Systems Biology and the Emergence of Systems Medicine

Nathan D. Price[1,2,3], Lucas B. Edelman[2,3], Inyoul Lee[4], Hyuntae Yoo[4], Daehee Hwang[5], George Carlson[6], David J. Galas[4,7], James R. Heath[8], and Leroy Hood[4,*]

[1]Department of Chemical and Biomolecular Engineering, [2]Department of Bioengineering, & [3]Institute for Genomic Biology, University of Illinois, Urbana-Champaign
[4] Institute for Systems Biology, Seattle, WA
[5]I-Bio Program and Department of Chemical Engineering, Pohang University of Science and Technology, Pohang, Korea
[6]McLaughlin Research Institute, Great Falls, MT
[7]Battelle Memorial Institute, Columbus, OH
[8]Department of Chemistry, California Institute of Technology

[*]To whom correspondence should be addressed: lhood@systemsbiology.org

Prepared for Genomic and Personalized Medicine: From Principles to Practice (Ginsburg, G. and Willard, H., editors). Elsevier, in press.

Introduction

A new approach to biology, that we call **systems biology**, has emerged over the past 10 years or so—an approach that looks at biology as an informational science, studies biological systems as a whole, and recognizes that biological information in living systems is captured, transmitted, modulated, and integrated by biological networks which pass this information to molecular machines for execution. This approach differs from early "systems approaches to biology" in that it attempts both a bottom-up approach (from large molecular datasets) and a top-down approach (from computational modeling and simulations) where there is an attempt to trace complex observations of phenotype back to the digital core encoded in the genome. New measurement and visualization technologies together with powerful computational and modeling tools have transformed systems biology by making it possible for the first time to execute the five uniquely defining features of contemporary systems biology: 1) measurements of the various

types of biological information that are **global** or comprehensive to the greatest extent possible (e.g., measure the digital code of all genes or, for example, the concentrations of all mRNAs, all proteins, all metabolites, etc.); 2) the different levels of information (DNA, RNA, protein, etc.) must be integrated to understand or capture how the environment has modified the basic digital information of the genome at each level of the biological information hierarchy (DNA, RNA, proteins, interactions, biological networks, cells, individuals and ecologies) and thereby to induce biological responses; 3) all biological systems (e.g. networks) must be studied dynamically as they capture, transmit, integrate and utilize the biological information necessary for the execution the two most fundamental responses of living organisms – development or physiological responses to the organism‘s environment, 4) all measurements must be quantitatively determined to the greatest extent possible, and 5) the global and dynamic data from the variety of information hierarchies must be integrated and modeled. These models essentially create hypotheses about biological functions and mechanisms of disease that can then be tested experimentally by systems perturbations—the enumeration of new data sets, their integration, modifications of models, in an iterative manner, until the working models reflect the reality of the experimental data. Once the models accurately fit the biological systems measurements, they can be used to predict the results of perturbing the system in new ways (e.g. for designing treatment strategies or for identifying likely underlying causes for disease).

The information revolution in biology will enable systems medicine to emerge in the next ten years. Key to this transformation will be harnessing computationally the vast amount of biological information becoming available through rapid advancement in measurement, visualization and computational technologies. Two fundamental types of biological information,

the digital information of the genome and interacting information from the environment, are integrated together to specify the five fundamental mechanisms of life—evolution, development, physiological responses, aging and the onset and progression of disease (Hood et al., 2004). Advances in high throughput DNA sequencing will enable each person's genome to be sequenced rapidly, and at a reasonable cost providing digital information for each person, and making possible the prediction of increasingly accurate probabilistic health futures. Also, the advance of high-throughput measurement technologies will enable the assessment of dynamic environmental information emerging from the integration of genome and environmental information from each individual, as reflected, for example, by the changing levels of proteins in the blood—thus providing a real time (current) heath assessment of the individual. These technologies will generate tremendously large, dynamical data sets about health states and hence about the states of relevant biological networks of the individual patients. This detailed information will arise through the use of perturbation experiments coupled with global experimental technologies, and progress in systems biology research is increasingly elucidating the functions and structures of these networks.

A systems approach to medicine argues that disease arises from disease-perturbed biological networks and that the dynamically changing, altered patterns of gene expression that are controlled by these perturbed networks give rise to the disease manifestations. Here, we present a systems view of biology and disease together with a discussion of some recent advances in state-of-the-art *in vitro* and *in vivo* diagnostics technologies, and we suggest how, as these technologies mature, they will move us towards a future of predictive, personalized, preventive, and participatory medicine **(P4 medicine)** (Hood et al., 2004).

I. Systems Science in Biology and Medicine

A. Systems biology

How do systems approaches to biology and medicine lead to a revolutionary view of medicine? We present this systems view in some detail because an understanding of P4 medicine and its implications for society are predicated upon understanding these systems principles. Two primary domains of biological information lend themselves readily to such systems-level analysis: the static, digital information of the genome, and the dynamic information arising from environmental interactions with the sub-cellular, cellular, and tissue levels of organization (Hood et al., 2004). Digital genome information encodes two types of biological networks — **protein interactions** and **gene regulatory networks**. Protein networks transmit and use biological information for development, physiology and metabolism. Gene regulatory networks — transcription factors, their regulatory binding sites and the small RNAs that regulate networks of other transcription factors and other RNAs interacting with one another — receive information from signal-transduction networks, integrate and modulate it, and convey the processed information to networks of genes or molecular machines that execute developmental and physiological functions. In biological systems these two types of networks are closely integrated. The organization of these networks can be inferred from various types of measurements including, for example, global measurements of dynamically changing levels of mRNAs and proteins during developmental and physiological responses, as well as large-scale measurements of protein–protein and protein–DNA interactions. There are multiple **hierarchical levels of organization and information** (for example, DNA, RNA and protein

networks, cellular and metabolic networks, and organization and responses of organ systems). To understand biological systems, information must be gathered from as many information levels as possible and integrated into models that generate hypotheses about how biology works. Let us now consider the logic of systems approaches to medicine.

B. Systems medicine

The central premise of systems medicine is that disease-perturbed networks alter patterns of expression in genes and proteins – and that these altered patterns encode the dynamic pathophysiology of the disease and necessarily results in altered molecular fingerprints that can be detected clinically. Through advances in measurement technologies and computational analysis tools for tissue and blood analyses, we will be able to read these signals to make a multitude of diagnoses to distinguish health from disease and to determine the nature of any pathology. Multi-parameter analyses of biological information will be key to tracking these altered patterns in disease-perturbed networks. Diagnostic methods of the past have been pauci-parameter in nature—usually measuring just a single parameter relevant to a specific disease state (e.g. PSA), and so our ability to track health and disease has been limited. Current and emerging technologies are creating a transition into a new era of predictive, preventative, and personalized medicine. A causal disease perturbation could be the result of specific, disease-causing DNA mutations, pathogenic organisms, or other pathological environmental factors such as toxins. Molecular fingerprints of pathological processes can take on many molecular forms, including the analyses of proteins (Wang et al., 2005), DNA (Papadopoulou et al., 2006), RNA (Scherzer et al., 2007), and metabolites (Solanky et al., 2003), as well as informative, post-translational modifications to these molecules such as protein phosphorylation.

Signals related to health and disease can be found in multiple sites. For instance, many bodily fluids such as the blood, urine, saliva, cerebral spinal fluid and so forth can be sampled to identify evidence of perturbed molecular fingerprints reflecting the altered expression patterns of genes and proteins in disease-perturbed biological networks. Of these, the blood is likely the most information rich organ (or fluid) in that it bathes all of the tissues in the body, and it is easily accessible for diagnostic procedures. In addition to the biomolecules secreted into the blood from cells and tissues throughout the body, the transcriptomes and proteomes of cells circulating in the blood (e.g. white blood cells) are also potentially an abundant source of biomedically important information (Buttner et al., 2007). Thus, the amount of information available in the blood about health and disease is enormous if we learn to read and interpret the molecular signals.

As an example of an environmental disease perturbation, we have studied the dynamic onset of the infectious prion disease in several strains of mice at the level of mRNA in the brain (the affected organ), and showed that a series of interlocking protein networks that surround the prion protein are significantly perturbed across the 150-day span from disease initiation to death. Figure 1 shows differential networks that were derived from comparisons of mRNA-expression patterns in normal and diseased brains at each of three time points after infection. They involve 67 proteins in the prion replication and accumulation network. The initial network changes occur well before the clinical signs of the disease can be detected and predict later widespread histopathological events.

