Animal Cells and Systems 12: 181-191, 2008

Animal Cells
and
Systems

http://www.AnimalCells.or.kr

Current Research Trends in Systems Biology

Do Han Kim*, Pradeep Kumar Shreenivasaiah, Seong-Eui Hong, Taeyong Kim and Hong Ki Song

Department of Life Science, GIST, Gwangju 500-712, Korea

Systems biology is a newly emerging biological field that aims to understand various complex life phenomena at a system level. The traditional biology has a tendency to break down the observable life phenomenon into a list of parts and for determining their interactions (reductionism), whereas system biology attempts to describe the complex and dynamic wiring of all the elements in a system and detect the emergent properties of it (holism). Systems biology has become realistic with the accumulation of large mass of biological data by use of the high-throughput omics technologies (e.g. genomics, transcriptomics, proteomics and metabolomics). This review provides an overview of major themes in the current research trends of systems biology, summarizing some of major strategies to generate, analyze and integrate the high-throughput data to make them useful information capable of predicting complex biological behaviors.

NETWORK ANALYSIS AND GENERATION OF HIGH THROUGHPUT DATA

Discovering design principle in cellular systems has been continuously studied in systems biology field since beginning of this century (Kitano, 2002). The scale-free topology of network has been reported in various cellular systems (Albert et al., 2000; Jeong et al., 2000) and its structural characteristics are extensively studied. Also discoveries in identification of network motifs (simple building blocks of network) such as feed-forward/feed-back loops in the cellular regulatory network have been made (Lee et al., 2002; Milo et al., 2002). Several studies have sought to investigate the design principle underlying dynamic network. Kwon and Cho (2008) found the various feedback loops interlinked coherently in signaling networks and suggested the coherently coupled feedback loop as the cellular design principles to stabilize and enhance the signals. Legewie et al. (2008) studied design principles of dynamic signaling pathway in mammalian systems and showed evidence that an asymmetric negative feedback regulation induced by a subset of signal inhibitor controls the signaling pathway, whereas positive feedback plays no role. Cellular systems have shown to be consisting of modular structures and few of studies have examined instances such as disease and condition-specific modularity (Bar-Joseph et al., 2003; Hartwell et al., 1999; Ihmels et al., 2002; Ravasz et al., 2002). Recently, microRNA (miRNA-small single-stranded ~22 nucleotides RNA) has been reported to regulate gene expression. These miRNA are now known to regulate cellular network such as transcription factor network, signaling network and metabolic networks. The role of miRNA in systems biology is elaborately reviewed by Edwin Wang (2008).

High-throughput data generation for systems biology is a continuously developing technology. The advent of sophisticated omics technologies has facilitated efficient handling of biological samples allowing measurement of large data points. These technologies have enormously reduced the time for generation and analysis of biological data (Hardiman, 2004). In the following section, the recent progress related to network analysis in various levels of biological systems and some examples of recently-developed instrumentation will be discussed.

Network analysis using transcriptomic data

Microarrays technology has been quite well established and is highly successful in obtaining large-scale transcriptional profiles since its emergence (Chee et al., 1996; Schena et al., 1995). Currently microarray chips can accommodate more than 200,000 spots. A single microarray chip can generate genome scale transcriptome data and it has become a core component of systems biology studies. Either microarray chips could be custom-made to accommodate the specific experimental requirements or commercially available chips (e.g. Amersham, Agilent and Affymetrix

*To whom correspondence should be addressed.
Tel: +82-62-970-2485; Fax: +82-62-970-3411
E-mail: dhkim@gist.ac.kr

chips) could be used for profiling mRNA, miRNA and SNP, or for ChIP assay. However like any other successful technology, microarrays also have potential problems associated with the poor specificity, reproducibility, inter-platform inconsistency. FDA in America has initiated MAQC project which enforces a basic framework to ensure high reliability of the technology (Shi et al., 2006).

There is a sign that an alternative method 'tag sequencing' could replace the microarray-based profiling soon. Using the Solexa Genome Analyzer, t'Hoen et al. (2008) reported that the transcript quantification by tag counting is more reliable than the five different microarray platforms. Illumina has recently announced that the number of tags produced with the existing and future instruments is increasing and the cost per sequence run will eventually go down.

With the rapid accumulation of high-throughput transcriptomic data, many ingenious algorithms to infer the genetic network have been also reported. Those algorithms are cumulatively listed and described in the recent review (Bansal et al., 2007). The research in transcriptomic network analysis has been progressed from its basic analysis of static networks to the advance analysis of the condition-specific networks including network dynamics and pathogenic networks. Recently, Busch et al. (2008) reported that transcriptomic network dynamics for keratinocyte migration. They combined the small-set of target genes for keratinocyte migration with the modeling and finally showed that the pulse-like activation of Met, the proto-oncogene receptor, is important for the responsive state and the EGF-receptor is required to initiate and maintain migration of keratinocyte.

The re-interpretation of non-coding sequences, which have been known to be a junk DNA before, has expanded the transcriptomic network to the post-transcriptional level. Some recent studies have shown the construction of post-transcriptomic network together with the miRNA datasets. Recently Brosh et al. (2008) reported that a family of 15 miRNAs play an important role in the post-regulation of E2F and p53 target genes in the proliferation networks. They emphasized that miRNAs are novel key players crucial for regulation of transcriptomic network.

Network analysis using proteomic data

Proteome network analysis has also been remarkably advanced with the rapid development of the various high-throughput technologies such as yeast two hybrid (Y2H) and affinity pull-down mass spectrometry (AP/MS). Also similar to gene microarrays, protein chips have proved their usefulness in identification of protein-protein interactions, protein-phospholipid interactions, small molecular targets and substrates for proteins kinases (Hall et al., 2007; Templin et al., 2003). Many researchers have continuously produced highly reliable datasets using the available methods. One recent study evaluated the high-quality yeast interactome across the multiple datasets, which were generated by the different methods (Yu et al., 2008). In this study, they suggested that Y2H outperform AP/MS at least for the binary interaction. However they emphasized that both methods could provide complementary information about the interactome and are vital to obtaining a complete picture of cellular protein-protein interaction networks.

Another study introduced the proteosome network investigated by the quantitative analysis of tandem-affinity purified cross-linked (x) protein complexes (QTAX) (Guerrero et al., 2006, 2008), which has an advantage to identify the weak or transient protein-protein interaction. Guerrero et al. (2008) analyzed the proteome network showing much increased sensitivity than previous reports. Boxem et al. (2008) also contributed to the increased sensitivity of protein interaction framed on domain-based interaction. They applied this strategy to identify proteome network of early embryonic cell division in *C. elegans* showing higher sensitivity of QTAX derived networks over the Y2H.

