OPINION

Predictive, personalized, preventive, participatory (P4) cancer medicine

Leroy Hood and Stephen H. Friend

Abstract | Medicine will move from a reactive to a proactive discipline over the next decade—a discipline that is predictive, personalized, preventive and participatory (P4). P4 medicine will be fueled by systems approaches to disease, emerging technologies and analytical tools. There will be two major challenges to achieving P4 medicine —technical and societal barriers—and the societal barriers will prove the most challenging. How do we bring patients, physicians and members of the health-care community into alignment with the enormous opportunities of P4 medicine? In part, this will be done by the creation of new types of strategic partnerships—between patients, large clinical centers, consortia of clinical centers and patient-advocate groups. For some clinical trials it will necessary to recruit very large numbers of patients—and one powerful approach to this challenge is the crowd-sourced recruitment of patients by bringing large clinical centers together with patient-advocate groups.

Hood, L. & Friend, S. H. *Nat. Rev. Clin. Oncol.* **8**, 184–187 (2011); doi:10.1038/nrclinonc.2010.227

Introduction

The term 'P4 medicine' was first discussed more than 5 years ago by one of us (Leroy Hood)[1,2] to denote an ongoing revolution in medicine—moving it from a reactive to a proactive discipline—where ultimately the objective is to maximize wellness for each individual rather than simply to treat disease. P4 medicine is used to describe a systems approach to medicine that includes predictive, personalized, preventive and participatory aspects.[3,4] This revolution is being fueled by several factors: first, an appreciation that medicine is an information science; second, systems or holistic approaches to studying the enormous complexities of disease; third, emerging technologies that will let us explore new dimensions of patient data space; and fourth, powerful new analytical technologies—both mathematical and computational—that will let us decipher the billions of data points associated with each individual. Patients integrate two general types of information—inherent 'digital' information from the genome, and environmental information coming from their surroundings. It is interesting to note that the microbiomes present in our gut and other epithelial surfaces can have a major role in the wellness–disease balance; of course, understanding the effect of these microbiomes on tumors will require analyzing their genomic as well as environmental information.[5] One should also point out that the germline DNA sequence of an individual is markedly different from the heterogeneous somatic genome sequences of cancer cells owing to the high mutation rates in neoplastic cells—hence cancer genomes must be distinguished from germline genomes.[6] The combined influences of these factors trigger fundamental life processes from embryonic development to physiological and behavioral responses and include disease initiation and progression. This digital and environmental information generates each individual's phenotype—molecular, cellular and descriptive features—through networks of biological pathways that capture, transmit, and integrate signals and, finally, send instructions to the molecular machines that execute the functions of life. The focus of systems (P4) medicine is ultimately to completely understand functions of these biological networks and molecular machines in disease.[3,4] Therefore, the P4 description is used to denote a revolution that extends far beyond what is usually covered by the term personalized medicine. The challenges in the implementation of P4 medicine are twofold: technical and societal. Ultimately, we believe that the societal challenges will prove more difficult to overcome.

P4 medicine is more than 'genomic medicine' because it requires the integration of diverse data from the various hierarchical levels of biological information—DNA (and epigenetic changes), RNA, proteins, metabolites, networks, cells, and tissues—to overcome the enormous signal to noise challenges of large data sets arising from both technical errors as well as reflecting many different aspects of biology that are irrelevant to the focus of particular studies,[7] and ultimately create effective predictive and actionable models that can be used to guide the treatment of the patient. Predictive medicine will employ the information arising from personal genome sequences and longitudinal molecular, cellular and phenotypic measurements to provide a baseline measurement that can be defined as healthy (or as wellness) and then employed to identify subsequent transitions to disease. One critical aspect of most complex diseases, for example breast cancer, is that the term encompasses many different types of the disease—as defined by differing combinations of disease-perturbed networks. Hence the stratification of disease is critical to choosing the correct and, presumably, different treatments for each disease type. Personalized medicine is a challenge because each of our genomes differ by approximately 6 million nucleotides, this large difference in individuals ultimately requires the use of the patient's own baseline healthy data as a 'control' that can then be used for subtractive analyses of longitudinal accumulations of molecular and cellular information that signal the transitions from wellness to disease. Preventive medicine can become a reality as the details of disease-perturbed molecular networks open the possibility of using drugs targeted at key nodal points to deter or stop disease progression. Indeed, applying the P4 medicine concept will facilitate a shift in health care from a focus on disease to an emphasis on wellness. Finally, the participatory aspect of P4 medicine indicates that there are cultural as well as technical challenges in its implementation—how do we educate patients, physicians, and the health-care community about the power of

