A survey on biomarker identification based on molecular networks

Guanghui Zhu, Xing-Ming Zhao* and Jun Wu*

Department of Computer Science and Technology, Tongji University, Shanghai 201804, China
* Correspondence: xm_zhao@tongji.edu.cn, wujun@tongji.edu.cn

Background: Identifying biomarkers for accurate diagnosis and prognosis of diseases is important for the prevention of disease development. The molecular networks that describe the functional relationships among molecules provide a global view of the complex biological systems. With the molecular networks, the molecular mechanisms underlying diseases can be unveiled, which helps identify biomarkers in a systematic way.

Results: In this survey, we report the recent progress on identifying biomarkers based on the topology of molecular networks, and we categorize those biomarkers into three groups, including node biomarkers, edge biomarkers and network biomarkers. These distinct types of biomarkers can be detected under different conditions depending on the data available.

Conclusions: The biomarkers identified based on molecular networks can provide more accurate diagnosis and prognosis. The pros and cons of different types of biomarkers as well as future directions to improve the methods for identifying biomarkers are also discussed.

Keywords: biomarker; molecular network; module; pathway

INTRODUCTION

In biomedicine, biomarkers generally refer to those biological characteristics that can describe the disease states very well, and thus can be used for diagnosis and prognosis. Recently, due to the rapid development of high throughput technologies, large amount of omics data is accumulated, e.g., transcriptomics and proteomics data, which provides opportunity to detect biomarkers more efficiently. For example, the omics data deposited in The Cancer Genome Atlas (TCGA) and International Cancer Genome Consortium (ICGC) can help detect gene biomarkers for diagnosis and prognostic markers for survival analysis. With these data, Akbani *et al.* identified protein-level signatures that are able to predict the survival time of cancer patients [1]. The gene biomarkers have also been detected that can stratify the ovarian cancers into two groups with significantly different survival time [2]. Later, with the gene expression data, Verhaak *et al.* refined the cancer subtypes defined by TCGA and predicted the survival time and therapeutic resistance of patients better based on the refined subtypes [3].

Although traditional methods can detect sets of genes as biomarkers, they generally investigated these genes independently and did not consider the interactions between the genes. Given that molecules in biological systems interact with each other in a complicated way and the rewiring of interactomes may lead to diseases, the molecular context should be taken into account when detecting molecular biomarkers. In the past decade, lots of molecular interactions have been detected. For example, physical interactions and functional interactions have been respectively found for human genes in the BioGRID and STRING databases. These interactomes can help understand how the molecules interact and how the biomarkers are involved in the pathogenesis processes. For instance, Wu and Stein identified module biomarkers from protein functional interaction networks for breast and ovarian cancers, which can stratify the patients into high-risk and low-risk groups with high accuracy on 4 independent data sets [44]. From microRNA (miRNA)-mRNA interaction network, Zhang *et al.* identified

miRNA biomarkers that are critical for prostate cancer [5]. By constructing a protein-protein interaction network (PPIN) for prostate cancer, Li *et al.* identified module biomarkers in the PPIN which can discriminate control from cancer samples [6]. By considering their molecular contexts, the candidate biomarkers detected generally perform better and robust.

In this survey, we will introduce the recent progress on computational approaches that have been developed for identifying biomarkers based on molecular networks. Based on the way they are detected as shown in Figure 1, the biomarkers identified based on molecular networks can be grouped into three categories, i.e., node biomarker, edge biomarker and network biomarker. The details on detecting the three types of biomarkers will be addressed in the following parts.

MOLECULAR NETWORKS

Right now, a lot of molecular networks have been built, including PPINs, gene regulation networks, miRNA-gene interaction networks, metabolic networks, and so on. These molecular networks can be described as graphs, where the nodes usually represent molecules while the edges represent the interactions between molecules. Since

these networks can describe various types of functional relationships among molecules, they are widely used for identifying biomarkers. Table 1 lists the popular resources about different types of human molecular networks.

IDENTIFICATION OF NODE BIOMARKERS

As described above, more and more molecular interactions are being collected, the molecular networks constructed accordingly can describe the biological systems very well. In general, a network can be described as a graph G(V, E), where V is the set of nodes representing the molecules and E is the set of edges representing the functional relationships among molecules. For example, the nodes in a protein-protein interaction network are proteins and the edges are interactions among proteins. The nodes in a miRNA-gene regulatory network respectively represent miRNAs and genes, and the edges represent the regulations of miRNAs on genes.

