

Review

Systems Approaches to Biology and Disease Enable Translational Systems Medicine

Leroy Hood*, Qiang Tian*

Institute for Systems Biology, Seattle, WA 98109-5234, USA

Received 31 July 2012; revised 9 August 2012; accepted 14 August 2012

Available online 23 August 2012

Abstract

The development and application of systems strategies to biology and disease are transforming medical research and clinical practice in an unprecedented rate. In the foreseeable future, clinicians, medical researchers, and ultimately the consumers and patients will be increasingly equipped with a deluge of personal health information, *e.g.*, whole genome sequences, molecular profiling of diseased tissues, and periodic multi-analyte blood testing of biomarker panels for disease and wellness. The convergence of these practices will enable accurate prediction of disease susceptibility and early diagnosis for actionable preventive schema and personalized treatment regimes tailored to each individual. It will also entail proactive participation from all major stakeholders in the health care system. We are at the dawn of predictive, preventive, personalized, and participatory (P4) medicine, the fully implementation of which requires marrying basic and clinical researches through advanced systems thinking and the employment of high-throughput technologies in genomics, proteomics, nanofluidics, single-cell analysis, and computation strategies in a highly-orchestrated discipline we termed translational systems medicine.

Keywords: Systems biology; P4 medicine; Family genome sequencing; Targeted proteomics; Single-cell analysis

Introduction

Systems biology strives to unravel the enormous complexity of biological systems through a holistic approach in the context of a cross-disciplinary environment. Since its founding in early 2000, the Institute for Systems Biology (ISB) has been pioneering systems strategies to biology and disease through the development of systems strategies and the application and/or development of cutting-edge high-throughput technologies to the investigation of model organisms and humans with varying degrees of complexity: from single-cell organisms (bacteria and yeast) [1–3] to experimental animal models (mouse) [4–7] and to human disorders [8–10]. Over the last decade, rapid advancements in genomic and proteomic technologies, computational

strategies and their applications in human diseases have demonstrated promising early success in genomic medicine. We discuss here our view of how systems approaches to biology and disease and emerging technologies are going to transform the medical practices by shaping up translational systems medicine for early diagnosis, disease progression, patient stratification, predicting recurrence, and therapeutic guidance.

Dealing with disease complexity—systems medicine and its 5 pillars

Human phenotypes are specified by two types of biological information: the digital information of the genome, and the environmental information that impinges upon and modifies the digital information. Two general biological structures connect the genotype and environment to phenotype: (1) biological networks capture, transmit, process and pass on information; these networks organize,

* Corresponding authors.

E-mail: Leroy.Hood@systemsbiology.org (Hood L), Qiang.Tian@systemsbiology.org (Tian Q).

integrate and model data to enormously increase the signal to noise; (2) simple and complex molecular machines execute biological functions. A systems view of disease postulates that disease arises from disease-perturbed networks. A ramification of this premise entails studies of disease pathogenesis at the network level through a systems approach so that better strategies for early diagnosis and therapeutics targeting these perturbed networks can be devised. We stipulate five pillars to address disease complexity upholding systems approach as follows.

- (1) Viewing biology and consequentially medicine as an informational science is one key to deciphering complexity.
- (2) Systems biology infrastructure and strategy—holy trinity of biology (*i.e.*, use biology to drive technology and computation development)—endorse cross-disciplinary culture and democratization of data-generation and data-analysis tools.
- (3) Holistic, systems experimental approaches enable deep insights into disease mechanisms and new approaches to diagnosis and therapy through analyzing the dynamics of disease processes.
- (4) Emerging technologies provide large-scale data acquisition and permit exploration of new dimensions of patient data space.
- (5) Transforming analytic tools will allow deciphering the billions of data points for each individual—sculpting in exquisite detail the wellness and disease landscapes.

These five fundamental principles will allow in-depth interrogation of diseased networks at unprecedented molecular resolution. Some disease events will occur well before the disease manifestation for early detection, whereas key nodal points amongst perturbed networks can be identified for diagnostic detection or therapeutic interventions. Both diseased organs/tissues and patient blood constitute excellent specimen reservoirs for systemic assessment of diseased conditions in multiple spatial and temporal measurements. Whole genome and whole tran-

scriptome sequencing, targeted proteomics via mass spectrometry and protein chips, single-cell analysis and a variety of targeted nucleic acid detection systems (*e.g.*, next-generation sequencing (NGS), DNA arrays, NanoString *n*-Counter [11], Fluidigm BioMark, *etc.*) will be the workhorse churning out enormous amount of data. We anticipate that in 10 years each individual will be surrounded by a virtual cloud of billions of data points. A key challenge is to fully integrate these diverse data type, correlate with distinct clinical phenotypes, extract meaningful biomarker panels for guiding clinical practice. We enumerate here some of the individual patient information-based assays of the present and future (**Table 1**).