Network analysis using metabolomic data

Metabolomics is an emerging field that complements other omics in systems biology field. A salient feature is that the metabolite in the body fluids (e.g. blood or urine) can be used as biomarker in noninvasive diagnostic tests. Similar to transciptomics and proteomics, typically metabolomic experiments will generate thousands of data points. Currently metabolomic data sets are generated using capillary electrophoresis (CE), Direct infusion of ESI (DIESI), Fourier transform ion cyclotron resonance mass spectrometry (FT-ICR-MS), Nuclear magnetic resonance (NMR). These technologies are reviewed extensively elsewhere (Kell, 2004; Van Dien and Schilling, 2006). Metabolites are the end-stage output of gene and protein-level processes. So in obtaining system-level understanding of global regulation of cellular metabolic networks, it becomes essential to adopt integrative approach, which spans across various biological levels involving transcriptomic and proteomic networks for analysis. Recently Duarte et al. (2007) investigated the global metabolic network in human which was manually curated from literatures. Using gene expression dataset with metabolic network, they showed evidence of influences of gastric bypass surgery on skeletal muscle metabolism. Ma et al. (2007) used a similar approach to elucidate a bow-tie architecture of the metabolic network. They found that an input of a wide range of nutrients could produce a large variety of products using a relatively few intermediates. Chechik et al. (2008) analyzed the influences of transcriptional network dynamics on metabolic network in *S.cerevisiae*. Their combined analysis suggested that cell system could choose the optimized activity motif (patterns in the dynamic use of a network) in response to the various conditions. For instances, forward activity motif is used to produce metabolic compounds more efficiently, whereas backward activity motif is used to rapidly stop the production.

Comprehensive understanding of the disease mechanism from network analysis

One of the most important emerging aspects of network analysis is deciphering the pathogenic mechanism of human disease. Recently, progress in this direction is reported by Yang et al. (2008). They developed a novel algorithm, multiple target optimal intervention (MTOI), to infer optimal intervention points in the disease network (inflammation related network). Their algorithm basically finds optimum perturbation conditions to reverse diseased state network into normal state network resulting in an identification of the effective points of intervention and combination of interventions. These intervention points can be promising for drug targets which they claim to be effective and safe. Another similar global network analysis approach to screen the human disease genes is reported by Wu et al. (2008). They developed CIPHER (Correlating protein Interaction network and PHEnotype network to pRedict disease genes) to discover the novel candidate disease genes and their network in human disease. In their published case study involving breast cancer, their approach showed a success in identifying many of the known breast cancer susceptibility genes such as BRCA1 and prediction of 15 novel breast cancer susceptibility genes.

STRUCTURAL SYSTEMS BIOLOGY

One of the major goals of systems biology is to deduce the behavior and emergent properties of a confined biological system on the basis of their components (Kitano, 2002; Levesque and Benfey, 2004; Rousseau et al., 2005; Pieper et al., 2004). Therefore the nature of binding between the components is an essential step for systems biology. "Structural systems biology" could be defined as a process of modeling protein interactions involved in the designated biological complexes on the basis of known or predicted molecular structures (Aloy and Russell, 2006). Previously various protein interaction networks were obtained from various species using both experimental and computational methods (e.g. Uetz et al., 2000). However, precise molecular details for the interactions are available for only limited number of cases (Rousseau et al., 2005; Pieper et al., 2004; Aloy and Russell, 2006). Precise 3D structures solved by X-ray crystallography and NMR spectroscopy provide the crucial atomic details of protein bindings. In light of the difficulties in obtaining large amounts of structural information regarding to protein assemblies, the computational prediction methods have become essential for structural systems biology. As of December 9, 2008, total 54,699 protein structures in various species have been reported. The majority of structures have been determined by X-ray crystallography and NMR spectroscopy (RCSB PDB). Only a part of them are related to complex structures.

'Structural genomics' consortiums have been organized for determination of the 3D structures of all proteins in a given organism by X-ray crystallography, NMR spectroscopy and computational methods. Structural genomics projects have emphasized high-throughput determination of protein structure. This has been performed in dedicated centers (e.g. SGC, http://www.sgc.utoronto.ca/, PSI, http://www.nigms.nih.gov/psi/, CESG, http://www.uwstructuralgenomics.org/, RIKEN SGPI, http://www.rsgi.riken.go.jp/, PSF, http://www.proteinstrukturfabrik.de/, OPPF, http://www.oppf.ox.ac.uk/, SPINE, http://www.spineurope.org/). As of December 9, 2008, 7225 protein structures are solved and the structural information is now available to the public. However, lack of the functional data is problematic to the proteins. It is interesting that RIKEN SGPI has been the most productive consortium so far (total 2,658 protein structures have been reported as of December 9, 2008).

Structural determination of protein assembly by use of cryo-EM and computational methods

As described above, the individual protein structure can be determined now in a relatively short time, with the help of sufficient material prepared using modern expression and purification techniques. However, a setting-up both purification and crystallization conditions for protein complexes are an extraordinary task, and it may require many years of efforts. Therefore, there will be an enormous time gap between the progress in developing protein networks from high-throughput data and the progress of complex 3D structures. In fact, the human intercome project is now in the fast pace and it will be soon obvious that lack of 3D complex structures will be a bottleneck for understanding the mechanisms of protein assemblies. The 3D structures of protein assembly could provide information on the strength of interactions. For instance, domain-domain interaction is stronger than domain-peptide interaction. If the peptide is phosphorylated, the interaction strength could be enhanced as compared with the unphosphorylated one. Computational approaches have become important to determine the protein interactions in a confined biological system (Marcott et al., 1999: Garvin et al., 2006; Lu et al., 2003). An example of the computational method is to predict the structure of binding motif by homology modeling and predict the binding partners by the method of 'Docking' (Gray, 2006). Some groups have developed methods predicting atomic details for pairs of interacting proteins and combined docking with chemical shift NMR experiments (e.g. HADDOCK, http://www.nmr.chem.uu.nl/haddock) (Lu et al., 2003). If the similar structural information is not available, then the domain structures are analyzed bioinformatically and the domain-domain interactions are investigated on the basis of already reported information (InterPReTS, http://interprets.embl.de) (Sprinzak et al., 2001). Peptides having a particular domain could share a consensus sequence pattern or linear motif for predicting binding partner (NetPhos, http://www.cbs.dtu.dk/services/NetPhos; PhosphoELM, http://phospho.elm.eu.org; iSPOT, http://cbm.bio.uniroma2.it/ispot) (Neduva et al., 2005).

