Chapter 23

Systems Medicine and the Emergence of Proactive P4 Medicine: Predictive, Preventive, Personalized and Participatory

Leroy Hood[1], Mauricio A. Flores[2], Kristin R. Brogaard[1], and Nathan D. Price[1]

[1]*Institute for Systems Biology, 401 N. Terry Ave, Seattle, WA 98121, USA*

[2]*P4 Medicine Institute, 401 N. Terry Ave, Seattle, WA 98121, USA*

Chapter Outline

INTRODUCTION

Medicine is undergoing a revolution that will transform the practice of healthcare in virtually every way. This revolution is emerging from the convergence of systems biology – a holistic approach to biology (and medicine) – and the digital revolution. Systems biology is opening what has historically been the black box of our individual biological systems as they change over time. The digital revolution is vastly expanding our capacity to generate and analyze 'big data' sets, deploy this information in business and social networks and create digital consumer devices to measuring personal information. Both of these new capabilities will be deployed in systems medicine (see below).

Systems biology emerged at least partly in response to the incredible complexity of biological systems, both normal and diseased [1,2]. This complexity arises from the natural process of Darwinian evolution – random mutations followed by natural selection generated in large part by the environment. Darwinian evolution is a random and chaotic process – building new complexities on top of previously evolved complexities. Indeed, biological systems resemble Rube–Goldberg devices. Consider a Rube–Goldberg device that attaches 14 different gadgets together to cool the temperature of soup. Deciphering the complexity of this 'soup-cooling system' requires (1) defining the components of the soup-cooling system, (2) determining how these components interact with one another, and (3) delineating the dynamics of these components in space and time that are necessary for carrying out their function of soup cooling. These are precisely the elements that systems biology attempts to define when deciphering the complexity of biological systems. To achieve actionable understandings of biological complexity, the analyses must be global (comprehensive), integrative and dynamic.

Systems medicine, the healthcare-focused derivative of systems biology, is beginning to alter the face of healthcare through (1) a systems approach to disease, (2) driving the emergence of technologies that permit the exploration of new dimensions of patient data space (e.g., sequencing the individual human genome), (3) the analyses of the quantized units of biological information (single genes, single molecules, single cells, single organs to provide disease-relevant information on health or disease for the individual), and (4) the resulting explosion of patient data that is transforming traditional biology and medicine into an information science [1–11].

Handbook of Systems Biology Concepts and Insights. http://dx.doi.org/10.1016/B978-0-12-385944-0.00023-X

The digital revolution has already transformed communications, finance, retail and information technology by harnessing big datasets through computational analyses and by creating powerful new business and social networks. The digital revolution is contributing to individualize healthcare in several important ways: (1) by providing tools and strategies for managing and analyzing large biological and environmental datasets; (2) by catalyzing the invention of personal monitoring devices that can digitalize biological and social information, thus enabling an assessment of wellness and disease for the individual (e.g., the 'quantified self'); and (3) by providing models for the creation of consumer (patient)-driven and consumer (patient)-participating social networks that focus on optimizing wellness and/or dealing with disease (Figure 23.1) [8–10,12].

The convergence of the digital revolution and systems approaches to wellness and disease is beginning to lead a proactive P4 medicine that is predictive, preventive, personalized and participatory [1,13–15]. Thus 'P4 medicine' is the clinical application of the tools and strategies of systems medicine to quantify wellness and demystify disease (Figure 23.2) for the wellbeing of the individual. The digital revolution has given scientists the ability to generate and analyze previously inconceivably large quantities of digital data. Using these new capabilities and employing the domain expertise of biology to direct the development of software, systems biologists have developed powerful new suites of tools for mining, integrating and modeling 'big data' sets of heterogeneous biological data to generate predictive and actionable models of health and disease for each patient [10,16–19]. 'Actionable' means that the data provide information that is useful for improving the health of the individual patient. Thus, systems biologists have transitioned from the reductive studies of traditional biology that focus on a few genes or proteins to the new holistic and comprehensive analyses of systems biology, analyzing how all of the components of biological systems interact.

Unlike the reactive, limited-data population-based hierarchical approach of contemporary evidence-based healthcare [20], P4 medicine is not confined to clinics and hospitals. It will be practiced in the home, as informed and networked consumers use new information, tools and resources such as wellness and navigation coaches and digital health information devices to better manage their health. Below we provide a brief picture of systems medicine and its role in the emergence of this proactive P4 medicine. We then will explore the multi-dimensionality of P4 medicine, with its many implications for the individual consumer (patient) and for society.

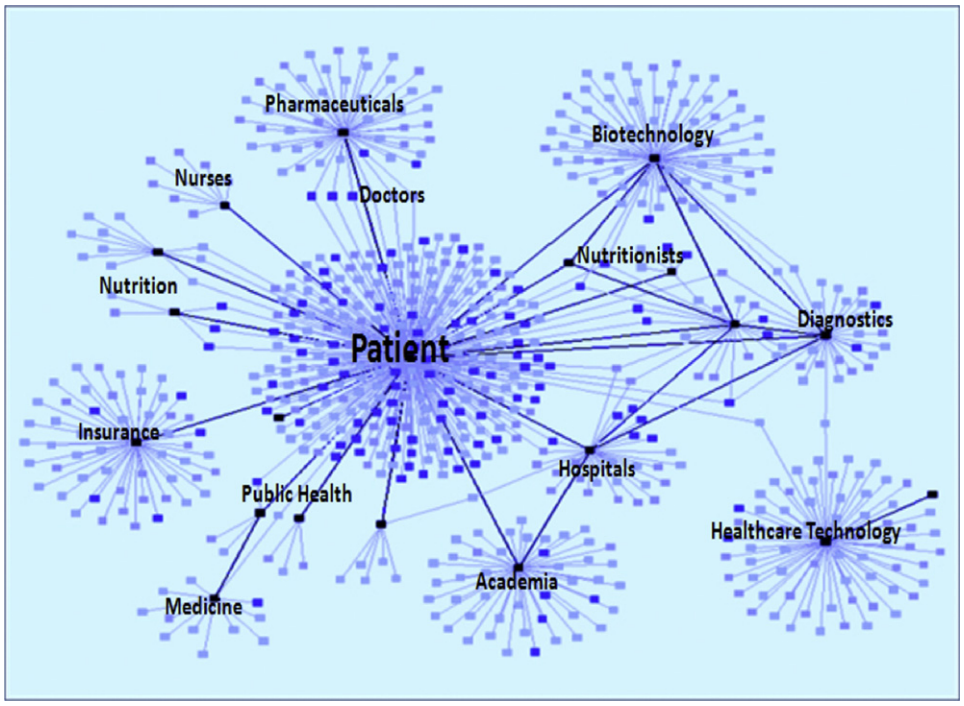


FIGURE 23.1 A network depicting the interacting components of the healthcare system indicating the dominant role patients will have in advancing P4 medicine through their consumer-driven social networks. Networks allow one to organize and model data and are important in dealing with the signal to noise problem of large data sets.

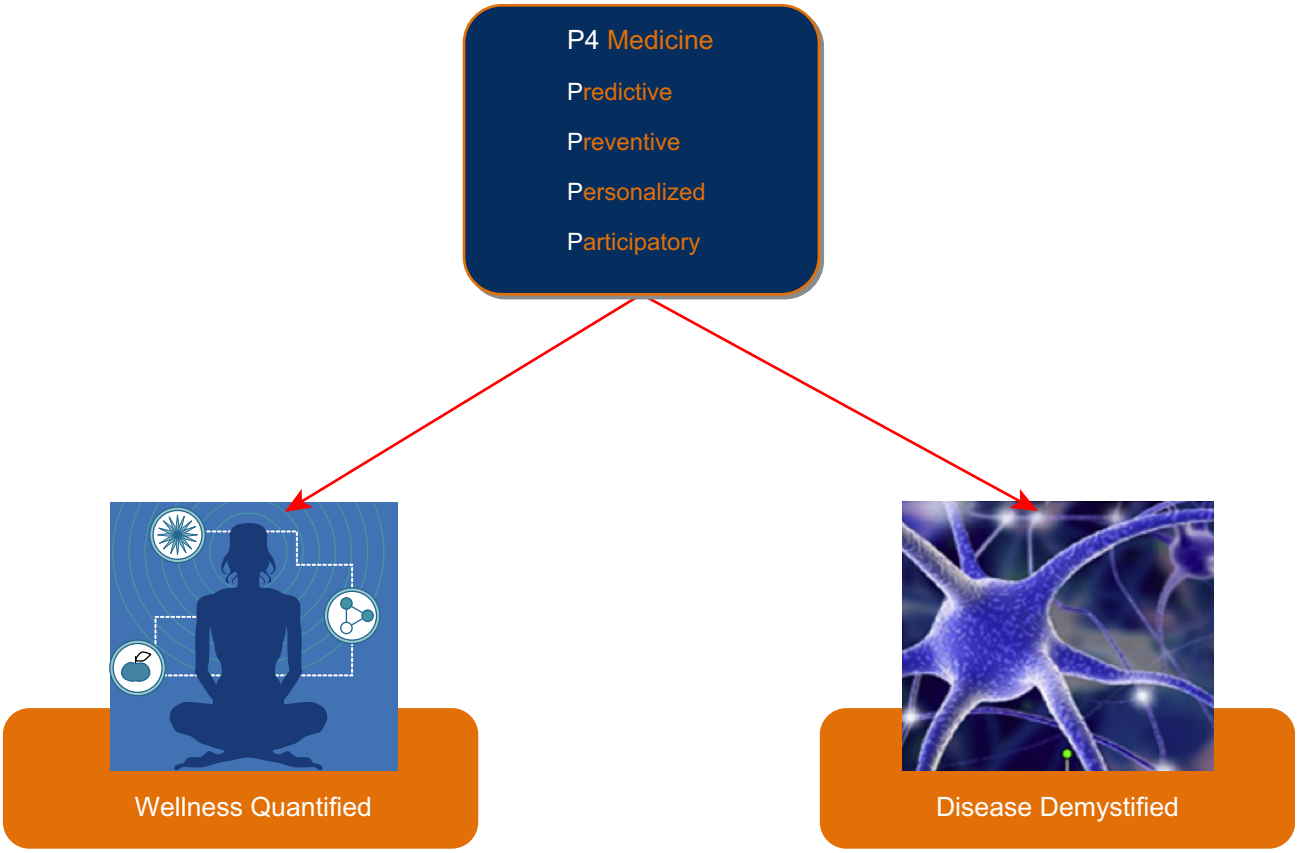


FIGURE 23.2 **A schematic representation of the two major objectives of P4 medicine: quantizing wellness and demystifying disease.**

SYSTEMS MEDICINE

We predict that within 10 years every healthcare consumer will be surrounded by a virtual cloud of billions of data points (Figure 23.3). These will range from molecular and cellular data, to conventional medical data, to enormous amounts of imaging, demographic and environmental data. Big data sets are essential to decipher 'signal' from the noise generated by the complexities of disease and wellness. As noted above, the complexity of biological systems arises from the random and chaotic processes of Darwinian evolution. The principal data analyzed by systems medicine relate to the following central concerns about human biological systems: the identification all of the system components, establishing their interactions, assessing the dynamics of those interactions – both temporal and spatial – and then attempting to understand how the system as a whole executes its biological functions and exhibits specific phenotypes.

