The Systeome Project

Hiroaki Kitano
Kitano Symbiotic Systems Project, ERATO, and
Sony Computer Science Laboratories, Inc.
E-mail: kitano@symbio.jst.go.jp

1 Introduction

Rapid progress of genome projects and proteome projects provide us abundant data of genes and proteins, so that their functions can be well understood. Nevertheless, these are components of systems, not the system itself. In order to fully understand biological system as a system, we need to investigate system-level behaviors. Although genome projects may provide us, for example, how single genetic mutation affects potential disease outbreak, diseases with multiple genetic disorders and other system-level disorders cannot be understood without understanding how a system of gene and metabolic network work as a system. While there are number of attempts for system-level understanding, there is a potential future needs for creating more comprehensive profiling of all sub-systems of a specific living system and their dynamics under various circumstances.

System-level understanding of biological system is an ultimate challenge in biology that is directly addressed in systems biology[Kitano, 00]. In order to promote scientific research of systems biology, it is critically important to create a comprehensive data resource that describes system's features that is analogous to the human genome project.

It is an enormous challenge that requires significant efforts to establish foundation for scientific research in the area, that is far beyond the capability of any single research group. Therefore, the author proposes "the systeome project" as a grand challenge project in the area of systems biology.

Systeome is an assembly of system profiles for all genetic variations and environmental stimuli responses. A system profile means a set of information on the properties of the system that includes structure of the system and their behaviors, analysis results such as phase portfolio, bifurcation diagrams. The structure of the system means a structure of gene and metabolic network and its associated constants, physical structures and their properties.

Systeome is different from a simple cascade map, because it assumes active and dynamic simulations and profiling of various system status, not a static en-

tity. The author claims that the project shall be established for a comprehensive efforts for profiling systeome of human, mouse, *Drosopila, C. elegans* and yeast.

The goal of the human systeome project shall be defined as "to complete a detailed and comprehensive simulation model of human cell at an estimated error margin of 20 percents by year 2020, and to finish the identification of system profile for all genetic variations, drug response, and environmental stimuli by year 2030."

Undoubtfully, this is an ambitious project, and need several milestones and pilot project leading to the final goal. Initial pilot projects can be set using yeast and *C. elegans* with five or seven years time frame after the full size budget approval. The human systeome project shall be commenced concurrent with such pilot projects.

2 Impacts of the systeome project

The impact of this project will be far reaching. It will be a standard asset for biological research as well as fundamental diagnostics and prediction basis for wide range of medical practices.

Systeome project is expected to contribute system-level understanding of life by providing exhaustive knowledge of system structure, dynamics, and their sensitivities against genetic variations and environmental stimuli. By the use of system profile, it is expected that more precise medical diagnosis and treatments can be accomplished due to quantitative understanding of metabolic state of the system. For example, a list of all possible feed back loops and their sensitivities, gain, time delay should be obtained, so that it can be used for drug design and clinical applications. Behaviors of feed back systems are often counter-intuitive and often eliminates, or compensates, effects of external stimuli. Understanding of complex circuit dynamics as signified by the feed back loops contribute to accurate prediction of effects of medical treatments.

The systeome project should maintain close link with genome and proteome data, particularly with various individual genetic variations, including single nucleotide polymorphisms (SNPs). SNPs is a typical example of an attempt to understand relationship between genetic variations and clinical observations.

It is predictable that there are cases effects of SNPs are masked by a mechanism that compensates such variations. In this case, corresponding SNPs does not seems to affect behaviors of the cell. However, if such a compensation mechanism is disrupted by SNP in locus that constitutes the compensation mechanism, effects of SNPs will directly show up in the cell's behavior. In such a case, it will be observed that for certain group of cells, SNPs does affects phenotype, but for the other group SNPs does not seems to affect phenotype.

While SNPs provides certain information on individual variation at the genetic level, it does not provides quantitative status of mRNA and proteins. Many biological phenomena has certain quantitative sensitivity. Cell cycle, for