Constructing protein assembly models with hybrids of high and low resolution techniques

Putative structural models of protein assemblies can be obtained by combination of high and low resolution techniques (Aloy and Russell, 2002; Aloy et al., 2004; Joshi-Tope et al., 2005). Previous studies have used the hybrid methods for determining the structure of small nuclear ribonucleoprotein particles (snRNPs) which bind pre-messenger RNA to form spliceosome (Aloy and Russell, 2006). For the structural analysis, various methods such as X-ray crystallography, cryo-electron microscope, chemical crosslinking were used. The hybrid approaches have been used to propose structures of other big molecular complexes (Aloy and Russell, 2006). Cryo-EM is a useful method to catch protein complexes in different conformational states (Muller et al., 2008). The ryanodine receptor (monomeric form: 550 kDa), an intracellular Ca^{2+} release channel is one of the hub proteins and bind various signaling proteins such as calmodulin and FKBP (Shreenivasaiah et al., 2008). We have attempted to build up an assembly model using various methods such as X-ray crystallography, cryo-EM and the computational docking methods (Lu et al., 2003). It is also useful to combine the signaling or pathway map with known 3D structures of participating molecules. We could get insights in the mechanisms regarding to how hub proteins interact with other proteins in the signaling cascade. The structural approaches could also lead to understanding the time-dependent changes of the interactions between the signaling proteins. Knowing strength of interactions in a particular system could help design therapeutic drugs to control the pathways on purpose.

Electron tomography-'Seeing is believing'

Electron tomography utilizes a tomography technique along with electron microscope to obtain 3D structures of macromolecules. In principle, electron beams are passed through the sample with increasing degrees of rotation around the sample. The collected information on the images is used to construct a 3D image of the molecules. The resolution of electron tomography systems has increased to 5-20 nm range which is suitable for examining multi-protein structures (Medalia et al., 2002; Nickell et al., 2006; Melo et al., 2008; Sali et al., 2003). Currently electron tomography study is going on various topics at the molecular resolution. For example, 3D structure of protein complexes such as ribosome can be visualized with other proteins nearby. It is possible in the near future that the hub proteins in the protein networks are directly compared with their 3D structures and the *in situ* locations in the different cells (Aloy and Russell, 2006). Since more realistic 3D cell images could be produced by this method, the physical properties such as diffusion constant and compartmentarization may be detected with the calculated dimensions (MCell, http://mecell.cnl.salk.edu).

DATA INTEGRATION

As described earlier, systems biology is basically an approach to produce useful hypothesis and theories using different levels of information pertaining to genes, mRNAs, proteins, and pathways. An enormous challenge that needs to be addressed is rigorous integration of heterogeneous data from the discrete levels. Figure 1 describes the importance of data integration and data management in context of systems biology research. Methodologies and computational tools that can use the multi-level data and generate comprehensive biological insights by integrating them is an immediate need. Technical challenges of data integration in systems biology are mainly along four lines: 1) data gathering, representation and validation; 2) tools and resources for data integration; 3) database integration and web technologies; and 4) data integration and analysis methodologies. Addressing these issues will assist in easy multi-level data integration and result in insightful biological discoveries.

Data gathering, representation and validation

Iterative cycles in systems biology process involve matching the data and model, which is rarely simple (Jaqaman and Danuser, 2006). The datasets from disparate (technologies and instrumentation) sources are often incomplete, not standardized, improperly annotated. Increased propensity of using poor quality data in work process could conduce in inconclusive or wrong results. An effective way would be to employ state of the art sensing, measurement and data processing technologies that could accurately capture precision experimental data. An emphasis should be more on context-based biological measurements than accumulation of biological data regardless of any intended use (Albeck et al., 2006; Waters et al., 2006). It is also essential that error with measurement must be determined and a metrics be put

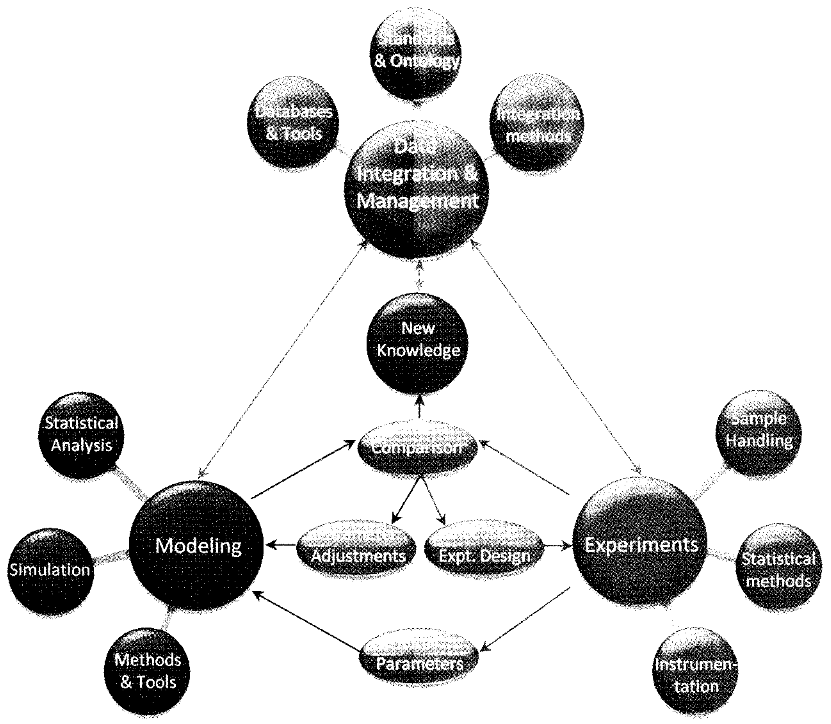

Fig. 1. Data Integration and data management is an integral part of iterative systems biology framework. Systems biology is an iterative approach requiring building an initial model with available experimental data. The model is used to arrive at meaningful predictions which could be compared to experimental data. In case of any inconsistencies either model could be refined to fit the data or new experiments could be designed, to generate suitable data for refining the model or to test the new resulting hypothesis by simulations. The above cycle is repeated until a reasonable agreement is achieved between model and experimental data leading to new knowledge. In each of these iterative steps (experiments, modeling and knowledge generation) data is used and also new data is generated which then may be subsequently utilized in other steps. Efficient and comprehensive management of the data thus is crucial for systematic research in systems biology.

in place for the validation of large data sets. Considering time and cost factors, successful economical strategies can be adopted in designing experiments (Albeck et al., 2006).

Supporting computational infrastructure should not only assist in efficient data capturing, but also should adopt consensus standards for the interpretation, handling and dissemination of data maximizing interoperability, accuracy and completeness. Several standardized formats for representation of data have been currently developed. Few examples of most widely used formats are: MIAME for microarray experiments, PSI-MI for protein-protein interaction, BIOPAX for pathways, MIRIAM for biochemical models and SBML/CellML for machine readable qualitative and quantitative model formats. Refer to review by Stromback and Lambrix (2005) for comparison and evaluation of several of these formats. Ontologies being structured vocabularies capturing the domain knowledge have proved to be highly advantageous in facilitating semantic-level data integration. Data fusion of heterogeneous datasets into self-consistent sets is crucial and ontologies proved to have tremendous advantage in this regard (Waters et al., 2006). Several analysis methodologies (e.g Gene enrichment analysis (Holden et al., 2008; Lin et al., 2008)) are using ontology's structure, relationship and associations to elucidate new biological knowledge from the existing data. Ontologies are in turn used to represent and annotate newly generated data thus maximizing reuse and knowledge utilization (Tu et al., 2008). Consortium such as Open Biomedical Ontologies (OBO) has played a major role in development and as a resource for distribution of bio-ontologies (Leontis et al., 2006; Smith et al., 2007). Few of the well known ontologies that are often used in systems biology research are: Gene Ontology (GO), MGED Ontology for microarray experiments, Protein Ontology (PO), Pathway Ontology and Systems biology Ontology. Several of the important biomedical ontologies are reviewed elsewhere (Bard and Rhee, 2004; Blake and Bult, 2006; Rubin et al., 2008).

