データ同化を使って全ゲノム転写動態シミュレーションを実現する

吉田　亮　　モデリング研究系　准教授

Rapid advance of biotechnology and data assimilation

Cutting-edge biotechnology has resulted in ever-increasing amounts of information disseminating within our society. The advent of high speed DNA sequencers is now bringing into reality a new era of personal genome and personalized medicine. Within the next several years, this technological breakthrough could cause significant changes in entire research fields of life science, involving every -omics studies, in terms of both quantity and quality. To uncover a complex world of cellular systems from such a vast amount of information, We aims to create a new research infrastructure of life science data assimilation systems involving experimental bioscience, bio-modeling, simulation, and state-of-art statistical science.

- **Human NSCLC-derived cell lines:**
  - ***PC9*** (Sensitive to treatments with *gefitinib*)
  - ***PC9GR*** (Resistance)
- **Gene expression arrays:**
  Agilent whole human genome (4×44K) oligo microarray (G4112F)
- **Time points:** 26 time points (24 hours)
- **Treatments:** **None**, **EGF**, **GFT** (*gefitinib*), **EGF+GFT**

Human Genome Center
Institute of Medical Science,University of Tokyo
Miyano and Gotoh lab @HGC

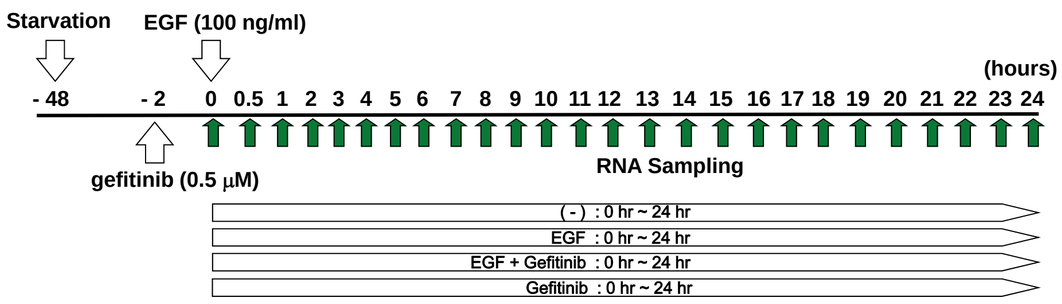

Fig1. DNA microarray experiments and normalized gene expression profiles. At the indicated 26 time points during 24 hr, transcriptional changes of parental PC9 and PC9GR (drug resistance) cells were measured by using a 44K Agilent Whole Human Genome Oligo Microarray. Following starvation for 48 hr before starting the expression profiling, each cell line was processed after being untreated (None), exposure to either EGF (EGF: 100 ng/ml treated at 0 hr) or gefitinib (GFT: 0.5 uM treated for 2 hr prior to the start of expression profiling), or coadministration of EGF and gefitinib (EGF+GFT).

Modeling transcription dynamics and statistical inferences

The biological circuit is modeled as a set of differential equations that defines rates of change in concentrations of p biological entities, $x(t)=(x_1(t), ..., x_p(t))$, over continuous times. Each variable (i th variable) is regulated by the parent variables Pa(i) with a rate equation having a set of kinetic parameters. Conduction of in vivo or in vitro time course experiments enables us to measure changes in concentrations of target molecules $y_n = (y_{in})$ during discrete time points. To proceed with the statistical learning, we here relate the differential equations to the experimental data using the state space model. Bayesian inversion analysis explores the unknown parameters in the model, involving initial state $x(0)$ and reaction kinetics θ, through the posterior distributions P($x(0)$, θ|Y, Pa) and P(Pa|Y) under which a circuit structure Pa, set of reactants for each variable, is specified or unknown, and a priori knowledge on reaction kinetics is expressed in a prior distribution.

- **Measurement model:**

$$\underbrace{\boldsymbol{y}_n}_{\sim 17000\text{ dimensional expression vector}} = \mathbf{H}\boldsymbol{x}_n + \boldsymbol{w}_n \quad \text{with} \quad \boldsymbol{x}_n = \boldsymbol{x}(n),\ \mathbf{H} \geq \mathbf{O} \quad \text{and} \quad \boldsymbol{w}_n \sim N(0, \mathbf{R})$$

H : (a) Block diagonal, and (b) non-negativity

- **System model:** **Interlocks of DE systems with each given as a basis expansion**

$$\frac{\mathrm{d}x_i}{\mathrm{d}t} = \sum_{k \in Pa(i)} \sum_{j \in \Phi} \beta_{ijk} \phi_{ij}(x_k) \quad \text{for } i = 1, \cdots, p$$

$$= \sum_{k=1}^{p} \mathbf{e}_{ik} \sum_{j \in \Phi} \beta_{ijk} \phi_{ij}(x_k) \quad \text{with edge: } \mathbf{e}_{ik} = \begin{cases} 1 & x_k \to x_i \\ 0 & \text{otherwise} \end{cases}$$

e.g. Dictionary Φ : Hill functions with a range of reaction parameters

Whole gene transcription simulation of human lung cancer systems

The invention of DNA microarray chip has enabled us to measure transcript levels of more than 20,000 genes in human genome, simultaneously. Our collaborators (Professor Miyano and his research team at Institute of Medical Science, the University of Tokyo) successfully monitored temporal change of all genes in two phenotypes of human lung cancers under the treatment with an anticancer agent; one demonstrating exquisite sensitivity to the drug treatment and the other being resistant. (Fig1 and Fig2) Recently, several studies have advocated relatively rapid acquisition of resistance within a few years after initiation of the anticancer drug treatment. The experimental data obtained now will certainly be vital to uncovering molecular basis of maintaining the viability of drug resistant cancer population. For the first time in the world, we succeeded in the development of whole gene transcription simulators that are highly reproducible to the observed gene expressions of the drug sensitive and resistant cancer cells (Fig4). The developed simulation models will be utilized in the discovery of key molecules that are promising for pharmacological treatment to improve drug efficacy, and a principle of action in maintaining the viability of the acquired drug resistance.