Traditionally, the molecular biomarkers can be identified by comparing their expressions in different conditions. For example, the genes that are differentially
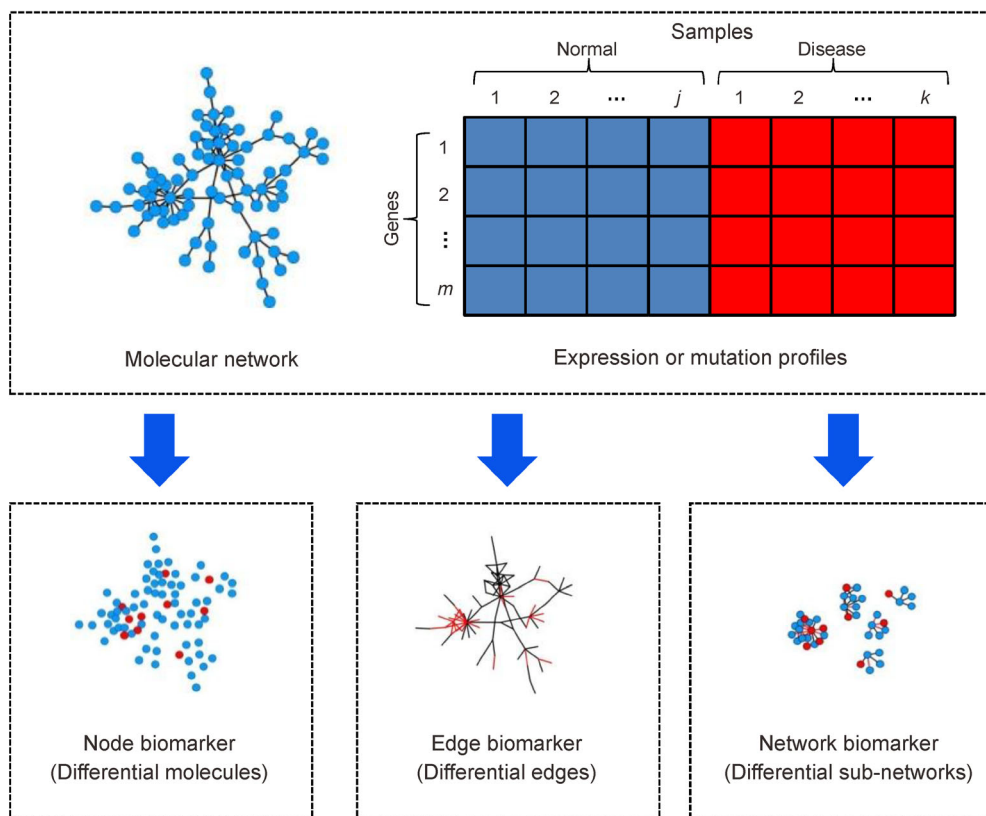

Figure 1.   Illustration of three types of network based biomarkers.

Table 1. Popular resources about different types of molecular networks.

Database	Website	Description	Interaction type	Interaction quality
HumanNet	http://www.functionalnet.org/humannet/	A probabilistic functional gene network of *Homo sapiens*	Interaction	Computational confidence is provided
PathwayCommons	http://www.pathwaycommons.org/	Collection of biological pathway information collected from public pathway databases	Pathway	The pathways are either experimentally determined or computationally predicted
BIND	http://bind.ca/	A biomolecular interaction network database	Interaction and pathway	The interactions are based on experiments and literature
MicroCosm	http://www.ebi.ac.uk/enright-srv/microcosm/htdocs/targets/v5/	A database contains computationally predicted targets for microRNAs across various species	MiRNA-target gene interaction	Computational confidence is provided
InnateDB	http://www.innatedb.com/	A database contains experimentally-verified interactions and signaling pathways involved in the innate immune response of humans, mice and bovines to microbial infection.	Interaction and signaling pathway	The interactions are experimentally verified
BioGRID	http://thebiogrid.org	Collections of physical and genetic interactions for major model organisms	Interaction	The interactions are curated from high-through-put datasets and individual focused studies
DIP	http://dip.doe-mbi.ucla.edu	The database contains experimentally determined interactions between proteins	Interaction	The interactions are experimentally determined
HPRD	http://www.hprd.org	A database about human proteins and their interactions	Interaction	The interactions are manually curated from literature
MINT	http://mint.bio.uniroma2.it/mint	Collection of molecular interactions extracted from literature	Interaction	The interactions are experimentally verified
STRING	http://string.embl.de/	A database about functional interactions for various organisms	Interaction	The interactions are either experimental determined or computationally predicted
KEGG	http://www.genome.jp/kegg	A resource about high-level functional information for distinct organisms	Pathway	The interactions are manually drawn and predicted