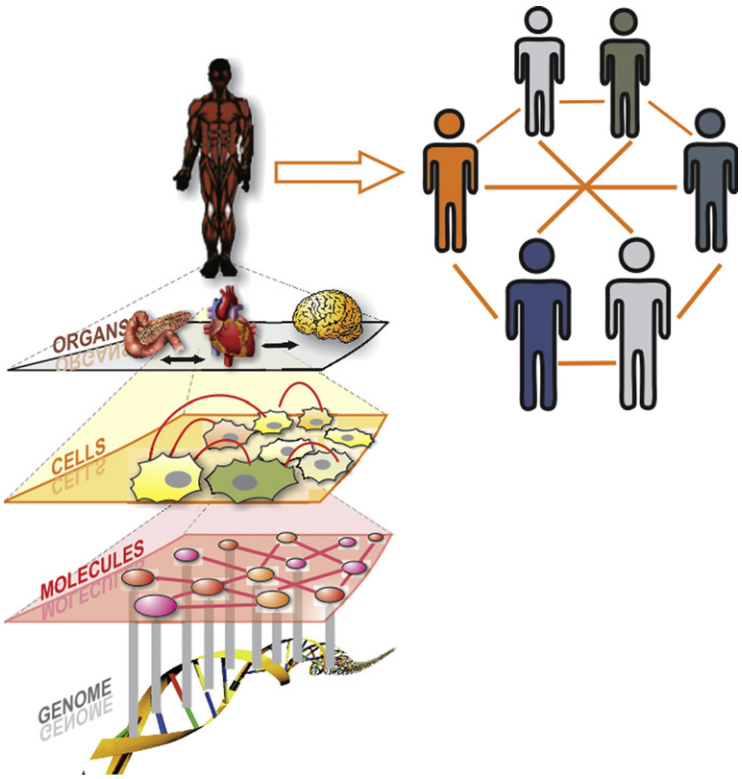
A systems approach to disease takes into account the myriad social and environmental factors that compound the innate complexity of human biology and which are crucial determinants of health. It does so by treating disease as a consequence of genetic and/or environmental perturbations of biological networks. These disease-perturbed networks all express altered information that changes dynamically across time (and often space) and arises from the perturbations. The amalgamation of this distorted information explains the pathophysiology of the disease, as discussed below in the context of a specific systems-disease example, neurodegeneration.

The above description is not intended to suggest that disease-perturbed networks are of a single type: there is a 'network of networks' that reflects the multi-dimensionality of both biological processes and disease processes (Figure 23.4). Thus we can describe genetic networks arising from the genome; molecular networks arising from protein and DNA interactions; cellular networks arising from cell interactions; organ networks arising from organ–organ interactions; and finally, social networks arising largely in part from the nature of an individual's environmental interactions [2,17,21]. The integration of all of these networks is necessary to understand their functioning in the context of the individual.

The ascertainment of these networks requires enormous amounts of data. For some of these measurements, the tools are just now being developed. Big data sets pose two

FIGURE 23.3 **In 10 years a virtual cloud of billions of data points will surround each patient.** These data will be of many different types and, accordingly, multistage. The challenge will be to convert these data into simple hypotheses about health and disease for the individual.

FIGURE 23.4 **A figure depicting the 'network of networks' that specifies the nature of some of the integrated networks that specify normal biology and disease.** The genetic, molecular, cellular, organ and individual networks are represented, and represent a fully integrated network of networks. Networks are powerful tools for integrating and modeling biological data. Networks also provide a powerful means for dealing with signal to noise problems. (*Figure adapted from [60].*)



significant problems. First, how to deal with the enormous signal-to-noise challenges intrinsic to all large data sets, and second, how to convert data into knowledge. Solving these problems is the role of systems medicine. The key for systems medicine in the future will be to ascertain and deconvolute the 'network of networks' for each individual and to be able to follow its dynamics in response to various types of biological information, providing fundamental insights into wellness and disease.

Looking back in history there were four paradigm changes that led directly to systems medicine: (1) automated and high-throughput biological technology, (2) the Human Genome Project, (3) the creation of cross-disciplinary biology institutes, and (4) the creation of systems biology as an area of research [22,23] We will spend some time focusing on one of these, the Human Genome Project, to describe the revolutionary effect that it has had on biology.

The first meeting on the Human Genome Project was held in the spring of 1985. Twelve 'experts' had been invited to Santa Cruz to consider whether sequencing the human genome was advisable. This committee came to two conclusions: (1) the project was feasible albeit technically difficult; and (2) the group was split 6 to 6 on whether it was a good idea. In the mid and late 1980s perhaps 90% of biologists were opposed to the project, as was the National Institutes of Health (for reasons such as: big science is bad; the genome is mostly junk, so why sequence it? no good scientists would participate in such a mindless endeavor, etc.). A committee of the National Academy of Sciences with both opponents and proponents was convened to consider this possibility, and their unanimously favorable report turned the tide. The project was initiated in 1990 and finished in 2004 – under budget and ahead of schedule.

Many have argued that the genome project did not fulfill its promise, but the truth is quite the opposite. The Human Genome Project has transformed both biology and medicine in several important ways:

1. It made systems biology possible by providing a complete parts list of all (most) of the genes (and by inference their corresponding proteins) in the human and several model organisms. This parts list was essential for global and integrative systems approaches, not to mention all subsequent human biology research.
2. It 'democratized' all genes (and indeed any region of the genome) by making all genes available to all biologists.
3. It pushed high-throughput biology to the next stage by driving technologies to increase the speed and accuracy of DNA sequencing to pioneer other genomic technologies, such as DNA arrays and parallel sequencing.
4. It made mass spectrometry-based proteomics possible by providing the sequences of proteins and their corresponding tryptic peptides.
5. It brought computer scientists, theoretical physicists, software engineers and mathematicians into biology to deal with the exponentially increasing data sets, thereby providing new software and mathematical tools for converting data into knowledge.
6. By requiring that genomic data be made public as soon as it was determined, it pioneered the idea of open data, rapidly published and available to all.
7. It also pioneered the importance of assessing data quality and the software necessary for this assessment.
8. It made the genomes of microbes, plants, and animals accessible to all biologists, transforming many fields of biology (e.g., microbiology, virology, immunology, etc.).
9. It revolutionized our understanding of molecular evolution.
10. For medicine, it made possible new approaches to genomic diagnoses; it enabled personalized medicine; and it is forging new approaches to assessing proper drugs (therapies) for differing kinds of cancer.
11. It brought biology into the realm of big science and initiated a big science/small science debate that continues even today. Big and small science can be beautifully integrated, and each plays an important role in deciphering biological complexity [1].

The Battelle Memorial Institute has recently estimated that the Human Genome Project has led to almost $800 billion in benefits for an initial investment of about $3.5 billion. It is clear from skimming though these benefits that the project enabled and enriched systems approaches to biology and medicine in many different ways that go far beyond genomics itself.

Many people use the term 'genomic medicine' to denote the medicine of the future, yet in principle genomic medicine is one-dimensional in nature, only encompassing nucleic acid information. Systems medicine, in contrast, is holistic and utilizes all types of biological information, including DNA, RNA, protein, metabolites, small molecules, interactions, cells, organs, individuals, social networks and external environmental signals – integrating them so as to lead to predictive and actionable models for health and disease.

FIVE SYSTEMS' STRATEGIES FOR DEALING WITH BIOLOGICAL COMPLEXITY

To develop predictive and actionable models we need to tease apart the complexity of health and disease. Systems medicine employs five strategies to deal with biological

complexity: (1) viewing medicine as an informational science, (2) creating a cross-disciplinary infrastructure in which to implement systems medicine, (3) employing experimental systems approach to disease that are holistic and integrative, (4) driving the development of new technologies that permit the exploration of new dimensions of patient data space, and (5) developing new analytical tools, both computational and mathematical, for capturing, validating, storing, mining, integrating and finally modeling data so as to convert data into knowledge.

These five pillars of systems medicine permit biological complexity to be deciphered by providing a path forward for both generating large amounts of data, integrating and modeling these data in ways that reduce noise and delineate biological mechanisms. They create the conceptual framework for converting data into knowledge.

1. Systems medicine views medicine as an informational science**,** providing an intellectual framework for dealing with complexity. Fundamentally, there are two types of biological information: the digital information of the genome and the environmental signals that come from outside the genome. Together these two types of information are integrated in the individual organism (e.g., a human) to produce its phenotype, healthy or diseased. These two types of information and the phenotypes they produce are connected through biological networks that capture, transmit, integrate signals and then pass the information to molecular machines that execute the functions of life. It is the dynamics of networks and molecular machines that constitute a major focus of systems studies. The 'network of networks' adds yet another multiscale challenge to organizing and integrating information (Figure 23.4).

   As noted above, systems medicine postulates that disease arises from disease-perturbed networks (perturbed by genetic changes and/or environmental signals). Altered molecular machinery encoded by the disease-perturbed networks leads to the pathophysiology of the disease. Thus following the dynamics of the disease-perturbed networks gives deep insights into disease mechanisms and provides a powerful tool for dealing with the signal to noise challenges of big data sets. The utility of this approach has been demonstrated in two mouse models – mouse neurodegeneration (prion infection) [16] and glioblastoma – from mice genetically engineered in a combinatorial manner with oncogenes and tumor suppressors). The prion model of neurodegenerative is discussed later in this chapter.

   To obtain the necessary information for systems medicine it is also critical to integrate and model the many diverse data types, including from animal models, that follow disease progression. The reasons for this are obvious. One cannot follow the disease from initiation to the end in humans; one cannot usually easily sample the diseased tissue at multiple different time points; nor can one experimentally perturb the system with environmental signals.

   The need for systems dynamics data to deal with noise and create models emphasizes the importance of experimental animal disease models where the starting point of the disease process can be known (e.g., by genetic activation of the disease process or the experimental initiation of disease such as an infection) and the dynamics followed until death. The key point is that animal models must closely mimic their human counterpart diseases. Scientists must clearly identify those aspects of the disease-perturbed systems that are orthologous to human disease and those that are unique to the animal – and use the former for gaining dynamical insights into human disease. When the disease process is translated into network dynamics, determining orthology between the animal model and human disease becomes much simpler. Indeed, one can draw inferences from model organism disease-perturbed networks that are orthologous to their human counterparts and ignore the disease-perturbed networks that are not orthologous. This approach enables animal studies to be powerfully informative about human disease.
2. Our belief is that a special infrastructure is required for practicing systems medicine. This belief is driven by the conviction that leading-edge biology must drive the development of new high-throughput technologies to explore new dimensions of patient data space. The data arising from these technologies in turn require the pioneering of new analytical tools for the integration and modeling of diverse data types. We have termed this the 'holy trinity' of biology – biology drives technology drives analytical tools – and integrated them together to revolutionize our understanding of medicine (Figure 23.5).

   This approach requires a cross-disciplinary environment where biologists, chemists, computer scientists, engineers, mathematicians, physicists and physicians all learn to speak the languages of the other disciplines and work together in biology-driven teams to achieve this holy trinity. To be effective, this cross-disciplinary environment requires the 'democratization' of data generation and data analysis tools; that is, it is essential to make these tools accessible to all individual scientists so that they may carry out either big science or small science projects.

   Thus the infrastructure of systems medicine consists of both the instrumentation to generate data for the diverse 'omic' technologies (genomics, proteomics, metabolomics, interactomics, cellomics, etc.) and a culture that encourages scientists to learn to speak the languages of

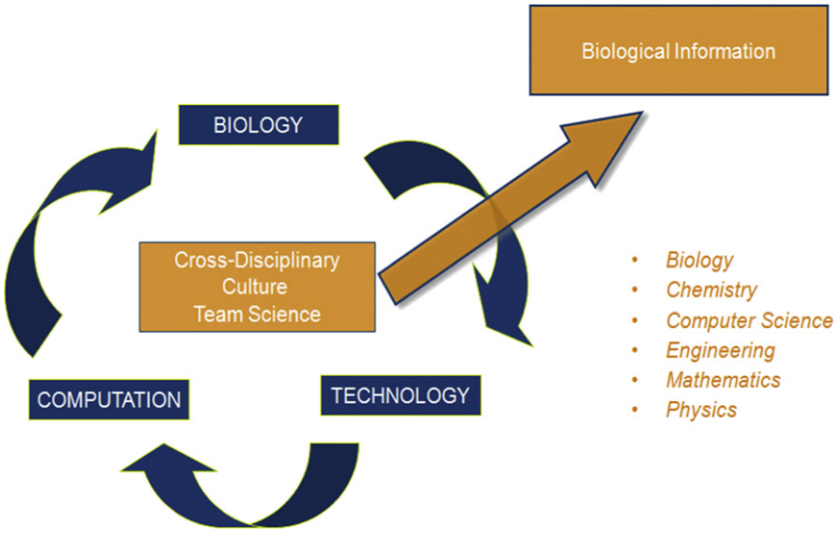


FIGURE 23.5 **The 'holy trinity of biology' where biology drives technology drives computational/mathematical tools.** Practicing this ideal requires a cross-disciplinary environment where scientists of many different disciplines (see lower right-hand side of figure) learn to speak the languages of the other scientists and to work together in teams. When the holy trinity is practiced effectively enormous amounts of biological information can be generated rapidly.