Fig2. Heatmap displays of the normalized time course data obtained from the conditioned microarray experiments. 17,654 genes appear in the same order across the eight images that were arranged according to a cluster analysis of the data in which all of the eight expression profiles were unified.

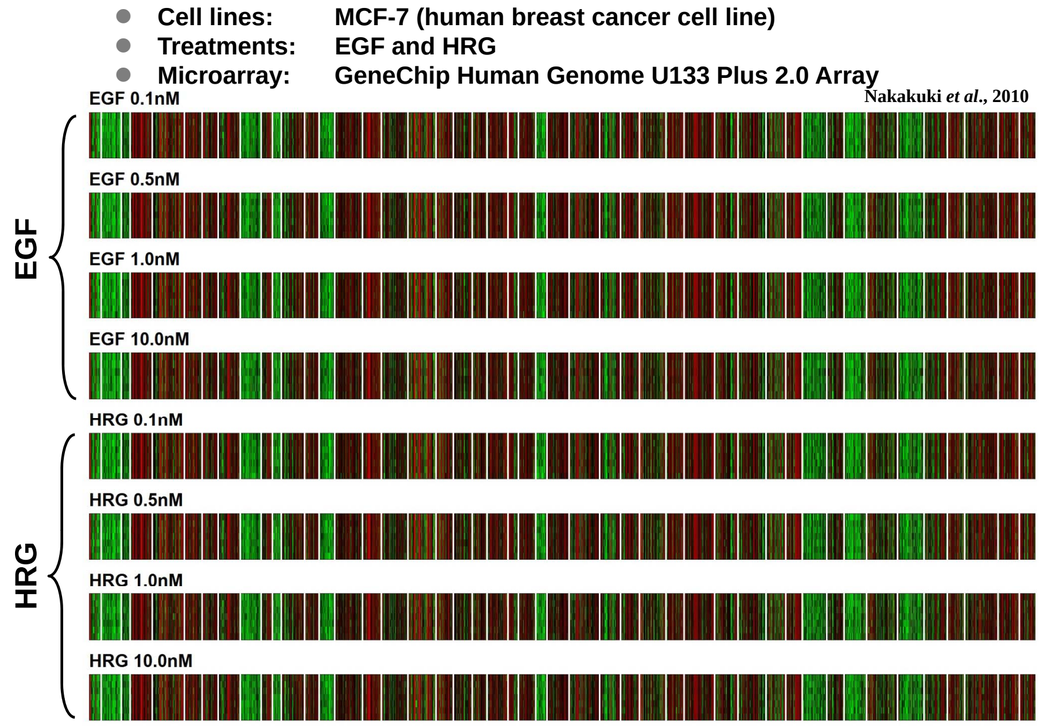

Fig3. The identified 46 expression clusters of PC9 and PC9GR would be universal, as they are shared by many other expression profiles, involving Different tissues, treatments, other cellular environments, experimental platforms, or organisms

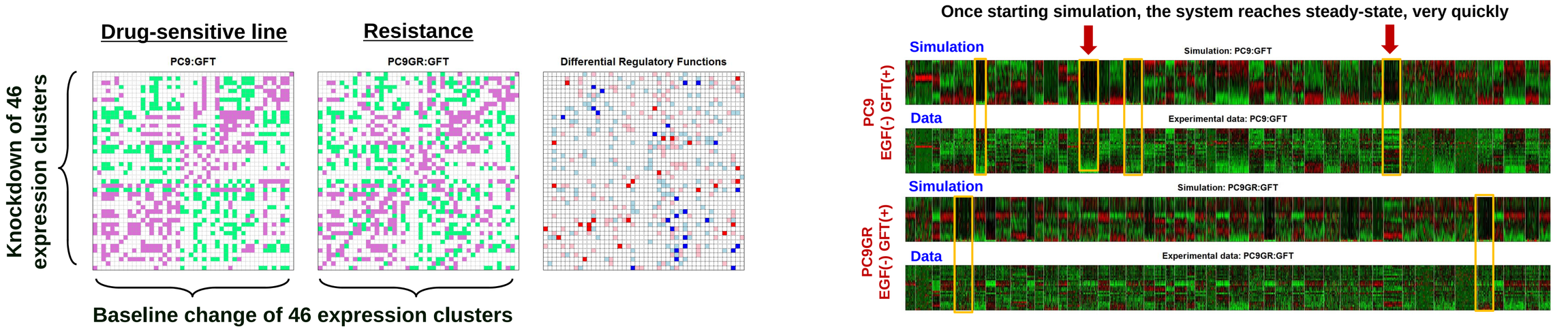

Fig4. The newly-derived whole gene transcription simulators could reproduce the observed expression dynamics of 17,654 genes in the drug sensitive and resistant NSCLC cells (PC9 and PC9GR) treated with gefitinib (right). Based upon further simulation experiments, such as RNA knockdown in *silico* (left), we can identify or predict key molecules that are promising for pharmacological treatment to improve drug efficacy, and a principle of action in maintaining the viability of the acquired drug resistance.