribution in this landscape.
• 1) Cancer EMT is three-staged process composed of epithelial, intermediate and mesenchymal states, relatively stable state forming respective potential basin.
• 2) From the structural change in gene network, major master regulators alternatively take a main role along the subsequent three stages.
• The major master regulators: TCF3 and TP63 in epithelial state, ZEB1 with relation to CTNNB1 in intermediate state, and SMAD2 and TWIST in mesenchymal state.
• 3) Intermediate state of the cancer EMT is considered to correspond to Chen’s critical transition state: key molecules TCF3, CTNNB1 and ZEB1 are highly correlated with increasing variances in the intermediate state, corresponding to the specific features of Chen’s critical transition state.
Future Work

- To establish the **three stage theory of cancer metastasis**
  - More detailed observation of “intermediate states: (6h,16h) ⇒ (4,6,8,10,12,14,16h)”
  - To determine “stem-cell like state” or “critical transition state”
- To develop the **diagnostic method for intermediate state**
- To develop the **strategy for drug discovery targeting the intermediate state; anti-metastatic drug**
Thank you for kind attention