expressed between normal and disease conditions are generally selected as candidate biomarkers. In a network, except for its own expression value, the importance of each node can be described by its topological properties. For example, the nodes with high degrees are generally regarded as important and have significant influences on other nodes in the network, where the degree of a node is the number of other nodes it connects to in the network. Therefore, the molecules that are differentially expressed between phenotypes of interest and have high degrees are more likely related to the phenotypes. Considering both the differential expression and the degrees of genes, Santiago and Potashkin identified two accurate biomarkers in blood samples for diagnosing early-stage Parkinson's disease [7]. In complex networks, the nodes with highest degrees are called hub nodes. By investigating the miRNAs with top 15% degrees in a miRNA-gene regulatory network, Li *et al.* found miRNA biomarkers that accurately predict the survival time of glioma patients with cox-regression [8]. Besides degree, the other kinds of centralities of a node are also widely used to measure its importance. Among different types of centralities, betweenness is one of the most widely used, which is defined as the number of the shortest paths going through the node of interest when connecting any pair of nodes in the network. Based on a prostate cancer (PC)-specific network consisting of known PC-related genes from the Prostate Gene DataBase (PGDB), Ozgur *et al.* ranked all genes with respect to their topological characteristics and found that betweenness yielded the highest precision in ranking PC associated genes [9]. When detecting the potential disease genes, it is still a challenge to set the number of top ranked genes to be chosen. Furthermore, distinct types of centrality indexes have different performance and it is a try and trial process to determine the appropriate centrality index to be used [9].

In general, the genes that harbor frequent mutations are often considered as candidate drivers of the pathological process and used as biomarkers. However, such genes are rarely available. Moreover, the mutated genes may not drive the pathogenetic process directly. Instead, the mutated genes will interact with other genes to initiate the pathogenetic process. Therefore, the effects of individual genes may be not significant, but their epistatic effect is evident in the disease state. Based on this idea, Bertrand *et al.* proposed a new approach to identify patient-specific driver gene [10]. They assumed that the mutated gene will lead to the abberent expression of its downstream genes, and the expression changes can be used to describe the impact of the mutated gene. With a gene interaction network, the mutated genes that can affect those differentially expressed genes through paths of limited length were identified and regarded as driver genes. When quantitatively describing the driver genes,

the sum of fold changes of those affected downstream genes were used. Later, Suo *et al.* proposed another method to identify driver genes predictive of breast cancer survival based on integration of mutation and transcriptome data [11]. For each mutated gene, they investigated those differentially expressed genes (DEGs) that directly interact with the gene, and defined the impact of the mutated gene as below.

$$z = \frac{d_{AF} - u_{AF}}{\sigma_{AF}}, \tag{1}$$

where d_{AF} is the number of DEGs, u_{AF} and σ_{AF} are respectively the mean and standard deviation of d_{AF} by randomizing the interactome. It has been found that the driver genes detected in this way are more accurate and meaningful compared with those detected based only on the mutation data [11].

In the methods described above, the influence of one gene is either described as its own expression change or the sum of expression changes of its downstream genes. Considering the interactions among genes, the network smoothing approaches are becoming popular, where the influence of each node in a molecular network will be refined based on its position in the network topology. For example, in Figure 2A, both genes A and B are mutated and will be regarded as disease genes by assigning weight of 1 to them. In the network, gene C interacts with both A and B, and may play important roles in pathogenesis even the gene has no mutation. However, gene C will be assigned a weight of 0 and is missed if only the mutation status is considered. After one-step smoothing, gene C will be assigned a weight of 1, which makes the important role of gene C clearer. Figure 2B shows another example with only gene D mutated, where only this gene will be considered (with weight of 1) based on the mutation data. Considering the interactions among genes, after one-step smoothing, gene C will be assigned a weight of 0.5. From these examples, we can see that the functional roles of genes can be better described with smoothing based on the molecular network. As one representative of the smoothing approaches, random walk has been widely used to detect disease related genes. For instance, the node weights in one network can be smoothed with the following p-step random walk kernel.