Competing interests
The authors declare no competing interests.

© 2011 Macmillan Publishers Limited. All rights reserved

P4 medicine, how do we obtain the data we need for predictive medicine as described below and how do we create an information technology system for health care that will permit us to deal with the exponentially accelerating accumulation of patient data that within 10 years will likely generate a 'virtual cloud' of billions of data points that will define the health (or disease) of each individual? This information technology for health care must reduce this enormous quantity of data to simple hypotheses about wellness and disease for the individual.

Our vision is that, in the not-too-distant future, each patient will be surrounded by a 'virtual cloud' of billions of data points that will uniquely define their past medical history and current health status. Furthermore, it will be possible to mine the billions of data points from hundreds of millions of individuals to generate algorithms to help predict the future clinical needs for each patient. A systems approach to disease will lead to powerful new diagnostics and therapeutics and provide invaluable new insights into prevention.[3,4] Key to the use of these data, and also the primary challenge, will be to mine and integrate these data in the context of the dynamics of biological networks and molecular machines and to construct models of wellness and disease that are both predictive and actionable—and therefore useful to patients and clinicians.

P4 medicine has the potential to catalyze a sharp turnaround in the ever-increasing costs of medical care and, in fact, reduce costs to the point where P4 medicine will be exported to the developing world. Indeed, in our opinion, P4 medicine will be the foundation of global health in the future. This reduction in cost would arise from a variety of different factors. The digitalization of medicine will arise from emerging transformational technologies (such as DNA sequencing with exponentially increasing throughput and decreasing costs) that will generate patient-relevant 'digital' information from a single molecule, single cell, or single tissue—indeed, any of the organized informational units in the hierarchy of life.[8] This digitalization will have the same implications as the digitalization of communications and information technology, that is, huge reductions in the costs of services in patient health care. One of the central challenges of modern medicine is taking complex diseases, such as breast cancer,[9] and stratifying them into distinct subtypes (each with differing combinations of disease-perturbed networks).[10] Systems approaches to blood, tissue and single-cell diagnostics will permit the stratification of these complex diseases, which will revolutionize the ability of the pharmaceutical industry to find novel drug targets through the analysis of disease-perturbed networks. In addition, current drugs can then be tested against the stratified types to identify the impedance match between stratified disease types and effective drugs. The use of trastuzumab in HER2-positive breast cancer patients is an excellent example of the effectiveness of this approach.[11] Furthermore, such systems approaches to the identification of drug targets may transform the pharmaceutical industry's approach to drug target identification; this arises from the simple idea that drugs can be identified to interact with key proteins in disease-perturbed networks to make their behavior more normal—or at least abrogate their most deleterious effects. Crucially, the use of P4 medicine will enable the focus of medicine to shift from disease to wellness—with enormous attendant cost savings to society resulting in a lower requirement for sick leave and a concurrent increase in national productivity. Additional arguments can be made in support of the idea that P4 medicine will turn around health-care costs, but the key point is that many factors will converge to bring the costs of health care down in a striking manner so that the benefits of P4 medicine can be shared by rich and poor nations alike.

How much data will be needed?

