CHAPTER

1

Foundations for Systems Biomedicine: an Introduction

Edison T. Liu
Genome Institute of Singapore, Singapore

OUTLINE

INTRODUCTION

Quantitative biology, mathematical biology and mathematical modeling have all been part of biological investigations in one form or another since the beginnings of investigative biology and medicine. Carl Linnaeus' creation of the binary nomenclature (*Systema Naturae*, Carolus Linnaeus, 1735) marked the origin of biologic taxonomy and provided the basis for phylogenic analysis. William Harvey (*An Anatomical Disquisition on the Motion of the Heart and Blood in Animals*, William Harvey, 1847) described the quantification of the amount of blood in the chambers of the heart, calculated the output of the heart by multiplying the volume by the number of heart beats per day, and noted that the output differed wildly from the volume of blood in an individual at any one time. With this information, he developed a model of circulating blood that could explain the blood volume discrepancies with supporting evidence from the anatomic presence of valves in veins. The mathematical tradition in biology therefore, runs long and deep.

However, systems biology, as we conceive of it, differs in scale and formalism from the

earlier quantitative traditions. As with any new field, there are many opinions as to what systems biology is. For the purposes of this book, systems biology can be described as a discipline that seeks to quantify and annotate complexity in biological systems in order to construct algorithmic models with which to predict outcomes from component input. Systems biomedicine is an extension of these strategies into the study of biomedical problems. We believe that this demarcation is relevant, given the challenges of the complexity of the human organism and the human impact of the results of these investigations.

This definition of systems biomedicine highlights the difference between quantitative data acquisition and systems biology. The scale of data acquisition in biology today is unparalleled in history. Analog and descriptive data such as cellular images are now digitalized and converted to discrete data points. Genomic- and proteomic-scale information is registered in the gigabyte scale per experiment. This reality also demands formal mathematical and algorithmic conversion of experimental data in biology in order for them to be simply understood by the investigator. The interposition of computers and their algorithms as an essential part of biological research immediately places, at least a rudimentary, mathematical formalism around all experiments performed in this fashion.

Although measuring outcomes is standard in day-to-day biological experiments, these earlier quantitative approaches do not scale. While detailed biochemical kinetics can be calculated for a single biochemical reaction, most commonly, we have tended to resort to descriptive generalizations when we ascend to physiological scales. With current technologies that can acquire precise, comprehensive and quantitative data, biological complexity can now be quantitatively analyzed. The challenge, however, is to identify the optimal mathematical approaches most suited for this scale and complexity of analysis.

Physiologists and pharmacologists have always sought to quantify inputs and outputs in complex organisms, and the later generations of William Harveys had rendered pulmonary and cardiac physiology into equations. To a large extent, this approach has been remarkably successful, and has brought us many of the medical advances in cardiopulmonary medicine and surgery. The cardiac diagnostics from angiography, to echocardiography, to telemetry in which patient physiologic output is monitored and automated alerts generated, represent a culmination of such research in cardiac function. In a sense, physiology was the systems science in medicine. However, these organ-level models do not parse with molecular realities, because their unit of measure is in average blood flow, for example, and not in the flow dynamics of the red corpuscle. Therefore, in the past, quantitative physiologic models could not be unified with cellular models and, by scale, to molecular models. Moreover, the need that assumptions be greatly simplified in order to arrive at computationally tractable models also limited the relevance of many physiologic models.

Now, however, medicine is becoming amenable to complexity analysis. The understanding of the cell and molecular biology of human disease has dramatically advanced in the past 25 years. Whereas the pathophysiology of most human diseases was previously limited to the analysis of organ failure, most diseases now have a cellular and molecular explanation. It is precisely this reduction to common units of measure—to the cell and the molecules within the cell—that allows systems analyses to be applied across the entire human condition. Therefore, the pump dynamics of the heart after myocardial infarction can be resolved at the same level as pancreatic beta-cell function in diabetes mellitus. There is convergence.

The current systems biology now includes two important new characteristics that distinguish it from historical physiology and mathematical biology. First, there is a focus on complexity; secondly, the fundamental unit of study resides in the DNA (and, by association, protein) sequence.

That the unit of measure can be the nucleotide now provides the *lingua franca* that permits the direct translation of experimental results from biochemistry to cell biology, to physiology and to population genetics. Moreover, the ultra-high-throughput and multiplex genomic technologies allow for the digitalization of experimental data of such precision and comprehensiveness that the true complexity of a biological system can actually be measured and dissected. In all aspects—biological and mathematical—the greatest advance has been the availability of computational capabilities that can match the systems complexity. This reliance on these genomic and computational technologies and datasets that can be transmuted across species has broadened significantly the applicability of systems approaches to very complex systems such as human medicine.