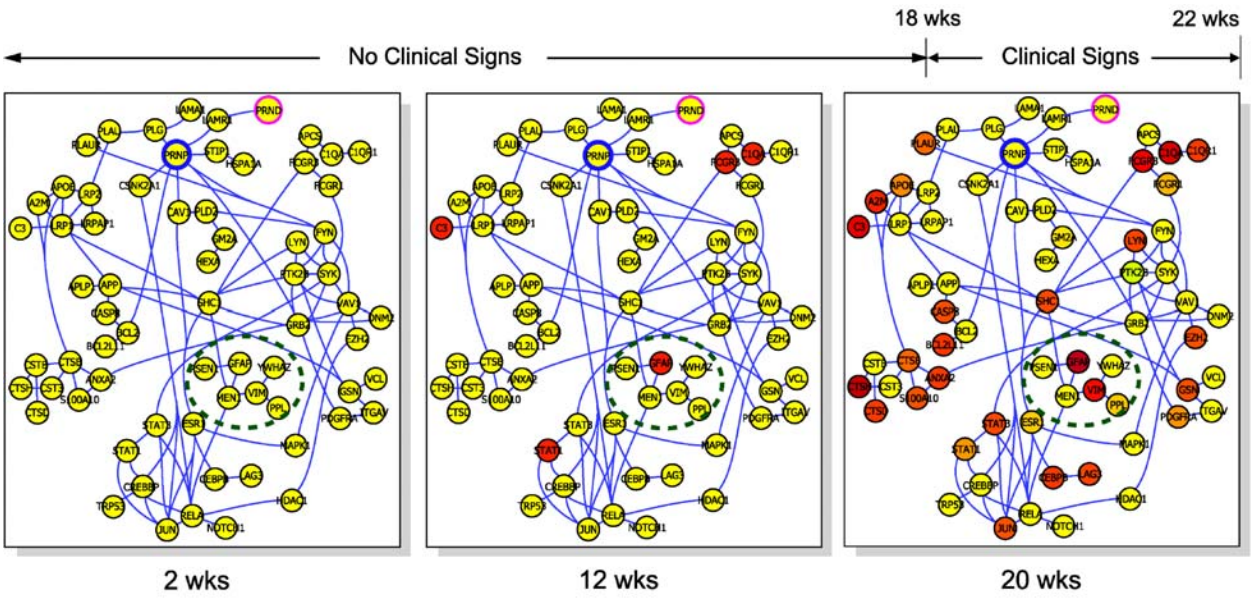

Figure 1. A central sub-network involved in neuronal pathology in mouse prion disease, shown at three times after inoculation. Red circles indicate increased levels of gene expression relative to controls.

These dynamically changing, disease-perturbed networks lead to two important conclusions. First, some significant network nodal points change before the related clinical or histological changes are evident. Therefore, labeled biomarkers that are specific for the changing nodes or the biological processes they regulate could be used for *in vivo* imaging diagnostics even before symptoms arise, as has already been shown in patients (Golub et al., 1999; Quackenbush, 2006; Ramaswamy et al., 2001; Wang et al., 2005). Alternatively, if some of these altered nodes encode secreted proteins, they could provide readily accessible *in vitro* diagnostic blood markers for early disease detection. Second, many of the sub-networks of proteins that change during the onset of disease affect changes in phenotypic traits that are consistent with the pathology of the disease. About 900 perturbed mRNAs appear to encode the core prion-disease response and during the progression of the prion-disease process, several hundred undergo significant gene expression changes well before the clinical signs of prion disease. Many of these potential early-

disease ‘sentinels’ are predicted to be secreted into the blood and therefore represent potential protein biomarkers for early disease diagnosis through blood protein analysis. One interesting new diagnostic method evaluates relative expression reversals in protein concentrations or gene expression levels. This procedure eliminates the need for data normalization or the establishment of population-wide thresholds. This approach has been successfully used to identify robust and accurate classifiers for prostate cancer (Xu et al., 2005), sarcomas (Price et al., 2007), and a variety of other cancers (Tan et al., 2005), as well as to predict treatment outcomes in breast cancer (Ma et al., 2004), demonstrating the efficacy of even relatively low-order systems analysis in medicine.

Another key issue that needs to be addressed is how to best decouple two primary dimensions of disease information: namely, disease stratification (e.g., which type of prostate cancer is present) and disease progression (e.g., stage of development of a particular prostate cancer stratification in time). Molecular signature changes can indicate the presence of different diseases (stratification) as well as their stage (progression). Thus, a key challenge in future will be to develop analysis tools that will enable us to differentiate the location of the physiologic state in regards to both of these critical clinical dimensions.

II. Multi-parameter Blood-Bourne Biomarkers

In principle, it is clear how blood samples containing secreted molecules from all of the tissues in the body can be used as a window into health or disease states (Anderson and Anderson, 2002; Fujii et al., 2004; Hood et al., 2004; Lathrop et al., 2003; Lee et al., 2006). In practice, however, the task of identifying markers of disease states in the vast array of secreted proteins can seem

daunting. There are millions of different proteins present in the blood and they are expressed at levels that probably differ by 12 orders of magnitude (10^{12}). Thus, there are indeed very significant challenges associated with defining relevant biomarkers in the plasma proteome, as is evidenced by the paucity of blood protein markers found thus far, despite significant efforts (Wilson, 2006). This fact highlights the need to develop well-founded systems approaches to diagnostics in order to hasten the identification of such markers and to harness the tremendous information potential of the blood. While the potential information available to diagnose heath and disease is enormous, it is also true that the challenges of separating signal from background noise (biological or measurement noise) are also very significant. Sources of noise include error in measurements, polymorphisms in the population, environmental variations, stochastic variations, as well as signal dilution through mixing and other processes of molecule transport. Thus, enhancing signal while reducing noise will be a primary theme of research in predictive medicine for the foreseeable future.

An important strategy for dealing with noise arising from genetic polymorphisms and health histories in the population are through dynamic, subtractive analyses carried out in the individual patient. In other words, measurements taken from individual patients at different time points (longitudinal data) can be used to perform subtractive analyses where only the differences observed are considered. Thus, each patient becomes their own control, which eliminates many of the sources of noise. Eventually, it will be important for each patient to have biannual blood analyses taken so that changes can be observed relative to the background of what is normal *in each individual* rather than relative to the population at large. Such databases of measurements

will be essential to enabling personalized and predictive medicine. Using each patient as their own control is one of the essential features of the emerging personalized medicine.

A. Organ-specific blood protein fingerprints

We discuss the organ-specific blood protein fingerprint approach to diagnostics because it lies at the very heart of the predictive medicine that will emerge over the next 10 years. Our feeling is that this will be one of the central data gathering strategies for predictive and personalized medicine and, accordingly, we will discuss it in some detail. The idea that disease arises from dynamically changing disease-perturbed networks, that the diseased organs will secrete proteins into the blood, and that if these proteins are encoded by disease-perturbed networks their levels of expression in the blood will be altered in a manner that reflects the specific nature of the disease. Indeed, this idea is the basis for the very broad range of blood biomarker studies that are being carried out by many scientific centers (Anderson and Anderson, 2002; Fujii et al., 2004; Hood et al., 2004; Lathrop et al., 2003; Lee et al., 2006). The major difficulty with this simple view is that if you identify multiple blood proteins whose changes are specific for a particular disease state (as compared against normal controls) and then examine the same blood markers for that disease in bloods drawn from individuals, with say 10 other pathologies, these marker proteins can change in unpredictable ways since multiple organs control the expression of most blood proteins and these organs respond differently to various environmental signals. The important point is that if a marker protein synthesized in five organs changes in the blood, we cannot be certain which organ(s) is the origin of the change. While that marker may sample a biological network relevant to disease diagnosis, since its origin is not clear its use as a disease biomarker may raise more questions than it answers. The solution to this dilemma is clear –

employ organ-specific blood protein biomarkers whose changes must therefore reflect changes only in the organ itself. If enough of these organ-specific blood proteins are sampled, they will represent a survey of many different biological networks in the organ of interest and will provide sufficient diagnostic information for any disease. Eventually, correlation of these organ specific biomarkers with more general biomarkers may prove to be the best long term strategy for developing diagnostic fingerprints.