Database integration, tools and web technologies

Another critical need for data integration is an integrated database holding raw and curated datasets. Such databases should enable rapid submission, meaningful data merging, efficient storage and speedy comprehensive retrieval of data. Centralized databases can hold global data as well as in-house research data can expedite access to information from various areas of research and facilitate selective mining, cross-correlation and analysis (Stein, 2003). A coherent contextual metadata must be curated and made accessible regarding the data stored in the databases. Several general databases have been successful in facilitating such data source integration (Flicek et al., 2008; Lenhard et al., 2003; Safran et al., 2002; Sayers et al., 2008; Sugawara et al., 2008). However some of these databases do have potential problems; they are biased to some biological processes and many important molecular entities are not yet sufficiently covered. Many of the general databases do not hold datasets that can be readily available for systems biology like networks, pathways, quantitative parameters and model information. In such scenario recommended approach would be development of context-dependent integrative databases which can focus on expediting rich data in specific context which could be readily consumed in systems research. Recently several such efforts have been initiated (Choi, 2007 ;Kahlem and Birney, 2007; Lynn et al., 2008; Zhang et al., 2007; CIDMS, http://cidms.org). Recent advances in internet technologies like web service can now support machine-to-machine interoperability. It can be exploited to integrate distributed specialized databases (Kahlem and Birney, 2007) and also modeling and analysis tools could use these services to get data directly into working environment minimizing time to data transfers and data format incompatibility issues (Cerami et al., 2006; Funahashi et al., 2007; Xia and Dickerson, 2008). Few other tools even provide the user functionalities for data management, retrieval, and visualization and integration capabilities (Cline et al., 2007; Shah et al., 2007; Toyoda et al., 2007; Wright and Wagner, 2008).

Data integration and analysis methodologies

Until now, the integration efforts have proved successful in each of the constituent levels under the preview of systems biology. Few of the most successful methods that are extensively used today are: 1) Pattern analysis of gene/protein expression data using microarray/mass spectrometry to identify differentially expressed genes/protein in a sample of interest yielding component data and/or interaction data; and 2) Clustering strategies to elucidate functional modules (functional states such as phenotypes) from the expression data with similar responses to perturbations. Often these gene expression analyses are combined with other advanced analysis strategies (e.g. promoter analysis: to establish a map of transcription binding sites in promoter regions in order enhance knowledge about mechanisms of gene expression regulation). Several other methodologies are extensively discussed in reviews elsewhere (De Keersmaecker et al., 2006; Joyce and Palsson, 2006). Recently Ishii et al (2007) generated system wide multi-level (including transcriptome, proteome, metabolome and interactome) quantitative datasets focused on central carbon metabolism of a bacterial cell. This is probably first quantitative dataset that includes parts and also functional state (interactions and interaction types) of a cell metabolomics. With this high quality data, it is possible to elucidate how static parts and topology of a network contribute to emergence of functional states in a dynamic network. Although individual analysis at each of these levels undoubtedly results into interesting finding, it is only by integrating each of these omics data will aid in gaining a system level insight into complex cellular behavior.

Given the abundant multi-level data from various experiments and databases, the most important task is development of the theoretical and experimental methodologies, which can efficiently integrate and analyze the data. However currently there are only few developed methodologies that can utilize combined data from multiple levels (Ishii et al., 2007; Liu and Zhao, 2004; Murray et al., 2007; Stemke-Hale et al., 2008). Murray et al (2007) combined metabolomic, transcriptional data and statistical analyses of transcriptional factor activity. They identified oscillatory parameters and constructed a large-scale yeast interaction network which was then used to identify crucial features that are important for respiratory oscillations. An integrative genomic and proteomic analysis conducted by Stemke-hale et al (2008) led to showing that PI3K pathway aberrations are common in breast cancer and PI3K targeted drugs are potentially important. Few other analytical approaches and methodologies for multi-level data integration are reviewed elsewhere (Cox et al., 2005; Joyce and Palsson, 2006; Sauer et al., 2007). Experimental data generated using different technologies and differing levels of measuring precession and coverage, results in inherent data discrepancies such as systemic biases and high false positive and false-negatives rates. But to assess the relationship between multiple biological levels in an effort to extract discernable biological meaning, such hurdles should be aptly addressed. Hwang et al. (2005a and b) have developed methodology that can integrate data from existing and future technologies ensuring selection of accurate (true-positive) data sets. They have also demonstrated applicability of their approach in systems biological context.

MODELING PROCESSES

Mathematical (Dynamic systems) modeling

In the post-genome era, theoretical approaches involving unambiguous representations and predictions of a novel biological mechanisms based on existing knowledge have become more important. Although huge amount of data is a necessity, as mentioned earlier, they themselves are often problematic, because of their sheer size, inconsistencies and less comprehensive (Bray, 2001). Since mathematical modeling is able to provide a framework to represent