Other thinkers have expounded on the new possibilities in integrating mathematics with biology. In an excellent essay, Joel E. Cohen (2004) noted that "mathematics is not only biology's next microscope, but in fact is better". He observed that, in biology, enormous complexity of up to 100 million species is built on just a few basic elements of carbon, nitrogen, hydrogen and oxygen. By contrast, the entire periodic table generates only several thousand kinds of minerals in the earth's crust. Thus the entire basis of biology is a complexity that produces ensemble or emergent properties of much greater function than the component parts. Cohen argued that mathematics can also benefit from attacking biological problems as it did in working through problems in physics. Calculus was developed in part to help solve the problems of celestial motion and of optics. Similarly, the multilayered complexity, interlocking control loops, distributed switch mechanisms and the differential use of the same components over developmental time challenges mathematical and computational solutions. It is likely that new mathematics will be required to deal with these ensemble properties and with the heterogeneity of the biological input that feeds into the organismic output.

Geneticists have already defined phenotypic interactions between genes or alleles as epistasis (Phillips, 2008). In many cases, new properties emerge: two white flowers that when crossed give a purple flower, or two genes that when individually mutated give no phenotype, but show a lethal outcome when both are mutated. The mathematical representation of epistasis can be:

$$W_{xy} = x + y + \delta \tag{1.1}$$

where W is the observed phenotype, x and y are the individual effects of each allele at loci x and y, δ is the deviation that is due to epistasis. Systems biology, however, examines the sum of all epistatic relationships and hopes to uncover the hierarchy. This, indeed, has been the direction of this line of genetic research. Tong et al. (2004) crossed mutations in 132 "query" genes into a set of 4700 viable yeast gene deletion mutants to develop a genetic interaction map containing more than 4000 functional gene interactions. Classical genetics converges on systems biology.

Kitano (2007) noted the importance of control theory in describing biological systems, and described the primacy of "robustness" in the design of biological systems. He differentiated robustness from homeostasis, in that homeostasis seeks to return the system to the original state, whereas robustness will accommodate migration to another state to achieve survivability. One characteristic of evolvable systems described by the Highly Optimized Theory (HOT) states that such systems are robust against common perturbations, but are fragile against unusual ones (Carlson and Doyle, 2000). A common example is the World Wide Web, which, despite being robust because of its high interconnectivity, has been brought down by specific attacks at hubs of activity. Thus systems robustness is a matter of "trade-offs." Mathematical descriptions of robustness have been attempted.

Kitano (2007) provides a representation of robustness in the following equation, but also

acknowledges that new mathematics may be necessary to accommodate these systems concepts in biology:

> "Robustness (R) of the system (s) with regard to function (a) against a set of perturbations (P):
>
> $$R^{s}_{a,P} = \int_{P} \psi(p) D^{s}_{a}(p) dp \qquad (1.2)$$
>
> The function ψ is the probability for perturbation 'p' to take place. P is the entire perturbation space, and D (p) is an evaluation function under perturbation (p)."

EXPERIMENTAL STRATEGIES IN SYSTEMS BIOLOGY

Systems approaches are characterized by several key attributes:

1. The measurement of quantitative and comprehensive data of an experimental system.
2. Assessment of the relationships between the component parts.
3. Perturbation of the system to detect response dynamics.
4. Intersection of orthogonal data to arrive at higher-order logic. (Orthogonal data are defined as datasets derived from different systems, perhaps addressing the same question in which the intersection of the two datasets can further resolve a problem: for example, the set of genes with binding sites of a transcription factor and the set of genes that are expressed with overexpression of the same transcription factor [*see* Chapter 4]).
5. Derivation of a model of the system that can be mathematical or qualitative.
6. Correct prediction of output based on the model.