We have several lines of preliminary evidence that suggest this organ-specific blood protein approach will be effective. In prostate cancer, for example, there are disease-mediated altered patterns of mRNA and protein expression in the prostate. Some of these genes are expressed primarily in the prostate (organ-specific products) and some of these organ-specific proteins are secreted into the blood, where they collectively constitute a protein molecular fingerprint comprised of say 100 or more proteins whose relative concentration levels probably report the status of the biological networks in the prostate gland. We have demonstrated that changes in the blood concentrations of several of these prostate-specific blood proteins reflect the various stages of prostate cancer and, as discussed above, various brain-specific blood proteins also reflect the progression of prion disease in mice. We propose that the distinct expression levels of the individual proteins in each fingerprint represent a multi-parameter (and therefore potentially information rich) diagnostic indicator reflecting the dynamic behavior of, for example, the disease-perturbed networks from which they arise. The analysis of 50 or so organ-specific proteins should allow us to both stratify diseases in the organ as well as determine their stage of progression. We have identified 10s to 100s of organ-specific transcripts in each of the 40 or so organs that we have examined in mouse and human.

We can envision a time over the next 5-10 years when 50 or so organ-specific blood proteins will be identified for each of the 50 or so major organs and tissues in humans—so that computational analyses of the relative concentrations of the protein components in these organ-specific fingerprints will enable blood to become the primary window into health and disease. When we analyze data from these blood indicators, we also may be in a position to identify the dynamically changing disease-perturbed network(s). The analysis of these dynamic networks will allow us to study in detail the pathophysiology of the disease response and hence be in a position to think of new approaches to therapy and prevention. To generate the ability to assess simultaneously all 50 organ blood protein fingerprints in patients, we ultimately need to develop the microfluidic or nanotechnology tools for making perhaps 2500 rapid protein measurements (e.g. 50 proteins from each of 50 human organs) from a droplet of blood. Nanotechnology is necessary because only this severe miniaturization can allow the necessary thousands of measurements from a single drop of blood. In order to reach this stage we will also have to create the appropriate computational tools to capture, store, analyze, integrate, model and visualize the information arising from this approach.

III. Emerging *in vivo* and *in vitro* technologies

A. In vitro *measurement technologies*

P4 medicine will require that *in vitro* measurements of thousands of blood proteins be executed rapidly, automatically and inexpensively on small blood samples. This will require miniaturization, parallelization, integration and automation of tissue and blood purification and measurement technologies. In short, it will require making *in vitro* measurements inexpensively

– perhaps for a penny or less per protein measured. This need for inexpensive *in vitro* measurements is driving the development of integrated microfluidics and nanotechnologies for *in vitro* diagnostics. We further argue that the complex changes in these protein levels must be analyzed computationally to search for the patterns that correlate with particular diseases—and indeed to stratify each disease as well as determine its stage of progression.

To meet the expanding requirements of systems medicine, *in vitro* measurements of cells, proteins, mRNAs, etc., whether for fundamental biological studies or for pathological analysis, must not just be inexpensive, but must also be sensitive, quantitative, rapid, and executed on very small quantities of tissues, cells, serum, etc. We are working towards developing chip platforms that can take a few microliters (a finger prick) of blood, separate the plasma from the serum, and then measure on the order of 100 or more proteins and/or mRNAs quantitatively, with high sensitivity and specificity, and in a few minutes time. We can envision a time 5-10 years hence when small hand-held devices will be able to make these 2500 measurements from a fraction of a droplet of blood, send them *via* wireless to a server for analysis and consequently inform the patient and physician as to the status of the patient. While many fundamental scientific challenges remain to be solved before this goal is achieved, none of those challenges appear insurmountable at this point.

One will also be able to use the blood cells as powerful diagnostic markers—either of infectious diseases or of genetic diseases. Microfluidic cell-sorting technologies for being able to sort blood cells into their 10 or so individual types for analysis are now available. Even more important is the emergence of single-cell analytic tools where DNA, mRNAs or small RNAs,

proteins and even metabolites can be analyzed rapidly from individual cells. The cells can also be perturbed with appropriate environmental stimuli to identify defects. It is possible that an appropriate analysis of immune cells (both innate and adaptive) from the blood will reveal important information about past antigenic history of the individual as well as current state of immunological responsiveness. Similarly, analysis of rare blood cells such as circulating cancer cells may also be utilized to guide therapies.

Another type of *in vitro* measurement will be a determination of the complete genome sequence of individuals with a nanotechnology approach to sequencing single strands of DNA on a massively parallel basis (billions of DNA strands simultaneously). A device like this will emerge over the next 5-10 years and will allow millions, if not billions, of individual human genome sequences to be determined rapidly, inexpensively and in a massively parallel manner. The 2500 organ-specific blood marker measurements and the complete individual human genome sequences will be the heart of the predictive and personalized medicine that will emerge over the next 10 years or so.

While transcriptomic approaches have proven useful for identifying informative molecular signatures for a number of diseases (Quackenbush, 2006), high throughput proteomic characterizations have lagged behind. The reason for this difference is clearly the more difficult challenge of measuring proteins compared to mRNA transcripts. Emerging proteomics technologies hold the promise of greatly improving our ability to make detailed disease assessments using protein-based molecular signatures. One key advantage of protein signatures relative to gene expression is that the proteins found in the blood and other accessible bodily

fluids exhibit slower degradation rates than their mRNA counterparts—and of course proteins are the direct agents mediating the disease process itself. Thus, protein concentration signatures represent an important class of molecular signature for disease diagnosis.

B. The challenges of proteomics and blood protein biomarkers.

The blood is an ideal organ for identifying biomarkers because it interacts with virtually every organ and major tissue type in the body, and each secretes proteins into the blood. Moreover, the blood is easily accessible for diagnostic studies—and presumably the analysis of changes in protein levels or protein modifications that could serve as surrogates or reflections of health and disease. As noted earlier, the blood is an enormously complex organ; it contains hundreds of thousands to millions of proteins whose concentrations range over 12 orders of magnitude. Since there is for proteins no equivalent to the polymerase chain reaction (PCR) amplification procedure, only the more abundant proteins can be seen by conventional protein analysis studies. It should also be pointed out that many different features of proteins ultimately must be characterized or quantified: identification, quantification, chemical modifications, alternative mRNA splicing products, cellular localizations, three-dimensional structures (and their dynamics), as well as their functions. We here are only concerned with the quantification and identification of proteins.

There are two general approaches to the analysis of protein mixtures. In one case protein-capture agents such as antibodies are used—and with appropriate controls these approaches can quantify the levels of proteins. These techniques include Western blot analyses, ELIZA assays, surface plasmon resonance and protein chips. A second approach is to use mass spectrometry. Since

proteins have large masses that are difficult to analyze accurately by mass spectrometry, most proteins are analyzed after converting them into tryptic peptides (or other proteolytic fragments) whose masses are far smaller and hence more easily analyzed. Because blood is such a complex mixture, proteins of interest are often enriched by some type of fractional procedure (charge, hydrophobility, size and/or the presence of modifications such as glycosylation or phosphorylation,) before (or after) their conversion into peptides and analysis by mass spectrometry. Isotopically labeled and synthesized peptides may be used to identify and quantify (relative or absolute) the same peptides from unknown samples (and hence one obtains a proxy for the quantification of their corresponding proteins) very effectively by mass spectrometry.

We discuss below examples of some of these techniques.

1. Antibody microarrays using surface plasmon resonance imagining (SPRI)

Protein chip methods hold great potential for broad quantitative screens of proteins, and a variety of techniques have been developed based on antibody binding (Haab et al., 2001; Olle et al., 2005; Song et al., 2007). Various types of antibody arrays have been used for biomarker discovery and protein profiling of serum from patients with prostate, lung, pancreas and bladder cancer (Gao et al., 2005; Miller et al., 2003; Orchekowski et al., 2005; Sanchez-Carbayo et al., 2006). One emerging approach with tremendous promise is SPRI (Hu et al., 2007; Huber and Mueller, 2006; Koga et al., 2006), which enables real-time, label-free measurement of protein expression by large numbers of different antibodies. SPR-based chips have a detection sensitivity of 10-100 times less than ELISA (Hu et al., 2007), but have a spatial resolution down to approximately 4 μm (Lyon et al., 1998). It is thus possible to print up to 800 unique

antibodies on Lumera Nanocapture Gold™ microarray slides and monitor the abundance of the target proteins in real time (Hu et al., 2007) even in complex samples such as blood. Because the same slide can be regenerated for reuse many times in 10 minutes or less (Z. Hu, C. Lausted, unpublished observations). We are in the process of automating the sample introduction procedures using microfluidic chips , which means that SPRI has the capacity necessary to screen rapidly through tens to hundreds of patient samples. Thus, this approach holds tremendous promise to be able to screen through large numbers of proteins, including secreted proteins and cell-surface markers and not only measure their presence, but also abundance and the dynamics of their binding. The limitation of this technique is its dependence on the affinity and specificity of the antibodies it employs for detection—cross reactivities in complex protein mixtures (such as blood) can pose significant problems.