In the current symptom-based practice of medicine, new diseases are often defined by a handful of patients who share a constellation of descriptive symptoms. The majority of the drugs in use today for diseases ranging from diabetes and depression to hypertension and macular degeneration were approved through trials in patients with these symptoms based disease classifications. The expectation was that all patients with diseases such as hypertension or lung cancer were part of a homogeneous cohort that could be reliably used as subjects when testing experimental therapies. It had been assumed that everyone within that disease group would have a roughly equal chance of responding to such a treatment. These clinical trials generally require a minimum of hundreds of patients to detect sufficient efficacy for drug approval. In situations where there is already an existing standard-of-care therapy, randomized clinical trials can require thousands or even tens of thousands of patients—because perhaps only one in 20 patients will respond to therapy. Such large trials can only be contemplated by a large medical institution or, more often, by collaborative trial groups and the key drivers of the most innovative trials have been the mega-cancer centers—first in the US, then in Europe, and more recently in Asia.

The failure of the current approach is becoming obvious. Systems medicine suggests that diseases such as lung cancer or diabetes are more properly stratified into multiple types by assessing their distinct combinations of disease-perturbed networks.[3,12] As we transition from pathologic-based descriptions of diseases to stratification of diseases using molecular and cellular data, the numbers of distinct diseases has exploded and will continue to do so. Nowhere has this evolution of disease stratification been more rapid than in oncology. Solid tumors, such as breast cancer,[13] and hematologic cancers, such as lymphoma,[14] have already been subdivided into five or more categories, on the basis of only one or a few molecular markers, such as receptor status or the presence of a genetic mutation. As the data begin to accumulate from whole-genome sequencing efforts of tumors from partnerships such as TCGA (The Cancer Genome Atlas)[15] and the ICGC (International Cancer Genome Consortium),[16] disease classifications are likely to become even more precise and be extended into more diverse cancer types. P4 medicine is providing the technical and computational tools for these stratifications. This emerging reality has a direct impact on how we need to perform future trials and one of the big challenges is recruiting sufficient patients to address diseases that may consist of 10 to 20 or more distinct subtypes. This will require a fundamental change in the business plans of the pharmaceutical industry.

In 2005, there was an unprecedented partnership set up between a drug discovery company, Merck and Co., and a hospital, The Moffitt Cancer Center in Tampa, Florida.[17] The agreement was based on the concept that every patient who entered the hospital would be asked to consent to having a portion of their tumor saved for research and that the analysis of the patient's sample might include studies on the tumor that could not be defined or anticipated at the time of the consent. This partnership ensured that most of these tumors would be analyzed for the expression levels of all genes by use of RNA expression arrays. At Moffitt, the patient tumor data were collected along with the clinical data from medical records so that if one wanted to ask a question about biomarkers—how many patients with breast cancer had a specific

© 2011 Macmillan Publishers Limited. All rights reserved

BRCA1 mutation or a *MYC* amplification?—one could look this up in a database and ask how this correlated with response to therapies. The arrangement benefited both parties. Moffitt was able to secure funding for analyzing more than 10,000 cancer patients and gained the remarkable ability to correlate alterations found in patient tumors to numerous other clinical and imaging parameters. Merck gained access to some of the best documented tumors in the world and had an opportunity to explore the pattern of mutations in many individual patient tumors to better understand the complexities of therapeutic responses. Importantly, Merck also had the ability to go though databases and identify those patients with alterations that might make them appropriate subjects for upcoming trials.[18] The agreement was designed so that all the data generated will enter the public domain following the end of the partnership. This arrangement has been successful for all involved: patients received better care, physicians gained valuable information about the molecular characteristics of their patients' tumors, and Merck was able to test therapies using molecular phenotypes. However, reality rarely lives up to expectation and the approach did not return all the anticipated value. An underappreciated problem was the lack of sufficient patients with particular molecular alterations to effectively conduct the desired trials. It is a simple numbers game; by the time a series of molecular criteria are established for a particular tumor type, the interval of time required to recruit the prerequisite patients becomes a barrier. The solution to identifying the needed patients for such focused molecular trials can be solved by the participation of multiple centers to greatly enlarge the available patient population.

