Early Diagnosis of Complex Diseases by Molecular Biomarkers, Network Biomarkers, and Dynamical Network Biomarkers

***Rui Liu,*[1] *Xiangdong Wang,*[2] *Kazuyuki Aihara,*[3] *and Luonan Chen*[3,4]**

[1]Department of Mathematics, South China University of Technology, Guangzhou, 510640, China
[2]Zhongshan Hospital, Fudan University, Shanghai, 200031, China
[3]Collaborative Research Center for Innovative Mathematical Modelling, Institute of Industrial Science, University of Tokyo, Tokyo, 153–8505, Japan
[4]Key Laboratory of Systems Biology, SIBS-Novo Nordisk Translational Research Centre for PreDiabetes, Shanghai Institutes for Biological Sciences, Chinese Academy of Sciences, Shanghai, 200031, China

Published online 17 June 2013 in Wiley Online Library (wileyonlinelibrary.com).
DOI 10.1002/med.21293

▼

Abstract: Many studies have been carried out for early diagnosis of complex diseases by finding accurate and robust biomarkers specific to respective diseases. In particular, recent rapid advance of high-throughput technologies provides unprecedented rich information to characterize various disease genotypes and phenotypes in a global and also dynamical manner, which significantly accelerates the study of biomarkers from both theoretical and clinical perspectives. Traditionally, molecular biomarkers that distinguish disease samples from normal samples are widely adopted in clinical practices due to their ease of data measurement. However, many of them suffer from low coverage and high false-positive rates or high false-negative rates, which seriously limit their further clinical applications. To overcome those difficulties, network biomarkers (or module biomarkers) attract much attention and also achieve better performance because a network (or subnetwork) is considered to be a more robust form to characterize diseases than individual molecules. But, both molecular biomarkers and network biomarkers mainly distinguish disease samples from normal samples, and they generally cannot ensure to identify predisease samples due to their static nature, thereby lacking ability to early diagnosis. Based on nonlinear dynamical theory and complex network theory, a new concept of dynamical network biomarkers (DNBs, or a dynamical network of

Contract grant sponsor: NSFC; *Contract grant numbers*: 61134013, 91029301, 61072149, and 11241002; *Contract grant sponsor*: Chief Scientist Program of SIBS of CAS; *Contract grant number*: 2009CSP002.

Correspondence to: Xiangdong Wang, Zhongshan Hospital, Fudan University, Shanghai 200031, China. E-mail: xiangdong.wang@clintransmed.org.

Kazuyuki Aihara, Collaborative Research Center for Innovative Mathematical Modelling, Institute of Industrial Science, University of Tokyo, Tokyo 153–8505, Japan. E-mail: aihara@sat.t.u-tokyo.ac.jp.

Luonan Chen, Key Laboratory of Systems Biology, SIBS-Novo Nordisk Translational Research Centre for PreDiabetes, Shanghai Institutes for Biological Sciences, Chinese Academy of Sciences, Shanghai 200031, China. E-mail: lnchen@sibs.ac.cn.

biomarkers) has been developed, which is different from traditional static approaches, and the DNB is able to distinguish a predisease state from normal and disease states by even a small number of samples, and therefore has great potential to achieve "real" early diagnosis of complex diseases. In this paper, we comprehensively review the recent advances and developments on molecular biomarkers, network biomarkers, and DNBs in particular, focusing on the biomarkers for early diagnosis of complex diseases considering a small number of samples and high-throughput data (or big data). Detailed comparisons of various types of biomarkers as well as their applications are also discussed.
Med. Res. Rev., 34, No. 3, 455–478, 2014

Key words: complex diseases; molecular biomarker; network biomarker; dynamical network biomarker; early diagnosis.

1. INTRODUCTION

Evidence suggests that there is usually a drastic change during disease progression, which results in the critical transition from a normal/stable state to a disease state of a living organism.[1–5] Therefore, as shown in Figure 1A, disease progression can generally be expressed by three stages, that is, a normal/stable state, a predisease state, and a disease state. A normal/stable state is a relatively "healthy" stage that includes the chronic inflammation period or the stable period in which the disease is under control, whereas a predisease state is actually the limit of the normal/stable state just before the critical transition. At this predisease stage, the state is considered to be reversible to the normal/stable state if an appropriate treatment is performed, and thus is unrobust. However, if the system moves over the critical point to the disease state, it becomes very difficult to be reversed to the normal/stable state even by advanced medical treatment. Therefore, it is important to distinguish the predisease state from the normal and diseased ones (or achieve early diagnosis) so as to take the prevention action at appropriate timing, which saves not only the human life but also medical resources. In the present study, we simply use a "normal state" to represent a "normal/stable state."

On the other hand, a biomarker is objectively measured and evaluated to indicate normal biological processes, pathogenic processes, or pharmacologic responses to a therapeutic intervention.[6] Specifically, a disease biomarker is an indicator to distinguish the disease state from the normal state, and its value is correlated with the disease-associated specificity, sensitivity, traceability, stability, repeatability, and reliability.[7–9] The research of biomarkers has already been one of the central topics in biological and medical fields. Translational and clinical research fundamentally depends on specific and accurate biomarkers. Discovery and identification of innovative biomarkers are also valuable and crucial for the successful development and validation of novel therapeutics. In particular, recent rapid advance of high-throughput technologies provides unprecedented rich information to characterize genotypes and phenotypes of diseases in a global and also dynamical manner, which significantly accelerates the study of biomarkers from both theoretical and clinical perspectives. Molecular biomarkers (e.g., genes, RNAs, proteins, and metabolites) are widely adopted in clinical practices due to the simplicity of measurement and implementation, for example, prostate-specific antigen (PSA),[10–12] BRCA mutations for breast cancer,[13,14] and expression profiles (such as serum protein electrophoresis) for detecting monoclonal gammopathies.[15] However, molecular biomarkers generally suffer from low coverage and high false-positive rates (or even high false-negative rates) due to complications and variations of genetic, epigenetic, and environment factors in the initiation and progression process of diseases, which seriously limit their further clinical applications to diagnosis and prognosis. To overcome those difficulties, network biomarkers (or module biomarkers)[16,17] attract much attentions and also achieve better performance because a network (or subnetwork)

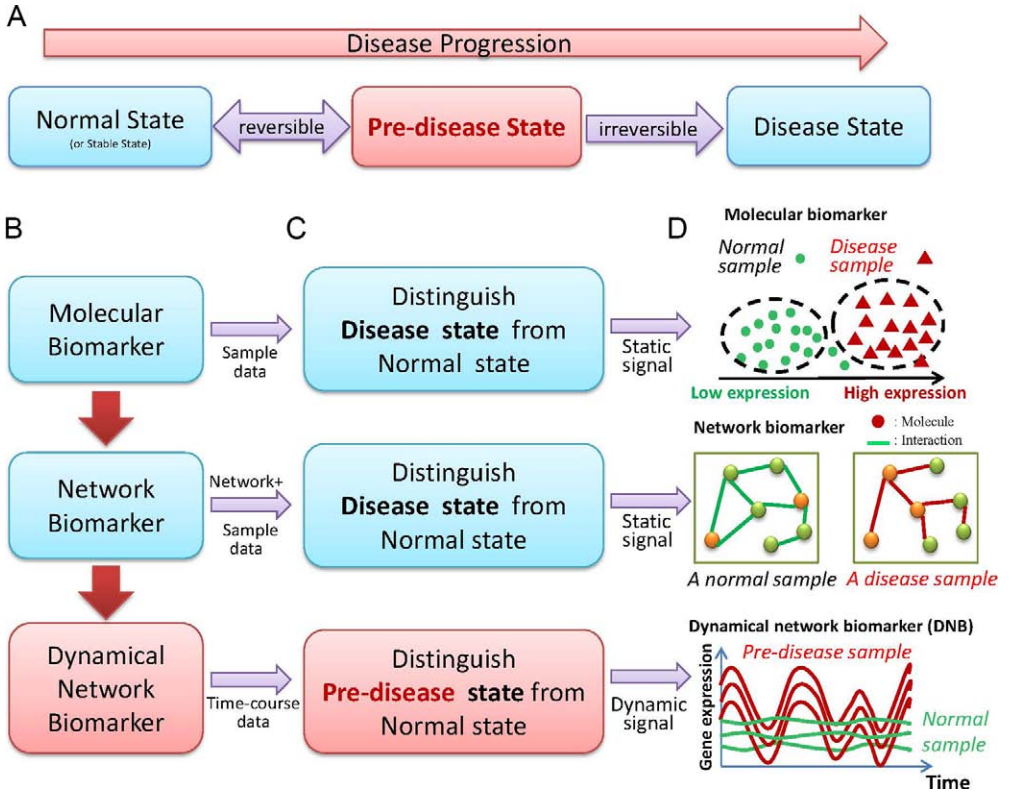

Figure 1. Disease states and biomarkers. (A) Three stages during disease progression, that is, a normal/stable state, a predisease state and a disease state. A normal/stable state is a relatively healthy stage including the chronic inflammation period or the stable period during which the disease is under control, whereas a predisease state is the limit of the normal/stable state just before the critical transition. At this stage, the predisease state is considered to be reversible to the normal/stable state if appropriately treated. However, if the system passes over the critical point to the disease state, it usually becomes irreversible to the normal/stable state. (B) Three types of biomarkers, that is, traditional molecular biomarkers, recent developed network biomarkers, and newly developed dynamical network biomarkers (DNBs). (C) Main purposes of the three types of biomarkers. Both molecular and network biomarkers are indicators on the disease state, whereas the DNBs signal the predisease state. (D) Major features of the three biomarkers. Clearly, both molecular and network biomarkers are static measurements on the disease, whereas DNBs are dynamical measurements on the predisease, thus, providing the early-warning signals for the predisease state. In this paper, a "normal state" means a "normal/stable state" for the purpose of simplicity.

is considered to be a more robust form to characterize diseases than individual molecules. But, both molecular and network biomarkers mainly distinguish disease samples from normal samples, which generally per se hardly identify predisease samples due to their static nature, thereby lacking ability to early diagnosis, as shown in Figure 1B–D. Unlike these traditional static approaches, a new concept of dynamical network biomarkers (DNBs, or a dynamical network of biomarkers)[18,19] has been developed on the basis of nonlinear dynamical theory and complex network theory. The DNB is able to fundamentally distinguish a predisease state from normal and disease states even by a small number of samples, and therefore has great potential to achieve "real" early diagnosis of complex diseases. Figure 1 illustrates major features of the three types of biomarkers. Note that a DNB is a group of molecules, which are highly fluctuating but strongly correlated without consistent values at the predisease stage, and thus it is a concept different from the conventional biomarkers, which are required to keep consistent values for the respective disease and normal samples. Also, a DNB has been shown to be the

leading network that makes the critical change first and thus drives the whole system into the disease state[18,19] and, therefore, is highly related to causal or driving factors (or genes) of the disease.