The most complete analyses that engage all these attributes have been made in lower organisms. Bonneau and colleagues (2007) reported the construction of a complete functional biological network map for *Halobacterium salinarum*, an Archaea species that thrives in conditions of high salinity. The final network map describes the regulatory functional relationships among 80% of its genes. The predictive power of this model was evident in its ability to predict the transcriptional responses to challenge with novel environmental conditions or disruption of transcription factors. The predictive capability of this genome-wide, whole-organism predictive model was significant. In order to achieve this, Bonneau and colleagues accomplished the following in order to achieve their goal:

1. The 2.6 Mb *Halobacterium salinarum* genome was sequenced and functions were assigned to each gene using protein sequence and structural similarities (*know all the components*).
2. Cells were perturbed by varying concentrations of environmental factors and / or gene knockouts (*perturbation analysis*).
3. The transcriptional changes of all genes using microarrays were determined after each perturbation (*genomic readout for perturbation analysis*).
4. Diverse data (mRNA levels, evolutionary conservation in protein structure, metabolic pathways, and *cis*-regulatory motifs) were integrated to identify subsets of genes that are co-regulated in certain environment (*data integration*).
5. A dynamic network model was constructed for the of influence environmental and transcription factor changes on the expression of co-regulated genes (*model building*).
6. The resulting network was explored using software visualization tools within an integrator that enables software interoperability and database integration. This allowed for manual exploration and generation of hypotheses used to plan additional iterations of the systems analysis (*model testing*).

Similar strategies have been applied to the eukaryotic model system, yeast, with less predictive success (Luscombe, 2004; Tong, 2004; Yu, 2008). Nevertheless, the strategy still requires the integration of heterogeneous datasets, such as transcription factor binding sites, transcriptional profiles and protein–protein interactions (Fig. 1.1).

SYSTEMS BIOMEDICINE

Systems biomedicine is the analysis of medical problems using systems approaches; therefore pertinence to the human condition is a prerequisite. Given the complexity of mammalian systems, are we ready to study the ensemble properties of the human model, and are we sufficiently clever to use these approaches to understand and to treat human disease? Before 2001, perhaps, it would have been difficult to answer affirmatively. If access to the complete human genome is a prerequisite for a systems analysis, only after the sequencing of the human genome could this goal be conceived (Lander et al., 2001). Together with the advent of expression arrays in 1996 (Shalon, 1996) and their stable use by 2000, these technologies launched the next phase of growth for systems approaches to complex organisms like mammals. Network analyses have been conducted primarily where the system is cell-based, such as immunology (Kitano, 2006) or cancer (Segal et al., 2005), or where the tissue is homogeneous such as the heart (Olson, 2006) or liver (Schadt, 2008). Interestingly, computer scientists have looked to the natural immune system to develop analogous artificial immune systems for computer system security (Forrest and Beauchemin, 2007). There is much to be learned from biological systems that have had the benefit of more than a billion years of evolutionary history.

The experimental systems approaches to studying a human problem will, by necessity, be different and potentially less complete than those appropriate for attacking a question using prokaryotes. Such reconstruction of a regulatory network has been difficult in higher organisms, owing to the dramatically increased complexity of the contributing subsystems. Thus the possible solution space is orders of magnitude

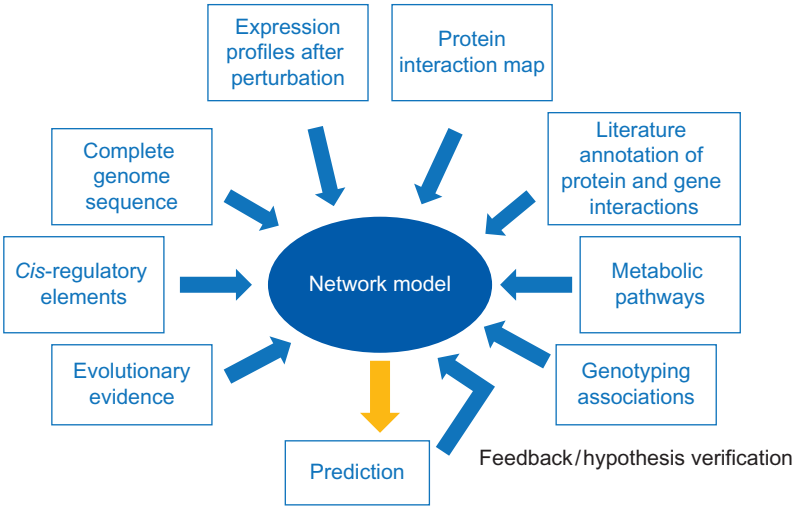

FIGURE 1.1 Input Information that can be Used to Construct a Network Model.

greater than that for lower organisms. Gene numbers increase in higher eukaryotes, but this is not the confounding factor: splice variants, transcription factors binding at great distances from the transcriptional start sites, gene duplication, post-transcriptional regulation by microRNAs and other non-coding RNA species, and complex post-translational modifications that change binding affinities all radically augment the complexity of the components.