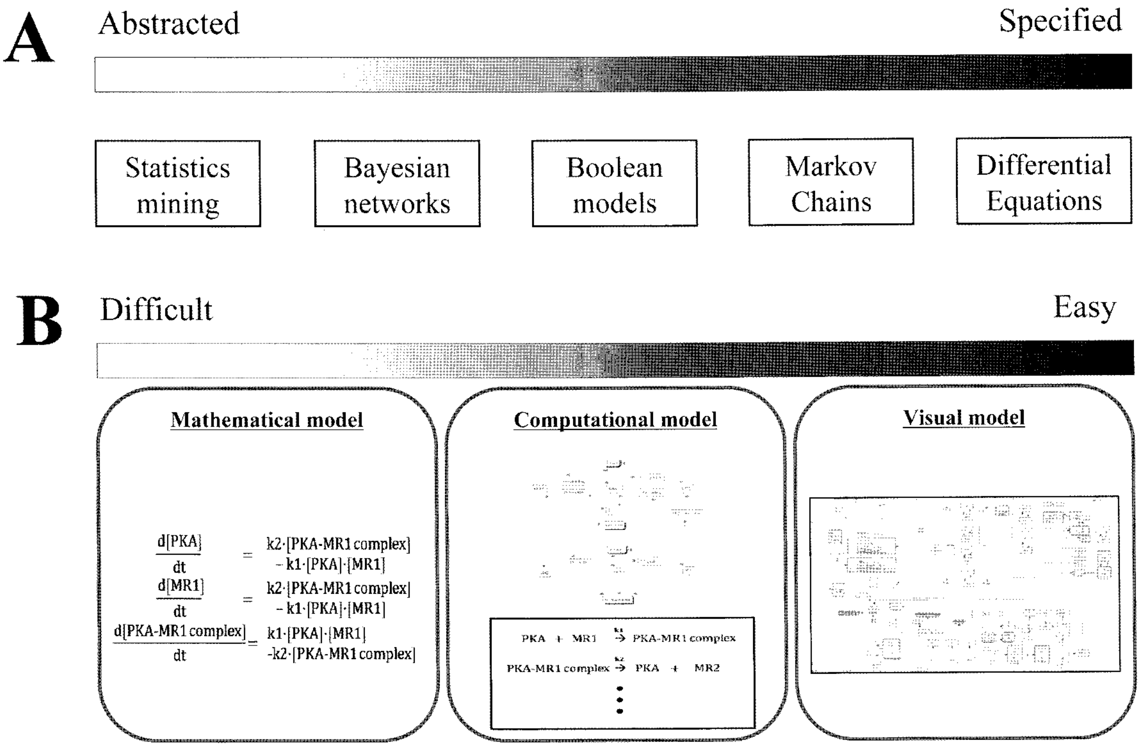
A
Abstracted
Specified
Statistics mining
Bayesian networks
Boolean models
Markov Chains
Differential Equations
B
Difficult
Easy
Mathematical model
d[PKA]/dt = k2·[PKA-MR1 complex] - k1·[PKA]·[MR1]
d[MR1]/dt = k2·[PKA-MR1 complex] - k1·[PKA]·[MR1]
d[PKA-MR1 complex]/dt = k1·[PKA]·[MR1] -k2·[PKA-MR1 complex]
Computational model
PKA + MR1 → PKA-MR1 complex
PKA-MR1 complex → PKA + MR2
Visual model

Fig. 2. Classifications of modeling in systems biology. Diverse spectrum of modeling based on availability of equation parameters. This classification was described in previous review paper by Ideker and Lauffenburger (2003). Diverse spectrum of modeling based on ease of modeling. At the difficult level of modeling, the abstract mathematical equations should be written representing biological processes by the modeler himself. Although there is more flexibility of what can be achieved, a substantial expertise in various mathematical methodologies is expected. Also model reusability is limited as in different set of equations have to be rewritten if one decided to change underlying mathematical methodology (e.g. deterministic to stochastic *vice versa*). The next level is the computational modeling. Researchers utilize specialized software packages to represent the biological processes through equations in a form that is more meaningful to them. The computer in turn will generate mathematical equations. The easiest level of modeling is visual modeling. Researchers use graphical notations to represent the biological knowledge and enter required parameters in a user friendly software environment. The visual modeling software package (e.g. CellDesigner) translates the graphical notations into mathematical equations and render it ready for simulations.

empirical observations in a physically and biologically realistic manner and generate novel and useful hypothesis, it play an instrumental role in system-level understanding of complex biological processes. Several comprehensible review papers are published focusing on mathematical modeling including model construction, model verification, model analysis, model regression and model validation that can serve as good guide for non-expert (Aldridge et al., 2006; Jaqaman and Danuser, 2006). Figure 2 lists few of the existing modeling methodologies and their classification. Furthermore computational tools have become integral part of mathematical modeling process and they can help non-mathematicians by assisting in daunting task of translating biochemical equations/diagrams into mathematical equations. Several tools exist which incorporates multitude of functionalities to suit to different modeling and analysis needs. We have compiled an exhaustive list of the useful software packages, databases and web sites previously (Shreenivasaiah et al., 2008). Alves *et al.* (2006) has reviewed comparisons and evaluation of 12 useful kinetic modeling software packages with respect to their functionality, reliability, efficiency, user-friendliness and compatibility (Alves et al., 2006). In the previous sections, we covered network modeling mainly focusing on topological design principles and structural characterization; here in this section general aspects of mathematical modeling useful from the perspective of dynamic systems modeling will be briefly discussed.

Trends in mathematical modeling efforts

A number of dynamic computational models have been developed and used as a power tool for testing hypotheses about biological system. Recently, Andersen et al. (2008) published manually reconstructed largest metabolic network model of *A. niger* (fungus) by integrating the genome, reactome and metabolome data from bibliome (totality of biological text corpus). This network along with the known data on fluxes, yields and transcription was used to construct the mathematic model, which they used to examine system-wide data in metabolic context. The model accurately predicted many of the experimental results already published pertaining to yields and flux distributions.

Further systematic perturbation analysis on model resulted in new information on physiological traits of *A. niger* which could be readily exploited for high yields in biotechnology industries.

Dynamic systems responses to various environmental stimuli can be elucidated by systems modeling of signaling pathways. Such model construction is feasible for sensitivity and stability analysis which can result in wealth of information such as behavior changes when stimuli and rate constants are modified (Kitano, 2002). Birtwistle et al. (2007) developed a computational ErbB signaling pathway (has key role in tumorigenesis) model composed of all four ErbB receptors, and corroborated the prediction results with traditional experiments. They compared the stabilities of different sub-networks by performing stability analysis and drew a conclusion that heregulin-dependent extracellular-signal-regulated kinase activation is more stable than epidermal growth factor-dependent activation. This model could help in gaining mechanistic insight into ligand-dependent response of ErbB signaling. Li et al. (2006) developed the stochastic model describing synthesis and uptake of AI-2 by bacteria, which is reported as a 'universal' signal molecule when bacteria are sensing the environmental cues. Using this mathematical model and testable hypothesis, they discovered the existence of an alternative pathway for AI-2 synthesis which was unable to be discovered by any of the reductionist approaches. All of these studies mentioned here nicely illustrate the value of quantitative modeling in guiding and interpreting experimental data collection. Similar results and conclusions could not have been drawn based solely on traditional approaches. However such comprehensive assessments and explanations come at a price that often these models need tens or even hundreds of parameters which are difficult to obtain, as high throughput methods for measuring biochemical parameters remain limited. Furthermore, since these parameters significantly affect model behavior, the values measured *in vitro* may produce inaccurate results in an *in vivo* application.