2. DNA-encoded antibody libraries (DEAL)

One recently developed technique that offers great potential for detailed analyses is DEAL. The primary advantage of DEAL is that it uses a single, robust chemistry – that of DNA hybridization – to spatially localize and detect proteins, mRNAs, and cells, all in a multiplexed fashion. Antibodies are typically too fragile to survive the fabrication procedures associated with assembling robust microfluidics chips, but DNA oligomers are significantly more robust. DEAL thus enables the detection of panels of protein biomarkers within a microfluidics environment and from very small quantities of biological material (100 nanoliters or so) (Yang et al., 2006). This amount of plasma can be readily separated from whole blood on-chip, thus allowing for the measurement of serum biomarkers from a fingerprick of whole blood. In addition, within the environment of flowing microfluidics, the rate-limiting step in performing a surface

immunoassay is the kinetics of the binding of the analyte to the surface-bound capture agent (Zimmermann et al., 2005). Thus, DEAL-based immunoassays can be executed very rapidly.

The versatility of DEAL also enables multiplexed cell sorting and localization, followed by few- or single-cell measurements of protein, RNA, and other biomolecules (Bailey et al., 2007). DEAL can be engineered into a highly sensitive and very rapid measurement technique, with a reported detection limit of 10 fM for the protein IL-2 – 150 times more sensitive than the corresponding commercial ELISA assay. This sensitivity can be applied to the isolation of rare cells based on combinations of cell-surface markers, enabling the isolation and addressing of individual cancer stem cells. DEAL can also be used to make single cell measurements of secreted proteins from each of these isolated single cells. Thus, DEAL offers superb sensitivity and the ability to perform spatially multiplexed detection for characterization of rare cell types, such as circulating cancer cells or cancer stem cells. These advantages still face the inherent limitations of antibodies, so the development of new approaches to generating protein-capture agents is a critical part of future development of comprehensive blood diagnostics.

D. Mass spectrometry based techniques

1. Isotopic tagging for relative and absolute quantification (iTRAQ) of proteins analyzed by mass spectrometry

Stable isotope labeling enables the quantitative analysis of protein concentrations through mass spectrometry (MS). One state-of-the-art technique for quantitative MS is iTRAQ (Ross et al., 2004), which uses stable isotope labeling of proteolytic peptides. This technique modifies

primary amino acid groups of peptides by linking a mass balance group, and a reporter group by forming an amide bond. When MS/MS is used for analysis with iTRAQ-tagged peptides, the mass balancing carbonyl moiety is released as a neutral fragment generating 4-8 distinct sets of peptides whose relative abundances can be determined quantitatively. Because eight different iTRAQ reagents are currently available, comparative analysis of a set of two to eight samples is feasible within a single MS/MS run (Hu et al., 2007). The, iTRAQ technology represents the state-of-the-art in quantitative proteomics and represents a promising technology for using proteomics to differentiate key differences in protein networks.

2. Glyco-peptide capture—front end enrichment of blood proteins containing sugar residues

Mass spectrometry-based methods will allow for the identification of proteins spanning approximately three orders of magnitude in concentration from a given sample. Therefore, methods that can select specified fractions of the proteome are important for simplifying the sample sufficiently to identify the proteins of interest. One recently developed approach is the shotgun glycopeptides capture approach (Sun et al., 2007). This approach selects for N-linked or O-linked glycosylated peptides by selectively coupling these peptides to beads—allowing the uncoupled peptides to be washed away—then the glycosylated peptides are be released and analyzed by mass spectrometry. Both secreted and cell-surface proteins are enriched for glycopeptides compared with their nuclear and cytoplasmic counterparts. Thus, this approach can be used to, for example, identify candidates for unique cell-surface markers to make identifications of clinically relevant cellular subpopulations.

D. In vivo *molecular diagnostics*

As it relates to personalized and predictive medicine, *in vivo* molecular diagnostics will also require the development of a diverse library of molecular imaging probes. These modular tools can be used to 1) identify the specific location of disease-perturbed networks in patients; 2) link *in vivo* molecular measurements in diseased tissue in patients to *in vitro* measurements; 3) rapidly assess the efficacy of personalized therapeutics; and 4) validate that a drug is hitting its target and inducing the desired pharmacological outcome. Although there are many *in vivo* imaging modalities, perhaps the best current method from the point of view of personalized and predictive medicine in patients is positron emission tomography (PET) molecular imaging (Czernin and Phelps, 2002). For PET, trace quantities of radiolabeled molecular probes are injected into the patient. As the probes circulate through the body and its various organ systems, they interact with target proteins to provide imaging assays for, as examples, the rate of metabolic processes, the concentration of receptors in signal transduction, enzyme activity, DNA-replication rates, hormone status, pharmacokinetics and pharmacodynamics.

IV. Computational and Mathematical Challenges in Systems Medicine

A. Molecular Signature Classifiers

Computational methods are needed to reduce the high degree of data dimensionality associated with global data sets to identify molecular signatures that can be used for disease diagnosis. Despite notable and significant challenges that remain (Dupuy and Simon, 2007; Simon, 2005), computational analyses to identify molecular signatures from global gene expression data sets that can be used for diagnosis and treatment selection (Quackenbush, 2006) is an area of research that has shown significant promise. These studies typically involve the collection of samples from two or more classes (e.g. cancer vs. normal, or responsive vs. non-responsive to treatment)

and the use of a set of data on which to train the classifier and another set on which to test. In the absence of a true test set, re-sampling methods (such as cross-validation) are generally used to estimate likely performance of the classifier on future data. The ability to generate an accurate classifier is a function of factors such as 1) the size of the training set relative to the number of features, 2) the computational method used, and 3) the inherent distinctness of the selected phenotypes. Typically, the number of samples is far less than the number of transcripts, and consequently over-fitting is a significant problem. This leads to the need for computational methods that avoid over-fitting when selecting a classifier. A variety of methods have been applied to cancer diagnoses including approaches based on support vector machines (Furey et al., 2000; Ramaswamy et al., 2001) and relative expression reversals (Geman et al., 2004; Price et al., 2007; Tan et al., 2005; Xu et al., 2005), among many others. Application of these methods has perhaps been applied most extensively to the study of cancer, leading to the discovery of molecular classifiers of varying degrees of accuracy to identify prognostic signatures for breast cancer (Adler and Chang, 2006; Buyse et al., 2006; Dai et al., 2005; Foekens et al., 2006; Glinsky et al., 2004; Goncalves et al., 2006; Ivshina et al., 2006; Liu et al., 2007; Ma et al., 2004; Nuyten et al., 2006; Park et al., 2007; Pawitan et al., 2005; Thomassen et al., 2007; van 't Veer et al., 2002; van de Vijver et al., 2002; Wang et al., 2005; Weigelt et al., 2005), ovarian cancer (De Cecco et al., 2004; Smith, 2002; Spentzos et al., 2004), colon cancer (Barrier et al., 2006; Giacomini et al., 2005), prostate cancer (Dhanasekaran et al., 2001; Glinsky et al., 2004; Halvorsen et al., 2005; Lin et al., 2005; Luo et al., 2002; Singh et al., 2002; Xu et al., 2005), and brain cancer (Fuller et al., 2005; Kim et al., 2002), among others.