Clinical trials are now expected to take advantage of the new, more precise, molecular characteristics of tumors; therefore, coordinated centers will need to be identified to collect and perform molecular analyses on samples in a coherent way that will seamlessly integrate the data from all the sites. As we now start looking to enroll focused subtypes of less-common diseases, we will need to form ever larger working consortia to capture sufficient patients for statistical validation of subtype results for the early clinical phase I and phase II trials. In the past, these numbers of patients have been required only for the large trials needed in phase III—a challenge that will be mitigated in the future by the potential effectiveness of stratified therapeutic trials that will hopefully lead to high percentage drug responses. As the number of patients initially screened at clinical trial sites increases to obtain the required cohort of an ever more specific subtype, there will be a point where certain trials will require the strategic partnership of large clinical centers.

An additional dimension that we must also acknowledge is that diseases occur along a temporal axis such that the samples collected at one time point may provide only minimal clues to the disease in the same patient at a later time. This reality demands a method to precisely describe the specific phases of progression of cancers. Therefore, it is reasonable to ask, what can be done to enable the evolving trials that will be needed as therapies become focused on individual patients with a particular type of disease? In each patient there are two dimensions of disease: first, which subtype of a particular disease the patient has (for example, which of five or more breast cancer types), and second, how far the cancer has progressed in the patient. To answer these questions we will have to use the systems approach to disease; to identify combinations of disease-perturbed networks that uniquely define each disease subtype and to integrate a variety of biological information to elucidate stages of progression in individual patients.

Patients to become partners

The complement to extensive clinical center partnerships is an obvious one; however, it involves moving beyond the current trial enrollment system as conducted by clinicians at academic hospital-based centers. The first part of the answer lies in the 'citizen-driven trials' that are beginning to occur. The term citizen-driven trials refers to trials that are initiated by contacting patients and enrolling them through an interest shown in participating by the patients themselves. Some of these citizen-driven trials use approved drugs and have minimal physician involvement. Other citizen-driven trials are initiated by patients and then are coordinated and documented in concert between the patient and their physician. Some powerful examples of these progressive cancer trials include those funded and recruited by the Multiple Myeloma Research Foundation,[19] and studies driven by Jeff Kauffman, at the Adenoid Cystic Carcinoma Research Foundation.[20] Even more novel are clinical trial designs where patients, organized by advocate communities, have begun to test the limits of 'crowd sourcing' clinical trials. Crowd sourcing refers to the ability to identify and enroll patients into clinical trials not by asking physicians or medical centers, but by reaching out directly to patients through e-mail or other means to determine who might want to enroll in a trial. Since 2008, the Dr Susan Love Research Foundation and the Avon Foundation has been recruiting a 'Love/Avon Army of Women' that, when fully recruited, will include one million women.[21] They have already enrolled more than 330,000 women who have been distributed among 34 studies (one half are already fully enrolled).[22] The strategy is that when women join the 'Army', they register sufficient data online so that when an investigator completes the forms to engage a new trial an 'e-blast' goes out to all the relevant enrolled women and they may opt into the study as desired.[22]

Another progressive approach has been taken by the cancer Biomedicine Informatics Grid (caBIG) and the City of Hope Medical Center in collaboration with the Army of Women.[23] This is a longitudinal study in breast cancer that uses a process of enrollment and tracking of new data that is done entirely online by patients and advocates. Two extensions of this citizen-driven approach that are already planned within the next 2 years include being able to collect patient samples and to enable patients to upload their own medical records.[24] These crowd-sourced trial networks have the obvious advantage of being very efficient at enrolling large numbers of patients. Importantly, such distributed patient access networks, because of their potentially large enrollment numbers, will be needed to usher in the era of personalized medicine because of their ability to identify and enroll patients who have distinct sets of molecular alterations (particular types of stratified disease) that no single clinical trial network of academic sites can match.