In this paper, we review the recent advances and developments on molecular biomarkers, network biomarkers, and DNBs of complex diseases, focusing on computational methods. Detailed comparisons of DNBs and traditional biomarkers as well as their applications are also presented. In addition to the three types of biomarkers shown in Figure 1, there are also other forms of biomarkers, for example, pathophysiologic status (such as patient performance status),[20] mammographic images,[21] and cell-based markers (e.g., circulating tumor cells),[22] which are beyond the scope of this paper, and the readers are suggested to refer to the related publications for the details on those biomarkers.

2. MOLECULAR BIOMARKER

Genes, RNAs, proteins, and metabolites are known as biological molecules, which are basic entities to interact with each other for performing various biological functions in a cell. With the rapid advance of high-throughput technologies at the molecular level, large amounts of data from genomics, proteomics, and metabolomics have been generated to tackle challenging problems in biomedical sciences, and further to provide new ways for studying diseases, by characterizing phenotypes, making early diagnosis, and developing niche-targeting drugs in a systematic manner.

A. Biomarkers at Molecular Level

Molecular biomarkers are quantifiable molecular measurements of biological homeostasis that aim to distinguish the disease state (which represents the stage of badly ill) from the normal state (which represents a relatively healthy stage, an incubation period, or a chronic inflammation stage), as shown in Figure 1. For example, the PSA, *kallikrein-3*, is used as an effective molecular biomarker to routinely screen for prostate inflammation and cancer.[23–25] Another example of molecular biomarker is *ERBB2*, which is a transforming cell growth factor, and the expression of ERBB2 is recognized to be associated with an aggressive phenotype of breast cancer. Molecular biomarkers are generally employed to indicate a specific disease state or a phenotype, based on the obviously different molecular features of the disease state from those of a normal state, which is also the basis of the diagnosis through molecular biomarkers.[26,27] Generally, the detection of new molecular biomarkers is based on their common property, that is, the expression profiles of the biomarkers should show distinct difference between a disease (or an abnormal) state and a normal state, which makes the classification or comparison of molecular expression profiles an important approach. Figure 2 shows a general framework to detect molecular biomarkers from high-throughput data. Obviously, the expressions of the molecular biomarkers should clearly reflect the severity or presence of the illness at a disease state such that the expression of a biomarker is significantly higher or lower in the disease state than that in a normal state (see Fig. 2D). Besides, from the viewpoint of clinical applications, the number of biomarkers for a specific disease is required to be as small as possible. Another important feature is that the molecular biomarkers need to be highly specific for each complex disease, since maintaining high specificity or low false-positive rate is of a high priority for disease sample screening.[28]

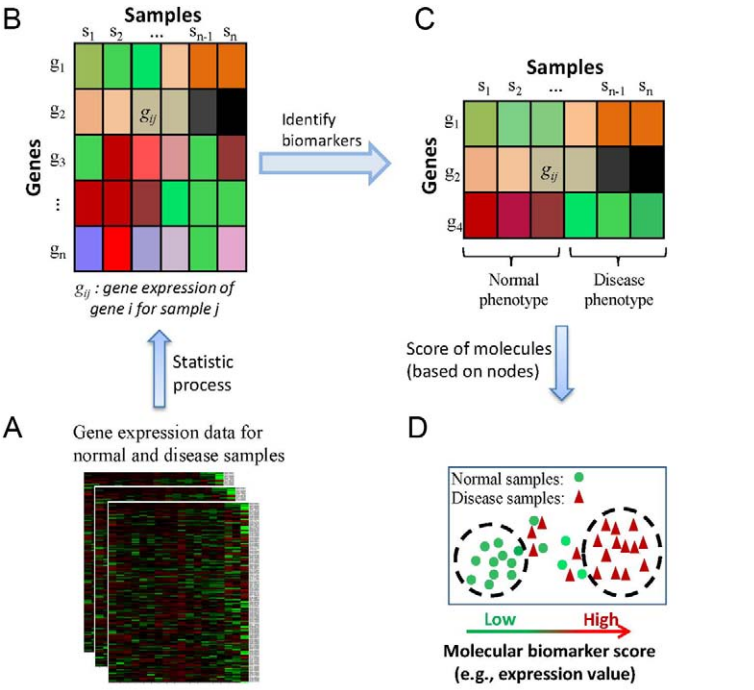

Figure 2. Molecular biomarkers. (A) Expression data of normal and disease samples (e.g., gene, protein, or metabolite expression data). (B) Expression profiles by statistical implementation (e.g., normalization) for both normal and disease samples. (C) Detection of molecular biomarkers, which show distinct difference between a disease state and a normal state. (D) Classification and validation of the detected molecular biomarkers.

B. Computational Methods for Identifying Molecular Biomarkers

Since a biomarker is a key indicator for specific diagnosis and reliable prognosis of a disease as well as therapy scheduling and monitoring, a great number of articles are published every year for the purpose of identifying novel biomarkers from both experimental and computational aspects.[29,30] Table I lists major computational methods and their important features for discovering molecular biomarkers based on high-throughput data. The main task of identifying molecular biomarkers is to find a number of molecules whose expressions can classify the disease and normal samples in a clear way, or can determine a clear boundary between the disease and normal samples. Specifically, multivariate logistic regression analysis is a classic approach to identify candidates of molecular biomarkers. The decision boundary derived from logistic regression is defined by an affine function of individual molecules, that is, a weighted sum plus a constant term,[12,31–40] which can identify the important candidate molecules or biomarkers responsible for the phenotypes. However, the effectiveness is highly dependent on the sample distribution, that is, variables should follow the multinomial distribution, which limits further clinical application. Classification and regression trees (CARTs) were also applied to detect the molecular biomarkers by many researchers.[41–48] CART is a training data-driven approach, by which the model needs not to assume any particular form for its decision boundary, and is considered as one of the powerful tools for developing nonlinear classification models on disease and normal samples. The limitation of this method comes from the complexity of computations during the construction of the tree, especially when there are a large number of

Table I. Overview of Computational Methods to Detect Molecular Biomarkers of Complex Diseases

Computational methods	Description	Advantages	Disadvantages	Representative references
Multivariate analysis (logistic regression)	Combine multiple measurements of the markers (a few markers or the same marker in different samples) into a single valued score	The method is easy to interpret, and possible to test global statistical significance	The effectiveness degrades when the distribution of sampling data deviates from the normal distribution	12,31–39
Classification and regression tree (CART)	Recursively partition the training dataset and construct groups of molecular biomarkers (multimarker) directly based on logical combinations of disease phenotype characteristics	It is unnecessary to assume any particular form of decision boundary, and easy to identify useful molecular groups that are based on simple combinations of clinical characteristics	It is required to perform hundreds of statistical comparisons during the construction of the tree, thus, sometimes the model is likely to diverge	41–48
Voting panel approach	Produce a positive or negative result by individually using the cutoff value for each of multiple clinical inputs	The method is very simple, and easy to operate	The method is not accurate by frequently using the voting scheme	49–53
Artificial neural network (ANN)	Construct a network of simple information-processing artificial neurons by arranging the elements in a particular interconnection pattern	It can be used to approximate any functional form from observed data with quick and parallel computation	It is sometimes difficult to select an appropriate ANN design with the right modeling capacity	54–69
Supporting vector machine (SVM)	Separate two classes of data in a high-dimensional space by hyperplanes based on some selected nonlinear functions (kernels)	The learning process often involves a straightforward solution to an optimization problem	The effectiveness is limited by the choice of the kernel functions	50
Genetic algorithm (GA)	Perform randomized search and optimization mimicking evolutional and natural genetics	It is a powerful tool for discovering a set of functions that best define properties of candidate genes	The method is less effective and has lower likelihood of convergence when there are a large number of genes	66–70

nodes. Besides, since CART is mainly based on heuristic algorithms, such as the greedy algorithm where locally optimal decisions are made at each node, it cannot guarantee to obtain the globally optimal decision tree with the best classification. Voting panel method is very simple and easy to be operated, and it can directly produce positive or negative results from control and disease samples by individually using the cutoff value for each of the multiple clinical inputs. Results from individual molecules are then combined using a mixture of logic "AND" or "OR" operators,[49–53] by which it can straightforwardly indicate the classification of samples. However, the classification is not accurate by frequently using voting scheme, if the sampling scale is large. Artificial neural networks (ANNs) also received much attentions as a nonlinear modeling tool from clinical diagnostics to theoretical understanding of the mechanics.[54–59] They are composed of simple information-processing elements, that is, artificial neurons, by

arranging them in particular interconnection patterns. This method provides information to rank the importance of molecules, and thus identifies the disease-related molecules. With the correct implementation, ANNs can be applied naturally in a large dataset. However, the classification process is not so straightforward and the robustness strongly relies on the appropriate selection of cost function and learning algorithm. The machine learning method, such as supporting vector machine (SVM), is widely applied on the engineering field. It is a supervised method that has been recently applied to derive molecular biomarkers for biomedical applications,[60–65] by separating the disease samples from normal ones in a high-dimensional space based on some selected nonlinear functions (i.e., kernels). The largest bottleneck of SVM is the parameters tuning, which is considered crucial in the procedure of training, and so experience dependent and fastidious that it would greatly influence the performances. Genetic algorithm (GA) performs randomized search and optimization mimicking evolutional and natural genetics, which is a powerful tool for discovering a set of functions that best define properties of genes, and thus has many applications in classifying disease samples and further finding disease-related molecules.[66–70] The shortcoming of GA is not so efficient with repeated evaluation process. In addition, classifying the samples by GA is often encumbered by exponentially increasing in search space size. Recently, the dimensionality reduction approach as a new algorithm was developed to discover biomarkers by classifying biomadical data, in order to distinguish a set of lower dimensional samples from higher dimensional expression datasets.[71,72] In addition, there are several methods or tools classifying a wide range of high-dimensional biomedical data, such as risk stratification approach[73–75] and heterogeneous expression profile analysis,[76,77] which are also potentially effective to biomarker discovery and assessment. For high-throughput data or big biological data, DNA microarray data at the gene level have been widely exploited to find molecular biomarkers on various diseases, including breast cancer,[78,79] brain cancer,[80] pancreatic cancer,[81] and acute leukemias.[82] On the other hand, at the protein and metabolite levels, proteomics and metabonomics data have been used to identify new biomarkers of many diseases,[83] such as bladder cancer,[84] diabetes mellitus,[85] and toxicity screening.[86] Recently in contrast to single-level data, multilevel data for a single sample (or disease; with DNA sequence, DNA methylation, gene expression, protein expression, and metabolite expression) become available (e.g., in TCGA database), which stimulates the study of the integration[87,88] of multilevel and multisource data to identify the biomarkers of complex diseases. The major features of those computational methods are shown in Table I.