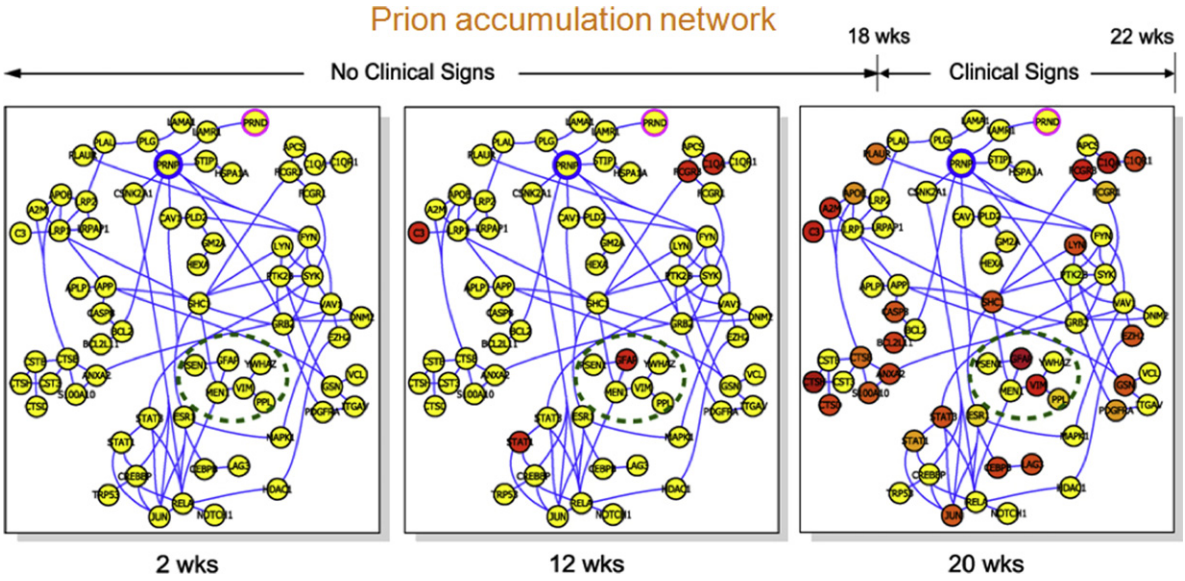


FIGURE 23.6 **A schematic of the prion accumulation and replication network in the prion-induced mouse neurodegenerative disease.** The red indicates transcript levels that have been increased in the brains from prion-infected animals compared with normal control brains. The yellow indicates transcripts that are the same in control and diseased animals. The three panels represent the network at 2, 12 and 20 weeks in animals that live about 22 weeks with this disease. The disease-perturbed networks appear about eight weeks before the clinical signs appear in these animals.

multiple scientific disciplines and how to work together in biology-driven teams practicing the holy trinity in the context of specific big or small science problems. This systems-driven infrastructure is what the Institute for Systems Biology (ISB) has spent the first 10 years of its existence creating [24].

3. Experimental systems approaches to disease and wellness are holistic. To decipher biological complexity, systems medicine depends on generating global and comprehensive data sets, following the dynamics of disease-perturbed networks across disease initiation and progression. Ultimately integrating diverse data types together to create predictive and actionable models (Figure 23.6) [2,24,25]. Thus systems medicine will give fundamental new insights into disease mechanisms — and open new opportunities for diagnosis, therapy and prevention.

The idea of systems biology can be illustrated by the example of how a radio converts radio waves into sound waves [26]. It is clear that merely understanding the function of each individual radio component would not provide us with an understanding of how radio waves are converted into sound waves. The same principle applies to

complex biological systems: understanding what individual genes or proteins do does not tell us how the biological systems in which they operate function. An engineer would connect the radio parts into their circuits and come to understand how the circuits worked individually and then collectively to convert radio into sound waves. So it is with living organisms: they employ biological circuits or networks to manage biological information and convert it into phenotype and function, and to understand these we must understand the dynamics of biological networks in handing information.

To generate holistic data systems biology (and therefore systems medicine) has three central elements. (1) It is hypothesis driven, where a model (which is a formally structured, precise and potentially complex) is formulated from existing data. Hypotheses from model predictions are then tested with systems perturbations and the high-throughput acquisition of data. The data (and metadata) are then reintegrated back into the model with appropriate modifications, and this process is repeated iteratively until new predictions from theory and experimental data are in agreement. (2) It is based on high-throughput data that should be (i) global (comprehensive), (ii) generated from different multi-scale data types (e.g., DNA, RNA, protein, metabolites, interactions, etc.), (iii) used to monitor networks dynamically, (iv) employed to provide deep insight into biology, and (v) integrated using proper statistics and bioinformatics to handle the enormous signal-to-noise problems. (3) Models may be descriptive, graphical or mathematical as dictated by the amount of available data, but they must be predictive. For medical use, predictions made must be actionable and useful for treating patients.

Boosting signal-to-noise in complex biology is essential for deciphering complexity. To reduce noise and to enhance statistical power, biologists have leveraged two fundamental ideas: filters and integrators [27]. Filters are used to winnow down the number of candidates based on the biological assumptions about complexity (e.g., modularity, hierarchical organization, complexity arising from evolution and inheritance). Integrators leverage the availability of complementary data of genome, transcriptome, miRNAome, proteome, metabolome, and interactome. Successful application of these strategies in disease will lead to a transformational understanding of disease and therapeutics.

The framework for approaching these studies in a holistic way is a systems approach to disease. As discussed above, the key idea is that disease arises as a consequence of the perturbation of one or more biological networks in the relevant organ. This perturbation alters the information the network encodes in a dynamic manner that changes during the progression of the disease (e.g., changing levels of mRNAs, miRNAs, or even proteins) — and these altered levels explain the pathophysiology of the disease and provide new insights into diagnosis and therapy.

A systems approach to a neurodegenerative disease in mice. We will illustrate this holistic systems approach as it applies to neurodegenerative disease (prion disease) in mice. This disease is initiated by the injection of 'infectious prion proteins' into the brains of mice. An important point is that we know precisely when the disease is initiated (at injection), allowing us to follow the dynamics of the disease process from initiation to termination. We analyzed the brain transcriptomes of the infected mice at 10 time points across the approximately 22 weeks of disease progression, in addition to the transcriptome of their healthy littermates. This procedure identified 7400 differentially expressed genes (DEGs), which represented a staggering signal-to-noise problem.

There are two types of noise: *technical noise*, which comes from the instrumentation/procedures for handling the data, and *biological noise*, which arises from biological processes other than neurodegeneration contributing to the phenotypic measurements. When you measure any aspect of phenotype in an organism, often those phenotypes are the sum of a number of different biologies. Hence one must use a deep understanding of biology to subtract from the biology of interest (neurodegeneration) the signals resulting from the other biologies.

To overcome the noise we carried out this study in eight different inbred-strain/prion-strain combinations of infected mice. With more data and a deep biological understanding of the disease process we were able to subtract away noise. For example, in the double-knockout mouse for the prion gene, after injection with infectious prions the animals never develop the disease. Thus, any changes in the brain transcriptomes of these animals were irrelevant to the prion neurodegeneration response and could be subtracted away. With seven additional subtractions, we identified a core of about 333 differentially expressed genes that encoded the basic prion neurodegeneration process. We mapped these DEGs on to four major biological networks of the prion disease process that had been defined by serial histopathology of the diseased brains. We then integrated the transcriptome data with (1) serial brain histopathological analyses of these animals, (2) serial sagittal brain sections stained for infectious prions, (3) clinical signs of the disease and (4) blood biomarker analyses. Figure 23.6 illustrates one of the major dynamically changing networks (prion replication and accumulation).

We drew the following conclusions from this study: (1) The disease starts with one or a few networks being

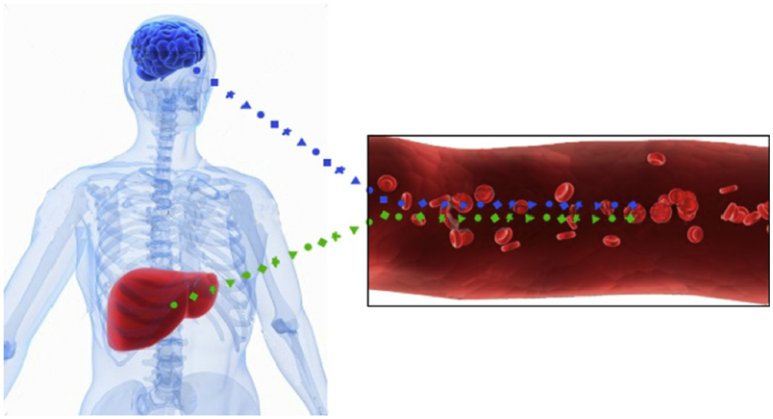


FIGURE 23.7 A diagram of organ-specific blood fingerprints (collections of organ-specific proteins) from the brain and the liver. For example, in a normal brain, each of the proteins in the brain-specific blood fingerprint will have one set of levels. In a diseased brain, the proteins whose cognate networks have become disease perturbed will change their levels. Since each disease leads to distinct combinations of disease-perturbed networks an analysis of the brain-specific protein fingerprints can distinguish healthy from diseased brains, and if diseased can stratify (e.g., distinguish from one another) the distinct types of brain disease. Thus organ-specific brain fingerprints can provide early detection, a stratification of different types of disease and the ability to follow the progression of the disease (not shown).

disease perturbed, and as the disease progresses more and more networks are recruited to become disease perturbed. (2) Two-thirds of the DEGs mapped into the four major networks – and the dynamics of these four networks explained virtually every aspect of known prion disease. (3) These four major networks were disease perturbed in a serial manner: first, prion replication and accumulation; second, glial activation; third, the degeneration of neuronal axons and dendrites; and finally neuron apoptosis. The importance of this observation is that if one is interested in new approaches to diagnostics and therapeutics the initial disease-perturbed network is a logical place to start. (4) The remaining one-third of the DEGs identified six new networks that were heretofore unknown to prion disease – the so-called 'dark genes and networks of prion disease' as identified by the global analyses of normal and diseased brain transcriptomes. These insights emphasize the importance of global analyses of the transcriptomes. (5) These studies suggested new approaches to blood diagnostics that are discussed below. (6) For therapy it is obvious that the first and most proximal prion-specific network should be re-engineered with drugs to make it behave in a more normal manner and hopefully abrogate the downstream consequences of this pathological progression. It is clear that multiple drugs will be required to re-engineer biological networks. We are now learning how to re-engineer biological networks with drugs in microbes, so that we can come to understand the basic logic of the approach and then extend it to higher model organisms (mice) before applying it to humans.

A holistic approach to disease will use blood as a window for monitoring health (wellness) and disease. A systems approach to blood diagnostics emerged from two ideas arising from the prion studies. First, some transcripts are expressed in their disease-perturbed networks 8 weeks or more before the first clinical signs (e.g., 10 weeks and 18 weeks, respectively). We were able to demonstrate that several of these differentially expressed gene (DEG) transcripts encoded proteins expressed in the blood, and we could see the altered protein levels in the blood. This was an example of presymptomatic diagnosis, a long-sought keystone of early disease detection. However, these DEGs were expressed in several different organs, so we could not be certain of the location of the disease-perturbed process directly from observing protein concentration changes in the blood. Second, in order to obtain blood markers with organ-specific addresses, we identified transcripts that were organ specific by deep comparative transcriptome analyses across 40+ different organs in humans and mice (Figure 23.7). From these analyses, through an examination of the human and mouse blood protein databases, and experimental mass spectrometry analyses, we were able to identify about 100 brain-specific proteins in humans and mouse. Of these, about 95% were orthologous between the two species (the presumption is that they will reflect similar activities in the two species), and these proteins collectively constituted a brain-specific blood fingerprint. We were able to show that some of these brain-specific proteins could also be used for presymptomatic diagnosis of prion disease in mice. Additionally, brain-specific blood proteins encoded by each of the four distinct networks exhibit concentration changes in the blood in a serial manner consistent with the order of disease perturbation of their cognate transcriptional networks. These data demonstrate that we will be able to assess both early disease detection and disease progression from the blood.