Visual modeling

Visual modeling is a rather unambiguous representation of system components and interactions using graphical notations for easy understanding and more efficient and accurate transmission of biological knowledge. Several graphical notations for molecular interactions and pathway diagram are proposed by different groups (Kohn, 2001; Pirson et al., 2000); among them most widely used notation are those proposed by Kitano which then later became basis for Systems Biology Graphical Notation (SBGN) initiative. Currently SBGN is considered as the standard for graphical notation used in diagrams of biochemical and cellular processes studied in systems biology. CellDesigner, a tool for visual process modeling is developed and maintained by Kitano's group (Oda et al., 2005) which implements SBGN. Other tools frequently used for visual modeling are Cytoscape, JDesigner, Ingenuity pathway analysis and Pathway studio. These software has facilitated construction of large scale signaling pathway maps, most of which are available online freely (Calzone et al., 2008; Oda and Kitano, 2006; Oda et al., 2005). Further, in order to equip the tools with additional functionalities often plug-ins are developed and distributed (Erhard et al., 2008; Funahashi et al., 2007). PANTHER pathway database is one of the largest pathway databases. It is equivalent to online/web version of CellDesigner (Mi et al., 2005). All of these user friendly tools enables development of mathematical models feasible for biologist without much modeling experience.

CONCLUDING REMARKS

Systems biology is one of the most widely discussed fields during the post-genomic era. Currently, the systems biological research efforts are focused on investigating the components of cellular networks and their interactions using the data obtained from various high-throughput technologies. A rapid progress has also been made in production of various computational algorithms that are used to examine the detailed network structures. In the future, continuous efforts will be made to develop different types of models to integrate the complex biological data, so that the complex life phenomena can be visualized and understood in a relatively comprehensive way. The information on the detailed protein complex structures and the way how to produce the assemblies will be determined not only by the biophysical methods such as X-ray crystallography, but also by the state of the art new technology such as electron tomography.

REFERENCES

Albeck JG, MacBeath G, White FM, Sorger PK, Lauffenburger DA, and Gaudet S (2006) Collecting and organizing systematic sets of protein data. *Nat Rev Mol Cell Biol* 7: 803-812.

Albert R, Jeong H, and Barabasi AL (2000) Error and attack tolerance of complex networks. *Nature* 406: 378-382.

Aldridge BB, Burke JM, Lauffenburger DA, and Sorger PK (2006) Physicochemical modelling of cell signalling pathways. *Nat cell biol* 8: 1195-1203.

Aloy P, Bottcher B, Ceulemans H, Leutwein C, Mellwig C, Fischer S, Gavin AC, Bork P, Superti-Furga G, Serrano L, and Russell RB (2004) Structure-based assembly of protein complexes in yeast. *Science* 303: 2026-2029.

Aloy P and Russell RB (2002) Interrogating protein interaction networks through structural biology. *Proc Natl Acad Sci U S A.* 99: 5896-5901.

Aloy P and Russell RB (2006) Structural systems biology:

modelling protein interactions. *Nat Rev Mol Cell Biol* 7: 188-197.

Alves R, Antunes F, and Salvador A (2006) Tools for kinetic modeling of biochemical networks. *Nat biotechnol* 24: 667-672.

Andersen MR, Nielsen ML, and Nielsen J (2008) Metabolic model integration of the bibliome, genome, metabolome and reactome of Aspergillus niger. *Mol syst biol* 4: 178.

Bansal M, Belcastro V, Ambesi-Impiombato A, and di Bernardo D (2007) How to infer gene networks from expression profiles. *Mol syst biol* 3: 78.

Bar-Joseph Z, Gerber GK, Lee TI, Rinaldi NJ, Yoo JY, Robert F, Gordon DB, Fraenkel E, Jaakkola TS, Young RA, and Gifford DK (2003) Computational discovery of gene modules and regulatory networks. *Nat biotechnol* 21: 1337-1342.

Bard JB and Rhee SY (2004) Ontologies in biology: design, applications and future challenges. *Nat Rev Genet* 5: 213-222.

Blake JA and Bult CJ (2006) Beyond the data deluge: data integration and bio-ontologies. *J Biomed Inform* 39: 314-320.

Boxem M, Maliga Z, Klitgord N, Li N, Lemmens I, Mana M, de Lichtervelde L, Mul JD, van de Peut D, Devos M, et al. (2008) A protein domain-based interactome network for C. elegans early embryogenesis. *Cell* 134: 534-545.

Bray D (2001) Reasoning for results. *Nature* 412: 863.

Brosh R, Shalgi R, Liran A, Landan G, Korotayev K, Nguyen GH, Enerly E, Johnsen H, Buganim Y, Solomon H, et al. (2008) p53-repressed miRNAs are involved with E2F in a feed-forward loop promoting proliferation. *Mol syst biol* 4: 229.

Busch H, Camacho-Trullio D, Rogon Z, Breuhahn K, Angel P, Eils R, and Szabowski A (2008) Gene network dynamics controlling keratinocyte migration. *Mol syst biol* 4: 199.

Cerami EG, Bader GD, Gross BE, and Sander C (2006) cPath: open source software for collecting, storing, and querying biological pathways. *BMC Bioinformatics* 7: 497.

Chechik G, Oh E, Rando O, Weissman J, Regev A, and Koller D (2008) Activity motifs reveal principles of timing in transcriptional control of the yeast metabolic network. *Nat biotechnol* 26: 1251-1259.

Chee M, Yang R, Hubbell E, Berno A, Huang XC, Stern D, Winkler J, Lockhart DJ, Morris MS, and Fodor SP (1996) Accessing genetic information with high-density DNA arrays. *Science* 274: 610-614.

Choi C, R Munch et al. (2007) SYSTOMONAS-an integrated database for systems biology analysis of Pseudomonas *Nucleic Acids Res* 35: D533-7.

Cline MS, Smoot M, Cerami E, Kuchinsky A, Landys N, Workman C, Christmas R, Avila-Campilo I, Creech M, Gross B, et al. (2007) Integration of biological networks and gene expression data using Cytoscape. *Nat Protoc* 2: 2366-2382.

Cox B, Kislinger T, and Emili A (2005) Integrating gene and protein expression data: pattern analysis and profile mining. *Methods* 35: 303-314.

De Keersmaecker SC, Thijs IM, Vanderleyden J, and Marchal K (2006) Integration of omics data: how well does it work for bacteria? *Mol Microbiol* 62: 1239-1250.

Duarte NC, Becker SA, Jamshidi N, Thiele I, Mo ML, Vo TD, Srivas R, and Palsson BO (2007) Global reconstruction of the human metabolic network based on genomic and bibliomic data. *Proc Natl Acad Sci U S A* 104: 1777-1782.

Flicek P, Aken BL, Beal K, Ballester B, Caccamo M, Chen Y, Clarke L, Coates G, Cunningham F, Cutts T, et al. (2008) Ensembl 2008. *Nucleic Acids Res* 36: D707-714.

Funahashi A, Jouraku A, Matsuoka Y, and Kitano H (2007) Integration of CellDesigner and SABIO-RK. *In silico biol* 7: S81-90.

Gray JJ (2006) High-resolution protein-protein docking. *Curr Opin Struct Biol.* 16: 183-193.