The ideal approach

The patient-driven trials discussed above will be a key component of the testing of novel therapies and also of re-examining the efficacy of current approved drugs for subpopulations of patients with specific molecular changes. That said, there are serious advantages to combining the breadth of the enrollments possible by patient-driven trials with the strength and robustness of clinician-driven trials. One obvious solution is to build a network of researchers at designated cancer centers that can coordinate their activities and link them to the patient-driven trial structures that are emerging. The Worldwide Innovative Networking (WIN) Consortium that has recently been formed is in a perfect position to provide such a

© 2011 Macmillan Publishers Limited. All rights reserved

network of investigators, cancer centers, and links with patient-driven trials.[25]

Sharing data

The fourth 'P' of P4 medicine is participatory; this includes a variety of aspects of participation including how we educate physicians, patients and other members of the medical community about the power of P4 medicine. Moreover, how do we design an information technology platform for health care that will provide the storage and computational capacity to mine and compare the billions of data points (both molecular data and medical records) for hundreds of millions of patients? One of the most perplexing challenges is that of acquiring all of the patient data—clinical, molecular, and phenotypic—in a manner that makes it accessible for all scientists to explore, mine and analyze. There are challenges of ethics (will the accessibility of these anonymized data lead to exploitation by employers or insurance companies), and security (how will the data be protected), as well as the technical challenges (how do we validate the quality of the data and get it into standard formats that permit it to be broadly used by the community of scientists). Finally, there are fascinating technical problems for mining, comparing and analyzing data sets of this magnitude. Not only are the facilities not readily available for the storage and analysis of data of this type but the software programs required to analyze these data will need to be developed and be open source for the full potential to be realized—a pathway many companies will be reluctant to go down. The issues of data ownership (by scientists and by institutions) and a reluctance to believe in open-source and open-data policies will have to be overcome if we are to mine the incredible potential of the exploding opportunities of patient data accumulation. Brand new types of strategic partnerships are going to be essential—for pioneering new systems approaches to disease, for developing new technologies and analytical tools, for bringing together complementary expertise, for employing new organizational structures capable of bringing together sufficient numbers of patients, and for creating new funding structures.

Conclusions

P4 medicine is catalyzing a transformation of medicine that promises to deal with the heretofore impenetrable barriers of incredible complexities of disease (and wellness) through systems approaches, emerging technologies and powerful analytical tools. The ultimate promise is that the focus of medicine will be shifted from disease to wellness and that the billions of data points for each individual will define with exquisite specificity the nature of their wellness—and any transitions into disease. Central to this view is the idea that the molecular, cellular and phenotypic data of eventually hundreds of millions of patients will be available for systems analyses—integration, mining and the development of predictive and actionable models. The availability of these data will be necessary to exploit the infinite potential of the P4 medicine of the future. In turn, it will be necessary to consider how to enroll trials of a magnitude that may not be possible unless they are crowd sourced. Finally, we will need a way to quality control, annotate, and curate data so that the person who generates the data can, and should, make it available to anyone else that could then benefit from the insights and models of human diseases developed. In addition, it will be necessary to 'version' the data and models in ways that are reminiscent of those approaches found to be highly efficient for building open access software.[26] This concept of enabling data and models to be shared as new maps of disease built in an open manner that includes researchers from around the world can fundamentally change not just the style, but also the speed of the discovery and development of new therapies. The scientific and societal barriers for these objectives are considerable. To solve them will require new ways for scientists to engage with each other, new relations between patients, scientists, and industry and , finally, will require new strategic partnerships with focused and coherent objectives.

*Institute for Systems Biology, 1441 North 34th Street, Seattle, WA 98103, USA (**L. Hood**). Sage Bionetworks, 1100 Fairview Avenue North, Seattle, WA 98109, USA (**S. H. Friend**).*