3. NETWORK BIOMARKER

Although molecules are basic components of a cellular machinery, a complex disease is generally caused not from the malfunction of individual molecules but from the interplay of a group of correlated molecules or a network.[89] In fact, a disease is the result of cell's or tissue's response to its microenvironment, and such response is usually not influenced by single molecules, but by complex interactions of many signaling pathways and molecular subnetworks.[90–92] In the past years, rapidly developing technology allows us to obtain gene (or protein) expressions and other high-dimensional profile data at the genome-wide scale,[93] that is, with over thousands of measurements in each sample including SNP (sequence data), gene expression (transcriptome), mass spectrum (proteome), and small molecules (metabolome) data in different levels. The availability of such high-throughput data has already driven the integrative research by describing complex phenomena to studying essential design principles, and by studying individual components to understanding functional modules or networks for biomolecular systems, such as cells, tissues, organs, and even the entire organism.[94–96] Therefore, to better diagnose a disease state, researchers proposed to study the combinations or a relatively large group of interacting

molecules to deeply understand the complex interplay and pathways of multiple molecules. From a network point of view, groups of interacting molecules with similar behavior, that is, network biomarkers or module biomarkers provide a quantifiable and also stable form to characterize biomedical phenotypes or diseases in contrast to individual molecular biomarkers, which has inspired the development of systems medicine in the network level.[93,96–98]

A. Biomarkers at Network Level

The network biomarker was first proposed in 2008.[16] The similar name of "subnetwork markers" was raised even earlier in 2007.[99] The concept of network biomarkers was established by the development of genomic high-throughput technology and the system-wide and multidimensional studies on molecule expression profiles of disease progression. Specifically, these techniques, such as microarray and mass spectrometry (MS) technology, can simultaneously screen the whole human genome in terms of RNA transcripts or proteins.[100] Based on the rapidly accumulated high-throughput datasets, the protein–protein interaction (PPI) networks have been constructed for many complex diseases, and played a central role in studying the pathway regulation of these diseases, which provides new perspectives to accurately and robustly classify disease samples on the basis of the rich information including both biomedical knowledge and topological structure. Therefore, the development of network biomarkers is mainly based on the available PPI networks and the signaling pathways. For instance, by applying to a cardiovascular related protein network, Jin et al.[16] identified some molecules, which compose a network with a set of high-confident interacting proteins that can classify two groups of patients more accurately than the former single molecules without consideration of biological molecular interaction. It is believed that some molecular interactions in such a network are activated under specific conditions, and thus may indicate the dysfunctional process underlying the corresponding disease phenotypes. Therefore, some key subnetworks with dysfunctional pathways of protein interactions (or gene regulations, or biochemical reactions) associated with certain disease are also called network biomarkers, which are able to distinguish the disease state in a more accurate manner. A typical procedure to identify network biomarkers is shown in Figure 3. Clearly, compared to Figure 2, network biomarkers focus on interacting molecules rather than individual molecules.

B. Network Biomarkers Based on Expression Data

The task to find reliable network biomarkers depends on the high-quality information of interactions among molecules as well as the available expression data for specific disease and control samples. Ideker et al. proposed that a type of active modules, which are connected subnetworks and whose genes show significant correlated changes in an mRNA-expression (or other) state under particular experimental conditions, can be taken as biomarkers.[101] Decomposing a whole network into active modules not only reduces network complexity, but also helps to find signaling pathways, based on which an open source software Cytoscape is developed and widely applied in studying protein–protein, protein–DNA, and genetic interactions that are increasingly available for humans and model organisms.[102] Some efficient methods were also proposed[103,104] to detect the active modules in a molecular interaction network. Combined with the benefits of high-throughput sampling, many research works[16,99,105–110] show that network biomarkers are potential candidates of clinical trials for complex diseases. Actually, many effective network biomarkers have been detected for complex diseases, for example, breast cancer[99,110] and gastric cancer.[111] Unlike conventional expression clustering or classification methods, network-based analyses could identify molecules that are not differentially expressed. Specifically, if the overall activity of a molecule is lowly expressed, it would not attract attentions through the

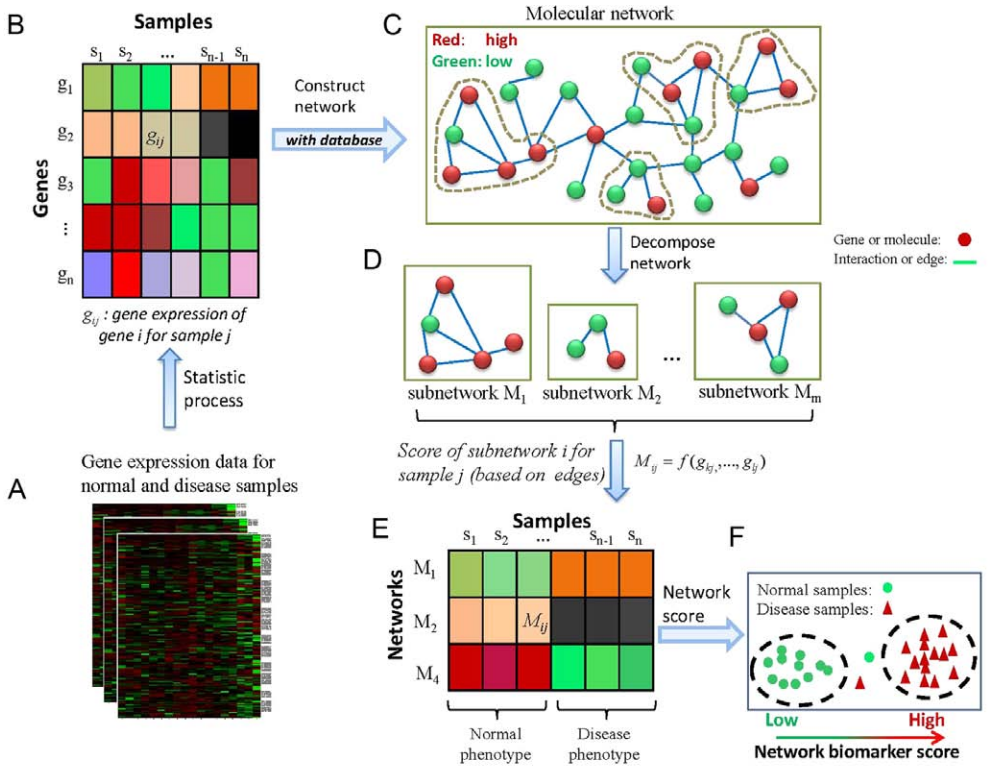

Figure 3. Network biomarkers. Schematic representation shows the main procedures to detect network biomarkers form high-throughput data, which distinguish distinct genotypes and phenotypes to indicate the disease state. From the high-throughput expression profiles (A), we obtain the normalized expression profiles (B) from which we can further construct the molecular interaction network by combining available interaction information and expression data (C). Some key subnetworks in (D) are identified for a certain disease by decomposing the molecular interaction network. (E) Those subnetworks are called network biomarkers if they can distinguish a disease phenotype from a normal phenotype to indicate the disease state. (F) The identified network biomarkers classifies the disease and normal samples in a more accurate manner comparing with individual molecular biomarkers selected without consideration of the network or interactions.

conventional expression comparison methods. However, if such molecules participate in a significant subnetwork that shows a distinct phenotype in the disease or abnormal stage, then they are essential for maintaining the module integrity to meet the requirement of interconnecting many higher scoring molecules (see Fig. 3). In such sense, this property is important for the discovery of disease-causing genes, because the phenotypic changes may be regulated not by individual expressions but by their collective behavior in the network. Recently, more and more works and methods are developed for studying the regulation pathways and their roles in the disease state. For instance, active subnetwork identification is a method for identifying active subnetworks by using existing PPI networks. Such active subnetworks, which are strongly connected regions of the whole network, show significant changes in expression over a particular subset of the conditions, and can classify the disease samples in a systematic way, thereby distinguishing the disease phenotype.[99] A similar method is referred to in Ref. 112, where Lee et al. proposed that the disease phenotype can be identified by pathway classifiers based on pathway activities, whose level is summarized from the gene expression levels of its condition responsive genes, defined as the subset of genes in the pathway whose combined expression provides

optimal discriminative power for identifying the disease phenotype. Another systems biology approach, that is, the disease-specific pathway identification method, was developed. It aims at extracting disease-specific subnetworks or pathways by using regression models or scoring modules, for example, carcinogenesis relevance values in Ref. 113 and modularity score in Ref. 114. This approach is effective in identifying network biomarkers and disease-specific dysregulated pathways based on the integration of PPIs, pathway knowledge, and graph information.[113–116] However, the regression model-based method is not suitable for the small sample cases, in which the parameters are biased. Classification of differential interactions is a new method based on the analyses of differential interactions between disease and normal samples, in contrast to the clustering method based on differential gene (or protein) expressions that are widely employed in conventional methods. This method was recently proposed and applied to discover module biomarkers for diseases, and the successful application on gastric cancer suggests that the differential interactions are effective on identifying dysfunctional modules from the molecular interaction network, which can be applied as network biomarkers.[111] Modeling the information flow is an approach for identifying dysfunctional modules in complex diseases, which models the information flow from source disease genes to targets of differentially expressed genes via a context-specific PPI network. The dysregulated pathways are sorted out as subnetworks from the pathway interaction networks. Such subnetworks effectively characterize the functional dependency or crosstalk between pathways, and thus are capable of distinguishing disease samples from normal ones.[118–120] The SVM-based method has been used for separating two groups of molecules, and recently is applied in detecting network biomarkers, by identifying a comprehensive key interaction map and integrating different types of interaction information of other species (heterogeneous data sources) within the SVM scheme.[121–123] The major shortcoming of classical SVM is its high computational cost for the constrained optimization programming. Classification of detecting topological changes in biological networks is an efficient and straightforward method in identifying a condition-specific local network under different biological conditions. This type of classification, for example differential dependency network (DDN) analysis, compares the topological differences between any two networks,[124,125] which is a powerful method for distinguishing disease samples when the topology of disease networks is significantly different from the topology of normal samples. In addition to detecting disease states, some recent works also identify network biomarkers in order to predict disease outcome, such as the research work by Taylor et al.,[126] in which functional module markers are picked out by observing substantial differences in the biochemical structure between different types of hubs that are usually associated with oncogenesis. The major features of computational methods for network biomarker are listed in Table II.