Family genome sequencing: integrating genetic and genomics

Complete human genome sequence is becoming increasingly affordable and will be a fundamental part of one's medical record in 10 years. While a great deal can be learned regarding one's predisposition to certain diseases from individual genome, sequencing of a family permit one to use the principles of Mendelian genetics to eliminate 70% sequencing error. This will greatly facilitate better identification of rare variants, determining chromosomal haplotypes and intergenerational mutation rate, and identification of candidate genes for simple Mendelian diseases. Moreover, knowledge of *cis* and *trans* linkage relationships of genes and control elements will be key for understanding biology and disease, and reducing the chromosomal search space for disease genes [9,12]. Recent developments by Complete Genomics Inc (CGI) employing long fragment reads (LFR) have demonstrated whole-genome sequencing from as few as 10–20 cells with three striking advances over typical NGS approach. These advances include (1) high accuracy with a genome error rate of 1 in 10 megabases; (2) assembly of diploid haplotypes from individual genome sequences; and (3) *de novo* assembly of individual genomes, which enables discovery of structural variations [13]. With this technology, comprehensive genetic studies and diverse clinical applications are within reach.

Table 1 Clinical assays and emerging technologies for exploring new dimensions of patient data space

Genomics

Complete individual genome sequences will be done by sequencing families—predictive health history

Complete individual cell genome sequences—cancer

Complete MHC chromosomal haplotypes in families—autoimmune disease and allergies

300 Actionable gene variants—pharmacogenetics-related and disease-related genes

Sequence 1000 transcriptomes—tissues and single cells—stratification disease

Analyze aging transcriptome profiles—tissues and single cells—wellness

Analyze miRNA profiles—tissues, single cells and blood—disease diagnosis

Proteomics

Organ-specific blood SRM protein assays

2500 Blood organ-specific blood proteins from 300 nanoliters of blood in 5 min—twice per year (50 proteins from 50 organs)—wellness assessment

New protein capture agents—D-amino acid peptides joined to create dimer or trimer capture agents

Array of 12,000 human proteins—against autoimmune or allergic sera—stratify—diseases that kill cells (neurodegenerative)

Single molecule protein analyses—blood organ-specific proteins and single cell analyses

SWATH™ analyses—global, dynamical analyses

Systems approach to blood biomarkers: making blood a window into health and disease

Since blood bathes all organs and receives their biomarkers, it shall reflect network disease-perturbations either directly or indirectly—a molecular fingerprint in the blood reflecting disease pathophysiology. We stress that organ-specific, cell-type specific or organelle-specific biomarkers are more informative since they inform as to the tissue, cell type or organelle sources of the disease. Moreover, blood biomarkers may also reflect general cell death or damage (*e.g.*, biomolecules released from nucleus or cytoplasm), secreted protein or membrane perturbations through proteolysis. Systems blood biomarkers shall include diverse types of biomolecules: proteins, mRNAs, non-coding RNAs (*e.g.*, microRNAs, long intergenic non-coding RNAs), metabolites, etc, while the combination of two or more types increases sensitivity and specificity of assay. These markers should be multiparameter consisting of many biomolecules of the same type, and even panels of multiple types of molecules so that multiple networks and features may be accessed. Ideally, blood biomarker panel shall assess all diseases in a given organ simultaneously. Another important point is that, given the vast individual variation, blood biomarkers should be analyzed in a longitudinal manner—so that the individual can be their own control against which change can be measured. Of note, another information-rich compartment in the blood includes the cellular component, *e.g.*, the peripheral blood mononuclear cells (PBMCs). These PBMCs contain mainly white blood cells (WBCs) for diagnosing inflammation, immunity and cell death; they also contain rare circulating tumor cells (CTCs) in cancer patients, indicative of tumor progression and recurrence [14,15].