Despite these challenges, network models of subsystems have been described, for example for the class of receptor tyrosine kinases (RTKs) (Amit et al., 2007a, 2007b; Katz, 2007). In these analyses, signaling hubs for the RTKs, such as RAF and the phosphoinositide 3′ kinase PI3K–AKT nodes, are noted to be frequent points of attack by oncogenic viruses, in addition to being sites of *de-novo* mutations in primary cancers. Such hubs, independently identified by both viruses and cancer mutations, also are effective targets for anticancer therapeutics.

Exploiting kinase networks, Sachs et al. (2005) pursued an interesting alternative strategy. Using multicolor / multiparameter flow cytometry in which up to 11 different features can be determined when labeled with different fluorophores, they quantitatively assessed the combinatorial presence of specific phosphorproteins indicative of activated kinases. Because flow cytometry assesses the biochemical state of individual cells, a large number of observations can be accumulated that would otherwise be an average of the population. In this manner, Sachs and colleagues were able to construct a Bayesian network from these data. Bayesian network models disclose the dependent effect of each biomolecule on the others, and therefore can infer causal relationships. Examining signaling in T cells, they could construct a network map that faithfully portrayed known and experimentally validated kinase–substrate relationships. In a similar fashion, they mapped the signaling profiles of acute myeloid leukemia cells after cytokine challenge and found 36 node states, following 6 stimulation conditions assessing 6 signaling molecules. These states could separate acute myeloid leukemia cells into signaling classes that corresponded to cytogenetic and clinical parameters (Irish, 2004).

It has been said that biology asks six kinds of question (Cohen, 2004): How is it built? How does it work? How did it begin? What is it for? The remaining two questions are more in the domain of medicine: What goes wrong? How is it fixed? So, systems biomedicine focuses, not only on human biology, but also human disease. Efforts to examine perturbations in gene and protein networks for clues to disease etiology have been pursued and will be described in subsequent chapters in this book. Most efforts are in the bench-to-bedside direction, but one approach that starts commonly from the patient and is validated at the bench is in human and population genetics of disease genes.

Human variations in the form of single nucleotide polymorphisms (SNPs) are used to identify genetic loci statistically associated with disease when compared with control populations. When assessed on a genome-wide basis, this has been a powerful, unbiased means of uncovering disease-associated genes. When expression arrays are coupled with genetic markers, expression quantitative trait loci (eQTL) can be assigned. In eQTL analysis, each transcript on the array is considered to be a quantitative phenotype and is correlated with the SNP configuration at each locus in the genome (Cheung et al., 2005; Sieberts and Schadt, 2007). *cis*-eQTL represent those SNPs adjacent to the measured gene of which the configuration is correlated with transcript levels, whereas *trans*-eQTLs are those associated with SNPs that are distant from the transcribed gene. eQTLs in humans have been used as proof of the genetic basis of gene expression in humans (Cheung et al., 2008; Spielman et al., 2007). When viewed on a genome-wide basis, a transcriptional network of regulatory "influence" can be discerned by statistical association between individual SNPs and expression of genes anywhere in

the genome. Schadt and his colleagues at Rosetta Pharmaceuticals have shown that combining genotypic data and expression data can increase precision of the discovery for disease-associated genes (Drake et al., 2005; Zhu et al., 2007).

CIRCADIAN CYCLES AS A RELEVANT MODEL FOR SYSTEMS BIOMEDICINE

An excellent example of a systems model that has medical importance is that of oscillators as regulators of the circadian rhythm. Oscillators are machines that cycle functions over time and are characterized by an automatic periodicity (*see* Chapter 4), and the best examples of biological oscillators are found in studies of circadian rhythm. The guiding motif for all living creatures is the ability to replicate, which imparts a cycling of functions. Over evolutionary time, there appeared to be an adaptive advantage to entrain such physiologic processes to an external clock defined by the day–night cycle. In order to do this, most organisms have found biochemical mechanisms to maintain this cycling, and mechanisms to sense the environment in order to modulate this periodicity.