C. Network Biomarkers Based on Sequence Data

Recent genome sequencing studies as well as genome-wide association studies (GWASs) have drastically expanded the knowledge on the relations between sequences and diseases, which enables us to integrate sequence data to discover new network biomarkers or functional module biomarkers. An example is the studies on integration of networks with single-nucleotide polymorphism (SNP) data for disease association, which becomes a rapidly growing trend and can be referred to a number of interesting works that develop various approaches of association tests and bridge the pathways and gene-oriented analyses of genome-wide association (GWA).[128] Analyzing GWA at the level of multiple SNPs enablesdetection of the cumulative effect of many variants within a biological pathway that may act additively to determine disease susceptibility.[129–131] Association tests by using imputed genotypes at many SNPs to facilitate comparisons with the results of other GWA scans allow geneticists to accurately evaluate the evidence for association at genetic markers that are not directly genotyped.[132,133] Applying

Table II. Overview of Computational Methods to Detect Network Biomarkers of Complex Diseases

Detection methods	Description	Advantages	Disadvantages	Representative references
Active subnetwork identification method	Identify active subnetworks by using available large-scale protein–protein interaction (PPI) networks. Both scoring for nodes and aggregate scoring for active subnetworks are used	The method can identify some disease-related genes that are not differentially expressed	The method is limited by the availability of interaction networks and can be used only when there is an activity *P*-value for every measurement	99, 112
Disease-specific module or pathway identification method	Extract disease-related subnetworks by identifying the key modules via regression models or scoring pathways	The method is effective in generating static models of disease-specific modules or signal transduction pathways	The method is time consuming due to exhaustive search procedure	113–115
Classification of differential interactions	Investigate differential interactions (between disease and control samples) and network rewiring between molecules related to pathogenesis	The method can identify crucial modules that cannot be detected in differential expression of genes by discovering information of the molecular interactions	The method is time consuming for large-scale interaction networks	111, 117
Modeling the information flow approach	Determine dysfunctional modules by modeling the information flow from source disease genes to targets of differentially expressed genes via a context-specific PPI network	The method is effective in finding dysregulated modules with dysfunctional pathways	The method is insensitive when the genes are not differentially expressed	118–120
Supporting vector machine (SVM)-based method	The kernel methods are employed to integrate network and expression data for the classification, and applied to identify disease states by detecting disease-related modules or disease-associated subnetworks	By using various biological knowledge and data sources (e.g., gene coexpression, regulatory networks, evolutionary relationship, and functional similarity), the effectiveness and efficiency are significantly improved	The effectiveness is limited by the choice of the kernel functions and the computational cost is high for real applications	14, 89, 103–105
Classification of detecting topological changes in biological network (network comparison method)	The network topology-based approaches (such as differential dependency network [DDN] analysis) to estimate statistically significant topological changes in the disease networks between different biological stages	The method can straightforwardly identify the local network under different biological conditions, compare the difference between any two networks, and thus is efficient when the network topology significantly differs	The method is not convenient in realistic applications, since this approach runs into the difficulty that the network structure learning can be inconsistent with a limited number of data samples	124, 125, 127

pathway-based association approaches, Wang et al.[134] identified most significant gene sets and pathways related to diseases, which is one of the first studies to propose the use of pathway information in GWA studies. Such pathway-based association approaches not only widen the application in GWA studies of complex diseases, but also suggest a new way to find molecular networks and cellular pathways that can mark disease states.

Using massive genome sequencing data, the pathway-based analysis also generates various methods and algorithms that can be employed to identify network biomarkers for complex diseases.[135] Vandin et al.[136] proposed an efficient algorithm for identifying significantly mutated pathways in cancer, which can detect the subnetworks in a genome-scale gene interaction network that are mutated in a statistically significant number of patients. Such pathway-based analysis is able to rescue true disease-related molecules from a list of nondifferential expression but involved in a high-scored pathway. Nguyen et al.[137] applied a new label-free quantitation method to assemble high-density temporal data and study cellular signaling pathways.

Figure 3 schematically illustrates how to detect network biomarkers from high-throughput expression data and available network information. In the procedure, a whole molecular network (Fig. 3B), which is constructed from high-throughput expression profiles and molecular interaction data, is decomposed into multiple subnetworks based on topological structures or other specific conditions, by network classification methods, for example, scoring each subnetwork according to its activity. From the scores of subnetworks on available disease and control samples, optimization or classification methods are generally used to identify candidate network biomarkers, which can detect distinct phenotypes between normal and disease samples. Many methods based on sequence data use the similar procedure shown in Figure 3 to identify biomarkers by further integrating sequence information.

4. DYNAMICAL NETWORK BIOMARKER

Both molecular biomarkers and network biomarkers aim to diagnose the disease state, rather than the predisease state before a critical transition, as shown in Figures 1 and 4. To achieve the early diagnosis on a complex disease, it is very important to detect early-warning signals of the predisease state so as to prevent the drastic deterioration, which is a common phenomenon in many complex diseases. The progression and development of a complex disease is often modeled as a nonlinear dynamical system, or a dynamical network. However, in contrast to the detection of the disease state for many diseases, it is usually a very difficult task to identify the predisease state as the state of the system may show little distinct change before the critical point or the predisease state is really reached. In other words, there may be no noticeable change between a normal and a predisease state. This is the reason that results in the failure of diagnosis based on traditional molecular biomarkers or static network biomarkers. To overcome this problem, a general theory and methodology to detect early-warning signal was recently proposed based on a model-free concept, that is, a dynamical subnetwork of biomarkers or a DNB, which can identify a predisease state[18] even with only a few of samples provided that high-throughput data are available for each sample. Figure 4 schematically illustrates the major dynamical features of DNB as well as its main differences from the traditional biomarkers.

A. Biomarkers at Dynamical Network Level

The DNB[18] was proposed as a general early-warning indicator based on a new concept, that is, a dynamical subnetwork of biomarkers, which appears only in the predisease stage and is proven to satisfy some measurable conditions.

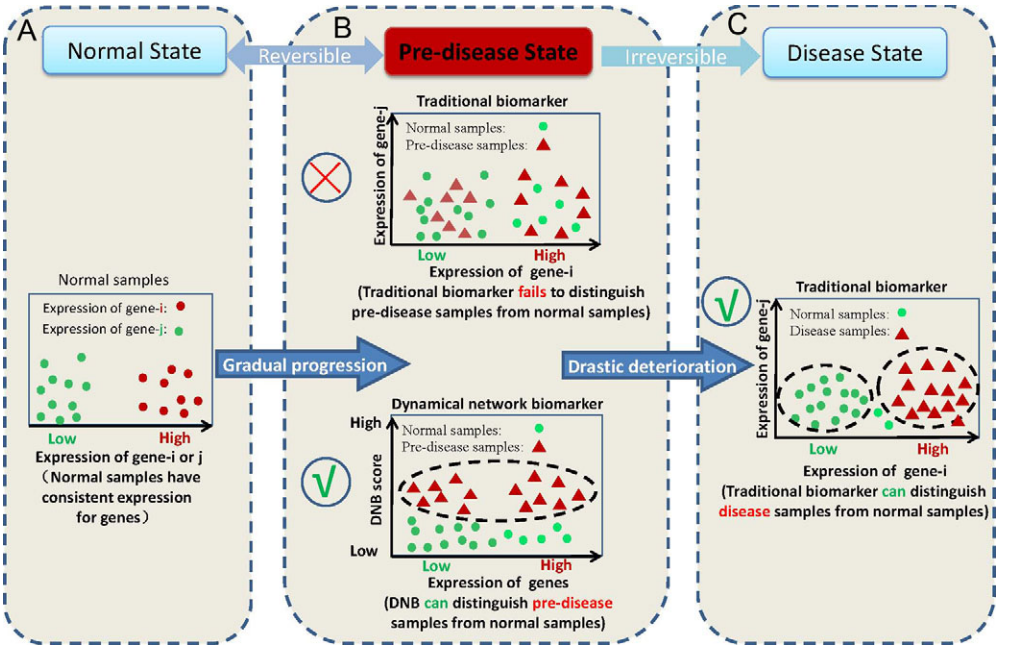

Figure 4. Dynamical network biomarkers. A schematic illustration of dynamical features for a disease progression from a normal state to a disease state through a predisease state. (A) The normal state is a steady state, where the system generally is stably and robustly regulated by its molecular network, and may be robust. In this state, the expression of genes are usually consistent. (B) The predisease state is defined a limit of the normal state and situated before the imminent phase transition point is reached. At this stage, traditional methods or biomarkers fail to distinguish the predisease samples from normal samples. However, DNB scorings/criteria are effective in distinguishing the predisease samples. DNB is dynamical measurements on the predisease, thus, providing the early-warning signals for the predisease state. (C) Beyond the critical point, the system abruptly deteriorates and enters the disease state. The disease state is the other steady state, during which the system is regulated by a disease network and also may be robust. The traditional biomarkers can distinguish disease samples from normal samples at this stage.

In particular, it can be theoretically proved that, when the system is near the critical point, there exists a dominant group or the so-called DNB, which is a group of molecules satisfying the following three conditions:

1. The correlation between any pair of members in DNB becomes very strong (e.g., the Pearson correlation coefficient [PCC] of their expression drastically increases).
2. The correlation between one member of DNB and any other molecule of non-DNB becomes very weak (e.g., PCC of their expression drastically decreases).
3. Any member of DNB becomes highly fluctuating (e.g., the standard deviation [SD] of its expression drastically increases).

In other words, DNB is an observable subnetwork of the original system, and composed of a special group of molecules that are strongly and dynamically correlated when the system is in a predisease state, according to the first condition. The second condition implies that DNB molecules behave almost independent of other non-DNB molecules although they are in the same system or network, that is, the DNB is indeed an isolated subnetwork or functional module, with all its members behaving dynamically in a strongly collective manner in the predisease state. The third condition implies that the expressions of these DNB molecules increasingly and strongly fluctuate as the system is approaching the critical state or point. It

is this dynamical property that makes the traditional molecular biomarkers or static network biomarkers fail to identify the predisease phenotype in the early stage. Therefore, regardless of disease types and personal variations, the three conditions are considered as essential criteria to identify the DNB that in turn indicates early-warning signals of the predisease state. In addition, these properties also hold in many complex diseases as well as many biological processes with sudden transition phenomena.[138]

In order to detect a reliable and clear signal of the predisease state, a composite index was proposed by combining the three criteria[18]:

$$I = \frac{SD_{\mathrm{d}} \cdot PCC_{\mathrm{d}}}{PCC_{\mathrm{o}}}$$

where PCC_{d} is the average Pearson's correlation coefficient among the molecules of DNB (or equivalently the dominant group) in absolute value; PCC_{o} is a factor representing the average Pearson's correlation coefficient of the molecules of DNB with the other molecules in absolute value; SD_{d} stands for the average standard deviation of the molecules of DNB. During any sampling interval in the predisease state, despite the stochastically fluctuation in the expression of each molecule, the composite index I is able to provide a reliable and also significant early-warning signal of a complex disease when the biological system approaches the predisease state/critical point, according to the three properties of DNB above.