In the organ-specific blood protein fingerprints each individual protein assesses the behavior of its cognate biological network, distinguishing normal functioning from disease-perturbed functioning by changes in their blood concentration levels. Because each disease perturbs different combinations of networks, the brain-specific blood fingerprints will be able to distinguish normal from disease and, if diseased, identify the

disease. This will enable the five holy grails of blood disease diagnosis: (1) presymptomatic diagnosis; (2) stratification of a disease into its different subtypes; (3) assessment of the progression of the disease; (4) following patient response to therapy; and (5) identifying recurrences. We are now applying this strategy to identify human organ-specific blood biomarkers for several cancer types. In addition to blood proteins as tumor biomarkers, circulating DNAs, mRNAs, and microRNAs, as well as circulating tumor cells have also been studied which can serve as surrogate disease biomarkers and for monitoring cancer recurrence [28–32]. The organ-specific blood fingerprints are powerful aids to disease diagnostics, assessing drug toxicities, validating the orthologies between human disease and animal models of that disease, assessing multiorgan responses to diseases (beginning to define the organ/organ communicating networks (Figure 23.4) and, as we shall discuss later, assessing longitudinally across time the wellness of individual patients).

4. The systems approach to disease mandates the need to develop new or emerging technologies that can explore new dimensions of patient data space as reflected in part by the dynamics of the network of networks. These technologies include new approaches to genomics, proteomics, metabolomics, interactomics, cellomics, organomics, in vitro and in vivo imaging, and other high-throughput phenotypic measurements [33–38]. Microfluidic and nanotechnology approaches are moving many of these assays towards further miniaturization, parallelization, automation and integration of complex chemical procedures [39–41]. The areas of in vitro imaging and high-throughput phenotypics assays are going to contribute enormously to expanding the data repertoires of a systems approach to disease. But these new technologies must be driven by the real needs of biology or medicine. The outcome is an exponentially increasing ability to generate enormous amounts of digitalized personal data – big data – that necessitates a mandate to translate these data into knowledge. Let us briefly consider several areas in which emerging technologies are or will transform a systems approach to medicine.

DNA sequencing. Automated DNA sequencing was the cornerstone technology for sequencing the genome [42]. Since the initial completion of the human genome sequence [43] there has emerged a series of 'next-generation sequencing (NGS)' technologies that have exponentially increased the throughput of DNA sequencing while bringing down the costs dramatically (through parallelization and miniaturization of the process). NGS generates short sequence reads (e.g., 50–100 base pair reads) which have enabled the rapid sequencing of human genomes, the direct sequence analyses of transcriptomes and miRNAomes, as well as the analyses of some epigenetic features such as methylation. DNA arrays remain useful for looking at genetic variation and at the quantification of RNA populations (transcriptomes and miRNAomes) – although it is clear that in time DNA sequencing will replace these DNA array analyses. There is now emerging a third generation of DNA sequencing instruments that employ nanopores for threading single DNA molecules through the pores to enable the electronic analysis of single-stranded DNA molecules [44]. These new techniques have the potential for extremely long sequence reads (50–100 kb or more) and the ability for such extensive parallelization of the sequencing runs (through threading many DNA molecules simultaneously through many nanopores) that one may imagine doing a sequence run for a complete human genome sequence in a fraction of an hour, rather than the day to week or more that is required by current NGS instruments.

NGS has enabled striking new strategies for generating data (e.g., RNAseq – the quantification of complete transcriptomes, exon sequencing – the sequence analyses of all (most) of the exons in a genome and family genome sequencing – determining the complete human genome sequences of all the members of a family).

Complete family genome sequencing integrates genetics with genomics and in doing so raises fascinating possibilities for delineating diverse chromosomal features. For example, in the sequence analysis of a family of four where the mother and father were healthy and the two children each had two recessive genetic diseases, we had hoped to identify a modest number of gene candidates to explain the two genetic diseases (Figure 23.8) [33]. To our surprise, family genome sequences enabled far more in the way of family analysis. First, we were able to correct about 70% of the DNA sequencing errors merely by a consideration of the principles of Mendelian genetics. Second, we were able to identify rare variants merely by asking whether two or more members of the family exhibited the variant (thus eliminating the possibility of DNA sequencing errors). Third, we were able to determine the sites of chromosomal recombination, and accordingly could determine complete chromosomal haplotypes for each member of the family. This turned out to be important, as it reduced the chromosomal space in which disease genes might reside (by asking which haplotype regions were shared by the diseased children compared to their healthy parents). In a family of four the two affected genes must reside within a defined quarter of

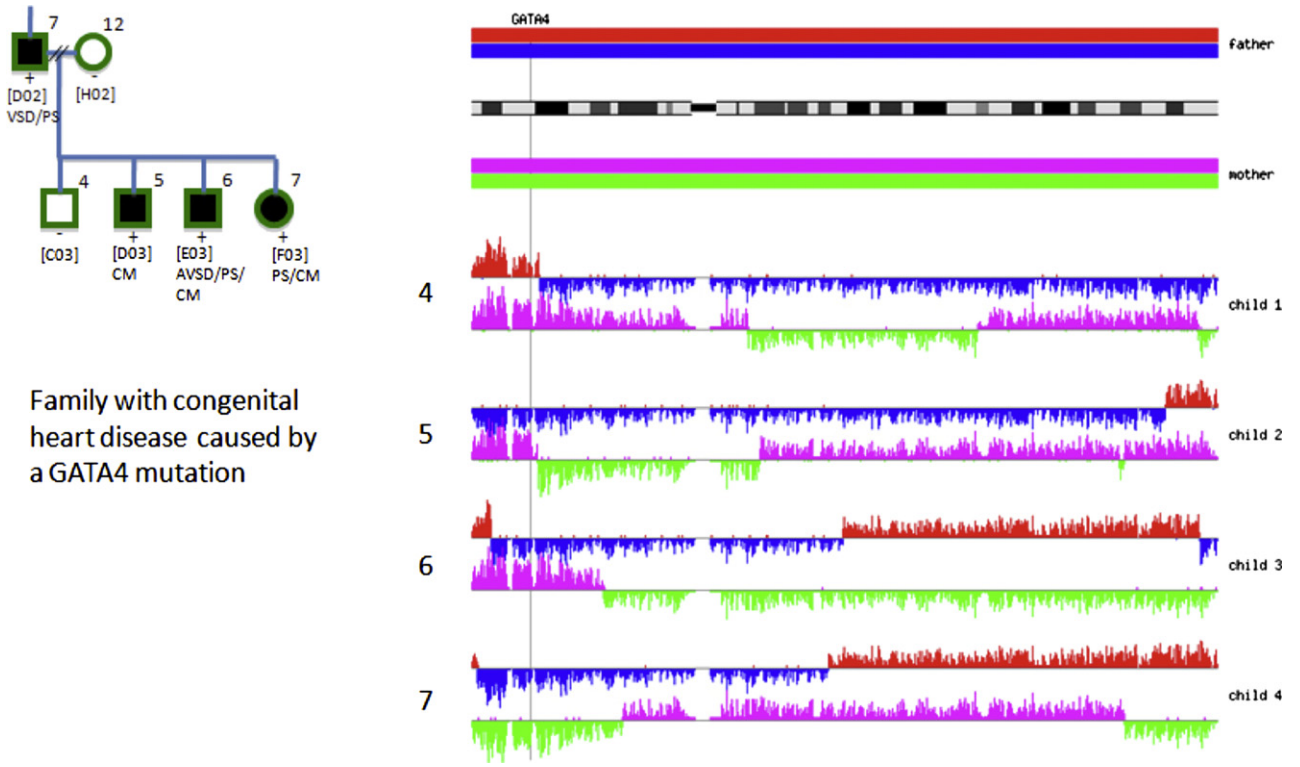


FIGURE 23.8 A schematic depicting the haplotypes of the members of a family of six. The family tree is indicated at the left. The four parental haplotypes (two for each parent) are indicated by four different colors. The portions of the parental haplotypes that are passed on to each child are indicated by the same colors. Each color change denotes a site of chromosomal recombination. Family genome sequencing permits one to determine these recombinational sites with great precision. The important point is that the genes that cause particular diseases must reside in areas of shared haplotype by those individuals in the family exhibiting the disease.

the genome, independent of any genetic models for the two diseases. Fourth, we were able to determine the intergenerational mutation rate for the two children (about 35 mutations per child). In this regard, it is interesting to note that because of intergenerational mutations there is no such thing as genetically identical twins. Finally, we were able to reduce the candidate gene list for the two diseases to just four possibilities. The correct defects could readily be associated with the diseases by using other genetic analyses of these defects. Thus whole-family genome sequencing is a powerful approach to enriching the signal-to-noise intrinsic to human genetic studies. Family genome studies constitute a powerful new approach to identifying the genome elements that are responsible for health or disease.

The technologies of genomics are becoming increasingly mature. This means that companies will be able to perform genomic analyses far more effectively than most academics, and these analyses will be increasingly outsourced to highly efficient vendors.

Mass spectrometry and the identification and quantification of proteins in proteomes. The analysis of proteins differs from that of their genomic counterparts in several ways. First, DNA is basically digital in nature, i.e., sequences and functions are specified by a digital four-letter language. In contrast, proteins are synthesized as linear structure, but fold into three-dimensional structures to execute their functions. We can, however, digitize the identification and quantification of proteins through analysis by mass spectrometry (see below). Second, proteins exhibit enormous complexity in structures due to many different procedures associated with their synthesis, including RNA editing, RNA splicing, protein modification, protein processing, etc. Indeed, some have estimated that the human genome may produce a million or more proteins [45]. Third, proteins are dynamic, changing their structures as they execute their functions and as they interact with other small and large molecules [46]. Fourth, proteins cannot be analyzed in a global or comprehensive manner (unlike genomic features) because of the enormous dynamic range of protein expression (10^6 in tissues and 10^{10} in blood), which exceeds the dynamic range of detection by the mass spectrometer. Finally, there is no protein amplification method equivalent to the polymerase chain reaction (PCR) of nucleic acids to enable the amplification and analysis of rare proteins.

A proteome is the collection of proteins that are present in a biological entity — a cell, an organ, the blood or an individual. The Human Genome Project has given us the sequences of most of the human proteins (and their tryptic peptides) and this has enable mass-spectrometry-based proteomics. Mass spectrometry can identify (and quantify) tryptic peptides of the proteins (those generated by the proteolytic enzyme trypsin cleaving either at the amino acid residues lysine or arginine). The mass spectrometer has the ability to separate and determine the mass to charge ratios of the tryptic peptides (and thus identify them). Quantification is achieved by determining the frequency of a particular tryptic peptide in a proteome mixture and averaging the frequencies of all the identifiable tryptic peptides shared by a particular protein.

Mass spectrometry has been used in two ways. A shotgun proteome analysis attempts to quantify all the tryptic peptides present in the given proteome. This procedure permits only the more dominant proteins to be quantified accurately because of the limited dynamic range and the fact that the peptides of rare proteins will be seen rarely, if at all. A targeted proteome analysis permits 100 proteins to be identified in a complex mixture by synthesizing isotopically labeled 'standard' peptides to be compared against their counterpart peptides in the proteome mixture. The triple quadrapole mass spectrometer has the ability to search out the peptides from the proteins that will be quantified by these assays. The targeted proteomic approach is called selective or multiple reaction monitoring (SRM or MRM) mass spectrometry. Recently, standard assays have been developed for most of the 20 000 or so human proteins [47]: just as the Human Genome Project 'democratized' all human genes by making them accessible to every biologist, so this proteome project has 'democratized' human proteins. Mass spectrometry can also be used to analyze the proteins present in organelles and can analyze those proteins interacting with one another (after pulling down and purifying the interacting protein complexes with specific antibodies). Mass spectrometry has also been used to look at translational chemical modifications and the protein forms arising from alternative RNA splicing.

The SRM assays have been used in assaying brain and liver organ-specific blood proteins, both in the organs and in the blood. As noted above in the section discussing 'blood as a window', one would like to contemplate the ability to analyze, say, 50 organ-specific proteins from each of 50 different human organs, and to follow for each patient's protein footprint across time to assess health vs. disease. As one contemplates the possibility of analyzing organ-specific blood proteins several times a year in the blood of the 340 million patients in the US, the mass spectrometer does not have the extendable throughput to manage 680 million samples. In the future these analyses will be done by microfluidic chips with ELISA (antibody) assays for each protein. A recent microfluidic protein chip has been designed with 50 ELISA assays for blood proteins that can be analyzed from 200 nL of plasma in about 5 minutes across a dynamic range of close to 10^5 and a sensitivity in the mid-atomole range [39]. We believe that the potential for this chip can be extended to thousands of protein measurements. Generating pairs of antibodies for this many proteins would be extremely expensive and time-consuming, and, moreover, antibodies are not very stable. Accordingly, one will have to develop more effective, stable and scalable protein-capture agents. Aptamers (for example, 60-mers of DNA) and trimeric or tetrameric peptide fragments (e.g., D amino acid 6-mers) appear to be interesting candidates as new protein-capture agents [48,49].