B. Biological Networks and the Interactome

Among the most promising opportunities presented by systems biology is the integrated assessment of multiple biological parameters to magnify our understanding of complex developmental or pathological states. The interaction between all individual biomolecules in complex regulatory, signal transduction, and feedback networks present in eukaryotic organisms forms the theoretical framework for this higher-order analysis and is referred to as the ***interactome***. These dynamic associations include protein-protein interactions, transcriptional regulation, and post-transcriptional gene silencing by short-interfering RNA and microRNA (Ruvkun, 2001). Multiple high-throughput methods have been developed to characterize the interaction between proteins, including the **yeast-two-hybrid system** (Ito et al., 2001) and **surface plasmon resonance** (Usui-Aoki et al., 2005). These methods are to determining ***interactome*** interactions what microarray studies are to determining relative mRNA concentrations. Identifying multi-protein and multi-gene interactions is a complex computational and statistical task. Current studies often examine pair-wise interaction data between biomolecules, such as whether two proteins are positively correlated or negatively correlated *in vivo*. These techniques have been used to profile portions of the yeast, fly, worm, and human interactomes, aided by automated data mining programs such as *Reactome* (Joshi-Tope et al., 2005) and curated databases including BioGRID (Stark et al., 2006). Computational appraisal of these biological networks promises multi-variate markers of pathology where traditional, single-parameter studies are insufficient, and represents a promising new domain for the characterization, diagnosis and potential treatment of human disease.

C. Quantitative predictive models of biological systems

Quantitative network models in biology can be grouped into two broad classes, statistical inference networks and biochemical reaction networks (Price and Shmulevich, 2007). Statistical inference networks represent one approach to modeling biological networks that holds tremendous potential for advancing knowledge of biology. One approach is the use of statistical associations in high-throughput data types to infer predictive network models. These statistical inference networks differ from interaction networks in that the connections between different network components is accompanied by mathematical functions precisely relating how perturbations to one (or a set of) component(s) appears to effect the others, independent of the state of interaction. In essence, statistical inference networks are predictive and computable, while interaction networks as a rule are not. Interaction networks can, however, be an important – even essential – information source for the reconstruction of predictive statistical inference networks. The growing amounts of high-quality microarray gene expression data, coupled with input from gene-protein interaction networks, are now making it possible to make robust statistical inferences about the structure and dynamics of biomolecular control systems, such as transcriptional regulatory networks. Many approaches attempt to infer relationships between gene expression measurements using deterministic or stochastic formalisms. The fundamental idea behind these approaches is that models that faithfully capture such relationships have predictive capacity as regards system behavior and can be used to gain insight about system-wide properties, such as steady-state behavior or responses to perturbations or specific stimuli. Accurate network models that can predict the results of perturbations hold promise for enabling personalized medicine by allowing for the rational design of therapies and the identification of likely disease-causing perturbations.

When the necessary biochemical detail is known, biochemical reaction networks can be reconstructed – providing the basis for quantitative network models. These networks represent the underlying chemistry of the system, and thus represent stoichiometric relationships between inter-converted biomolecules. Biochemical reaction networks are directly based on chemistry, rather than a reflection of statistical associations, as are statistical inference networks. These stoichiometric reconstructions have been most commonly applied to small-molecule interconversions in metabolic networks, but this formalism can easily incorporate biological transformations of all types, including for metabolic, signaling, protein-protein interactions, and gene regulatory networks (Palsson, 2004). One enabling resource for large-scale quantitative modeling of metabolic networks to study human disease is the recent stoichiometric reconstruction of known human metabolism at the genome-scale (Duarte et al., 2007). These biochemical reaction networks can be used to make numerous quantitative simulations that have been shown previously to match well with experimental data in model organisms (Price et al., 2004). These successes with model organisms have also been extended to models of simple human systems such as the erythrocyte (Mulquiney and Kuchel, 2003; Price et al., 2004) and mitochondria (Thiele et al., 2005), with the global metabolic reconstruction poised to allow for larger human metabolic networks to now be modeled. Genome-scale models based on biochemical networks provide a comprehensive, yet concise, description of cellular functions, and provide means for quantitatively interrogating the effects on heath and disease of specific network perturbations and/or defined environmental changes.

D. Personalized and Predictive Genomics and Proteomics

The striking change in medical practice over the next 5–20 years will be the exponential increase

in the amount of biological data that will be gathered on individual patients at the DNA, RNA, protein, and cellular levels, all of which will be available for multi-parameter and multi-scale analyses. Our prediction is that billions of measurements will eventually be made over the lifetime of each individual. For example, in the future, each human will have his or her genome sequenced with 6 billion DNA bases (spanning both the maternal and paternal contributions to the individual's genome). The computational challenges associated with comparing billions of individual human genomes one to another to extract relevant differences and then to associate these differences with health and disease will pose a striking series of computational challenges. Moreover, perhaps 2,500 blood-protein measurements will be taken twice a year on each individual; and *in vivo* imaging analyses will be carried out, albeit less frequently. These data will be assimilated to generate a highly detailed picture of the evolving health/disease states of each individual.

Conclusions and Perspectives

Nanotechnology and microfluidic sensing chips with integrated microfluidic delivery devices will be developed over the next 5–10 years so that we reach a point at which thousands of proteins and/or RNAs can be analyzed quantitatively from only a fraction of a droplet of blood. This will allow physicians to assess the health/disease status of virtually every organ in the body at 6-month intervals and will become the first critical element of predictive medicine. We also predict that nanotechnology approaches to the sequencing of single DNA strands, will over the next 5-10 years lead to instrumentation that can rapidly, cheaply and simply sequence individual human genomes. This will be a second foundation of predictive medicine – assessing the variability in individual human genomes to generate probabilistic future health histories for each

individual human, and integrating it with diagnostics. Similarly, *in vivo* imaging agents will permit the functional visualization of informational molecules and drugs in humans, which will provide powerful and informative *in vivo* diagnostics approaches. These measurement technologies, together with both new computational approaches to extract information from these data and the systems view of medicine, will lead to three important revolutions in medicine.

First, over the next 10–20 years, a medicine that is predictive, personalized, preventive and participatory will emerge. Obtaining individual human genome sequences will lead to individual predictive health histories, whereas biannual blood measurements of thousands of proteins from the molecular fingerprints of human organs will give us a real-time assessment of individual health. Predictive medicine will result in a personalized medicine that focuses on the illnesses of an individual and eventually their wellness, preventing disease rather than treating it. The systems approach to disease will in time permit the identification of the key disease-perturbed sub-networks and the identification of important nodal points, which, if perturbed by drugs, could make the network behave in a more normal fashion or at least delete the more deleterious effects of the disease-perturbed network. This capability will provide a new approach for drug-target discovery, and a powerful and rapid new approach for developing drugs. Predictive, personalized and preventive medicine, if appropriately orchestrated with patient-oriented interpretations, will also enable patients to understand more deeply and actively participate in personal choices about illness and wellness. Participatory medicine will necessitate the development of powerful new approaches to handling enormous amounts of personal information in a secure manner, and to a new form of medical education of individual patients as well as their physicians. Over the next 5-20 years medicine will become predictive, personalize,

preventive and participatory (P4 medicine).

Another very exciting idea is that, in the long-term, systems analysis of blood offers the potential to open up a new avenue for studying human biology – when we learn to read and interpret the information inherent in secreted organ-specific protein patterns. The key patterns in these secreted proteins represent different network perturbations, and hence, different diseases. When secreted proteins enter the blood, they can provide a novel means to study biology in higher organisms and to identify drug targets through linking blood measurements to perturbations in underlying biological networks. Importantly, learning to study perturbations in underlying networks through secreted protein patterns (to whatever resolution is possible) will provide access to studying biological networks *in vivo* that are not amenable to direct experimentation. Developing the capacity to identify *in vivo* network perturbations through secreted protein measurements in the blood will open up a new avenue to drug target identification and will provide a novel means to discover the perturbed sub-networks. Since taking blood is relatively non-invasive it has less potential than biopsies or other invasive techniques of greatly distorting the system being studied. Thus, this approach has the potential to strongly complement existing approaches to study human biology. The effect of drugs, toxicity, human development and even aging may all be amenable to study through the blood if we can learn to read and interpret the signals in the proteins, which has the potential to be very significant in human biology.

Second, as the sensitivity of measurements increases (both *in vitro* and *in vivo*), we will achieve a digitalization of biology and medicine — that is, the ability to extract relevant information content from single individuals, single cells and ultimately single molecules — with its attendant

economies of scale. Just as Moore's law led in time to the widespread digitalization of information technologies and communications, the exponentially increasing ability to extract quantized biological information from individual cells and molecules will transform biology and medicine in ways that we can only begin to imagine.