$$\boldsymbol{K} = (\alpha \boldsymbol{I} - \boldsymbol{L}^{norm})^p, \tag{2}$$

where the $\boldsymbol{L}^{norm}$ is the normalized Laplacian matrix of the network, α is a constant parameter, and $\boldsymbol{I}$ is a diagonal matrix with each entry equals to 1 [12]. After smoothing the weights of both mRNAs and miRNAs in a network consisting of PPIs and miRNA-target gene interactions, Cun and Frohlich identified biomarkers that better predict the survival of cancer patients compared with those before
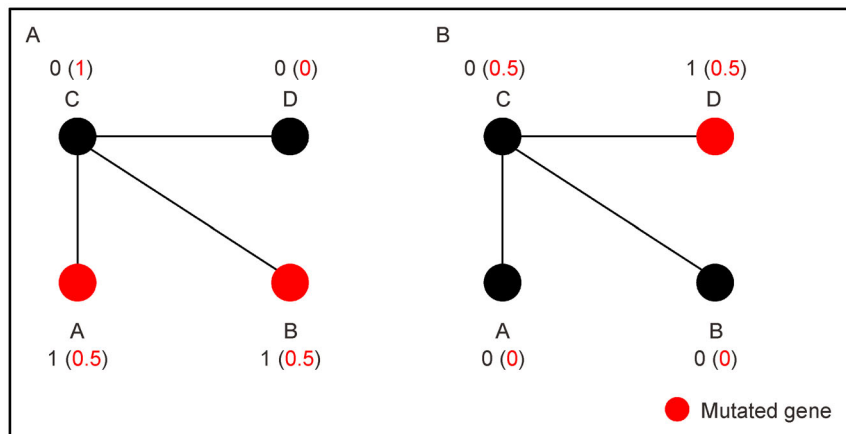
Figure 2. The example of the smoothing process. The black and red values respectively represent the weights before and after smoothing.

smoothing [13]. With smoothing over the network based on the mutation data, Hofree *et al.* stratified patients into different subtypes and identified genes that predict survival of individual patients better with Cox-proportional hazards model [14]. We have used smoothing in domain-domain interaction networks to extend the disease-domain associations, which leads to detecting more disease associated domains [15].

Since the node biomarkers are generally single genes or proteins, it is more easy to interpret the roles of those genes in diseases. Furthermore, the node biomarkers are more applicable in clinics. However, many diseases happen due to the coordination of a set of functionally related genes instead of single genes. The node biomarkers may not work well in these cases.

IDENTIFICATION OF EDGE BIOMARKERS

The node biomarkers described above generally assume that they have significant characteristics in different conditions, e.g., aberrant expression change in diseases. However, in many cases, the diseases are not driven by single genes. Instead of significant changes of nodes, it is the rewiring of the interactome that makes diseases happen. The rewired interactions, i.e., the differential edges, can be directly used as features to quantify the biomarker, and these differential edges are called edge biomarkers here since the edges were used as potential biomarkers. For example, Zhang *et al.* developed Edge-Marker to identify the edge biomarkers among the non-differentially expressed genes from the co-expression network [16]. With the edge biomarkers detected, the phenotypes can be classified with higher accuracy than the traditional DEGs. We have previously proposed a differential interaction approach to identify the edge biomarkers for diseases [17]. With a protein-protein interaction network as the background network, we detected the interactions that were rewired in diseases based on the gene expression data. The genes involved in differential interactions were found to be significantly related to diseases. When applied to a real gastric cancer dataset, the cancers samples can be separated from normal ones in a more accurate way compared with established gene biomarkers. Moreover, the edge biomarkers can help unveil how the interactions are rewired during the development of diseases. With the similar idea, Ben-Hamo *et al.* developed a PhenoNet approach that identifies changes in regulation under different phenotypes. With the most changed pathway as baseline, PhenoNet identified the differential regulations with changes above the baseline. The edge biomarkers identified by PhenoNet have been successfully applied to the diagnosis of breast cancer across multiple datasets [18].

Except for PPIN, the gene regulatory networks consist of regulations between genes, e.g., regulations between transcription factors (TFs) and genes, are also widely used for edge biomarker identification. For instance, Ma *et al.* developed an approach, namely DRAGEN, to detect the altered regulations in diseases [19]. The regulation strengths were defined as the regression coefficients derived from a linear regression model, and the regulations with altered regulation strengths in different conditions were identified and used as edge biomarkers. DRAGEN was proved to have better performance in identifying phenotype-associated genes than the popular gene set enrichment analysis method GSEA which does not consider network information. With a probabilistic miRNA-mRNA interaction network, Li *et al.* identified the altered miRNA-mRNA regulations in thyroid cancer

samples [20]. Compared with either miRNAs or mRNAs detected differentially expressed, the edge biomarkers are found to perform best.