The master circadian regulator in mammals is in the suprachiasmatic nucleus (SCN) in the brain. The molecular mechanism underpinning this oscillator has been elucidated. The basic helix–loop–helix containing transcription factor CLOCK interacts with BMAL1 to activate transcription of the *Per* and *Cry* genes. The Period (PER) and Cryptochrome (CRY) protein products heterodimerize and undergo negative feedback to inhibit their own transcription, and that of BMAL1. The PER–CRY repressor complex is degraded during the night, and *Clock-Bmal1* are de-repressed and can then induce transcription. There is a secondary feedback loop that involves the induction of a nuclear hormone receptor, REV-ERBα, by BMAL1/CLOCK. When REV-ERBα accumulates to threshold levels, it represses BMAL1/CLOCK. This secondary regulatory loop is not essential for the establishment of the circadian cycle, but it appears to be involved in stabilizing the regulatory framework. The oscillator function can be explained by a time delay in PER/CRY feedback inhibition of BMAL/CLOCK establishing a composite negative network motif with asymmetric timing. This oscillator is also affected by enzyme-families such as casein kinase 1 (CSNK1ε and CSNK1δ) that regulate the degradation of critical components like the PER protein (Fig. 1.2). (Takahashi et al., 2008).

Peripheral tissues also exhibit autonomous circadian rhythms but are subservient to and are entrained by the SCN. The SCN coordinates the peripheral clocks through humoral and neural signals that are not well understood, and by indirect means such as body temperature, wakefulness and food intake. Thus the entire circadian system is a hierarchy of subnetworks that extend from the molecular and biochemical level to the physiological level.

Components of the circadian clock are deeply involved in human physiology and disease. The most obvious association is with sleep disorders. Familial advanced sleep-phase syndrome (FASPS) is an autosomal dominant circadian rhythm disorder characterized by an abnormal phasing of the circadian cycle relative to the desired sleep–wake schedule. Here sleep onset and awakening times are 3–4 hours ahead of the desired times. Through linkage analysis, individuals with the syndrome were found to harbor a missense mutation, S662G, in the human *PER2* gene. This S662G mutation disrupts a phosphorylation site within a casein kinase 1 (CSNK1)-binding domain of *PER2*, resulting in the increased turnover of nuclear PER2. As evidence that FASPS has heterogeneous genetic origins, a mutation in a casein kinase isoform, CSNK1δ, was also found in FASPS.

Such sleep disorders are rare; however, there is the cumulative evidence that molecular

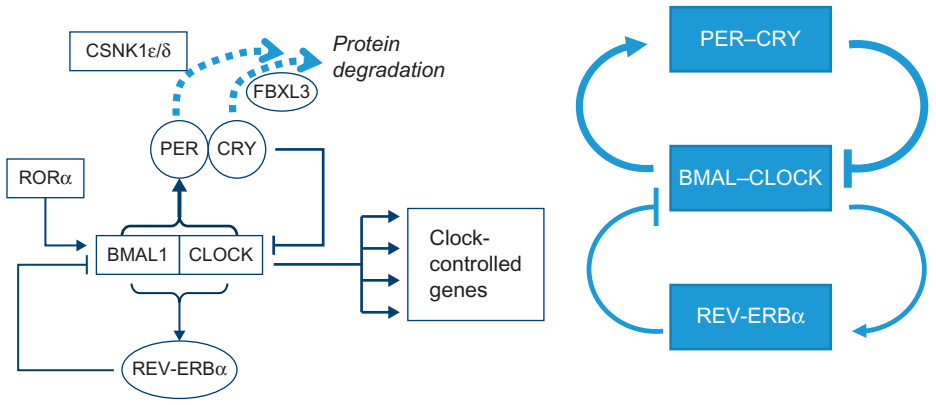

FIGURE 1.2 Schematic Representation of the Control Network for Circadian Cycling.
The core organization of the regulatory module is comprised of two interlocking negative-feedback network motifs with asymmetric timing. BMAL1, brain and muscle Arnt (aryl hydrocarbon receptor nuclear translocator)-like protein-1; CLOCK, Circadian Locomotor Output Cycles Kaput gene; CSNK1ε/δ, casein kinase 1 ε and δ; CRY, Cryptochrome; FBXL3, an E3 ubiquitin ligase; PER, Period; REV-ERBα: a retinoid-acid-related orphan nuclear hormone receptor; RORα, a retinoid-acid-related orphan nuclear hormone receptor.

components of the circadian oscillator may be involved in many common disorders. Gene profiling experiments demonstrated that up to 10% of the testable transcriptome shows circadian periodicity, and that the attributes of these clock-regulated genes are highly enriched for metabolic functions. Recall that the nuclear hormone receptors, RORα and REV-ERBα, are integral parts of the oscillator loop. Extending this analysis further, Yang and colleagues (2006) examined the detailed gene expression the 49 nuclear receptors in mice, and found that 28 display tissue-specific circadian rhythms. Given the function of nuclear receptors in metabolic regulation, their circadian control provides one explanation for the diurnal behavior of glucose and lipid metabolism.