Our method of choice for evaluating blood protein biomarkers is targeted proteomics employing selective reaction monitoring (SRM) mass spectrometry (MS) [3]. This technology allows the analysis of 100–200 proteins quantitatively in 1 h. ISB has developed SRM assays for most of the known 20,333 human proteins. In particular, we have validated SRM assays for 100 brain-specific and 100 liver-specific proteins for human and mouse [16]. These protein panels have been applied in mouse disease models and patient blood samples for successful identification of biomarkers for the diagnosis of liver injury, liver fibrosis/cirrhosis, prion and other neurological diseases. For instance, we identified a panel of 15 brain-specific blood proteins that indicate the initiation and progression of disease-perturbation of networks (prion accumulation, glial activation, synaptic degeneration, and neuronal cell death) in a mouse model of prion disease [4]. A panel of three liver-specific proteins successfully stratify liver cirrhosis patients from patients with various degree of liver fibrosis and normal controls [16]. The same strategy is being actively pursued for the identification of brain tumor cell membrane protein biomarker in the blood (unpublished data).

While it is conceivable to set up a SRM-MS infrastructure to provide blood diagnostics to serve clinical needs for a variety of diseased conditions as discussed above, this requires highly-sophisticated expertise in MS instrumentation and supporting informatics capacities. The company Integrated Diagnostics is pursuing a systems approach to diagnostics for selected disease applications. An alternative is to develop targeted protein and antibody chips or chips of protein-catalyzed capture (PCC) agents. The latter demonstrates advantages since it is chemically-stable, low cost, and requires relatively little input of blood samples. In addition, we are developing a protein Elisa assay on the NanoString *n*-Counter instrument, in conjunction with their capacity to detect mRNA and miRNA molecules, to generate an assay that combines multiple analytes (mRNA, miRNA, and protein) in a single platform with no loss in sensitivity. We envision that in a 10-year future, an integrated nanotech/microfluidics platform, consisting of 50 organ-specific blood proteins from each of 50 major human organs, will measure 2500 blood proteins using a fraction of droplet of blood in 5 min at the mid amol level of sensitivity. The prototype of this nanochip has already been tested in hospitals [17,18].

Single-cell analysis allows interrogation of heterogeneous cell populations at unprecedented resolution

Most of the current global molecular profiling studies measure mixed diseased cell populations for averaged signals. However, there are distinct cell types in any given diseased tissues each with its own distinct perturbed genomic and proteomic profiles. Although global genome and transcriptome sequencing for single cell is still challenging, early efforts have already revealed important population heterogeneity in tumor cells [19,20]. We envision that more single-cell analysis will be applied clinically. For instance, one can analyze 10,000 B cells and 10,000 T cells for the functional regions of their immune receptors to inform past and present immune responsiveness, follow vaccinations, and identify autoimmune antibodies. Single-cell analysis can also be applied concomitantly with various technologies for separating epithelial cells from WBCs in blood, for identifying and monitoring of CTCs. Single-cell transcriptome analysis can also be applied to quantize cell populations in cancer tissues and differentiating progenies of stem cells.

Systems medicine is transforming healthcare leading to predictive, preventive, personalized and participatory (P4) medicine

Systems medicine provides fundamental insights into disease network mechanisms to enable diagnosis, therapy and prevention for the individual patient (Figure 1). Family genome sequencing reveals disease and wellness genes and actionable genes. Transforming blood into a window to distinguish health from disease opens up new way for

Systems medicine: network of networks

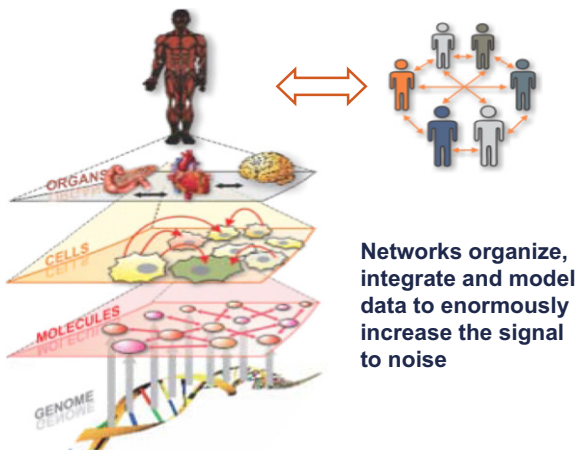


Figure 1 Networks organize and integrate information at different levels to create biologically meaningful models

Networks formulate hypotheses about biological function and provide temporal and spatial insights into dynamical changes.

disease diagnostics, and assessment of drug toxicity and wellness. Molecular profiling stratifies diseases into their distinct molecular subtypes for impedance match with appropriate drugs. New approaches to drug target discovery are being devised—re-engineer disease-perturbed networks with drugs for faster and cheaper drug development.