The mass spectrometer is one of the most powerful approaches for analyzing metabolites [35,50]. The challenges for metabolite analysis are generally similar to those of proteins (distinguishing the enormous number of different metabolites, facilitating their identification dealing with their broad dynamic ranges of expression, etc.).

Single-cell analyses. Most of our understanding of development, physiological responses and the initiation and progression of disease comes from studies that assess populations of cells. In the future the analyses of large numbers of *individual* cells will become important and standard practice — analyses that allow the genomes, epigenomes, transcriptomes, miRNAomes, proteomes, metabolomes, and interactomes within single cells to be determined. A variety of microfluidic and nanotechnologic approaches are being applied to these problems. Single-cell analyses will allow us to answer two fundamental questions. First, do discrete, quantized populations of cells exist within given organs? Preliminary results suggest that the answer to this question is yes. The fundamental issue is: what are the biological roles of these quantized populations? For example, they may represent a series of transition intermediates on their way to a final end stage. Alternatively, they may represent discrete populations of cells each with a separate function within the organ. Second, the expression of some of the information molecules within individual cells may behave in a stochastic manner. Single-cell analyses, properly executed, will permit us to distinguish between quantized cell populations and stochastic variability. We believe that single-cell analyses will be a critical tool in the future for deciphering biological complexity.

As indicated in the introduction to this section, many additional technologies are emerging that will open up the exploration of new dimensions of patient data space.

5. The 'data explosion' requires that new analytic tools be created for capturing, validating, storing, mining, integrating and finally modeling all of these biological data sets, thus helping to convert them into knowledge. A critical point is that these software solutions must be driven by the needs of leading-edge biology and medicine — and by biological domain expertise. One big revolution in medicine is that we will create massive amounts of digital data for the 'quantified self' of each individual that will transform our ability to monitor and optimize our own wellness. The following sections discuss in detail the transformation of big data sets to medically relevant information

Computational integration of 'quantified self' data will revolutionize health. Information on the quantified self provides enormous potential for the future of P4 medicine, as we are able to harness this information productively through powerful data analysis and large-scale computation (Figure 23.9). The key issue is how such large repositories of data can be turned into actionable knowledge. The potential of this endeavor is enormous, as we will gain unprecedented detail about how our bodies work, what brings about disease, and how wellness can be maintained. Interpreting multifaceted biological data deeply for each individual — and integrating it broadly across populations — will open new vistas of biological knowledge and clinical power. The pace of the technological changes will be quick, driven by exponentially rising computational power to take advantage of the exponentially rising amounts of high-throughput biological data. This is P4 medicine's heritage from the digital revolution.

Four factors will be important for dealing with the striking signal-to-noise issues of large data sets: (1) the integration of similar data types from different laboratories to enormously enlarge the data sets analyzed (see below); (2) the integration of data of different types — including molecular, cellular, conventional medical and phenotypic data; (3) the transformation of these data into the 'network of networks' for each individual patient — and following the dynamics of the 'network of networks'; and (4) the use of subtractive biological analyses to eliminate various forms of biological noise, as described in the prion discussion (see above). Each of these approaches represents significant computational/mathematical, technical and biological challenges, some of which are illustrated in the following discussion.

Realizing the power of high-dimensional diagnostics requires overcoming very significant data analysis challenges. This exciting future will only be realized as we address very significant computational and data analysis challenges. The human body is an enormously complex dynamic system interacting with an ever-changing and diverse environment. As the capacity to make molecular measurements continues to increase in scope and precision, the challenge of finding the relevant signals amidst the sea of observations can be daunting. As with any complex system, causality is often difficult to find and there are many ways that systems can break down and result in disease. Our bodies have enormously intricate and beautiful approaches for dealing with disease, for example via the immune system, and thus the residual medical problems we must solve must consider the consequences of the highly adaptive protective immune responses: both those that have been successful and those that were unsuccessful (such as the failure to check malignant cancers). Thus, these problems are often highly challenging, including from an informational point of view.

The primary challenge of big data in biology is to separate relevant signal from noise, including both technical noise (from measurements) and biological noise (from other biological factors besides those of interest). The number of measurements that come from the quantified self present significant hazards for proper interpretation. Having increasing ability to make precise measurements is exciting because so much new information is available, but care must be taken not to build overly complex models that appear very good on initial data assessment, but which fail when moved forward towards potential clinical use. Using an overly complex model that fits the already observed data really well, but then does not maintain accuracy when applied to new data for the same phenomenon (i.e., the model is fitting noise rather than the true underlying relationship), is called overfitting. In biological and clinical studies with 'omics' data, we are typically in what statisticians refer to as the small samples size regime. That is, we have very many more variables than we do observations. For example, there are tens of thousands of different transcripts measured in a human transcriptome (the variables), but generally only of the order of 100 or so samples (the observations) in a given study. This is exactly the opposite situation of what is desired to reliably use measurements to distinguish classes and establish reliable relationships among the variables (e.g., transcripts): one would like to have very many observations relative to the number of variables that are being used. Because the number of variables is

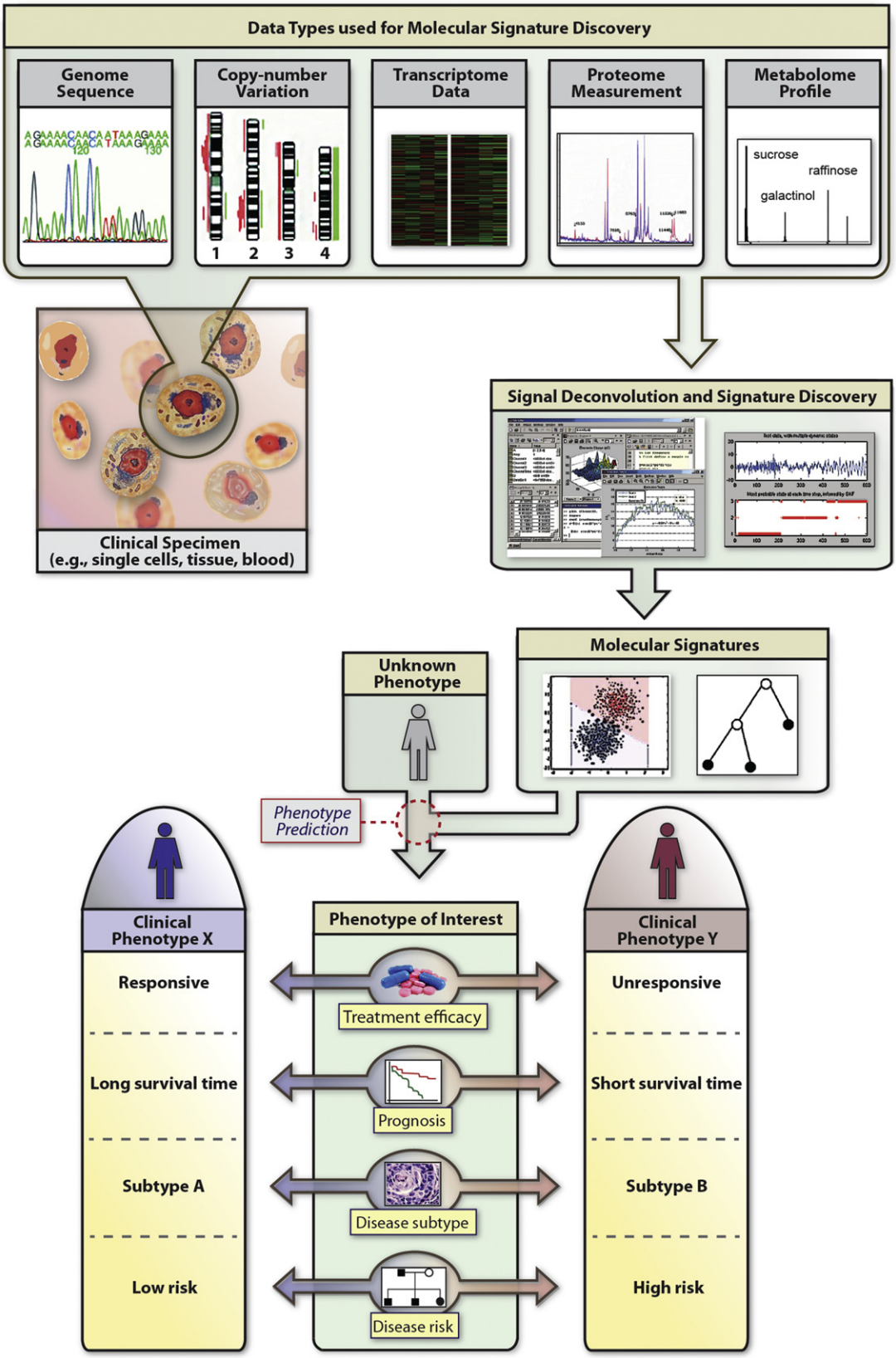


FIGURE 23.9 **Overview of the discovery and application of molecular signatures for phenotypic assessment, including disease diagnosis, treatment selection, prognosis, and risk assessment.** *(Figure adapted from [51].)*

so high, this increases the chances that overfitting to the observed data will become an significant issue, leading to results that appear promising but that in reality will not hold up for clinical use. For the quantified self, transcriptomics will be only one of the data types, and thus the number of measurements means that statistical approaches to control for overfitting will be essential [36,51]. Fortunately, the continued dramatic reduction in cost for many 'omics' technologies will help to mitigate some of these challenges by making it possible to analyze larger sample numbers, but for what is contemplated in the near future for P4 medicine we are still very much in the small sample regime, and will be for quite some time to come.

This high dimensionality of data exacerbates issues related to data reproducibility, which can be difficult to reproduce in detail from study to study, or even from batch to batch [52]. For example, in order to statistically expect overlap of 50% in identified differentially expressed genes between two different studies comparing breast cancer to normal tissue, one would need of the order of thousands of samples [53]. Almost all individual studies today have fewer than that, and so even if executed with the highest possible experimental rigor one would not expect differentially expressed gene lists between studies to be very similar — even if everything is done correctly by both laboratories doing the studies. The same is true for identified molecular signatures, where a number of molecular measurements are coupled with a computational algorithm to differentiate phenotypes (e.g., make a disease diagnosis) [54].

Importantly, when one study is used to train a molecular signature to differentiate between different phenotypes (e.g., cancer vs. control) and then tested on a separate study of the same phenotypes, very often the classifier will fail, or at best the signature performance degrades severely. A primary reason for this drop in performance comes from heterogeneity between studies, due both to underlying variance in the biology of the patients studied and to technical variations in precisely how the data were measured, normalized, and analyzed. Whenever two individual studies are compared it is very often the case that the differences that we will refer to here as laboratory effects are greater than the differences due to phenotype (e.g., cancer vs. normal). One powerful means for making signatures much more robust is to integrate their identification across multiple studies at multiple sites [55]. In such integrated studies, the signal associated with the phenotype difference is amplified, while the laboratory effects are damped out. Signatures learned across multiple different studies from multiple different laboratories perform much better on average on yet additional studies than do signatures learned from one study alone. This fact argues strongly for the need to [1] build consortiums that enable the integration of large amounts of data from multiple sites [14] and [2] make data publicly available so that they can be aggregated together in meta-analyses. Such data integration is essential to enable P4 medicine.