Studies in animal models also continue to uncover associations between clock genes and metabolic phenotypes: homozygous *Clock*-mutant mice are hyperphagic, obese and exhibit a metabolic syndrome with hyperlipidemia, fatty liver, high circulating glucose concentrations and low circulating insulin concentrations (Turek et al., 2005). *Bmal1*$^{-/-}$ knockout mice not only have abnormal sleep patterns, but also show low body weight and sensitivity to insulin shock. Fibroblasts from these *Bmal1* knockout mice also cannot undergo adipocyte differentiation (Shimba et al., 2005). Clinically, a link between circadian cycles and metabolism has been observed. Epidemiologic studies in shift workers have shown an increase in body mass index, and in the rates of incidence of metabolic disorders and cardiovascular events (Ellingsen et al., 2007). It is also well understood that the specific sensitivity to exogenous insulin exhibited by diabetic patients changes over the time of day. Thus the circadian clock mechanisms are inextricably linked to metabolic functions, and may represent an adaptive evolutionary response to maximizing energy utilization that is dependent on a consistent environmental change—the planetary reality of the day / night cycle (Green et al., 2008).

An intriguing side observation that now has significant ramifications for cancer therapeutics is that liver detoxifying genes also show significant circadian oscillations and have been shown to be regulated by clock mechanisms. Doses of the chemotherapeutic agent, cyclophosphamide, given at different times of the circadian cycle can

result in differences in mortality rates—from 20% to 100% (Gorbacheva et al., 2005). Exploring this phenomenon further, the investigators found that *Clock* and *Bmal1* knockout mice are sensitive to the toxic effects of cyclophosphamide, but *Cry1* and *Cry2* double-knockout mice are resistant. This resistance was not caused by pharmacokinetic differences, but appeared to be correlated with cellular insensitivity of B lymphocytes to the lymphotoxic effects of this drug. These experiments validate the clinical observations that timing of chemotherapeutic administration has an effect on drug toxicity and drug effectiveness (reviewed by Takahashi et al., 2008).

The growing body of knowledge of the mechanisms around circadian clocks and their impact on health has provided opportunities for the development of drugs targeting these molecules. Many of the clock-associated genes are amenable to the action of drugs or represent biochemical classes amenable to small-molecule modulation: the melatonin receptors are G-protein-coupled receptors; GSK3β is a kinase that modifies PER, and REV-ERBβ casein kinase 1 is another class of kinases; REV-ERB and ROR are nuclear hormone receptors (the ligand for REV-ERBα has been identified as heme). All these targets have candidate small-molecule modifiers. This has led companies to explore the use of cell-based screens to identify molecules that would disrupt or alter the circadian clock. Cell systems with luciferase reporter genes controlled by clock-dependent regulatory elements can be used to screen libraries of small molecules. The readout would be disruption of the periodicity (reviewed by Liu et al., 2007). Thus, starting from a simple oscillator, explanations of human physiology and identification of targets for therapy can be explored.

CONCLUSION

How is systems biomedicine different from other forms of systems studies? In my opinion, the differences are only ones of scale and experimental access. Clearly, the human genome and proteome is more complex than those of yeast and bacteria, and human genetic studies are more complex than those in mice. Moreover, the complexity of a multicellular and multi-organ system has yet to be configured into the equation. To date, the comparative extent of that complexity remains not quite known; therefore, how much more data and how much more computing will be necessary to achieve the same coverage as that described for *Halobacterium salinarum* is unclear, but will undoubtedly be more than the ratio of the size of our genome to that of this microbe. However, the approaches and the opportunities are the same.

Of course, in the final analysis, systems biomedicine, by directly benefiting human health, will be a significant endeavor. So any increment in improvement in prediction will help medicine and benefit society. The challenges, however, are logistical, computational and organizational. Logistical because first, for obvious ethical reasons, experimentation in human systems is slower and more ponderous; secondly, human variation will make initial estimates less generalizable; and thirdly, the further division into organ systems linked by circulation and endocrine factors will increase the number of studies needed in order to complete the human organism. The computational challenges have been alluded to, and are most critical: massive amounts of data requiring integration and iterative analysis of high computational complexity. The new technologies in sequencing, genotyping, proteomics and imaging are generating a hyper-exponential growth in data acquisition that is quickly outstripping the capabilities of most biological laboratories and departments. The physical sciences have pioneered the use of supercomputers with the capability of handling this challenge. However, the porting of the all biological, genetic and genomic algorithms to these new platforms and their continued development will be a prodigious task. Lastly, the simple fact is that our data standards do not routinely allow for

cross-platform comparisons. Manual curation is still required for most high-level systems integration. There is a need for integration of heterogeneous data (e.g. protein–protein interaction, RNA expression information, biochemical pathways, genomic data and literature-based connections) and for visualization tools that will enable the presentation of large-scale data that are interpretable to bench biologists.