The convergence of the digital revolution and systems medicine leads to deciphering of complexity and P4 medicine

- (1) Predictive: the probabilistic health history is revealed by DNA sequence and regular multi-parameter (blood) measurements.
- (2) Preventive: design of therapeutic and preventive drugs and vaccines via systems approaches; emphasis on wellness.
- (3) Personalized: unique individual human genetic variation mandates individual treatment and that patient will be their own control for data analyses.
- (4) Participatory: patient-driven social networks for disease and wellness will be a driving force in P4 medicine. Society must access patient data and make it available to biologists for pioneering predictive medicine of the future. How does one educate patients, physicians and the healthcare community about P4? The answer is IT for healthcare.

P4 medicine differs from evidence-based medicine in that it is proactive, individualized, with an emphasis not only on disease, but also on wellness. It involves generation, mining and integration of enormous amounts of data on individual patients to produce predictive and actionable models of wellness and disease. Large patient populations will be analyzed at single individual level (not population averages) to generate quantized stratification of patient populations and

create the predictive medicine of the future. It entails patient-driven social networks.

There are several societal implications for P4 medicine. It forces a revision of business plans of almost every sector of healthcare industry, producing enormous economic opportunity. Digitalization of medicine for the individual patients is a larger revolution than the digitization of information technologies and communication in that it is patient-driven medicine and wellness. It turns sharply around escalating costs of healthcare—democratization of healthcare through (1) early blood diagnosis; (2) benefits of wellness—*e.g.*, survey biannually 2500 blood organ-specific protein measurements (50 from each of the 50 major organs) for global early detection of the transition from health to disease; (3) digital technologies exponentially increasing in measurement potential and decreasing in cost—sculpt for individuals the dimensions of health/disease while dramatically decreasing measurement costs, *e.g.*, sequencing a human genome cost about \$300 million dollars in 2000 but only about \$3000 in 2012—a 100,000-fold decrease in cost—for digitalization of medicine. Eventually, P4 medicine will create significant wealth.

Translational systems medicine should practice proactive P4 medicine

A core mission of ISB is to disseminate systems approaches to biology and medicine to the society by and large. ISB has formed strategic partnerships with Ohio State University, Peace Health, and the State of Luxembourg to promote the practice of P4 medicine. We propose that any institutions wishing to establish a translational systems medicine program shall adopt the five pillars of a systems approach to disease: an informational view of biology and disease, a cross-disciplinary infrastructure, global experimental systems approaches to capture the dynamics of disease, employ of emerging technologies to search new areas of patient data space and powerful novel analytical tools to handle all the new data generated. They shall partner with institutions who have systems biology, systems medicine and P4 medicine expertise who can guide, teach and help recruit leadership who understands systems medicine and translational opportunities. Committed political and scientific leadership at both local and national levels are also indispensable.

Competing interests

The authors have declared that no competing interests exist.

Acknowledgements

We gratefully acknowledge funding from the Grand Duchy of Luxembourg, NIH/NCI NanoSystems Biology Cancer Center (Grant No. U54 CA151819A), NIH/NIGMS

Center for Systems Biology (Grant No. P50GM076547) and NIH/NIAMSD (Grant No. RC2AR059010).