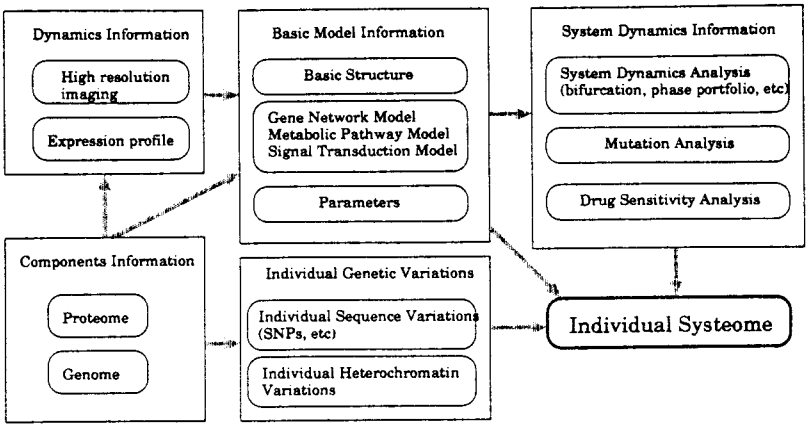

Figure 1: Genome, Proteome, and Systeome

example, is expected to take place when cyclin synthesis and degradation rate are within a certain range. SNPs and other existing genetic analysis cannot provide insights into quantitative aspects of such phenomena.

Scientifically, exhaust understanding of circuits and their dynamics contribute to deeper understanding of the biological systems, that has already been discussed elsewhere.

Identification of metabolic and signal transduction circuits in various model systems provides an interesting opportunities to compare evolutionary conserved genetic information not only at gene-level, but also at a circuit-level.

Evolutionary conserved circuits will be an important concept that may be widely used in the study of gene and metabolic network behaviors. Numbers of circuits that may be found in yeast and *C. elegans* may be used also in mouse and human, similar to the idea of homologue genes.

Some of the feed back circuits, for example, may be so essential that it was conserved through the course of evolution. At the same time, a certain circuit may be duplicated and revised version is applied for other part of the system. With the progress of systeome project in various model systems, such a comparative study and homology search at the circuit level would be made possible.

There are numbers of scientific opportunities once the systeome project is commenced and its data is made available to scientific research.

3 Measurements

Although broad range of features of the living systems need to be measured, an initial list of items that need to be measured in the systeome project provides a clear view on the scope of the project. A set of basic items that should be measured is as follows:

- Complete identification of gene and protein networks
- Comprehensive measurement of parameters, such as reaction rate, degradation rate, binding constant, and diffusion speed.
- High precision measurement of protein movement and localization within and between cells.
- High precision measurement of time series expression profiles at both mRNA and protein level.

In addition, technologies shall be developed to measure following items:

- A detailed measurement of heterochromatin structures that includes physical localization, genomic position, and their temporal changes.
- A detailed measurement of physical structures of the cell and tissues, that includes numbers and position of organella, legand, receptors, microtubles, cell shape, etc.

These measurement efforts requires extensive invetiment in the research and development of new measurement and experimental systems. For example, measurements of highly accurate localization of protein may require development of new type of microscopy and constructs to be used with a new imaging system.

4 Software Platform

A set of software and data/model representation standard need to be developed to incorporate experimental results for scientific research. Software platform includes (also shown in Figure 2):

- database for storage of experimental data
- cell and tissue simulator
- parameter optimization software
- bifurcation and systems analysis software
- hypotheses generator and experiment planning advisor software

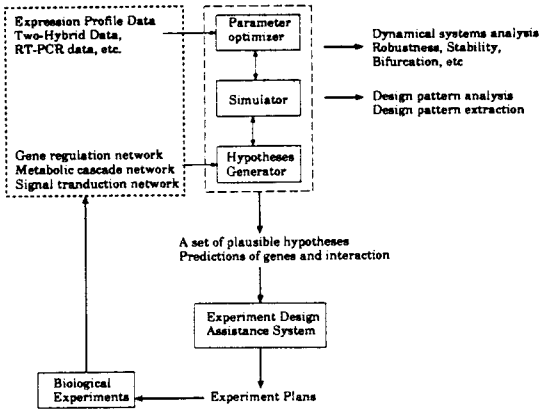

Figure 2: Software Tools for Systems Biology

In addition, standard for representation of models and data exchange need to be establish at the earliest stage, but it should be up-dated to cope with improvement of experimental methods and software features. Already, we have taken the first step of such a standard formation as Systems Biology Mark-up Language (SBML) as an extension of XML [Hucka, et al., 00].

5 Conclusion

The systeome project will be a major commitment. However, it is indispensable for promoting systems biology at its fastest speed and to contribute to better understanding of living systems and for medical practice. The systeome project has a major engineering project for measurement and software platform development. The best way to proceed with this project is to initiate this as an international joint project at the scale comparable to the human genome project.

References

[Kitano, 00] Kitano, H., "Perspectives on Systems Biology," *New Generating Computing*, Vol. 18, 199-216, 2000.

[Hucka, et al., 00] Hucka, M., Sauro, H., Finney, A., and Bolouri, H., *An XML-Based Model Description Language for Systems biology Simulations*, Working Draft, ERATO Kitano Project - CALTECH group, 2000.