Computational challenges of blood as a window. The computational challenges associated with maintenance of wellness and the pre-symptomatic diagnosis and prevention of disease are particularly challenging. For example, the envisioned blood diagnostics of the future must be able to distinguish not just one disease from normal but rather must differentiate any possible disease against the background of normal that can be affected by many conditions — including even mundane changes such as diet, exercise, time of day, sleep cycle and so forth. As is well appreciated in machine learning, accuracies tend to degrade quickly as more potential phenotypes need to be separated simultaneously. Deciphering signal from noise against such a dynamic and multifaceted background is a daunting challenge indeed. A number of strategies will therefore be important to harness the information content of the blood as a window to health and disease. 1) It is unlikely that any one data platform will be sufficient to achieve the accuracies that will be needed for clinical practice across the wide range of possible disease states. Therefore, to achieve the predictive and preventive aspects of P4 medicine will require multifaceted data analysis, including multiple sources of molecular data from the blood. As is described here, the blood contains enormous numbers of different molecular information sources, including not just the proteins, but also metabolites, miRNAs, mRNAs, circulating cells, antibodies and so forth. It will also be important to link these molecular data with clinical data as well as input from activated and digitally networked patients (such as changes in lifestyle, environmental exposures). Patient activation refers to a person's willingness and ability to manage their health and healthcare, as measured by the Patient Activation Measure [56]. 2) Another key to address this challenge is to build coarse-to-fine hierarchies, where coarse overall assessments are made initially and then followed by tests of increasingly finer levels of specificity, for disease diagnosis and wellness monitoring. For example, organ-specific blood proteins can be used to first answer the question of what organ system is being perturbed. Following this assessment, more specific molecular markers of finer resolution that differentiate different diseases of the

organ system can then be used to narrow down the disease possibilities further. Once a disease is distinguished, yet finer resolution molecular signatures can be used to help determine perturbed networks and hence the best therapy options. 3) Considering the plethora of computational challenges above, there are many instances where it is hard to imagine amassing sufficient statistical power to address all the relevant states of wellness and disease. The key in these prevalent cases will be to leverage deep biology and knowledge of mechanisms. In this case, the mapping of molecular networks through systems biology approaches — and the interplay between genomics and the environment — will be crucial to deciphering what signals of the quantified self really matter to disease treatment and health maintenance.

Health information systems of the future. Looking forward, the development of P4 medicine based on 'omics' data will require extremely large repositories of data in minable health information systems where signatures are constantly evaluated, updated, locked down and then re-evaluated for efficacy. Such systems will be based largely on data in the 'real world' of patient treatment and clinical outcomes — since this is where the vast majority of medically relevant data will come from in the future. Learning what factors most affect patient outcomes by broadly measuring new data sources and linking these back to activated patient communities will serve as a powerful paradigm for developing new tests to move forward iteratively through clinical evaluation. Importantly, these systems will be unbiased in the sense that they will record both positive and negative outcomes as seen in the clinic equally (a key problem with current literature practices, where essentially only good outcomes are widely published and reported in the development phases, and bad outcomes are selected against). Such systems will need to be very expansive in terms of numbers of samples and measurements to be integrated — and there will thus be institutional barriers to sharing data that will need to be overcome through collaborative models. Whether the data will be formed as raw data or processed into metadata for storage is a critical question, and almost certainly we will move toward the storage of metadata to reduce the data dimensionality. As an example, the 6 billion nucleotides of the human genome can be compared against a reference genome — and since humans differ by about 0.1% of their genomes we could store only the differences, reducing the data dimensionality by three orders of magnitude. There will be a strong financial incentive to collaborate, as the clear winners in this area will come only from those who can achieve sufficient sample numbers across a sufficient breadth of the population to identify the most robust signatures. Thus, there will be enormous commercial opportunities that will form the basis for emerging health information companies that will mine this data and produce content that is directly usable by consumers (patients) as well as physicians. While the number of signatures that are translated currently is small relative to the number that have been reported in papers, reasons for this from a statistical point of view — given that we are deeply in the small sample regime — are clear. There is every reason to believe that as the data are integrated at the scale that is necessary, with the rigor that is necessary, and with the connection to biological networks and mechanisms that is necessary, these approaches will indeed transform the practice of medicine. And they will enable all of the ideal features of diagnostics — early detection, assessment of the stage of disease progression, stratification of disease, following the response to therapy and detecting reoccurrences of disease.

Systems medicine provides powerful approaches for dealing with signal-to-noise issues and biological complexity. Systems medicine allows us to reduce enormously the dimensionality of the search space for accurate and robust biomarkers. For example, systems approaches have led to organ-specific, cell type-specific and organelle-specific biomarkers reflecting the functioning of key disease-perturbed networks. Such approaches have also been used to identify biomarkers in the blood, using secreted proteins, proteins cleaved from the membranes of the disease-perturbed cells, cytoplasmic and nuclear proteins reflecting the death of cells, etc. The important point going forward is to start with a narrow and targeted set of biomarkers to search for molecular fingerprints that can reflect the disease, including early diagnosis, stratification of the disease types, and assessing the progression of the disease. Accordingly much smaller populations of patients can be used to identify valid biomarker signatures with focused studies that leverage biological knowledge. Exactly this approach has been applied successfully to several mouse model diseases.

These five pillars of systems medicine together with the digital revolution have given rise to the medical opportunities embodied in P4 medicine — prediction, prevention, personalization and participation.

P4 MEDICINE

Systems medicine is focused on developing biological, technical and computational tools to decipher the complexities of disease. P4 medicine employs the strategies and tools of systems medicine for quantifying wellness and

demystifying disease for the benefit of the individual, as well as dealing with the societal opportunities and challenges created by this revolution in medicine.

A key component of P4 medicine is the 'activated' (continually informed patients possessing the knowledge, skill and confidence necessary to manage their health and healthcare and that of their families) and 'networked' (patients connected with other patients and other members of their community, working together to enhance personal and community wellbeing) patients [56]. The convergence of systems approaches to wellness and disease with activated and networked patients will result in a P4 medicine that integrates discovery science with clinical practice and health management by networked and activated patients. This integration will generate the virtual clouds of billions of data points for each patient (Figure 23.3). Analyses of this data cloud will decipher the 'network of networks' for each patient and lead to discoveries and the optimization of wellness and disease emerging from relevant patient social networks. P4 medicine is now pioneering something that never existed before – actionable understandings of disease and wellness as a continuum of network states unique in time and space to each individual human being that can be perturbed by various means, including but by no means limited to drugs, to restore and maintain health.

These databases of patient information and social networks will provide physicians and other healthcare providers with the information they need to deliver care tailored to the circumstances of each individual. We are beginning to see this in the case of certain cancers, where the DNA sequencing of the tumors provides insights into the mutations in signal transduction pathways that inform the choice of therapy for the individual patient [57, 58]. In addition, data from clinical encounters, and from networked patients and consumers actively involved in managing their health and that of their families, will provide millions of data points that will enable systems biologists to decipher signal from noise in complex biological networks. These data will be funneled to scientific research centers and emerging health information companies to fuel large-scale studies. Eventually 100 000s or even millions of patients' data on physiological, cellular, and molecular markers will be used to increasingly demystify disease and quantify wellness by augmenting the relevant biological signals.

An exciting new cycle of accelerating biomedical innovation will emerge as systems-medicine discoveries are routed back to patients and consumers, thereby generating more data to fuel further advances in actionable insights into individual biological systems. P4 medicine includes novel techniques and paradigms for facilitating new relationships between scientists, care providers, patients and consumers, as well as for dealing with the social opportunities and challenges that inevitably will arise from these new relationships. Here is a 10-year glimpse into the envisioned future of the 4 Ps.

Predictive. In 10 years nearly everyone will have his or her genome sequenced. 'Actionable genetic variants' – those whose identification opens the door to a course of action that will improve physical health or relieve anxiety for the individual – will drive forward the acceptance of the complete genome sequence as a part of the individual's medical record. While most of the medically relevant variation catalogued to date occurs in the coding regions of genes, it is becoming clear that variations in non-coding regions, copy number, structural variations, and other features of chromosomal architecture play a role in disease etiology. We believe that complete genome sequencing should be done in families. Family sequencing enables the correction of a significant fraction of the DNA sequencing errors, thus generating very accurate sequences. It also provides a deep understanding of the one-dimensional organization of the many genetic variants in the chromosomes of each individual (this is denoted haplotype determination) (Figure 23.8), thereby enormously facilitating the discovery of disease genes or loci. We have identified more than 300 of these actionable variants and many more are being identified with every passing year. Indeed, every person's genome will be reviewed yearly for new actionable variants – and these will provide powerful insights to optimize the wellness of the individual. The genomes of individuals will provide an investment in individual information that will permit a yearly optimization of wellness for the rest of the person's life.

P4 medicine is making blood a window for assessing health and disease. Organ-specific proteins found in the blood will be analyzed in a longitudinal manner across the individual's life (Figure 23.7). Within 10 years, we envision a hand-held device that can prick your finger, take a fraction of a droplet of blood and quantify several thousand organ-specific proteins in 5 minutes. This device will permit the health or impairment status for each of your major 50 organ systems to be followed in a longitudinal manner over time [41]. Moreover, these 'microfluidic protein assays' will permit hundreds of millions of patients to be analyzed routinely, e.g., biannually. Thus transitions from health to disease may immediately be identified and acted upon. The organ-specific blood fingerprints will allow us to stratify diseases into their distinct subtypes (for an impedance match against proper therapies) and to follow the progression of a disease. In the future different drugs will be effective against different stages of a disease. In addition, organ-specific blood protein fingerprints will enable one to analyze the responses of multiple organs to a given disease – and thereby extend the analysis of the 'networks of networks' from DNA, molecules and cells to organs (Figure 23.4). Finally, organ-specific blood protein will

enable one to rapidly and precisely identify off-target reactions of drugs, thereby providing a powerful approach to assessing drug toxicities.

Preventive. Systems analyses provide insights into the dynamics of disease-perturbed networks. A new 'network-centric' rather than 'gene-centric' approach to choosing drug targets will employ multiple drugs to 're-engineer' a disease-perturbed network to make it behave in a more normal manner. We are now exploring this possibility in microorganisms to learn the fundamental principles of network re-engineering before applying this strategy to the more complex requirements of human disease therapies [59]. Our conviction is that the treatment of disease in the future will often require the combination of two or more drugs. Tools are now being developed to explore the combinatorial analyses of the hundreds of drugs that have met the safety requirement – both those that were efficacious for a particular disease and those that were not. Hence for the effective and non-useful drugs perhaps combinations could be identified to more effectively attack a wide variety of diseases. This could make drugs more effective and far less expensive, because there is a rationale for the choice of drug targets and their perturbations by drugs. Through longitudinal multi-biomarker (blood protein, miRNA, mRNA, metabolites etc.) analyses, P4 medicine will be able to predict the potential future emergence of disease-perturbed networks in patients and then design 'preventive drugs' that will block the emergence of these disease-perturbed networks and their cognate diseases. A systems approach to the immune response will, in time, give us a deeper understanding of how to create effective cellular as well as humoral immune responses, permitting us to create effective vaccines for scourges such as AIDS. Clearly, stem cells will provide powerful possibilities in the future for replacing damaged cellular and even organ components (as well as being powerful tools for understanding disease mechanism and stratifying disease). Finally, the digitized data defining the quantified self will provide powerful new insights into optimizing wellness for the individual. Indeed, the focus of P4 medicine will increasingly move from disease to wellness.

Personalized. On average, humans differ from one another by about 6 million nucleotides in their genomes – hence we, individually, are genetically unique. Even identical twins each may exhibit 35 different 'intergenerational mutational nucleotide differences' from each parent and from one another [33]. Each person must be treated as a unique individual and not as a statistical average. We must take account of the fact individuals vary in ways that significantly affect effective treatment. Individuals should each serve as their own controls to determine when their own data reflect transitions from health to disease. Moreover, there is a growing sentiment that observations on single individuals may collectively provide fundamental new insights into the disease (or wellness) process. The so-called experiments where patient N equals 1 may open up powerful new approaches to more effectively dealing with the individual patient and aggregating the useful data they generate. Indeed, the first molecularly detailed tracking of two individuals (the 'quantified self') is already yielding fascinating insights [12,25]. Imagine, in 10 years, 340 million Americans each with billions of data points: this will potentially create a powerful aggregated data source from which to infer the predictive medicine of the future.

In this regard, we believe that it will be critical to champion the idea that it is essential that all patients' data (with appropriate privacy measures including anonymization) be made available through an appropriately constituted entity for qualified researchers and physicians to mine for the predictive medicine of the future. After all, this contribution will enable us to revolutionize the healthcare for our children and grandchildren – a point to which most patients are responsive. Moreover, laws must be passed to protect the individual against the exploitation of their medical data by other elements of society, such as employers or insurance companies. It is interesting to note that all of us make available our entire financial histories to three credit agencies for the convenience of having a credit card. Surely patients will recognize the benefits of being able to mine their collective data to pioneer the future of P4 medicine for the benefit of their families.