Finally, all of these changes — a systems approach to medicine with its focus on disease prevention and more efficient drug discovery, the introduction of increasingly inexpensive nanotechnology based diagnostics and *in vivo* measurement technologies, the highly accurate and specific molecular characterization of the systems biology of disease, and the digitalization of medicine — will start to reduce the inexorably increasing costs of health care. This can, in principle, enable us to afford to provide for the more than 45 million people in the United States who currently lack health insurance, to reduce the crushing costs of healthcare to society, and to export our digital predictive and preventive medical approach to the developing world. Just as the mobile phone has become a fundamental communication mode in developing countries, and has changed the lives of much of the world's population, so digital medicine of the 21st century can bring to the world's citizens a global and strongly improved state of human health and healthcare. In our view P4 medicine will become the foundational framework for global medicine.

References

Adler, A.S. and Chang, H.Y. (2006) From description to causality: mechanisms of gene expression signatures in cancer, *Cell Cycle*, 5, 1148-1151.

Anderson, N.L. and Anderson, N.G. (2002) The human plasma proteome: history, character, and diagnostic prospects, *Mol Cell Proteomics*, 1, 845-867.

Bailey, R.C., Kwong, G.A., Radu, C.G., Witte, O.N. and Heath, J.R. (2007) DNA-encoded antibody libraries: a unified platform for multiplexed cell sorting and detection of genes and proteins, *J Am Chem Soc*, 129, 1959-1967.

Barrier, A., Boelle, P.Y., Roser, F., Gregg, J., Tse, C., Brault, D., Lacaine, F., Houry, S., Huguier, M., Franc, B., Flahault, A., Lemoine, A. and Dudoit, S. (2006) Stage II colon cancer prognosis prediction by tumor gene expression profiling, *J Clin Oncol*, 24, 4685-4691.

Buttner, P., Mosig, S., Lechtermann, A., Funke, H. and Mooren, F.C. (2007) Exercise affects the gene expression profiles of human white blood cells, *J Appl Physiol*, 102, 26-36.

Buyse, M., Loi, S., van't Veer, L., Viale, G., Delorenzi, M., Glas, A.M., d'Assignies, M.S., Bergh, J., Lidereau, R., Ellis, P., Harris, A., Bogaerts, J., Therasse, P., Floore, A., Amakrane, M., Piette, F., Rutgers, E., Sotiriou, C., Cardoso, F. and Piccart, M.J. (2006) Validation and clinical utility of a 70-gene prognostic signature for women with node-negative breast cancer, *J Natl Cancer Inst*, 98, 1183-1192.

Czernin, J. and Phelps, M.E. (2002) Positron emission tomography scanning: current and future applications, *Annu Rev Med*, 53, 89-112.

Dai, H., van't Veer, L., Lamb, J., He, Y.D., Mao, M., Fine, B.M., Bernards, R., van de Vijver, M., Deutsch, P., Sachs, A., Stoughton, R. and Friend, S. (2005) A cell proliferation signature is a marker of extremely poor outcome in a subpopulation of breast cancer patients, *Cancer Res*, 65, 4059-4066.

De Cecco, L., Marchionni, L., Gariboldi, M., Reid, J.F., Lagonigro, M.S., Caramuta, S., Ferrario, C., Bussani, E., Mezzanzanica, D., Turatti, F., Delia, D., Daidone, M.G., Oggionni, M.,

Bertuletti, N., Ditto, A., Raspagliesi, F., Pilotti, S., Pierotti, M.A., Canevari, S. and Schneider, C. (2004) Gene expression profiling of advanced ovarian cancer: characterization of a molecular signature involving fibroblast growth factor 2, *Oncogene*, 23, 8171-8183.

Dhanasekaran, S.M., Barrette, T.R., Ghosh, D., Shah, R., Varambally, S., Kurachi, K., Pienta, K.J., Rubin, M.A. and Chinnaiyan, A.M. (2001) Delineation of prognostic biomarkers in prostate cancer, *Nature*, 412, 822-826.

Duarte, N.C., Becker, S.A., Jamshidi, N., Thiele, I., Mo, M.L., Vo, T.D., Srivas, R. and Palsson, B.O. (2007) Global reconstruction of the human metabolic network based on genomic and bibliomic data, *Proc Natl Acad Sci U S A*, 104, 1777-1782.

Dupuy, A. and Simon, R.M. (2007) Critical review of published microarray studies for cancer outcome and guidelines on statistical analysis and reporting, *J Natl Cancer Inst*, 99, 147-157.

Foekens, J.A., Atkins, D., Zhang, Y., Sweep, F.C., Harbeck, N., Paradiso, A., Cufer, T., Sieuwerts, A.M., Talantov, D., Span, P.N., Tjan-Heijnen, V.C., Zito, A.F., Specht, K., Hoefler, H., Golouh, R., Schittulli, F., Schmitt, M., Beex, L.V., Klijn, J.G. and Wang, Y. (2006) Multicenter validation of a gene expression-based prognostic signature in lymph node-negative primary breast cancer, *J Clin Oncol*, 24, 1665-1671.

Fujii, K., Nakano, T., Kawamura, T., Usui, F., Bando, Y., Wang, R. and Nishimura, T. (2004) Multidimensional protein profiling technology and its application to human plasma proteome, *J Proteome Res*, 3, 712-718.

Fuller, G.N., Mircean, C., Tabus, I., Taylor, E., Sawaya, R., Bruner, J.M., Shmulevich, I. and Zhang, W. (2005) Molecular voting for glioma classification reflecting heterogeneity in the continuum of cancer progression, *Oncol Rep*, 14, 651-656.

Furey, T.S., Cristianini, N., Duffy, N., Bednarski, D.W., Schummer, M. and Haussler, D. (2000) Support vector machine classification and validation of cancer tissue samples using microarray expression data, *Bioinformatics*, 16, 906-914.

Gao, W.M., Kuick, R., Orchekowski, R.P., Misek, D.E., Qiu, J., Greenberg, A.K., Rom, W.N., Brenner, D.E., Omenn, G.S., Haab, B.B. and Hanash, S.M. (2005) Distinctive serum protein profiles involving abundant proteins in lung cancer patients based upon antibody microarray analysis, *BMC Cancer*, 5, 110.

Geman, D., d'Avignon, C., Naiman, D.Q. and Winslow, R.L. (2004) Classifying gene expression profiles from pairwise mRNA comparisons, *Stat Appl Genet Mol Biol*, 3, Article19.

Giacomini, C.P., Leung, S.Y., Chen, X., Yuen, S.T., Kim, Y.H., Bair, E. and Pollack, J.R. (2005) A gene expression signature of genetic instability in colon cancer, *Cancer Res*, 65, 9200-9205.

Glinsky, G.V., Glinskii, A.B., Stephenson, A.J., Hoffman, R.M. and Gerald, W.L. (2004) Gene expression profiling predicts clinical outcome of prostate cancer, *J Clin Invest*, 113, 913-923.

Glinsky, G.V., Higashiyama, T. and Glinskii, A.B. (2004) Classification of human breast cancer using gene expression profiling as a component of the survival predictor algorithm, *Clin Cancer Res*, 10, 2272-2283.

Golub, T.R., Slonim, D.K., Tamayo, P., Huard, C., Gaasenbeek, M., Mesirov, J.P., Coller, H., Loh, M.L., Downing, J.R., Caligiuri, M.A., Bloomfield, C.D. and Lander, E.S. (1999) Molecular classification of cancer: class discovery and class prediction by gene expression monitoring, *Science*, 286, 531-537.

Goncalves, A., Esterni, B., Bertucci, F., Sauvan, R., Chabannon, C., Cubizolles, M., Bardou, V.J., Houvenaegel, G., Jacquemier, J., Granjeaud, S., Meng, X.Y., Fung, E.T., Birnbaum, D., Maraninchi, D., Viens, P. and Borg, J.P. (2006) Postoperative serum proteomic profiles may

predict metastatic relapse in high-risk primary breast cancer patients receiving adjuvant chemotherapy, *Oncogene*, 25, 981-989.

Haab, B.B., Dunham, M.J. and Brown, P.O. (2001) Protein microarrays for highly parallel detection and quantitation of specific proteins and antibodies in complex solutions, *Genome Biol*, 2, RESEARCH0004.