Guerrero C, Milenkovic T, Przulj N, Kaiser P, and Huang L (2008) Characterization of the proteasome interaction network using a QTAX-based tag-team strategy and protein interaction network analysis. *Proc Natl Acad Sci USA.* 105: 13333-13338.

Guerrero C, Tagwerker C, Kaiser P, and Huang L (2006) An integrated mass spectrometry-based proteomic approach: quantitative analysis of tandem affinity-purified in vivo cross-linked protein complexes (QTAX) to decipher the 26 S proteasome-interacting network. *Mol Cell Proteomics* 5: 366-378.

Hall DA, Ptacek J, and Snyder M (2007) Protein microarray technology. *Mech Ageing Dev* 128: 161-167.

Hardiman G (2004) Microarray platforms--comparisons and contrasts. *Pharmacogenomics* 5: 487-502.

Hartwell LH, Hopfield JJ, Leibler S, and Murray AW (1999) From molecular to modular cell biology. *Nature* 402: C47-52.

Holden M, Deng S, Wojnowski L, and Kulle B (2008) GSEA-SNP: applying gene set enrichment analysis to SNP data from genome-wide association studies. *Bioinformatics* 24: 2784-2785.

Hwang D, Rust AG, Ramsey S, Smith JJ, Leslie DM, Weston AD, de Atauri P, Aitchison JD, Hood L, Siegel AF, and Bolouri H (2005a) A data integration methodology for systems biology. *Proc Natl Acad Sci U S A* 102: 17296-17301.

Hwang D, Smith JJ, Leslie DM, Weston AD, Rust AG, Ramsey S, de Atauri P, Siegel AF, Bolouri H, Aitchison JD, and Hood L (2005b) A data integration methodology for systems biology: experimental verification. *Proc Natl Acad Sci U S A* 102: 17302-17307.

Ideker T and D Lauffenburger (2003) Building with a scaffold: emerging strategies for high- to low-level cellular modeling. *Trends Biotechnol* 21(6): 255-62.

Ihmels J, Friedlander G, Bergmann S, Sarig O, Ziv Y, and Barkai N (2002) Revealing modular organization in the yeast transcriptional network. *Nat Genet.* 31: 370-377.

Ishii N, Nakahigashi K, Baba T, Robert M, Soga T, Kanai A, Hirasawa T, Naba M, Hirai K, Hoque A, et al. (2007) Multiple high-throughput analyses monitor the response of E. coli to perturbations. *Science* 316: 593-597.

Jaqaman K and Danuser G (2006) Linking data to models: data regression. *Nat Rev Mol Cell Biol* 7: 813-819.

Jeong H, Tombor B, Albert R, Oltvai ZN, and Barabasi AL (2000) The large-scale organization of metabolic networks. *Nature* 407: 651-654.

Joshi-Tope G, Gillespie M, Vastrik I, D'Eustachio P, Schmidt E, de Bono B, Jassal B, Gopinath GR, Wu GR, Matthews L, et al. (2005) Reactome: a knowledgebase of biological pathways. *Nucleic acids res* 33: D428-432.

Joyce AR and Palsson BO (2006) The model organism as a system: integrating ‘omics’ data sets. *Nat Rev Mol Cell Biol* 7: 198-210.

Kahlem P and Birney E (2007) ENFIN a network to enhance integrative systems biology. *Ann N Y Acad Sci* 1115: 23-31.

Kell DB (2004) Metabolomics and systems biology: making sense of the soup. *Curr Opin Microbiol.* 7: 296-307.

Kitano H (2002) Systems biology: a brief overview. *Science* 295: 1662-1664.

Kwon YK and Cho KH (2008) Coherent coupling of feedback loops: a design principle of cell signaling networks. *Bioinformatics* 24: 1926-1932.

Lee TI, Rinaldi NJ, Robert F, Odom DT, Bar-Joseph Z, Gerber GK, Hannett NM, Harbison CT, Thompson CM, Simon I, et al. (2002) Transcriptional regulatory networks in Saccharomyces cerevisiae. *Science* 298: 799-804.

Legewie S, Herzel H, Westerhoff HV, and Bluthgen N (2008) Recurrent design patterns in the feedback regulation of the mammalian signalling network. *Mol syst biol* 4: 190.

Lenhard B, Wahlestedt C, and Wasserman WW (2003) GeneLynx mouse: integrated portal to the mouse genome. *Genome Res* 13: 1501-1504.

Leontis NB, Altman RB, Berman HM, Brenner SE, Brown JW, Engelke DR, Harvey SC, Holbrook SR, Jossinet F, Lewis SE, et al. (2006) The RNA Ontology Consortium: an open invitation to the RNA community. *RNA* 12: 533-541.

Lin R, Dai S, Irwin RD, Heinloth AN, Boorman GA, and Li L (2008) Gene set enrichment analysis for non-monotone association and multiple experimental categories. *BMC Bioinformatics* 9: 481.

Liu Y and Zhao H (2004) A computational approach for ordering signal transduction pathway components from genomics and proteomics Data. *BMC Bioinformatics* 5: 158.

Lynn DJ, Winsor GL, Chan C, Richard N, Laird MR, Barsky A, Gardy JL, Roche FM, Chan THW, Shah N, et al. (2008) InnateDB: facilitating systems-level analyses of the mammalian innate immune response. *Mol Syst Biol* 4.

Lu L, Arakaki AK, Lu H, Skolnick J (2003) Multimeric threading-based prediction of protein-protein interactions on a genomic scale: application to the Saccharomyces cerevisiae proteome. *Genome Res.* 13:1146-1154

Ma H, Sorokin A, Mazein A, Selkov A, Selkov E, Demin O, and Goryanin I (2007) The Edinburgh human metabolic network reconstruction and its functional analysis. *Mol Syst Biol* 3: 135.

Marcotte EM, Pellegrini M, Thompson MJ, Yeates TO, and Eisenberg D (1999) A combined algorithm for genome-wide prediction of protein function. *Nature* 402: 83-86.

Medalia O, Weber I, Frangakis AS, Nicastro D, Gerisch G, and Baumeister W (2002) Macromolecular architecture in eukaryotic cells visualized by cryoelectron tomography. *Science* 298: 1209-1213.

Melo RC, Dvorak AM, and Weller PF (2008) Electron tomography and immunonanogold electron microscopy for investigating intracellular trafficking and secretion in human eosinophils. *J Cell Mol Med* 12: 1416-1419.

Milo R, Shen-Orr S, Itzkovitz S, Kashtan N, Chklovskii D, and Alon U (2002) Network motifs: simple building blocks of complex networks. *Science* 298: 824-827.

Muller SA, Aebi U, and Engel A (2008) What transmission electron microscopes can visualize now and in the future. *J Struct Biol* 163: 235-245.

Murray DB, Beckmann M, and Kitano H (2007) Regulation of yeast oscillatory dynamics. *Proc Natl Acad Sci USA* 104: 2241-2246.