Correspondence to: L. Hood
lhood@systemsbiology.org

1. Weston, A. D. & Hood, L. Systems biology, proteomics, and the future of health care: toward predictive, preventative, and personalized medicine. *J. Proteome Res.* **3**, 179–196 (2004).
2. Hood, L. A. in *Physical Biology: From Atoms to Medicine* (ed. Zewail, A. H.) 337–366 (Imperial College Press, London, 2008).
3. Price, N. D. *et al.* in *Essentials of Genomic and Personalized Medicine* (eds Ginsburg, G. S. & Willard, H. F.) 131–141 (Elsevier Academic Press, Maryland Heights, 2009).
4. Auffray, C., Charron, D. & Hood, L. Predictive, preventive, personalized and participatory medicine: back to the future. *Genome Med.* **2**, 57 (2010).
5. Zhao, L. Genomics: The tale of our other genome. *Nature* **465**, 879–880 (2010).
6. Greenman, C. *et al.* Patterns of somatic mutation in human cancer genomes. *Nature* **446**, 153–158 (2007).
7. Hwang, D. *et al.* A systems approach to prion disease. *Mol. Syst. Biol.* **5**, 252 (2009).
8. Hood, L. & Galas, D. The digital code of DNA. *Nature* **421**, 444–448 (2003).
9. Carey, L. A. Through a glass darkly: advances in understanding breast cancer biology, 2000–2010. *Clin. Breast Cancer* **10**, 188–195 (2010).
10. Ivshina, A. V. *et al.* Genetic reclassification of histologic grade delineates new clinical subtypes of breast cancer. *Cancer Res.* **66**, 10292–10301 (2006).
11. Piccart-Gebhart, M. J. *et al.* Trastuzumab after adjuvant chemotherapy in HER2-positive breast cancer. *N. Engl. J. Med.* **353**, 1659–1672 (2005).
12. Schadt, E. E. *et al.* An integrative genomics approach to infer causal associations between gene expression and disease. *Nat. Genet.* **37**, 710–717 (2005).
13. Cleator, S. & Ashworth, A. Molecular profiling of breast cancer: clinical implications. *Br. J. Cancer* **90**, 1120–1124 (2004).
14. Monti, S. *et al.* Molecular profiling of diffuse large B-cell lymphoma identifies robust subtypes including one characterized by host inflammatory response. *Blood* **105**, 1851–1861 (2005).
15. [No authors listed] Comprehensive genomic characterization defines human glioblastoma genes and core pathways. *Nature* **455**, 1061–1068 (2008).
16. Hudson, T. J. *et al.* International network of cancer genome projects. *Nature* **464**, 993–998 (2010).
17. Dalton, W. S. & Friend, S. H. Cancer biomarkers—an invitation to the table. *Science* **312**, 1165–1168 (2006).
18. Marx, V. Merck–Moffitt personalized medicine partnership builds out information and data integration pipeline. *GenomeWeb* [online], http://www.genomeweb.com/informatics/merck-moffitt-personalized-medicine-partnership-builds-out-information-and-data- (2009).
19. Multiple Myeloma Research Foundation. *Multiple Myeloma Research Consortium (MMRC) activates clinical trials 30–40 percent faster than industry standard*. *MMRF*[SM] [online], http://www.themmrf.org/archive/outdated-content/press-releases-061809.html (2009).
20. Adenoid Cystic Carcinoma Research Foundation [online], http://www.accrf.org/ (2010).
21. Love, S. M., Mills, D., Eraklis, E. & Hulbert, M. The Love/Avon Army of Women: a "just in time" resource to encourage research in women [abstract]. *J. Clin. Oncol.* **27** (15 Suppl.), a1535 (2009).
22. Love/Avon *Army of women* [online], http://www.armyofwomen.org/ (2010).
23. Bouchie, A. Coming soon: a global grid for cancer research. *Nat. Biotechnol.* **22**, 1071–1073 (2004).
24. Love/Avon. *Health of women study* [online], http://www.armyofwomen.org/HOW_Study (2010).
25. Mendelsohn, J., Tursz, T., Schilsky, R. L. & Lazar, V. Focus on personalized medicine: WIN Consortium—challenges and advances. *Nat. Rev. Clin. Oncol.* **8**, 133–134 (2011).
26. Sonnenburg, S. *et al.* The need for open source software in machine learning. *J. Mach. Learn. Res.* **8**, 2443–2466 (2007).

Author contributions
Both authors contributed to researching the data for the article, discussion of content, and writing and reviewing the manuscript.

© 2011 Macmillan Publishers Limited. All rights reserved