Halvorsen, O.J., Oyan, A.M., Bo, T.H., Olsen, S., Rostad, K., Haukaas, S.A., Bakke, A.M., Marzolf, B., Dimitrov, K., Stordrange, L., Lin, B., Jonassen, I., Hood, L., Akslen, L.A. and Kalland, K.H. (2005) Gene expression profiles in prostate cancer: association with patient subgroups and tumour differentiation, *Int J Oncol*, 26, 329-336.

Hood, L., Heath, J.R., Phelps, M.E. and Lin, B. (2004) Systems biology and new technologies enable predictive and preventative medicine, *Science*, 306, 640-643.

Hu, Z., Hood, L. and Tian, Q. (2007) Quantitative Proteomic Approaches for Biomarker Discovery, *Proteomics*, In review.

Huber, W. and Mueller, F. (2006) Biomolecular interaction analysis in drug discovery using surface plasmon resonance technology, *Curr Pharm Des*, 12, 3999-4021.

Ito, T., Chiba, T., Ozawa, R., Yoshida, M., Hattori, M. and Sakaki, Y. (2001) A comprehensive two-hybrid analysis to explore the yeast protein interactome, *Proc Natl Acad Sci U S A*, 98, 4569-4574.

Ivshina, A.V., George, J., Senko, O., Mow, B., Putti, T.C., Smeds, J., Lindahl, T., Pawitan, Y., Hall, P., Nordgren, H., Wong, J.E., Liu, E.T., Bergh, J., Kuznetsov, V.A. and Miller, L.D. (2006) Genetic reclassification of histologic grade delineates new clinical subtypes of breast cancer, *Cancer Res*, 66, 10292-10301.

Joshi-Tope, G., Gillespie, M., Vastrik, I., D'Eustachio, P., Schmidt, E., de Bono, B., Jassal, B., Gopinath, G.R., Wu, G.R., Matthews, L., Lewis, S., Birney, E. and Stein, L. (2005) Reactome: a knowledgebase of biological pathways, *Nucleic Acids Res*, 33, D428-432.

Kim, S., Dougherty, E.R., Shmulevich, I., Hess, K.R., Hamilton, S.R., Trent, J.M., Fuller, G.N. and Zhang, W. (2002) Identification of combination gene sets for glioma classification, *Mol Cancer Ther*, 1, 1229-1236.

Koga, H., Kyo, M., Usui-Aoki, K. and Inamori, K. (2006) A chip-based miniaturized format for protein-expression profiling: the exploitation of comprehensively produced antibodies, *Electrophoresis*, 27, 3676-3683.

Lathrop, J.T., Anderson, N.L., Anderson, N.G. and Hammond, D.J. (2003) Therapeutic potential of the plasma proteome, *Curr Opin Mol Ther*, 5, 250-257.

Lee, H.J., Lee, E.Y., Kwon, M.S. and Paik, Y.K. (2006) Biomarker discovery from the plasma proteome using multidimensional fractionation proteomics, *Curr Opin Chem Biol*, 10, 42-49.

Lin, B., White, J.T., Lu, W., Xie, T., Utleg, A.G., Yan, X., Yi, E.C., Shannon, P., Khrebtukova, I., Lange, P.H., Goodlett, D.R., Zhou, D., Vasicek, T.J. and Hood, L. (2005) Evidence for the presence of disease-perturbed networks in prostate cancer cells by genomic and proteomic analyses: a systems approach to disease, *Cancer Res*, 65, 3081-3091.

Liu, R., Wang, X., Chen, G.Y., Dalerba, P., Gurney, A., Hoey, T., Sherlock, G., Lewicki, J., Shedden, K. and Clarke, M.F. (2007) The prognostic role of a gene signature from tumorigenic breast-cancer cells, *N Engl J Med*, 356, 217-226.

Luo, J., Dunn, T., Ewing, C., Sauvageot, J., Chen, Y., Trent, J. and Isaacs, W. (2002) Gene expression signature of benign prostatic hyperplasia revealed by cDNA microarray analysis, *Prostate*, 51, 189-200.

Lyon, L.A., Musick, M.D. and Natan, M.J. (1998) Colloidal Au-enhanced surface plasmon resonance immunosensing, *Anal Chem*, 70, 5177-5183.

Ma, X.J., Wang, Z., Ryan, P.D., Isakoff, S.J., Barmettler, A., Fuller, A., Muir, B., Mohapatra, G., Salunga, R., Tuggle, J.T., Tran, Y., Tran, D., Tassin, A., Amon, P., Wang, W., Wang, W., Enright, E., Stecker, K., Estepa-Sabal, E., Smith, B., Younger, J., Balis, U., Michaelson, J., Bhan, A., Habin, K., Baer, T.M., Brugge, J., Haber, D.A., Erlander, M.G. and Sgroi, D.C. (2004) A two-gene expression ratio predicts clinical outcome in breast cancer patients treated with tamoxifen, *Cancer Cell*, 5, 607-616.

Miller, J.C., Zhou, H., Kwekel, J., Cavallo, R., Burke, J., Butler, E.B., Teh, B.S. and Haab, B.B. (2003) Antibody microarray profiling of human prostate cancer sera: antibody screening and identification of potential biomarkers, *Proteomics*, 3, 56-63.

Mulquiney, P.J. and Kuchel, P.W. (2003) *Modelling metabolism with Mathematica detailed examples including erythrocyte metabolism*. CRC Press, Boca Raton, Fla.

Nuyten, D.S., Kreike, B., Hart, A.A., Chi, J.T., Sneddon, J.B., Wessels, L.F., Peterse, H.J., Bartelink, H., Brown, P.O., Chang, H.Y. and van de Vijver, M.J. (2006) Predicting a local recurrence after breast-conserving therapy by gene expression profiling, *Breast Cancer Res*, 8, R62.

Olle, E.W., Sreekumar, A., Warner, R.L., McClintock, S.D., Chinnaiyan, A.M., Bleavins, M.R., Anderson, T.D. and Johnson, K.J. (2005) Development of an internally controlled antibody microarray, *Mol Cell Proteomics*, 4, 1664-1672.

Orchekowski, R., Hamelinck, D., Li, L., Gliwa, E., vanBrocklin, M., Marrero, J.A., Vande Woude, G.F., Feng, Z., Brand, R. and Haab, B.B. (2005) Antibody microarray profiling reveals

individual and combined serum proteins associated with pancreatic cancer, *Cancer Res*, 65, 11193-11202.

Palsson, B. (2004) Two-dimensional annotation of genomes, *Nat Biotechnol*, 22, 1218-1219.

Papadopoulou, E., Davilas, E., Sotiriou, V., Georgakopoulos, E., Georgakopoulou, S., Koliopanos, A., Aggelakis, F., Dardoufas, K., Agnanti, N.J., Karydas, I. and Nasioulas, G. (2006) Cell-free DNA and RNA in plasma as a new molecular marker for prostate and breast cancer, *Ann N Y Acad Sci*, 1075, 235-243.

Park, E.S., Lee, J.S., Woo, H.G., Zhan, F., Shih, J.H., Shaughnessy, J.D. and Frederic Mushinski, J. (2007) Heterologous tissue culture expression signature predicts human breast cancer prognosis, *PLoS ONE*, 2, e145.

Pawitan, Y., Bjohle, J., Amler, L., Borg, A.L., Egyhazi, S., Hall, P., Han, X., Holmberg, L., Huang, F., Klaar, S., Liu, E.T., Miller, L., Nordgren, H., Ploner, A., Sandelin, K., Shaw, P.M., Smeds, J., Skoog, L., Wedren, S. and Bergh, J. (2005) Gene expression profiling spares early breast cancer patients from adjuvant therapy: derived and validated in two population-based cohorts, *Breast Cancer Res*, 7, R953-964.

Price, N.D., Reed, J.L. and Palsson, B.O. (2004) Genome-scale models of microbial cells: evaluating the consequences of constraints, *Nat Rev Microbiol*, 2, 886-897.

Price, N.D., Schellenberger, J. and Palsson, B.O. (2004) Uniform sampling of steady-state flux spaces: means to design experiments and to interpret enzymopathies, *Biophys J*, 87, 2172-2186.

Price, N.D. and Shmulevich, I. (2007) Biochemical and statistical network models for systems biology, *Curr Opin Biotechnol*, 18, 365-370.