In the network described above, the nodes are molecules. Recently, Yu *et al.* constructed an edge network to identify the edge biomarkers [21]. In the edge network, each node is an edge described as the correlation between a pair of genes and the correlation between two nodes was defined as below.

$$PCC(x_i, x_j, x_k, x_l) = \frac{C(x_i, x_j)C(x_k, x_l) + C(x_i, x_k)C(x_j, x_l) + C(x_i, x_l)C(x_j, x_k)}{\sqrt[3]{V(x_i)V(x_j)V(x_k)V(x_l)}}, \quad (3)$$

$$C(x_i, x_j) = E((x_i - u_i)(x_j - u_j)), \quad (4)$$

$$V(x_i) = E(x_i - u_i)^2, \quad (5)$$

where x_i denotes the expression of the i-th molecule, and u_i denotes the mean of x_i. After the edge-network was constructed for every patient, the dynamical network biomarker (DNB) approach was employed to detect the early signals of disease state. The early signal of disease state was defined as below.

$$CI = \frac{SD_d \cdot PCC_d}{PCC_o}, \quad (6)$$

where PCC_d and SD_d respectively denotes the mean and standard deviation of the correlation coefficients between nodes within DNB, and PCC_o is the average of the PCC correlation coefficients between nodes within and outside of DNB. When applied to influenza infection, the edge biomarkers yield high accuracy of 90%.

Compared with node biomarkers, the edge biomarkers take into account the functional relationships among genes and achieve better accuracy. However, the edge biomarkers are generally gene pairs, and different gene pairs are considered independently, which is not the case in the biological systems. In addition, the correlation coefficients are usually adopted to quantify the edges, and more indexes can be used to characterize the edges.

IDENTIFICATION OF NETWORK BIOMARKERS

Despite both node and edge biomarkers perform well, the molecular biomarkers detected are generally independent and the functional relationships among molecules are not considered. In the biological systems, there are complex interactions among molecules instead of pairwise interactions. Therefore, when detecting the biomarkers, the functional context of the biomarkers in the molecular circuit should be taken into account. The molecular pathways have important roles in the biological systems, the dysregulation of which leads to diseases. Recently, some dysfunctional pathways have been detected and used as biomarkers. For example, we have proposed a feature selection approach to identify the dysfunctional pathways based on a pathway interaction network, where

each node in the pathway interaction network denotes a pathway and the links among nodes represent the crosstalks among pathways [22]. Compared with the gene biomarkers, the pathway biomarkers we detected perform better and robustly on multiple cancer datasets. Wang *et al.* found that the pathway biomarkers are more robust across nine individual microarray datasets [23]. Recently, we further developed a new approach to identify both pathways and miRNAs that have important roles in cancer based on the miRNA-pathway interactions [24]. The detected dysfunctional miRNA-pathway interactions help understand how the miRNAs are involved in cancer.

The pathway biomarkers detected can interpret the molecular mechanisms underlying diseases. Unfortunately, our knowledge about pathways is far from complete. Therefore, some approaches have presented to detect modules from molecular networks, where a module refers to a set of nodes that are more tensely connected to each other rather than to the other nodes outside of the module in a network. Given the fact that molecules tend to interact with each other to perform their functions, it is more reasonable to consider network modules instead of single genes when detecting biomarkers. Some tools have been developed to detect modules from molecular networks. For example, MCODE has been developed to detect protein complexes in protein-protein interaction networks [25], and ClusterOne was presented to detect modules from a network by employing greedy search [26]. The modules can complement very well with molecular pathways and can be regarded as mega-nodes to some extent, where the modules can be quantitatively characterized with the nodes in the module. If one module shows different activities in disease and normal conditions, the module can be used as a biomarker. For example, Zhang *et al.* identified a module biomarker containing 32 genes with high accuracy in discriminating type 2 diabetes mellitus samples from normal samples [27], where the module biomarkers were defined to be able to minimize the intersection area of the module expression distribution curves in disease samples and normal samples. Chuang *et al.* detected module biomarkers that are most associated with the phenotype of interest based on mutual information [28]. Zeng *et al.* developed the module network rewiring-analysis (MNR) approach to investigate drug response [29], where

consistent modules were detected based on the integration of time-course gene expression data, PPIs and TF-target interactions. Based on the module network with the modules as nodes, MNR was successfully used to discriminate the protected vaccines from non-protected vaccines based on the gene expression data from a malaria vaccine trial [30].