Participatory. P4 medicine relies greatly on the positive contributions of activated patients and consumers. Our existing healthcare system is not well adapted to exploit the new capabilities of P4 medicine. Physicians as well as pharmaceutical and medical device companies are compensated solely for the delivery of specific procedures and products, hence they have limited financial incentive to deploy new innovations to predict or prevent disease or to maintain wellness. Moreover, the healthcare industry is locked into financial and regulatory models based on large-scale population studies that ignore crucial genetic and environmental exposure differences among individuals. Pressure for change is beginning to be felt as the medical profession faces the looming challenge of increasingly being compensated for outcomes as opposed to service delivery. However, the most important source of pressure for change will be newly activated and networked patients and consumers. Collectively, they will constitute a vital new stakeholder in P4 medicine very different from the passive recipients (patients) of expert advice characteristic of pre-digital medicine. Activated and networked consumers will do more than demand more effective healthcare – they will help direct the changes to achieve it.

Activated and networked consumers are beginning to push for healthcare that is adapted to their own particular circumstances, including their individual genome (which is static and need be sequenced only once) and dynamic measurements such as from blood (which change over time,

must be measured longitudinally in time, and can track changing wellness and disease states) [10]. They are also beginning to push for new ways in which to engage with our science-based healthcare system to maintain wellness and achieve life goals as well treating disease. Because of the reactive nature of the existing healthcare system, which is more accurately described as the 'disease management industry', the vast bulk of health management in areas such as nutrition, exercise and sleep takes place in the home, without the assistance of physicians or other professionally trained care providers. The writers of bestsellers about the latest diet and purveyors of unregulated health products operate largely outside the constraints of science-based healthcare. Today's educated consumers are increasingly conscious of this fact and are beginning to demand that science-based healthcare address their need for assistance in managing these other areas of their own health. They are stimulating the growth of a new market for devices that deliver increasingly real-time digital data about every aspect of their health (quantifying their wellness), ranging from activity levels to vital signs, and many of them are starting to come to their physicians for help in interpreting these data. The digital revolution is beginning to help fashion a new dimension for healthcare — the wellness arising from the quantified self.

P4 medicine responds to these growing demands by providing patients and consumers with actionable information that they can use to improve their health. Clinical institutions using wellness coaches, genetic counselors and physicians can provide this information cost-effectively. It will be conveyed largely through digitally linked social networks, the most important of which will be family networks. One effective strategy may be to identify family members who are the most active in setting familial health-related standards and in caring for members with health problems, and then to work with those individuals to help them do a better job. Medicine today systemically approaches patients as statistical abstractions, relying on the efforts of time-pressed physicians to achieve some degree of personalized care. Working with family and social networks will allow P4 medicine to systemically and more effectively deal with the reality of the social context in which patients and consumers are embedded and which largely determines how they eat, exercise and sleep. Activated patents and consumers working effectively within their family and social networks to utilize the increasingly real-time flow of personalized health-related data will be able to reduce the incidence of and to better manage complex diseases such as type II diabetes, which account for a huge percentage of total healthcare costs and, of course, to optimize their wellness.

To achieve this goal we need to develop and continually update a 'gold standard' of reliable data, information and explanations of disease and health that will meet the needs of both physicians and patients. Developing and maintaining this gold standard will require a close working relationship between clinical, systems-based scientists and patients. In addition, we need to develop ways to actively counter misinformation that might begin to spread through social networks and to correct misinformation often found on the current medical websites and in other sources of medical information. While significant, these challenges are outweighed by benefits to be gained by wide dissemination of actionable 'gold standard' of personalized health data reviewed for consistency with the standards of our science-based healthcare system.

Ultimately patients will be recognized as not only a source of disease problems to be solved but as a source of disease and wellness solutions as revealed by their data. Creative new forms of engagement with networked and activated consumers as active participants in healthcare, as opposed to passive recipients of expert advice, will become a major source of value tapped by the healthcare revolution. Networked and activated participants will find new ways to adjust diet and exercise to move their biomarkers in the direction of better health. Crowd sourcing these problems will yield many benefits. For example, researchers will be able to correlate their behavior changes with biomarker fluctuations, their genome, their medical histories and other key parameters. Such data from millions or even tens of thousands of patients would provide researchers with deep insights into the effects of nutrition and exercise that have never before been possible. These large-scale personalized data sets would be the basis for the quantification of wellness, providing society with a far more effective understanding of the effects of diet, exercise and sleep on highly stratified population sectors.

The digitization of P4 medicine enables its distribution to all citizens of the world — both developed and less developed. For example, we remember those who thought the initial large brick-like cell phones of the early 1990s were ridiculous and could not imagine their widespread acceptance (there are now more than 4 billion cell phones worldwide). Today, a woman in a rural village in India can make a living for her family with a cell phone thanks to the digitization of communications, with its potential for transforming the economic conditions of even the poor. So we will see a 'democratization of P4 medicine' throughout the world as inexpensive and digitized P4 medicine becomes available (see below).

To summarize, we argue that P4 medicine has two major objectives for each participant: to quantify their wellness and to demystify their disease (Figure 23.2). Our feeling is that the quantification of wellness will become increasingly important over time, ultimately dominating as the concern of most individuals. Table 23.1 provides a striking comparison of proactive P4 medicine with contemporary, reactive evidence-based medicine.

TWO BIG CHALLENGES: EDUCATION AND INFORMATION TECHNOLOGY FOR HEALTHCARE

One big challenge for P4 medicine is the education of consumers, patients, physicians, and the members of the broader medical community, including the principal stakeholders in the healthcare industry. This education will present an enormous challenge and will ultimately require the effective exploitation of social networks for education integrated with new effective information technology teaching strategies. Many individuals will initially want to remain 'old-fashioned patients' letting the doctor tell them what is best. However, once individuals see the power of consumer-driven medicine to improve individual health, that will change (just as skepticism has disappeared with the widespread acceptance of cell phones).

The Institute for Systems Biology has successfully developed modules for teaching systems biology to insert leading-edge biology into high-school biology courses, and is currently developing similar modules for P4 medicine. Early education of consumers/patients is key. Another interesting idea is the suggestion that there be a commercial TV program, hopefully with very broad coverage, along the lines of the forensic CSI TV program to explore solving the problems of P4 medicine in a well-written and compelling manner that brings a knowledge of P4 medicine to the average viewer, just as CSI has brought insights in crime forensics to a broad audience. Another possibility is that one could use computer-game-like strategies to bring the principles of P4 medicine to patients, physicians and members of the healthcare community – at least for those who are comfortable with the digital revolution.

Another challenge is how to produce an information technology (IT) for healthcare that can handle the enormous multi-scale data dimensionality that will arise from P4 medicine – for in the end P4 medicine is defined by the interconnected 'network of networks' – genetic networks connected to molecular networks, to cellular networks, to organ networks, to the networks of individuals in society, for each provide unique insights into the complexities of disease (Figure 23.4). We must understand the individual in the context of all of these integrated networks, as this is the only way to capture both the digital information of the genome and all of the diverse environmental signals impinging on the individual from many different sources. This requirement places enormous demands on the need to develop an effective IT for healthcare. Healthcare IT must be comprehensive, interoperable, data-driven (e.g., bottom-up), biology-driven and, we believe, fundamentally open source. It is probably beyond the capacity of any single organization to fashion a comprehensive IT for healthcare that goes beyond medical records to encompass the collection and distribution of the entire heterogeneous data cloud at the heart of P4 medicine. Yet that is what is required, and an effectively orchestrated open-source approach could transform IT for healthcare. We must be able to capture the deep insights that will come from various patient social networks.

IMPACT OF P4 MEDICINE ON SOCIETY

P4 medicine will have an enormous impact on society and healthcare.

1. P4 medicine will transform the practice of healthcare in virtually every way. Table 23.2 provides a summary of some of the powerful new strategies and technologies P4 medicine will create and employ.
2. P4 medicine will require that all healthcare companies rewrite their business plans in the next 10 years or so. Many will not be able to do so and will become 'industrial dinosaurs'. There will be enormous economic opportunities for the emergence of new companies tailored to the needs and opportunities of P4 medicine.
3. P4 medicine will at some time in the future turn around the ever-escalating costs of healthcare and will in fact reduce these costs to the point where P4 medicine can be exported to the developing world, enabling a 'democratization of healthcare' unimaginable even 5 years ago. These savings will arise from many of the features described in Table 23.2: the early diagnosis and hence more effective treatment of disease; the stratification of each major disease into its major subtypes to achieve a proper impedance match for each individual against a drug effective for a particular subtype of disease; the ability to identify genetic variants that cause drugs to be metabolized in a manner dangerous to the patient (this is termed pharmacogenomics, and more than 50 such variants have been identified to date); the ability to 're-engineer' disease-perturbed networks with drugs to generate a powerful and less-expensive rationale for drug-target selection; an increasing focus on wellness for each individual; and the emergence of striking near-term advances in modern medicine. These include an increasing ability to deal effectively with cancer, and to use stem cells for replacement therapy. Additionally, these advances will lead to new approaches to diagnostics and understanding disease mechanisms, an understanding of aging that will allow individuals to optimize and extend their effective mental and physical health routinely into their 80s and 90s, an understanding of the metagenome (e.g., population of microbes) of the gut and other body surfaces that will provide deep insights into one incredibly important manner in which the microbes of our environment influence our health, and finally the emergence

of a deep understanding of neurodegeneration to avoid the personal and societal tragedies of diseases such as Alzheimer's and Parkinson's.

4. P4 medicine, through its driving of the emergence of new technologies and computational techniques, is pioneering the digitization of medicine. The 'quantified self' will provide the data that will enable each individual to optimize his or her own health. This will also provide the data to empower P4 medicine to revolutionize healthcare through consumer-driven social networks (Figure 23.1). The digitization of medicine through the generation of big data sets for each individual – allowing one to sculpt with exquisite specificity wellness and appropriate responses to emerging diseases – is one of the transforming aspects of P4 medicine. Another important implication arising from the digitization of medicine is the fact that personal data will become incredibly inexpensive (e.g., it is estimated the first human genome sequence finished in 2003 cost about $1 billion; today it costs a few thousand dollars, and in a few years genome sequences will cost perhaps $100) – thus digital technologies and their exponentially declining costs will contributing significantly to reversing the escalating costs of healthcare.
5. P4 medicine will bring increased wealth to the healthcare systems, communities and nations that practice it. The decreasing costs of healthcare have been mentioned above. Many economic opportunities will evolve from the knowledge of P4 medicine through the transformation of the healthcare industry. We predict that there will be a 'wellness industry' that will emerge over the next 10–15 years that will in time far exceed the size of the healthcare industry. P4 medicine is an area replete with economic opportunities for those who are practicing it at the leading edge.
6. The patient (consumer), through social networks, will drive the emergence of P4 medicine. Because of intrinsic conservatism and sclerotic bureaucratic systems, physicians, healthcare specialists and the healthcare industry will take a back seat to the power of patient-driven social networks in bringing change to the healthcare system. Indeed, patients may be the only driving force capable of truly changing our contemporary healthcare system to the proactive P4 mode.

HOW TO BRING P4 MEDICINE TO PATIENTS

The challenges of bringing P4 medicine to patients and consumers have two critical dimensions, technological and societal. The latter is far more complex. A variety of efforts are focused on bringing personalized medicine to healthcare and to dealing with some of the societal issues of this new medicine (including the Personalized Medicine Coalition, established to advance the future of personalized medicine by connecting the scientists, clinicians, the media and the general community in a common goal). We will discuss below only our own efforts to bring P4 medicine to patients.

At the Institute for Systems Biology, in conjunction with Ohio State Medical School and PeaceHealth clinics, we have created the P4 Medicine Institute (P4MI), a nonprofit organization that is committed to creating a network of six or so clinical centers with the ISB to employ the strategies and tools of P4 medicine in pilot projects to prove the power of P4 medicine. Pilot project success will be critical in convincing conservative physicians, a skeptical medical community and an often bureaucratic and herd-driven healthcare industry as to the potential of P4 medicine. P4MI is also interested in bringing relevant industrial partners to this clinical network. In addition, P4MI has established a Fellows Program to begin delving into some of the societal challenges of P4 medicine, eventually expressing these issues through 'white papers' on economics, the 'gold standard of healthcare information', ethics, regulations, etc. Strategic partnerships are a critical component for bringing P4 to patients worldwide and gaining its widespread acceptance in the national and international medical communities.