Price, N.D., Trent, J., El-Naggar, A.K., Cogdell, D., Taylor, E., Hunt, K.K., Pollock, R.E., Hood, L., Shmulevich, I. and Zhang, W. (2007) Highly accurate two-gene classifier for differentiating

gastrointestinal stromal tumors and leiomyosarcomas, *Proc Natl Acad Sci U S A*, 104, 3414-3419.

Quackenbush, J. (2006) Microarray analysis and tumor classification, *N Engl J Med*, 354, 2463-2472.

Ramaswamy, S., Tamayo, P., Rifkin, R., Mukherjee, S., Yeang, C.H., Angelo, M., Ladd, C., Reich, M., Latulippe, E., Mesirov, J.P., Poggio, T., Gerald, W., Loda, M., Lander, E.S. and Golub, T.R. (2001) Multiclass cancer diagnosis using tumor gene expression signatures, *Proc Natl Acad Sci U S A*, 98, 15149-15154.

Ross, P.L., Huang, Y.N., Marchese, J.N., Williamson, B., Parker, K., Hattan, S., Khainovski, N., Pillai, S., Dey, S., Daniels, S., Purkayastha, S., Juhasz, P., Martin, S., Bartlet-Jones, M., He, F., Jacobson, A. and Pappin, D.J. (2004) Multiplexed protein quantitation in Saccharomyces cerevisiae using amine-reactive isobaric tagging reagents, *Mol Cell Proteomics*, 3, 1154-1169.

Ruvkun, G. (2001) Molecular biology. Glimpses of a tiny RNA world, *Science*, 294, 797-799.

Sanchez-Carbayo, M., Socci, N.D., Lozano, J.J., Haab, B.B. and Cordon-Cardo, C. (2006) Profiling bladder cancer using targeted antibody arrays, *Am J Pathol*, 168, 93-103.

Scherzer, C.R., Eklund, A.C., Morse, L.J., Liao, Z., Locascio, J.J., Fefer, D., Schwarzschild, M.A., Schlossmacher, M.G., Hauser, M.A., Vance, J.M., Sudarsky, L.R., Standaert, D.G., Growdon, J.H., Jensen, R.V. and Gullans, S.R. (2007) Molecular markers of early Parkinson's disease based on gene expression in blood, *Proc Natl Acad Sci U S A*.

Simon, R. (2005) Roadmap for developing and validating therapeutically relevant genomic classifiers, *J Clin Oncol*, 23, 7332-7341.

Singh, D., Febbo, P.G., Ross, K., Jackson, D.G., Manola, J., Ladd, C., Tamayo, P., Renshaw, A.A., D'Amico, A.V., Richie, J.P., Lander, E.S., Loda, M., Kantoff, P.W., Golub, T.R. and

Sellers, W.R. (2002) Gene expression correlates of clinical prostate cancer behavior, *Cancer Cell*, 1, 203-209.

Smith, D.I. (2002) Transcriptional profiling develops molecular signatures for ovarian tumors, *Cytometry*, 47, 60-62.

Solanky, K.S., Bailey, N.J., Beckwith-Hall, B.M., Davis, A., Bingham, S., Holmes, E., Nicholson, J.K. and Cassidy, A. (2003) Application of biofluid 1H nuclear magnetic resonance-based metabonomic techniques for the analysis of the biochemical effects of dietary isoflavones on human plasma profile, *Anal Biochem*, 323, 197-204.

Song, S., Li, B., Wang, L., Wu, H., Hu, J., Li, M. and Fan, C. (2007) A cancer protein microarray platform using antibody fragments and its clinical applications, *Mol Biosyst*, 3, 151-158.

Spentzos, D., Levine, D.A., Ramoni, M.F., Joseph, M., Gu, X., Boyd, J., Libermann, T.A. and Cannistra, S.A. (2004) Gene expression signature with independent prognostic significance in epithelial ovarian cancer, *J Clin Oncol*, 22, 4700-4710.

Stark, C., Breitkreutz, B.J., Reguly, T., Boucher, L., Breitkreutz, A. and Tyers, M. (2006) BioGRID: a general repository for interaction datasets, *Nucleic Acids Res*, 34, D535-539.

Sun, B., Ranish, J.A., Utleg, A.G., White, J.T., Yan, X., Lin, B. and Hood, L. (2007) Shotgun glycopeptide capture approach coupled with mass spectrometry for comprehensive glycoproteomics, *Mol Cell Proteomics*, 6, 141-149.

Tan, A.C., Naiman, D.Q., Xu, L., Winslow, R.L. and Geman, D. (2005) Simple decision rules for classifying human cancers from gene expression profiles, *Bioinformatics*, 21, 3896-3904.

Thiele, I., Price, N.D., Vo, T.D. and Palsson, B.O. (2005) Candidate metabolic network states in human mitochondria. Impact of diabetes, ischemia, and diet, *J Biol Chem*, 280, 11683-11695.

Thomassen, M., Tan, Q., Eiriksdottir, F., Bak, M., Cold, S. and Kruse, T.A. (2007) Prediction of metastasis from low-malignant breast cancer by gene expression profiling, *Int J Cancer*, 120, 1070-1075.

Usui-Aoki, K., Shimada, K., Nagano, M., Kawai, M. and Koga, H. (2005) A novel approach to protein expression profiling using antibody microarrays combined with surface plasmon resonance technology, *Proteomics*, 5, 2396-2401.

van 't Veer, L.J., Dai, H., van de Vijver, M.J., He, Y.D., Hart, A.A., Mao, M., Peterse, H.L., van der Kooy, K., Marton, M.J., Witteveen, A.T., Schreiber, G.J., Kerkhoven, R.M., Roberts, C., Linsley, P.S., Bernards, R. and Friend, S.H. (2002) Gene expression profiling predicts clinical outcome of breast cancer, *Nature*, 415, 530-536.

van de Vijver, M.J., He, Y.D., van't Veer, L.J., Dai, H., Hart, A.A., Voskuil, D.W., Schreiber, G.J., Peterse, J.L., Roberts, C., Marton, M.J., Parrish, M., Atsma, D., Witteveen, A., Glas, A., Delahaye, L., van der Velde, T., Bartelink, H., Rodenhuis, S., Rutgers, E.T., Friend, S.H. and Bernards, R. (2002) A gene-expression signature as a predictor of survival in breast cancer, *N Engl J Med*, 347, 1999-2009.

Wang, X., Yu, J., Sreekumar, A., Varambally, S., Shen, R., Giacherio, D., Mehra, R., Montie, J.E., Pienta, K.J., Sanda, M.G., Kantoff, P.W., Rubin, M.A., Wei, J.T., Ghosh, D. and Chinnaiyan, A.M. (2005) Autoantibody signatures in prostate cancer, *N Engl J Med*, 353, 1224-1235.

Wang, Y., Klijn, J.G., Zhang, Y., Sieuwerts, A.M., Look, M.P., Yang, F., Talantov, D., Timmermans, M., Meijer-van Gelder, M.E., Yu, J., Jatkoe, T., Berns, E.M., Atkins, D. and Foekens, J.A. (2005) Gene-expression profiles to predict distant metastasis of lymph-node-negative primary breast cancer, *Lancet*, 365, 671-679.

Weigelt, B., Hu, Z., He, X., Livasy, C., Carey, L.A., Ewend, M.G., Glas, A.M., Perou, C.M. and Van't Veer, L.J. (2005) Molecular portraits and 70-gene prognosis signature are preserved throughout the metastatic process of breast cancer, *Cancer Res*, 65, 9155-9158.

Wilson, J.F. (2006) The rocky road to useful cancer biomarkers, *Ann Intern Med*, 144, 945-948.

Xu, L., Tan, A.C., Naiman, D.Q., Geman, D. and Winslow, R.L. (2005) Robust prostate cancer marker genes emerge from direct integration of inter-study microarray data, *Bioinformatics*, 21, 3905-3911.

Yang, S., Undar, A. and Zahn, J.D. (2006) A microfluidic device for continuous, real time blood plasma separation, *Lab Chip*, 6, 871-880.

Zimmermann, M., Delamarche, E., Wolf, M. and Hunziker, P. (2005) Modeling and optimization of high-sensitivity, low-volume microfluidic-based surface immunoassays, *Biomed Microdevices*, 7, 99-110.