Finally, the organizational challenges, although man- made and therefore surmountable by man, are also daunting (Liu et al., 2005). These organizational challenges are rooted in the sometimes contradictory requirements of systems biology research and the operational intentions of our academic and funding institutions.

In systems research, scientists with very different skills (biology, mathematics, engineering, medicine) must be working closely together and have proximity with one another in what might almost be scientific collectives (Liu, 2009). Traditionally, bioinformatics resided in a computer science or biostatistics department, biology in a biochemistry department and a genomics center that was functionally dissociated from the previous two. However, the scale of this interaction requires coordinated resources from the funding agencies, much akin to the supercomputing program of the US National Science Foundation. This unfortunate disconnect would benefit from some conceptual realignment. Regarding data presentation, there is a need to provide more natural interfaces between humans and computers to service the non-expert user. There will be a demand for simplified interfaces specifically designed for biologists. This does not detract from the important need to train the next generation of biologists who are mathematically and computationally literate, and the next generation of mathematicians, computer scientists and engineers who are steeped in the nuances of biology.

Grants management is often at odds with collective efforts. Funding for critical technology platforms is too often bypassed as lacking scientific content. By discounting participation in collaborative projects and focusing exclusively on individual effort, University promotion processes historically encourage faculty insularity. Graduate student training, restrained by classical departmental boundaries and focused on individual faculty projects, is not responsive to the educational requirements for success in integrative and systems biology. Systems biology is deeply cross-disciplinary.

Daunting as these challenges are, the stakes are high. I believe that systems approaches in biology will become as common as molecular technologies are in current biological investigations. Molecular biology, which was a new creature in the 1970s and early 1980s and which spawned biotechnology companies and institutes and departments with "molecular biology" in their title, is now commonplace and integrated into the fabric of biological teachings. Current medical investigations are all molecular medicine. The same will be true of systems approaches.

Systems Biomedicine, indeed, is here to stay.

References

Amit, I., Citri, A., Shay, T., et al., 2007a. A module of negative feedback regulators defines growth factor signaling. Nat. Genet. 39, 503–512.

Amit, I., Wides, R., Yarden, Y., 2007b. Evolvable signaling networks of receptor tyrosine kinases: relevance of robustness to malignancy and to cancer therapy. Mol. Syst. Biol. 3, 151.

Bonneau, R., Facciotti, M.T., Reiss, D.J., et al., 2007. A predictive model for transcriptional control of physiology in a free living cell. Cell 131, 1354–1365.

Carlson, J.M., Doyle, J., 2000. Highly optimized tolerance: robustness and design in complex systems. Phys. Rev. Lett. 84, 2529–2532.

Cheung, V.G., Spielman, R.S., Ewens, K.G., Weber, T.M., Morley, M., Burdick, J.T., 2005. Mapping determinants of human gene expression by regional and genome-wide association. Nature 437, 1365–1369.

Cheung, V.G., Bruzel, A., Burdick, J.T., Morley, M., Devlin, J.L., Spielman, R.S., 2008. Monozygotic twins reveal