B. Detecting the Leading Network by High-Throughput Data

DNB is also the leading network of the critical transition, which makes the critical change first and thus drives the whole system into a disease state through a predisease state.[19] That is, the leading network can be viewed as the subnetwork that first moves over the critical point into the disease state, and thus has strong relationship with the causal genes (or driving factors) or with the disease network in contrast to those consequent differentially expressed genes resulting from the disease. Therefore, identifying this leading network during a critical transition can not only indicate the emergence of the predisease state, but also reveal the underlying pathogenesis and mechanism of the disease initiation as well as progression at the network level.[138–140]

Generally, reliable identification of the leading network and the predisease stage from a large number of genes as well as from many stages by high-throughput data is, however, a very difficult problem because of widely existing noise in data and usually a small number of samples. In addition, it is a costly computational work to find the leading network by satisfying the three criteria due to a large number of variables of high-throughput data. Therefore, effective computational methods are demanded to detect DNB in a reliable and efficient manner, so as to accurately identify the predisease state and further elucidate the mechanism of the sudden deterioration, which is still an open problem. In order to demonstrate how DNB can reveal the imminent catastrophic transition and achieve the early diagnosis of complex diseases, in Figure 5, we show a successful application of DNB on a specific complex disease, that is, the acute lung injury driven by carbonyl chloride inhalation.[141] By applying the three criteria and DNB classification scheme on the microarray data (GSE2565), a group of observable molecules were screened out, which form a strong correlated subnetwork just before the occurrences of sudden deterioration and thus provide a reliable early-warning signal.[18] Specifically, we use time-course (from 0 to 72 hr) microarray data of mice for the lung injury with carbonyl chloride inhalation exposure. Clearly, DNB and its members show little distinct differences comparing with other genes at all sampling points except at 8 hr in Figure 5. At 8 hr, there is a strong signal from DNB (Fig. 5D), which indicates the imminent deterioration of the lung injury (or the predisease state) due to the exposure to the gas (although there is no injury in the lung at this

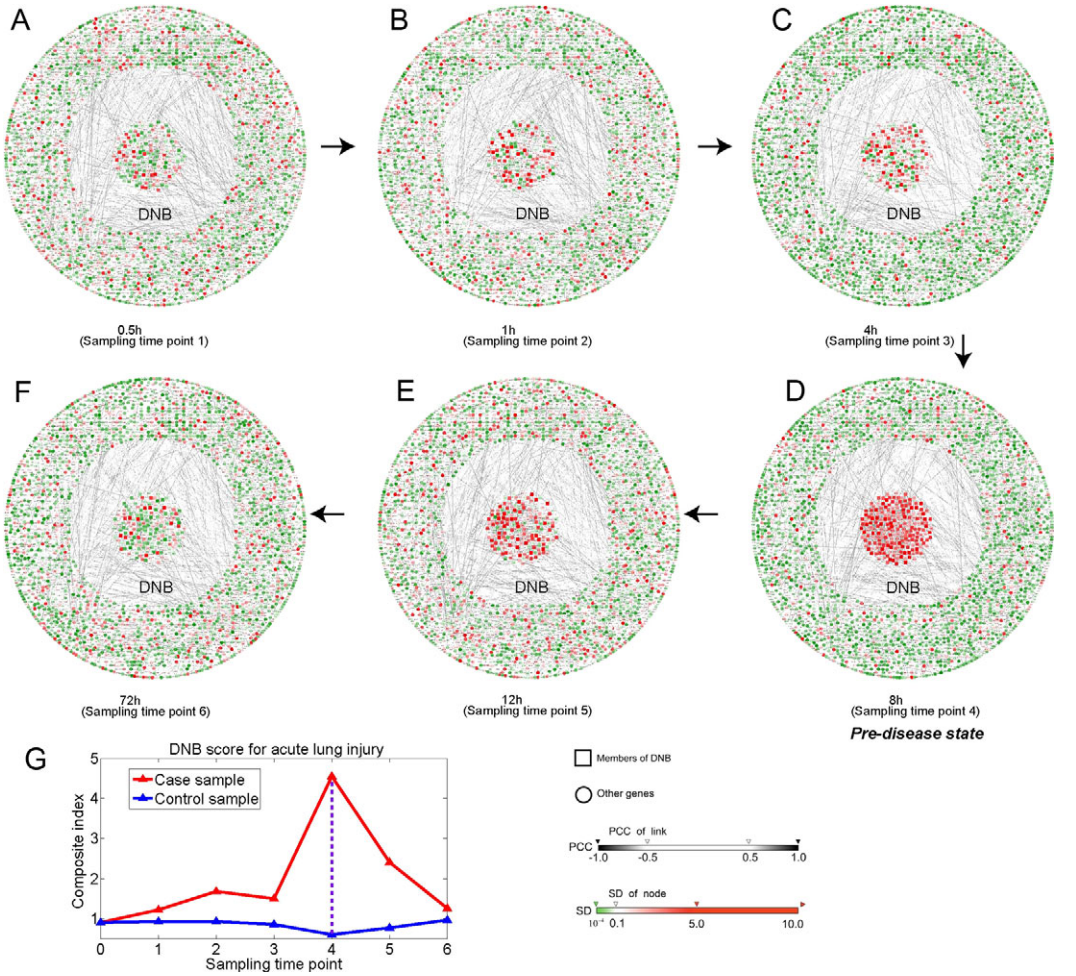

Figure 5. Detection of early-signal for acute lung injury by dynamical network biomarker (DNB). (A–F) Show the dynamical evolution of the whole mouse network (3452 genes and 9238 links) including the detected DNB during the disease progression (from 0.5 to 72 hr). The network was constructed from the whole mapped mouse molecular network (PPIs and transcription factors (TF)-target regulations) based on the expression data. The thickness of each edge represents the PCC value between a pair of molecules, while the color of each node represents the SD value of a molecule. The identified DNB is placed in the center of the whole network. The predisease state or DNB was detected at 8 hr, at which there is a strong signal (D) to indicate the imminent deterioration of the disease. (G) Shows the composite index of DNB. The dotted purple line indicates the predisease period (at sampling time point 4, i.e., 8 hr). The red curve represents the case group, while the blue one represents the control group. The composite index increase drastically from sampling time point 3 (4 hr), and reaches the peak at sampling time point 4 (8 hr). This fact strongly suggests that the predisease state is near sampling time point 4, and the system is driven into the disease state after sampling time point 4, which is consistent with the experimental results.[18,141] Clearly, DNB is not only wildly fluctuated, but also strongly correlated subnetwork at the predisease or critical state. Also, see the reference[18] for the original analysis.

timing). Actually, it was confirmed that the phenotypic changes (injury) for the lung occurred at the next time point (12 hr; Fig. 5E), which validated the effectiveness of DNB for the early diagnosis on the predisease state. Interestingly, however, after the system passes the critical point and is in a disease state, for example, at 72 hr (Fig. 5F), the members in DNB behave similarly to other genes again, without any significant difference from other genes. In other words, the DNB distinguishes not the disease state from the normal state, but the predisease state from the

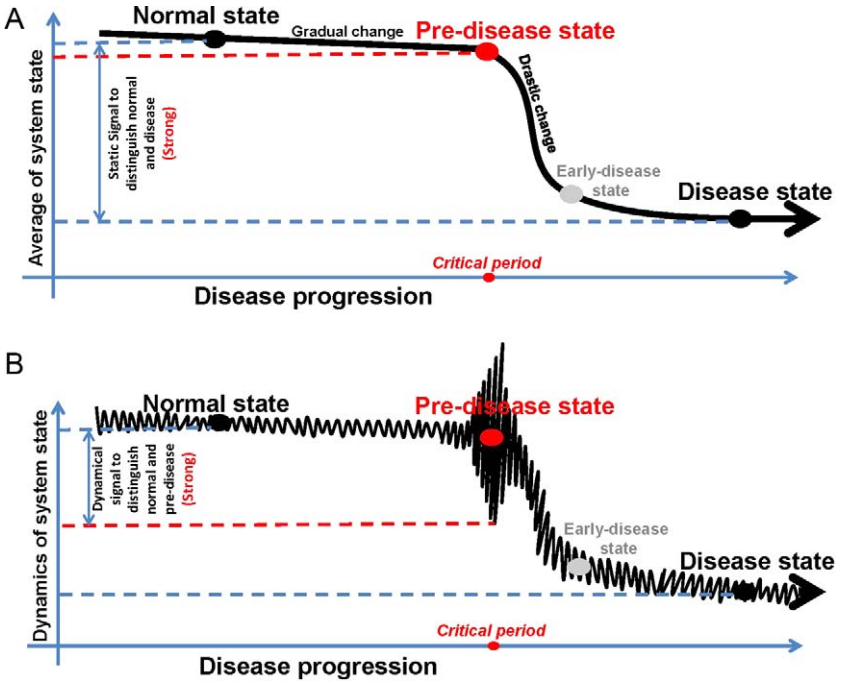

Figure 6. Static and dynamical signals during disease progression. (A) The average value of a molecule (e.g., gene or protein expression) at each stage during disease progression is considered as a static signal, which is used in the traditional molecular or network biomarkers. Such a signal is able to distinguish a disease state from a normal state but cannot clearly distinguish a predisease state from a normal state. (B) The dynamics of a molecule at each stage during disease progression is a dynamical signal, which is used in dynamical network biomarker (DNB). This signal is able to clearly identify a predisease state from a normal state, and therefore can be used for early diagnosis of a disease. Note that the static signal (A) is the average value of the dynamical signal (B) at each stage.

normal state. Figure 5G shows the strong signal of the DNB based on the composite index I at 8 hr (or sampling time point 4). More examples for demonstrating the DNB for early diagnosis of diseases can be found in Refs. 18, 19, 139, 140.

Figure 6 illustrates the static and dynamical signals during the progression of a disease to clearly describe the difference between the traditional biomarkers and DNB. Figure 6A shows a curve of the average value of a molecule (e.g., gene or protein expression) at each stage during the disease progression, which is considered as a static signal. Such a signal is able to identify a disease state, and thus is used in the traditional molecular or network biomarkers. But the static signal cannot clearly tell the difference between a predisease state and a normal state so as to make early diagnosis of the illness or disease. Figure 6B demonstrates dynamics of a molecule at each stage during the disease progression, which is a dynamical signal and used in DNB. This signal is able to clearly discriminate a predisease state from a normal state, and thereby can be used for early diagnosis of a disease. Figure 6A is the average value of the dynamical signal (i.e., Fig. 6B) at each stage. It is noteworthy that due to individual variations, each patient may not have exactly the same leading network or DNB even for the same disease, that is, some molecule members in the DNB may differ from person to person.[18] Hence, unlike molecular biomarkers and network biomarkers, a DNB does not always contain a group of fixed members even for the same disease but might have different molecules depending on individual variations that can be identified by high-throughput data of each individual. Comparing with the traditional molecular and network biomarkers, DNB has obvious advantages. First, DNB is used for

detecting the predisease state instead of the disease state, and thus provides the early signal of a disease. Second, since DNB is based on a model-free method and further can be obtained by a small number of samples, it is relatively easy to be implemented clinically. In addition, although a DNB is now used for detecting the predisease state, theoretically it can be used to any biological process to detect the critical transition as well as the leading network of the related phenotype provided that there is a drastic change during the process, for example, switching behavior of cell-differentiation processes, aging processes, and phase changes of cell cycle or circadian rhythm. It also opens a new way to explore the information from big biological data to understand the underlying mechanism of complex biological behaviors. Moreover, in addition to complex diseases, since DNB is conceptually a strongly correlated but also wildly fluctuated subnetwork (or group), the concept of a DNB would be generally applicable for detecting early-warning signals of critical transitions or bifurcations to a wide class of complex networks/systems.