References

- [1] Ideker T, Thorsson V, Ranish JA, Christmas R, Buhler J, Eng JK, et al. Integrated genomic and proteomic analyses of a systematically perturbed metabolic network. *Science* 2001;292:929–34.
- [2] 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.
- [3] Picotti P, Bodenmiller B, Mueller LN, Domon B, Aebersold R. Full dynamic range proteome analysis of *S. cerevisiae* by targeted proteomics. *Cell* 2009;138:795–806.
- [4] 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.
- [5] Park H, Li Z, Yang XO, Chang SH, Nurieva R, Wang YH, et al. A distinct lineage of CD4 T cells regulates tissue inflammation by producing interleukin 17. *Nat Immunol* 2005;6:1133–41.
- [6] Tian Q, Feetham MC, Tao WA, He XC, Li L, Aebersold R, et al. Proteomic analysis identifies that 14–3-3zeta interacts with beta-catenin and facilitates its activation by Akt. *Proc Natl Acad Sci USA* 2004;101:15370–5.
- [7] Tian Q, Stepaniants SB, Mao M, Weng L, Feetham MC, Doyle MJ, et al. Integrated genomic and proteomic analyses of gene expression in Mammalian cells. *Mol Cell Proteomics* 2004;3:960–9.
- [8] Majeti R, Becker MW, Tian Q, Lee TL, Yan X, Liu R, et al. Dysregulated gene expression networks in human acute myelogenous leukemia stem cells. *Proc Natl Acad Sci USA* 2009;106:3396–401.
- [9] 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.
- [10] Yan X, Ma L, Yi D, Yoon JG, Diercks A, Foltz G, et al. A CD133-related gene expression signature identifies an aggressive glioblastoma subtype with excessive mutations. *Proc Natl Acad Sci USA* 2011;108:1591–6.
- [11] Geiss GK, Bumgarner RE, Birditt B, Dahl T, Dowidar N, Dunaway DL, et al. Direct multiplexed measurement of gene expression with color-coded probe pairs. *Nat Biotechnol* 2008;26:317–25.
- [12] Roach JC, Glusman G, Hubley R, Montsaroff SZ, Holloway AK, Mauldin DE, et al. Chromosomal haplotypes by genetic phasing of human families. *Am J Hum Genet* 2011;89:382–97.
- [13] Peters BA, Kermani BG, Sparks AB, Alferov O, Hong P, Alexeev A, et al. Accurate whole-genome sequencing and haplotyping from 10 to 20 human cells. *Nature* 2012;487:190–5.
- [14] Lin EH, Hassan M, Li Y, Zhao H, Nooka A, Sorenson E, et al. Elevated circulating endothelial progenitor marker CD133 messenger RNA levels predict colon cancer recurrence. *Cancer* 2007;110:534–42.
- [15] Iinuma H, Watanabe T, Mimori K, Adachi M, Hayashi N, Tamura J, et al. Clinical significance of circulating tumor cells, including cancer stem-like cells, in peripheral blood for recurrence and prognosis in patients with Dukes' stage B and C colorectal cancer. *J Clin Oncol* 2011;29:1547–55.
- [16] Qin S, Zhou Y, Lok AS, Tsodikov A, Yan X, Gray L, et al. SRM targeted proteomics in search for biomarkers of HCV-induced progression of fibrosis to cirrhosis in HALT-C patients. *Proteomics* 2012;12:1244–52.
- [17] Fan R, Vermesh O, Srivastava A, Yen BK, Qin L, Ahmad H, et al. Integrated barcode chips for rapid, multiplexed analysis of proteins in microliter quantities of blood. *Nat Biotechnol* 2008;26:1373–8.
- [18] Ma C, Fan R, Ahmad H, Shi Q, Comin-Anduix B, Chodon T, et al. A clinical microchip for evaluation of single immune cells reveals high functional heterogeneity in phenotypically similar T cells. *Nat Med* 2011;17:738–43.
- [19] Navin N, Kendall J, Troge J, Andrews P, Rodgers L, McIndoo J, et al. Tumour evolution inferred by single-cell sequencing. *Nature* 2011;472:90–4.
- [20] Dalerba P, Kalisky T, Sahoo D, Rajendran PS, Rothenberg ME, Leyrat AA, et al. Single-cell dissection of transcriptional heterogeneity in human colon tumors. *Nat Biotechnol* 2011;29:1120–7.



Leroy Hood, MD, PhD, president and co-founder of the Institute for Systems Biology in Seattle, is a pioneer in systems approaches to biology and medicine. He and others developed the DNA sequencer and synthesizer, and the protein synthesizer and sequencer—four instruments that paved the way for the successful mapping of the human genome. Dr Hood's research has focused on the study of molecular immunology, biotechnology and genomics. He is a member of the National Academy of Sciences, the National Academy of Engineering, and the Institute of Medicine.



Qiang Tian, MD, PhD, Cancer & Stem Cell Group Leader of Institute for Systems Biology, is interested in applying the powerful systems approach with the enabling genomics, proteomics, and single cell analysis technologies to address some of the most pressing issues pertaining to human health. He has led the development of gene signature panels for cancer patient stratification, and has elucidated protein interaction networks for potential therapeutic targeting. He also contributed to the molecular characterization of multiple Th cell subsets.