germline contribution to allelic expression differences. Am. J. Hum. Genet. 82, 1357–1360.
Cohen, J.E., 2004. Mathematics is biology's next microscope, only better; biology is mathematics' next physics, only better. PLoS Biol. 2, e439.
Drake, T.A., Schadt, E.E., Davis, R.C., Lusis, A.J., 2005. Integrating genetic and gene expression data to study the metabolic syndrome and diabetes in mice. Am. J. Ther. 12, 503–511.
Ellingsen, T., Bener, A., Gehani, A.A., 2007. Study of shift work and risk of coronary events. J. R. Soc. Health 127, 265–267.
Forrest, S., Beauchemin, C., 2007. Computer immunology. Immunolog. Rev. 216, 176–197.
Green, C.B., Takahashi, J.S., Bass, J., 2008. The meter of metabolism. Cell 134, 728–742.
Gorbacheva, V.Y., Kondratov, R.V., Zhang, R., et al., 2005. Circadian sensitivity to the chemotherapeutic agent cyclophosphamide depends on the functional status of the CLOCK/BMAL1 transactivation complex. Proc. Natl. Acad. Sci. USA 102, 3407–3412.
Irish, J.M., Hovland, R., Krutzik, P.O., et al., 2004. Single cell profiling of potentiated phospho-protein networks in cancer cells. Cell 118, 217–228.
Katz, M., Amit, I., Citri, A., et al., 2007. A reciprocal tensin-3-cten switch mediates EGF-driven mammary cell migration. Nat. Cell Biol. 9, 961–969.
Kitano, H., 2006. The B-cell interactome. Available from http://amdec-bioinfo.cu-genome.org/html/BCellInteractome.html#Publication
Kitano, H., 2007. Towards a theory of biological robustness. Mol. Syst. Biol. 3, 137.
Kitano, H., Oda, K., 2006. Robustness trade-offs and host-microbial symbiosis in the immune system. Mol. Syst. Biol. 2, 2006–2022.
Lander, E.S., Linton, L.M., Birren, B., et al., 2001. For the international human genome sequencing consortium. Initial sequencing and analysis of the human genome. Nature 409, 860–921.
Liu, E.T., 2005. Systems biology, integrative biology, predictive biology. Cell 121, 505–506.
Liu, E.T., 2009. Integrative biology—a strategy for systems biomedicine. Nat. Rev. Genet. 10, 64–68.
Liu, A.C., Lewis, W.G., Kay, S.A., 2007. Mammalian circadian signaling networks and therapeutic targets. Nat. Chem. Biol. 3, 630–639.
Luscombe, N.M., Babu, M.M., Yu, H., Snyder, M., Teichmann, S.A., Gerstein, M., 2004. Genomic analysis of regulatory network dynamics reveals large topological changes. Nature 431, 308–312.
Olson, E.N., 2006. Gene regulatory networks in the evolution and development of the heart. Science 313, 1922–1927.
Phillips, P.C., 2008. Epistasis—the essential role of gene interactions in the structure and evolution of genetic systems. Nat. Rev. Genet. 9, 855–867.
Sachs, K., Perez, O., Pe'er, D., Lauffenburger, D.A., Nolan, G.P., 2005. Causal protein-signaling networks derived from multiparameter single-cell data. Science 308, 523–529.
Schadt, E.E., Molony, C., Chudin, E., et al., 2008. Mapping the genetic architecture of gene expression in human liver. PLoS Biol. 6, e107.
Segal, E., Friedman, N., Kaminski, N., Regev, A., Koller, D., 2005. From signatures to models: understanding cancer using microarrays. Nat. Genet. 37 (Suppl), S38–S45.
Shalon, D., Smith, S.J., Brown, P.O., 1996. A DNA microarray system for analyzing complex DNA samples using two-color fluorescent probe hybridization. Genome Res. 6, 639–645.
Shimba, S., Ishii, N., Ohta, Y., et al., 2005. Brain and muscle Arnt-like protein-1 (BMAL1), a component of the molecular clock, regulates adipogenesis. Proc. Natl. Acad. Sci. USA 102, 12071–12076.
Sieberts, S.K., Schadt, E.E., 2007. Moving toward a system genetics view of disease. Mamm. Genome 18, 389–401.
Spielman, R.S., Bastone, L.A., Burdick, J.T., Morley, M., Ewens, W.J., Cheung, V.G., 2007. Common genetic variants account for differences in gene expression among ethnic groups. Nat. Genet. 39, 226–231.
Takahashi, J.S., Hong, H.K., Ko, C.H., McDearmon, E.L., 2008. The genetics of mammalian circadian order and disorder: implications for physiology and disease. Nat. Rev. Genet. 9, 764–775.
Tong, A.H., Lesage, G., Bader, G.D., et al., 2004. Global mapping of the yeast genetic interaction network. Science 303, 808–813.
Turek, F.W., Joshu, C., Kohsaka, A., et al., 2005. Obesity and metabolic syndrome in circadian Clock mutant mice. Science 308, 1043–1045.
Yang, X., Downes, M., Yu, R.T., et al., 2006. Nuclear receptor expression links the circadian clock to metabolism. Cell 126, 801–810.
Yu, H., Braun, P., Yildirim, M.A., Lemmens, I., et al., 2008. High-quality binary protein interaction map of the yeast interactome network. Science 322, 104–110.
Zhu, J., Wiener, M.C., Zhang, C., et al., 2007. Increasing the power to detect causal associations by combining genotypic and expression data in segregating populations. PLoS Comput. Biol. 3, e69.