5. CONCLUSION

Several decades of intensive research have discovered many molecular biomarkers on various diseases that are useful in the diagnosis, characterization, and therapy selection of complex disease. Considering the biomarker at a system level, that is, the network biomarker. Further this biomarker in a dynamical manner, that is, the DNB, would greatly advance the understanding of complex diseases by identifying the dynamical relationships of molecules with biological behavior associated with these diseases, and thus not only achieve accurate early diagnosis, but also provide deep insight into the clinicopathologic features during the disease initiation and progression.

ACKNOWLEDGMENTS

This work was supported by NSFC under Nos. 61134013, 91029301, 61072149, 91230204, H0108, and 11241002, and by the Chief Scientist Program of SIBS of CAS with No. 2009CSP002. This work was also supported by Shanghai Pujiang Program by The Shanghai Committee of Science and Technology (12JC1402200, 12431900207, 11410708600), by the National Center for Mathematics and Interdisciplinary Sciences of CAS, and by the FIRST program from JSPS initiated by CSTP.

Conflict of Interest

The authors declare no competing financial interests.

REFERENCES

1. Venegas JG, Winkler T, Musch G, Vidal Melo MF, Layfield D, Tgavalekos N, Fischman AJ, Callahan RJ, Bellani G, Harris RS. Self-organized patchiness in asthma as a prelude to catastrophic shifts. Nature 2005;434:777–782.
2. McSharry PE, Smith LA, Tarassenko L. Prediction of epileptic seizures: Are nonlinear methods relevant? Nat Med 2003;9:241–242.

3. Roberto PB, Eliseo G, Josef C. Transition models for change-point estimation in logistic regression. Stat Med 2003;22:1141–1162.
4. Paek SH, Chung HT, Jeong SS, Park CK, Kim CY, Kim JE, Kim DG, Jung HW. Hearing preservation after gamma knife stereotactic radiosurgery of vestibular schwannoma. Cancer 2005;104:580–590.
5. Liu JK, Rovit RL, Couldwell WT. Pituitary apoplexy. Semin Neurosurg 2001;12:315–320.
6. Biomarkers Definitions Working Group. Biomarkers and surrogate endpoints: Preferred definitions and conceptual framework. Clin. Pharmavcol Ther 2001;69:89–95.
7. Wang XD, Ward PA. Opportunities and challenges of disease biomarkers: A new section in the journal of translational medicine. J Transl Med 2012;10:220, 1–4.
8. Ptolemy AS, Rifai N. What is a biomarker? Research investments and lack of clinical integration necessitate a review of biomarker terminology and validation schema. Scand Clin Lab Invest 2010;70(Suppl 242):6–14.
9. Ransohoff DF. How to improve reliability and efficiency of research about molecular markers: Roles of phases, guidelines, and study design. J Clin Epidemiol 2007;60:1205–1219.
10. Rhodes DR, Sanda MG, Otte AP, Chinnaiyan AM, Rubin MA. Multiplex biomarker approach for determining risk of prostate-specific antigen-defined recurrence of prostate cancer. J Natl Cancer Inst 2003;9:661–668.
11. Mikolajczyk SD, Song Y, Wong JR, Matson RS, Rittenhouse HG. Are multiple markers the future of prostate cancer diagnostics? Clin Biochem 2004;37:519–528.
12. Huang Y, Isharwal S, Haese A, Chun FK, Makarov DV, Feng Z, Han M, Humphreys E, Epstein JI, Partin AW, Veltri RW. Prediction of patient-specific risk and percentile cohort risk of pathological stage using continuous PSA measurement, clinical stage and biopsy Gleason score. BJU Int 2011;107(10):1562–1569.
13. Berchuck A. Biomarkers in the ovary. J Cell Biochem 1995;23:223–226.
14. Rhodes DR, Sanda MG, Otte AP, Chinnaiyan AM, Rubin MA. Shaping genetic alterations in human cancer: The p53 mutation paradigm. Cancer Cell 2007;12:303–312.
15. Landgren O, Weiss BM. Patterns of monoclonal gammopathy of undetermined significance and multiple myeloma in various ethnic/racial groups: Support for genetic factors in pathogenesis. Leukemia 2009;23:1691–1697.
16. Jin G, Zhou X, Wang H, Zhao H, Cui K, Zhang X-S, Chen L, Hazen AL, Li K, Wong STC. The knowledge-integrated network biomarkers discovery for major adverse cardiac events. J Proteome Res 2008;7:4013–4021.
17. Ideker T, Sharan R. Protein networks in disease. Genome Res 2008;18:644–652.
18. Chen L, Liu R, Liu Z, Li M, Aihara K. Detecting early-warning signals for sudden deterioration of complex diseases by dynamical network biomarkers. Sci Rep 2012;2:342, 1–8. doi: 10.1038/srep00342.
19. Liu R, Li M, Liu Z, Wu J, Chen L, Aihara K. Identifying critical transitions and their leading networks for complex diseases. Sci Rep 2013;2:813, 1–9. doi: 10.1038/srep00813.
20. Cook IA, Hunter AM, Abrams M, Siegman B, Leuchter AF. Midline and right frontal brain function as a physiologic biomarker of remission in major depression. Psychiatry Res Neuroimag 2009;174:152–157.
21. Atkinson C, Warren R, Bingham SA, Day NE. Mammographic patterns as a predictive biomarker of breast cancer risk: Effect of tamoxifen. Cancer Epidemiol Biomark Prev 1999;8:863–866.
22. Cristofanilli M, Budd GT, Ellis MJ, Stopeck A, Matera J, Miller MC, Reuben JM, Doyle GV, Allard WJ, Terstappen L, Hayes DF. Circulating tumor cells, disease progression, and survival in metastatic breast cancer. N Engl J Med 2004;351:781–791.
23. Hernandez J, Thompson IM. Prostate-specific antigen: A review of the validation of the most commonly used cancer biomarker. Am Cancer Soc 2004;894–904.

24. Allhoff EP, Proppe KH, Chapman CM, Lin CW, Prout GR, Jr,. Evaluation of prostate specific acid phosphatase and prostate specific antigen in identification of prostatic cancer. J Urol 1983;129:315–318.
25. Hirata Y, Bruchovsky N, Aihara K. Development of a mathematical model that predicts the outcome of hormone therapy for prostate cancer. J Theor Biol 2010;264:517–527.
26. Poste G, Carbone DP, Parkinson DR, Verweij J, Hewitt SM, Jessup JM. Leveling the playing field: Bringing development of biomarkers and molecular diagnostics up to the standards for drug development. Clin Cancer Res 2012;18(6):1515–1523.
27. Duffy MJ, Crown J. A personalized approach to cancer treatment: How biomarkers can help. Clin Chem 2008;54(11):1770–1779.
28. Pepe MS, Etzioni R, Feng Z, John D, Potter Thompson ML, Thornquist M, Winget M, Yasui Y. Phases of biomarker development for early detection of cancer. J Natl Cancer Inst 2001;93(14):1054–1061.
29. Pepe MS, Feng Z, Janes H, Bossuyt PM, Potter JD. Pivotal evaluation of the accuracy of a biomarker used for classification or prediction: Standards for study design. J Natl Cancer Inst 2008;100:1432–1438.
30. Betsou F, Barnes R, Burke T, Coppola D, DeSouza Y, Eliason J, Glazer B, Horsfall D, Kleeberger C, Lehmann S, Prasad A, Skubitz A, Somiari S, Gunter E. Human biospecimen research: Experimental protocol and quality control tools. Cancer Epidemiol Biomarkers Prev 2009;18(4):1017–1025.
31. Rouzier R, Perou CM, Symmans WF, Ibrahim N. Breast cancer molecular subtypes respond differently to preoperative chemotherapy. Clin Cancer Res 2005;11:5678–5685.
32. Hosmer DW, Lemeshow S. Applied Logistic Regression. New York: Wiley; 1989.
33. Lette J, Colletti BW, Cerino M, MeNamara D, Eybalin MC, Levasseur A, Nattel S. Artificial intelligence versus logistic regression statistical modeling to predict cardiac complications after non-cardiac surgery. Clin Cardiol 1994, 17:609–614.
34. Skates SJ, Xu FJ, Yu YH, Sjovall K, Einhorn N, Chang Y, Bast RC, Jr., Knapp RC. Toward an optimal algorithm for ovarian cancer screening with longitudinal tumor markers. Cancer 1995, 76:2004–2010.
35. Bhatavdekar JM, Patel DD, Shah NG, Vora HH, Suthar TP, Chikhlikar PR, Ghosh N, Trivedi TI. Prognostic significance of immunohistochemically localized biomarkers in stage II and stage III breast cancer: A multivariate analysis. Ann Surg Oncol 2000;7:305–311.
36. Diamandis EP. Tumor Markers: Physiology, Pathobiology, Technology, and Clinical Applications. Washington, DC: American Association for Clinical Chemistry Press; 2002.
37. Etzioni R, Kooperberg C, Pepe M, Smith R, Gann PH. Combining biomarkers to detect disease with application to prostate cancer. Biostatistics 2003;4(4):523–538.
38. Potter SR, Miller MC, Mangold LA, Jones KA, Epstein JI, Veltri RW, Partin AW. Genetically engineered neural networks for predicting prostate cancer progression after radical prostatectomy. Urology 1999;54:791–795.
39. Veltri RW, Manisha C, Miller MC, Poole EC, O'Dowd GJ, Partin AW. Comparison of logistic regression and neural net modeling for prediction of prostate cancer pathologic stage. Clin Chem 2002;48(10):1828–1834.
40. Xue X, Kim MY, Gaudet MM, Park Y, Heo M, Hollenbeck AR, Strickler HD, Gunter MJ. A comparison of the polytomous logistic regression and joint cox proportional hazards models for evaluating multiple disease subtypes in prospective cohort studies. Cancer Epidemiol Biomarkers Prev 2013;22(2):275–285. doi:10.1158/1055-9965.EPI-12-1050.
41. Mullera R, Mockel M. Logistic regression and CART in the analysis of multimarker studies. Lancet Neurol 2008;394:1–6.
42. Vlahou A, Schorge JO, Gregory BW, Coleman RL. Diagnosis of ovarian cancer using decision tree classification of mass spectral data. J Biomed Biotechnol 2003;5:308–314.
43. Breiman L, Friedman JH. Estimating optimal transformations for multiple regression and correlation. J Am Statist Assoc 1985;80:590–598.