As the P4MI pilot programs become successful and begin demonstrating the power of P4 medicine, we would like next to persuade a small country to build a P4 medicine/healthcare system. This country could play a key leadership role in pioneering the new medicine of the 21st century, just as Johns Hopkins Medical School in the early 1900s adopted some recommendations of the Flexner Report on the future of medicine (sponsored by the Carnegie Foundation), which included the recommendation that medicine should integrate basic research and clinical medicine. Thus Johns Hopkins propelled itself from a mediocre medical trade school into a world leader in US and world medicine. Now there is now a similar opportunity for institutions (and countries) that pioneer P4 medicine/healthcare to become world leaders. It will take courage, leadership, resources and an effective communication of the vision to the stakeholders to catalyze the revolution in P4 medicine/healthcare. It goes without saying that any nation that assumes a leading-edge leadership role in catalyzing the emergence of P4 healthcare will be in a unique position to transform the healthcare of its citizens, to revolutionize its medical research agenda and to take advantage of the associated economic opportunities associated with the newly emerging world of P4 medicine.

ACKNOWLEDGMENTS

Thanks to Lee Rowen for advice and counsel in preparing this paper and Jaeyun Sung for help with Figure 23.9. This paper was in part

adapted from a paper now in press – L Hood and M Flores; Systems Medicine and the Emergence of Proactive P4 Medicine: Predictive, Preventive, Personalized and Participatory; New Biotechnology, in press. LH would like to acknowledge the support of the Luxembourg Centre for Systems Biomedicine and the University of Luxembourg, the General Medical Sciences Center for Systems Biology GM076547 and a Department of Defense contract on Liver Toxicity W911SR-09-C-0062. NDP acknowledges funding from an NIH-NCI Howard Temin Pathway to Independence Award in Cancer Research, a Roy J. Carver Young Investigator Grant, and the Camille Dreyfus Teacher–Scholar program.

REFERENCES

[1] Hood L. Deciphering Complexity: A personal view of systems biology and the coming of 'Big' science. Genet Eng News 2011;31:131.

[2] Barabasi AL, Oltvai ZN. Network biology: understanding the cell's functional organization. Nat Rev Genet 2004;5:101–13.

[3] Hood L, Heath JR, Phelps ME, Lin B. Systems biology and new technologies enable predictive and preventative medicine. Science 2004;306:640–3.

[4] Patrinos GP, Brookes AJ. DNA, diseases and databases: disastrously deficient. Trends Genet 2005;21:333–8.

[5] Loscalzo J, Kohane I, Barabasi AL. Human disease classification in the postgenomic era: a complex systems approach to human pathobiology. Mol Syst Biol 2007;3:124.

[6] Auffray C, Chen Z, Hood L. Systems medicine: the future of medical genomics and healthcare. Genome Med 2009;1:2.

[7] Price N, Edelman L, Lee I, Yoo H, Hwang D, Carlson G, et al. Systems biology and the emergence of systems medicine. Genomic and Personalized Medicine: From Principles to Practice 2009;1:131–41.

[8] Buell J. The Digital Medicine Revolution in Healthcare. American College of Healthcare Executives, Chicago 2011.

[9] Garg V, Arora S, Gupta C. Cloud computing approaches to accelerate drug discovery value chain. Comb Chem High Throughput Screen 2011;14:861–71.

[10] Topol E. The Creative Destruction of Medicine: How the Digital Revolution Will Create Better Health Care, a Member of the Perseus Books Group. New York: Basic Books; 2012.

[11] Collino S, Martin FP, Rezzi S. Clinical metabolomics paves the way towards future healthcare strategies. Br J Clin Pharmacol 2012. doi: 10.1111/j.1365-2125.2012.04216.x.

[12] Smarr L. Quantified Health: A 10-year detective story of the digitally enabled genomic medicine. Strateg News Lett 2011;14:1–33.

[13] Weston AD, Hood L. Systems biology, proteomics, and the future of health care: toward predictive, preventative, and personalized medicine. J Proteome Res 2004;3:179–96.

[14] Hood L, Friend SH. Predictive, personalized, preventive, participatory (P4) cancer medicine. Nat Rev Clin Oncol 2011;8:184–7.

[15] Tian Q, Price ND, Hood L. Systems cancer medicine: towards realization of predictive, preventive, personalized and participatory (P4) medicine. J Intern Med 2012;271:111–21.

[16] Hwang D, Lee IY, Yoo H, Gehlenborg N, Cho JH, Petritis B, et al. A systems approach to prion disease. Mol Syst Biol 2009;5:252.

[17] Hidalgo CA, Blumm N, Barabasi AL, Christakis NA. A dynamic network approach for the study of human phenotypes. PLoS Comput Biol 2009;5:e1000353.

[18] Kinross JM, Darzi AW, Nicholson JK. Gut microbiome–host interactions in health and disease. Genome Med 2011;3:14.

[19] Williams A, Smith JR, Allaway D, Harris P, Liddell S, Mobasheri A. Applications of proteomics in cartilage biology and osteoarthritis research. Front Biosci 2012;17:2622–44.

[20] Janecka I. Is U.S. health care an appropriate system? A strategic perspective from systems science. Health Res Policy Syst 2009;7.

[21] Christakis NA, Fowler JH. The spread of obesity in a large social network over 32 years. N Engl J Med 2007;357:370–9.

[22] Hood L. A personal journey of discovery: developing technology and changing biology. Annu Rev Anal Chem (Palo Alto Calif) 2008;1:1–43.

[23] Hood L. Acceptance remarks for Fritz J. and Delores H. Russ Prize. NAE Journal The Bridge 2011;41:46–9.

[24] Hood L, Rowen L, Galas DJ, Aitchison JD. Systems biology at the institute for systems biology. Brief Funct Genomic Proteomic 2008;7:239–48.

[25] Chen R, Mias GI, Li-Pook-Than J, Jiang L, Lam HY, Chen R, et al. Personal omics profiling reveals dynamic molecular and medical phenotypes. Cell 2012;148:1293–307.

[26] Lazebnik Y. Can a biologist fix a radio? Or, what I learned while studying apoptosis. Biochemistry 2004;69:1403–6.

[27] Ravasi T, Suzuki H, Cannistraci CV, Katayama S, Bajic VB, Tan K, et al. An atlas of combinatorial transcriptional regulation in mouse and man. Cell 2010;140:744–52.

[28] Keller A, Backes C, Leidinger P, Kefer N, Boisguerin V, Barbacioru C, et al. Next-generation sequencing identifies novel microRNAs in peripheral blood of lung cancer patients. Mol Biosyst 2011;7:3187–99.

[29] Pinzani P, Salvianti F, Zaccara S, Massi D, De Giorgi V, Pazzagli M, et al. Circulating cell-free DNA in plasma of melanoma patients: qualitative and quantitative considerations. Clin Chim Acta 2011;412:2141–5.

[30] Schwarzenbach H, Hoon DS, Pantel K. Cell-free nucleic acids as biomarkers in cancer patients. Nat Rev Cancer 2011;11:426–37.

[31] Chapman MH, Sandanayake NS, Andreola F, Dhar DK, Webster GJ, Dooley JS, Pereira SP. Circulating CYFRA 21–1 is a specific diagnostic and prognostic biomarker in biliary tract cancer. J Clin Exp Hepatol 2011;1:6–12.

[32] Matsusaka S, Suenaga M, Mishima Y, Kuniyoshi R, Takagi K, Terui Y, et al. Circulating tumor cells as a surrogate marker for determining response to chemotherapy in Japanese patients with metastatic colorectal cancer. Cancer Sci 2011;102:1188–92.

[33] Roach JC, Glusman G, Smit AF, Huff CD, Hubley R, Shannon PT, et al. Analysis of genetic inheritance in a family quartet by whole-genome sequencing. Science 2010;328:636–9.

[34] Chiu CL, Randall S, Molloy MP. Recent progress in selected reaction monitoring MS-driven plasma protein biomarker analysis. Bioanalysis 2009;1:847–55.

[35] Yao M, Ma L, Duchoslav E, Zhu M. Rapid screening and characterization of drug metabolites using multiple ion monitoring dependent product ion scan and postacquisition data mining on a hybrid triple quadrupole-linear ion trap mass spectrometer. Rapid Commun Mass Spectrom 2009;23:1683–93.

[36] Ma S, Funk CC, Price ND. Systems approaches to molecular cancer diagnostics. Discov Med 2010;10:531–42.

[37] Bartfai T, Buckley PT, Eberwine J. Drug targets: single-cell transcriptomics hastens unbiased discovery. Trends Pharmacol Sci 2012;33:9–16.
[38] Liu L, Ye Q, Wu Y, Hsieh WY, Chen CL, Shen HH, et al. Tracking T-cells in vivo with a new nano-sized MRI contrast agent. Nanomedicine 2012; http://dx.doi.org/10.1016/j.nano.2012.02.017.
[39] Heath JR, Davis ME, Hood L. Nanomedicine targets cancer. Sci Am 2009;300:44–51.
[40] Chakraborty M, Jain S, Rani V. Nanotechnology: emerging tool for diagnostics and therapeutics. Appl Biochem Biotechnol 2011;165:1178–87.
[41] Shi Q, Qin L, Wei W, Geng F, Fan R, Shin YS, et al. Single-cell proteomic chip for profiling intracellular signaling pathways in single tumor cells. Proc Natl Acad Sci U S A 2012;109:419–24.
[42] Hodgson J. Gene sequencing's industrial revolution. Spectrum 2000;37:36–42.
[43] International human genome sequencing consortium. Finishing the euchromatic sequence of the human genome. Nature 2004;431:931–45.
[44] Oxford Nanopore Press release. DNA 'Strand Sequencing' on the High-Throughput GridION Platform and Presents MinION, a Sequencer the Size of a USB Memory Stick. Oxford: Oxford Nanopore; 2012.
[45] American Medical Association (AMA) Proteomics, http://www.ama-assn.org; 2012.
[46] Juritz EI, Alberti SF, Parisi GD. PCDB: a database of protein conformational diversity. Nucleic Acids Res 2011;39:D475–479.
[47] Moritz R. Institute for systems biology. Pers Commun 2012.
[48] Connor AC, McGown LB. Aptamer stationary phase for protein capture in affinity capillary chromatography. J Chromatogr A 2006;1111:115–9.
[49] Zichel R, Chearwae W, Pandey GS, Golding B, Sauna ZE. Aptamers as a sensitive tool to detect subtle modifications in therapeutic proteins. PLoS One 2012;7:e31948.
[50] Rakhila H, Rozek T, Hopkins A, Proudman S, Cleland L, James M, et al. Quantitation of total and free teriflunomide (A77 1726) in human plasma by LC-MS/MS. J Pharm Biomed Anal 2011;55:325–31.
[51] Sung J. Molecular signatures from omics data: from chaos to consensus. Biotechnol J 2012;7(8)946–57.
[52] Leek JT, Scharpf RB, Bravo HC, Simcha D, Langmead B, Johnson WE, et al. Tackling the widespread and critical impact of batch effects in high-throughput data. Nat Rev Genet 2010;11:733–9.
[53] Ein-Dor L, Zuk O, Domany E. Thousands of samples are needed to generate a robust gene list for predicting outcome in cancer. Proc Natl Acad Sci U S A 2006;103:5923–8.
[54] Quackenbush J. Microarray analysis and tumor classification. N Engl J Med 2006;354:2463–72.
[55] Sirota M, Dudley JT, Kim J, Chiang AP, Morgan AA, Sweet-Cordero A, et al. Discovery and preclinical validation of drug indications using compendia of public gene expression data. Sci Transl Med 2011;3:96ra77.
[56] Hibbard JH, Stockard J, Mahoney ER, Tusler M. Development of the Patient Activation Measure (PAM): conceptualizing and measuring activation in patients and consumers. Health Serv Res 2004;39:1005–26.
[57] Pleasance ED, Cheetham RK, Stephens PJ, McBride DJ, Humphray SJ, Greenman CD, et al. A comprehensive catalogue of somatic mutations from a human cancer genome. Nature 2010;463:191–6.
[58] Dancey JE, Bedard PL, Onetto N, Hudson TJ. The genetic basis for cancer treatment decisions. Cell 2012;148:409–20.
[59] Bonneau R, Facciotti MT, Reiss DJ, Schmid AK, Pan M, Kaur A, et al. A predictive model for transcriptional control of physiology in a free living cell. Cell 2007;131:1354–65.
[60] Barabási A-L. N Engl J Med 2007;357(4):404–7.