Neduva V, Linding R, Su-Angrand I, Stark A, de Masi F, Gibson TJ, Lewis J, Serrano L, and Russell RB (2005) Systematic discovery of new recognition peptides mediating protein interaction networks. *PLoS biology* 3: e405.

Nickell S, Kofler C, Leis AP, and Baumeister W (2006) A visual approach to proteomics. *Nat Rev Mol Cell Biol* 7: 225-230.

Pieper U, Eswar N, Braberg H, Madhusudhan MS, Davis FP, Stuart AC, Mirkovic N, Rossi A, Marti-Renom MA, Fiser A, et al. (2004) MODBASE, a database of annotated comparative protein structure models, and associated resources. *Nucleic acids res* 32: D217-222.

Ravasz E, Somera AL, Mongru DA, Oltvai ZN, and Barabasi AL (2002) Hierarchical organization of modularity in metabolic networks. *Science* 297: 1551-1555.

Rousseau F and Schymkowitz J (2005) A systems biology perspective on protein structural dynamics and signal transduction. *Curr Opin Struct Biol* 15: 23-30.

Rubin DL, Shah NH, and Noy NF (2008) Biomedical ontologies: a functional perspective. *Brief Bioinform* 9: 75-90.

Safran M, Solomon I, Shmueli O, Lapidot M, Shen-Orr S, Adato A, Ben-Dor U, Esterman N, Rosen N, Peter I, et al. (2002) GeneCards 2002: towards a complete, object-oriented, human gene compendium. *Bioinformatics* 18: 1542-1543.

Sali A, Glaeser R, Earnest T, and Baumeister W (2003) From words to literature in structural proteomics. *Nature* 422: 216-225.

Sauer U, Heinemann M, and Zamboni N (2007) Genetics. Getting closer to the whole picture. *Science* 316: 550-551.

Sayers EW, Barrett T, Benson DA, Bryant SH, Canese K, Chetvernin V, Church DM, Dicuccio M, Edgar R, Federhen S, et al. (2008) Database resources of the National Center for Biotechnology Information. *Nucleic Acids Res.*

Schena M, Shalon D, Davis RW, and Brown PO (1995) Quantitative monitoring of gene expression patterns with a complementary DNA microarray. *Science* 270: 467-470.

Shah AR, Singhal M, Klicker KR, Stephan EG, Wiley HS, and Waters KM (2007) Enabling high-throughput data management for systems biology: the Bioinformatics Resource Manager. *Bioinformatics* 23: 906-909.

Shi L, Reid LH, Jones WD, Shippy R, Warrington JA, Baker SC, Collins PJ, de Longueville F, Kawasaki ES, Lee KY, et al. (2006) The MicroArray Quality Control (MAQC) project shows inter- and intraplatform reproducibility of gene expression measurements. *Nat Biotechnol* 24: 1151-1161.

Shreenivasaiah PK, Rho SH, Kim T, and Kim DH (2008) An overview of cardiac systems biology. *J Mol Cell Cardiol* 44: 460-469.

Smith B, Ashburner M, Rosse C, Bard J, Bug W, Ceusters W, Goldberg LJ, Eilbeck K, Ireland A, Mungall CJ, et al. (2007) The OBO Foundry: coordinated evolution of ontologies to support biomedical data integration. *Nat Biotechnol* 25: 1251-1255.

Sprinzak E and Margalit H (2001) Correlated sequence-signatures as markers of protein-protein interaction. *J Mol Biol* 311: 681-692.

Stein LD (2003) Integrating biological databases. *Nat Rev Genet* 4: 337-345.

Stemke-Hale K, Gonzalez-Angulo AM, Lluch A, Neve RM, Kuo

WL, Davies M, Carey M, Hu Z, Guan Y, Sahin A, et al. (2008) An integrative genomic and proteomic analysis of PIK3CA, PTEN, and AKT mutations in breast cancer. *Cancer Res* 68: 6084-6091.

Sugawara H, Ogasawara O, Okubo K, Gojobori T, and Tateno Y (2008) DDBJ with new system and face. *Nucleic Acids Res* 36: D22-24.

’t Hoen PA, Ariyurek Y, Thygesen HH, Vreugdenhil E, Vossen RH, de Menezes RX, Boer JM, van Ommen GJ, and den Dunnen JT (2008) Deep sequencing-based expression analysis shows major advances in robustness, resolution and inter-lab portability over five microarray platforms. *Nucleic Acids Res* 36: e141.

Templin MF, Stoll D, Schwenk JM, Potz O, Kramer S, and Joos TO (2003) Protein microarrays: promising tools for proteomic research. *Proteomics* 3: 2155-2166.

Toyoda T, Mochizuki Y, Player K, Heida N, Kobayashi N, and Sakaki Y (2007) OmicBrowse: a browser of multidimensional omics annotations. *Bioinformatics* 23: 524-526.

Uetz P, Giot L, Cagney G, Mansfield TA, Judson RS, Knight JR, Lockshon D, Narayan V, Srinivasan M, Pochart P, et al. (2000) A comprehensive analysis of protein-protein interactions in Saccharomyces cerevisiae. *Nature* 403: 623-627.

Van Dien S and Schilling CH (2006) Bringing metabolomics data into the forefront of systems biology. *Mol Syst Biol* 2: 2006 0035.

Wang E (Edwin Wang) MicroRNA systems biology. RNA Technologies in Cardiovascular Medicine and Research, Springer-Verlag: 69-80.

Waters KM, Pounds JG, and Thrall BD (2006) Data merging for integrated microarray and proteomic analysis. *Brief Funct Genomic Proteomic* 5: 261-272.

Wright J and Wagner A (2008) The Systems Biology Research Tool: evolvable open-source software. *BMC Syst Biol* 2: 55.

Wu X, Jiang R, Zhang MQ, and Li S (2008) Network-based global inference of human disease genes. *Mol Syst Biol* 4: 189.

Xia T and Dickerson JA (2008) OmicsViz: Cytoscape plug-in for visualizing omics data across species. *Bioinformatics* 24: 2557-2558.

Yang K, Bai H, Ouyang Q, Lai L, and Tang C (2008) Finding multiple target optimal intervention in disease-related molecular network. *Mol Syst Biol* 4: 228.

Yu H, Braun P, Yildirim MA, Lemmens I, Venkatesan K, Sahalie J, Hirozane-Kishikawa T, Gebreab F, Li N, Simonis N, et al. (2008) High-quality binary protein interaction map of the yeast interactome network. *Science* 322: 104-110.

Zhang W, Zhang Y, Zheng H, Zhang C, Xiong W, Olyarchuk JG, Walker M, Xu W, Zhao M, Zhao S, et al. (2007) SynDB: a Synapse protein DataBase based on synapse ontology. *Nucleic Acids Res* 35: D737-741.

[Received November 18, 2008; accepted December 15, 2008]