44. Li Y, John M, Zhou X, Kim Y, Sinha U, Jordan R, Eisele D, Abemayor E, Elashoff D, Park NH, Wong DT. Salivary transcriptome diagnostics for oral cancer detection. Clin Cancer Res 2004;10:8442–8450.
45. Veltri RW, Isharwal S, Miller MC, Epstein JI, Mangold LA, Humphreys E, Partin AW. Long-term assessment of prostate cancer progression free survival: Evaluation of pathological parameters, nuclear shape and molecular biomarkers of pathogenesis. Prostate 2008;68:1806–1815.
46. Zlobec I, Steele R, Nigam N, Compton CC. A predictive model of rectal tumor response to preoperative radiotherapy using classification and regression tree methods. Clin Cancer Res 2005;11:5440–5443.
47. Yu Y, Chen S, Wang LS, Chen WL, Guo WJ, Yan H, Zhang WH, Peng CH, Zhang SD, Li HW, Chen GQ. Prediction of pancreatic cancer by serum biomarkers using surface-enhanced laser desorption/ionization-based decision tree classification. Oncology 2005;68:79–86.
48. Barnholtz-Sloan JS, Guan X, Zeigler-Johnson C, Meropol NJ, Rebbeck TR. Decision tree-based modeling of androgen pathway genes and prostate cancer risk. Cancer Epidemiol Biomarkers Prev 2011;20(6):1146–1155.
49. Oluwadara O, Chiappelli F. Biomarkers for early detection of high risk cancers: From gliomas to nasopharyngeal carcinoma. Bioinformation 2009;3:332–339.
50. Robin X, Turck N, Hainard A, Lisacek F, Sanchez JC, Muller M. Bioinformatics for protein biomarker panel classification: What is needed to bring biomarker panels into in vitro diagnostics? Expert Rev Proteomics 2009;6:675–689.
51. Woolas RP, Xu FJ, Jacobs IJ Yu YH, Daly L, Berchuck A, Soper JT, Clarke-Pearson DL, Oram DH, Bast RC, Jr., Elevation of multiple serum markers in patients with stage I ovarian cancer. J Natl Cancer Inst 1993;85:1748–1751.
52. Woolas RP, Conaway MR, Xu F, Jaeobs IJ, Yu Y, Daly L, Davies AP, O'Briant K, Berchuck A, Soper JT, Clarke-Pearson DL, Rodriguez G, Oram DH, Bast Jr., RC. Combinations of multiple serum markers are superior to individual assays for discriminating malignant from benign pelvic masses. Gynecol Oncol 1995;59:111–116.
53. Burczynski ME, Peterson RL, Twine NC, Zuberek KA, Brodeur BJ, Casciotti L, Maganti V, Reddy PS, Strahs A, Immermann F, Spinelli W, Schwertschlag U, Slager AM, Cotreau MM, Dorner AJ. Molecular classification of Crohn's disease and ulcerative colitis patients using transcriptional profiles in peripheral blood mononuclear cells. J Mol Diagn 2006;8:51–61.
54. Mian S, Ball G, Hornbuckle J, Holding F, Carmichael J, Ellis I, Ali S, Li G, McArdle S, Creaser C, Ree R. A prototype methodology combining surface-enhanced laser desorption/ionization protein chip technology and artificial neural network algorithms to predict the chemoresponsiveness of breast cancer cell lines exposed to Paclitaxel and Doxorubicin under in vitro conditions. Proteomics 2003;3:1725–1737.
55. Ball G, Mian S, Holding F, Allibone RO, Lowe J, Ali S, Li G, McCardle S, Ellis IO, Creaser C, Rees RC. An integrated approach utilizing artificial neural networks and SELDI mass spectrometry for the classification of human tumours and rapid identification of potential biomarkers. Bioinformatics 2002;18:395–404.
56. Wei JT, Zhang Z, Barnhill SD, Madyastha KR, Zhang H, Oesterling JE. Understanding artificial neural networks and exploring their potential applications for the practicing urologist. Urology 1998;52:161–172.
57. Rodvold DM, McLeod DG, Brandt JM, Snow PB, Murphy GP. Introduction to artificial neural networks for physicians: Taking the lid off the black box. Prostate 2001;46:39–44.
58. Babaian RJ, Zhang Z. Computer-assisted diagnostics: Application to prostate cancer. Mol Urol 2001;5:175–180.
59. Rumelhart DE, McCelland JL. Parallel Distributed Processing: Explorations in the Microstructure of Cognition. Cambridge, MA: MIT Press; 1995.
60. Su AI, Welsh JB, Sapinoso LM, Kern SG, Dimitrov P, Lapp H, Schultz PG, Powell SM, Moskaluk CA, Frierson Jr., HF, Hampton GM. Molecular classification of human carcinomas by use of gene expression signatures. Cancer Res 2001;61:7388–7393.

61. Vapnik VN. In: Haykin S, Ed. Statistical Learning Theory. The support vector method for estimating indicator functions. New York: John Wiley & Sons; 1998. p 401–441.
62. Burges CJC. A tutorial on support vector machines for pattern recognition. *Data Min Knowl Discov* 1998, 2:121–167.
63. Chow ML, Moler EJ, Mian IS. Identifying marker genes in transcription profiling data using a mixture of feature relevance experts. Physiol Genomics 2001;5:99–111.
64. Sanchez-Carbayo M, Socci ND, Lozano J, Li W, Charytonowicz E, Belbin T, Prystowsky MB, Ortiz AR Childs G, Cordon-Cardo C. Gene discovery in bladder cancer progression using cDNA microarrays. Am J Pathol 2003;163:505–516.
65. Lee Y, Lee C. Classification of multiple cancer types by multicategory support vector machines using gene expression data. Bioinformatics 2002;19:1132–1139.
66. Dolled-Filhart M, Ryden L, Cregger M, Jirstrom K, Harigopal M, Camp RL, Rimm DL. Classification of breast cancer using genetic algorithms and tissue microarrays. Clin Cancer Res 2006;12:6459–6468.
67. Liu JJ, Cutler G, Li W, Pan Z, Peng S, Hoey T, Chen L, Ling XB. Multiclass cancer classification and biomarker discovery using GA-based algorithms. Bioinformatics 2005;21:2691–2697.
68. Bailey WJ, Ulrich R. Molecular profiling approaches for identifying novel biomarkers. Expert Opin Drug Saf 2004;3:137–151.
69. Petricoin EF, Ardekani A, Hitt B, Levine P, Fusarob VA, Steinberg S, Mills G, Simone C, Fishman D, Kohn E, Liotta L. Use of proteomic patterns in serum to identify ovarian cancer. Lancet 2002;359:572–577.
70. Adam B, Qu Y, Davis JW, Ward MD, Clements MA, Cazares LH, Semmes OJ, Schellhammer PF,Yasui Y, Feng Z, Wright G. Serum protein fingerprinting coupled with a pattern-matching algorithm distinguishes prostate cancer from benign prostate hyperplasia and healthy men. Cancer Res 2002;62:3609–3614.
71. Hilario M, Kalousis A. Approaches to dimensionality reduction in proteomic biomarker studies. Brief Bioinform 2008;9(2):102–118. doi: 10.1093/bib/bbn005.
72. Viswanath S, Madabhushi A. Consensus embedding: Theory, algorithms and application to segmentation and classification of biomedical data. BMC Bioinformatics 2012;13:26, 1–20.
73. Janes H, Pepe MS, Gu W. Assessing the value of risk predictions by using risk stratification tables. Ann Intern Med 2008;149:751–760.
74. Apple FS, Wu AHB, Mair J, Ravkilde J, Panteghini M, Tate J, Pagani F, Christenson RH, Mockel M, Danne O, Jaffe AS. Future biomarkers for detection of ischemia and risk stratification in acute coronary syndrome. Clin Chem 2005;51(5):810–824.
75. Taylor JMG, Ankerst DP, Andridge RR. Validation of biomarker-based risk prediction models. Clin Cancer Res 2008;14(19):5977–5983.
76. Xu QW, Zhao W, Wang Y, Sartor MA, Han DM, Deng J, Ponnala R, Yang JY, Zhang QY, Liao GQ, Qu YM, Li L, Liu FF, Zhao HM, Yin YH, Chen WF, Zhang Y, Wang XS. An integrated genome-wide approach to discover tumor-specific antigens as potential immunologic and clinical targets in cancer. Cancer Res 2012;72(24):6351–6361.
77. Xu QW, Zhang Y, Wang XS. HEPA and PARSE: Systematic discovery of clinically relevant tumor-specific antigens. OncoImmunology 2013;2(3):e23249, 1–3.
78. van't Veer LJ, Dai H, van de Vijver MJ, He YD, Hart AAM, Mao M, Peterse HL, van der Kooy K, Marton MJ, Witteveen AT, Schreiber GJ, Kerkhoven RM, Roberts C, Linsley PS, Bernards R, Friend SH. Gene expression profiling predicts clinical outcome of breast cancer. Nature 2002;415:530–536.
79. Hedenfalk I, Duggan D, Chen Y, Radmacher M, Bittner M, Simon R, Meltzer P, Gusterson B, Esteller M, Raffeld M, Yakhini Z, Ben-Dor A, Dougherty E, Kononen J, Bubendorf L, Fehrle W, Pittaluga S, Gruvberger S, Loman N, Johannsson O, Olsson H, Wilfond B, Sauter G, Kallioniemi OP, Borg Å, Trent J. Gene-expression profiles in hereditary breast cancer. N England J Med 2001;344:539–548.

80. Mischel PS, Cloughesy TF, Nelson SF. DNA-microarray analysis of brain cancer: Molecular classification for therapy. Nat Rev Neurosci 2004;5:782–792.
81. Jimeno A, Hidalgo M. Molecular biomarkers: Their increasing role in the diagnosis, characterization, and therapy guidance in pancreatic cancer. Mol Cancer Ther 2006;5:787–796.
82. Golub TR, Slonim1 DK, Tamayo1 P, Huard1 C, Gaasenbeek M, Mesirov JP, Coller H, Loh ML, Downing JR, Caligiuri MA, Bloomfield CD, Lander ES. Molecular classification of cancer: Class discovery and class prediction by gene expression monitoring. Science 1999;15:531–537.
83. Thomas CE, Sexton W, Benson K, Sutphen R, Koomen J. Urine collection and processing for protein discovery and quantification. Cancer Epidemiol Biomarkers Prev, 2010;19(4): 953–959.
84. Adam BL, Vlahou A, Semmes OJ, Wright GL, Jr., Proteomic approaches to biomarker discovery in prostate and bladder cancers. Proteomics 2001;1:1264–1270.
85. Xu XZ, Lu X, Yang X, Yin P, Kong H. Comprehensive two-dimensional gas chromatography/time-of-flight mass spectrometry for metabonomics: Biomarker discovery for diabetes mellitus. Anal Chim Acta 2009;633:257–262.
86. Robosky LC, Robertson DG, Baker JD, Rane S, Reily MD. In vivo toxicity screening programs using metabonomics. Comb Chem High Throughput Screen 2002;5:651–662.
87. Ilyin SE, Belkowski SM, Plata-Salamán CR. Biomarker discovery and validation: Technologies and integrative approaches. Trends Biotechnol 2004;22:411–416.
88. Martínez-Cruz LA, Rubio A, Martínez-Chantar ML, Labarga A, Barrio I, Podhorski A, Segura V, Sevilla Campo JL, Avila1 MA, Mato JM. GARBAN: Genomic analysis and rapid biological annotation of cDNA microarray and proteomic data. Bioinformatics 2003;19:2158–2160.
89. Schadt EE. Molecular networks as sensors and drivers of common human diseases. Nature 2009;461(7261):218–223.
90. Wang X. Role of clinical bioinformatics in the development of network-based biomarkers. J Clin Bioinform 2011;1:28, 1–3.
91. Barabási AL, Gulbahce N, Loscalzo J. Network medicine: A network-based approach to human disease. Nat Rev Genet 2011;12:56–68.
92. Garraway LA, Lander ES. Lessons from the cancer genome. Cell 2013;153:17–37.
93. Chen L, Wang R, Zhang X. Biomolecular Networks: Methods and Applications in Systems Biology. Hoboken, NJ: John Wiley & Sons; 2009.
94. Barabasi AL, Albert R. Emergence of scaling in random networks. Science 1999;286:509–512.
95. Albert R. Scale-free networks in cell biology. J Cell Sci 2005;118:4947–4957.
96. Liu Z-P, Wang Y, Zhang XS, Chen L. Network-based analysis of complex diseases. IET Syst Biol 2010;1–12.
97. Kitano H. Systems biology: A brief overview. Science 2002;295:1662–1664.
98. Aryee MJ, Liu W, Engelmann JC, Nuhn P, Gurel M, Haffner MC, Esopi D, Irizarry RA, Nelson WG, Luo J, Xu J, Isaacs WB, Bova GS, Yegnasubramanian S. DNA methylation alterations exhibit intraindividual stability and interindividual heterogeneity in prostate cancer metastases. Sci Transl Med 2013;5:169ra10, 1–10.
99. Chuang HY, Lee E, Liu YT, Lee D, Ideker T. Network-based classification of breast cancer metastasis. Mol Syst Biol 2007;3:140, 1–10.
100. Wang P, Tang H, Zhang H, Whiteaker J, Paulovich AG, McIntosh M. Normalization regarding non-random missing values in high-throughput mass spectrometry data. Pac Symp Biocomput 2006;315–326.
101. Ideker T, Ozier O, Schwikowski B, Siegel AF. Discovering regulatory and signalling circuits in molecular interaction networks. Bioinformatics 2002;18(Suppl 1):S233–S240.
102. Shannon P, Markiel A, Ozier O, Baliga NS, Wang JT, Ramage D, Amin N, Schwikowski B, Ideker T. Cytoscape: A software environment for integrated models of biomolecular interaction networks. Genome Res 2003;13:2498–2504.

103. Scott J, Ideker T, Karp RM, Sharan R. Efficient algorithms for detecting signaling pathways in protein interaction networks. J Comput Biol 2006;13:133–144.

104. Qiu Y, Zhang S, Zhang X, Chen L. Detecting disease associated modules and prioritizing active genes based on high throughput data. BMC Bioinformatics 2010;11:26–37.

105. Liu M, Liberzon A, Kong SW, Lai WR, Park PJ, Kohane IS, Kasif S. Network-based analysis of affected biological processes in type 2 diabetes models. PLoS Genet 2007;3:e96, 958–972.

106. Jiang BB, Wang JG, Xiao JF, Wang Y. Gene prioritization for type 2 diabetes in tissue-specific protein interaction networks. Lect Notes Oper Res 2009;11:319–328.

107. Jin G, Zhou X, Cui K, Zhang XS, Chen L, Wong ST. Cross-platform method for identifying candidate network biomarkers for prostate cancer. IET Syst Biol 2009;3:505–512.

108. Krauthammer M, Kaufmann CA, Gilliam TC, Rzhetsky A. Molecular triangulation: Bridging linkage and molecular-network information for identifying candidate genes in Alzheimer's disease. Proc Natl Acad Sci USA 2004;101:15148–15153.

109. Liu Z-P, Wang Y, Wen T, Zhang XS, Xia W, Chen L. Dynamically dysfunctional protein interactions in the development of Alzheimer's disease. Proceedings of the IEEE International Conference on Systems, Man and Cybernetics, San Antonio, USA 2009. p 4262–4267.

110. Dao P, Wang K, Collins C, Ester M, Lapuk A, Sahinalp SC. Optimally discriminative subnetwork markers predict response to chemotherapy. ISMB 2011;27:i205–i213.

111. Liu X, Liu Z-P, Zhao XM, Chen L. Identifying disease genes and module biomarkers by differential interactions. J Am Med Inform Assoc 2012;19:241–248.

112. Lee E, Chuang HY, Kim JW, Ideker T, Lee D. Inferring pathway activity toward precise disease classification. PLoS Comput Biol 2008;4:e1000217, 1–9.

113. Wang Y-C, Chen B-S. A network-based biomarker approach for molecular investigation and diagnosis of lung cancer. BMC Med Genom 2011;4:2, 1–15.

114. Xu M, Kao MCJ, Nunez-Iglesias J, Nevins JR, West M, Zhou XJ. An integrative approach to characterize disease-specific pathways and their coordination: A case study in cancer. BMC Genom 2008;9(Suppl 1):S12, 1–12.

115. Huang Y, Sun X, Hu G. An integrated genetics approach for identifying protein signal pathways of Alzheimer's disease. Comput Methods Biomech Biomed Engin 2011;14:1–7.

116. Bonnet E, Michoel T, Van de Peer Y. Prediction of a gene regulatory network linked to prostate cancer from gene expression, microRNA and clinical data. Bioinformatics 2010;26:i638–i644.

117. Zhan X, Desiderio D. Signaling pathway networks mined from human pituitary adenoma proteomics data. BMC Med Genom 2010, 3:13, 1–26. doi: 10.1186/1755-8794-3-13.

118. Liu KQ, Liu Z-P, Hao JK, Chen L, Zhao XM. Identifying dysregulated pathways in cancers from pathway interaction networks. BMC Bioinformatics 2012;13:126–131.

119. He D, Liu Z-P, Chen L. Identification of dysfunctional modules and disease genes in congenital heart disease by a network-based approach. BMC Genomics 2011;12:592–607.

120. Kreeger PK, Lauffenburger DI. Cancer systems biology: A network modeling perspective. Carcinogenesis 2010;31(1):2–8.

121. Lavi O, Dror G, Shamir R. Network-induced classification kernels for gene expression profile analysis. J Comput Biol 2012;19(6):694–709.

122. Zhu Y, Shen X, Pan W. Network-based support vector machine for classification of microarray samples. BMC Bioinformatics, 2009;10(Suppl 1):S21, 1–11.

123. Chen L, Xuan J, Riggins RB, Clarke R, Wang Y. Identifying cancer biomarkers by network-constrained support vector machines. BMC Syst Biol 2011;5:161, 1–20.

124. Zhang B, Li H, Riggins RB, Zhan M, Xuan J, Zhang Z, Hoffman EP, Clarke R, Wang Y. Differential dependency network analysis to identify condition-specific topological changes in biological networks. Bioinformatics 2009;25(4):526–532.

125. Zhang B, Tian Y, Jin L, Li H, Shih I-M, Madhavan S, Clarke R, Hoffman EP, Xuan J, Hilakivi-Clarke L, Wang Y. DDN: A caBIG® analytical tool for differential network analysis. Bioinformatics 2011;27(7):1036–1038.
126. Taylor IW, Linding R, Warde-Farley D, Liu Y, Pesquita C, Faria D, Bull S, Pawson T, Morris Q, Wrana JL. Dynamic modularity in protein interaction networks predicts breast cancer outcome. Nat Biotechnol 2009;27:199–204.
127. Liu Y, Shin S, Zeng X, Zhan M, Gonzalez R, Mueller F-J, Schwartz CM, Xue H, Li H, Baker SC, Chudin E, Barker DL, McDaniel TK, Oeser S, Loring JF, Mattson MP, Rao MS. Genome wide profiling of human embryonic stem cells (hESCs), their derivatives and embryonal carcinoma cells to develop base profiles of U.S. Federal government approved hESC lines. BMC Dev Biol, 2006;6:20, 1–16.
128. Wang K, Li M, Hakonarson H. Analysing biological pathways in genome-wide association studies. Nat Rev Genet 2010;11:843–854.
129. Li M, Wang K, Grant SF, Hakonarson H, Li C. ATOM: A powerful gene-based association test by combining optimally weighted markers. Bioinformatics 2009;25:497–503.
130. Eleftherohorinou H, Wright V, Hoggart C, Hartikainen A-L, Jarvelin M-R, Balding D, Coin L, Levin M. Pathway analysis of GWAS provides new insights into genetic susceptibility to 3 inflammatory diseases. PLoS ONE 2009;4:e8068, 1–11.
131. Gauderman WJ, Murcray C, Gilliland F, Conti DV. Testing association between disease and multiple SNPs in a candidate gene. Genet Epidemiol 2007;31:383–395.
132. Li Y, Willer C, Sanna S, Abecasis G. Genotype imputation. Annu Rev Genomics Hum Genet 2009;10:387–406.
133. Marchini J, Howie B. Genotype imputation for genome-wide association studies. Nat Rev Genet 2010;11:499–511.
134. Wang K, Li M, Bucan M. Pathway-based approaches for analysis of genomewide association studies. Am J Hum Genet 2007;81:1278–1283.
135. Gentles AJ, Gallahan D. Systems Biology: Confronting the complexity of cancer. Cancer Res 2011;71(18):5961–5964.
136. Vandin F, Upfal E, Raphael BJ. Algorithms for detecting significantly mutated pathways in cancer. J Comput Biol 2011;18:507–522.
137. Nguyen V, Cao L, Lin JT, Hung N, Ritz A, Yu K, Jianu R, Ulin SP, Raphael BJ, Laidlaw DH, Brossay L, Salomon AR. A new approach for quantitative phosphoproteomic dissection of signaling pathways applied to T cell receptor activation. Mol Cell Proteomics 2009;8:2418–2431.
138. Liu R, Aihara K, Chen L. Dynamical network biomarkers for identifying critical transitions and their driving networks of biologic processes. Quant Biol 2013. doi:10.1007/s40484-013-0008-0.
139. Liu XP, Liu R, Zhao X-M, Chen L. Detecting early-warning signals of type 1 diabetes and its leading biomolecular networks by dynamical network biomarkers. BMC Med Genom 2013;6(Suppl 2):S8, 1–10. doi:10.1186/1755-8794-6-S2-S8.
140. Li M, Tao Z, Liu R, Chen L. Detecting tissue-specific early-warning signals for complex diseases based on dynamical network biomarkers: Study of type-2 diabetes by cross-tissue analysis. Brief Bioinform 2013. doi: 10.1093/bib/bbt027.
141. Sciuto AM, Phillips CS, Orzolek LD, Hege AI, Moran TS, Dillman JF. Genomic analysis of murine pulmonary tissue following carbonyl chloride inhalation. Chem Res Toxicol 2005;18:1654–1660.