With the mutated genes as seeds, Leung *et al.* proposed a new approach to detect network modules with mutated genes that are close in PPIN with the assumption that interacting mutated genes have more important roles in diseases compared against single mutated genes [31]. The module biomarkers they detected can predict the survival time of patients very well. Recently, the mutations were found to be exclusive in cancer. With this prior information, Kim *et al.* developed the MEMCover algorithm to detect modules with mutated genes that are mutually exclusive across various types of cancer samples based on the Module Cover algorithm [32,33]. Unlike traditional molecular networks with edges describing the interactions among molecules, they built a mutual exclusivity and functional network, where the edges not only describe the functional relationships among genes but also the exclusivity among genes. The module biomarkers they detected were found to cover many known cancer driver genes, implying the effectiveness of the module biomarkers.

As the pre-disease state is usually reversible to the normal state, the biomarkers that can diagnose the pre-disease state is important for the prevention of disease. The DNB approach has developed to detect early-warning signals of the complex disease [34] based on the assumption that groups of molecules can reveal the early-warning signals when the system approaches the critical transition from normal to disease states. The pre-disease state was assumed to be reached when the following three conditions were satisfied:

(i) The correlation between any pair of nodes within DNB becomes stronger.

(ii) The correlation between any nodes within DNB and any nodes out of DNB becomes weaker.

(iii) The nodes within DNB start to fluctuate drastically.

These 3 conditions were quantified to calculate the DNB score as shown in Equation(6). The DNB score quantifies the dynamic nature of the leading networks and indicates important information about the critical transition, which cannot be detected by conventional biomarkers. The DNBs have been used to accurately detect the pre-disease states for lung injury disease, liver cancer, and lymphoma cancer [34]. Considering the computational cost of identifying DNBs from all nodes in the network, Liu *et al.* adopted the state-transition-based entropy (SNE) to detect the pre-disease state [35]. As SNE quantitatively

measures the structural stability and robustness of the dynamical network, the drastic decrease of SNE indicates the dynamical network is approaching the critical transition. The SNE accurately predicted the pre-disease states of HCV-induced dysplasia and hepatocellular carcinoma. The original DNB approach and their improved versions identified network biomarkers based on co-expression networks, and the biomarkers may not be applicable. Li *et al.* utilized ordinary differential equation models to combine time-course gene expression data and prior knowledge of PPIs to construct dynamic networks, based on which DNBs identified were smaller and could be used in clinical applications [36]. With a module-module network, the module biomarkers identified by Zeng based on DNB criterions can classify the pre-disease stage from the advanced one [37].

Both the module and pathway biomarkers are actually subnetworks of genes, and are therefore called network biomarkers here. Although the network biomarkers take into account the functional relationships among genes, it is still challenging to quantify a network biomarker. The most adopted average value over members of either a module or a pathway has been proved not effective enough [38]. In addition, the module biomarkers are generally defined as those most determinative modules, which increases the risk of over-fitting and decreasing the robustness of the biomarkers across distinct datasets for the same disease [38]. Recently, Allahyar and de Ridder developed a method, namely FERAL, to identify the module biomarkers by using the Sparse Group Lasso, where appropriate multiple operators were utilized to quantify each module [38]. The biomarkers identified by FERAL achieved high accuracy and perform robustly across individual data sets. With more data integrated with network, more appropriate quantification indexes should be defined in the future.

DISCUSSION

The identification of biomarkers is important for precision medicine. However, it is a challenging task to develop biomarkers with high accuracy and robustness considering the complexity of the biological systems and the variability among patients [39]. The molecular networks that describe the functional relationships among molecules provide a global view of the complex biological systems. With the molecular networks, the molecular mechanisms underlying diseases can be unveiled. Accordingly, many computational approaches have been developed to identify biomarkers that can be used for diagnosis or prognosis based on molecular networks [40,41]. In this survey, we reported the recent progress on computational approaches that have been developed for identifying biomarkers based on the topology of mole-

cular networks.

Lots of efforts have been made to classify the biomarkers as the way the network information utilized in the process of identification [30,42,43]. For instance, depending on how network information was used, Zeng *et al.* classified the biomarkers into node biomarker, network-based biomarker, edge biomarker, and dynamical network biomarker [42]. Based on the way they were detected, we grouped the biomarkers detected from molecular networks into three general categories, i.e., node biomarkers, edge biomarkers and network biomarkers. The node biomarkers used single molecules as biomarkers which are easy to interpret and can be used directly in clinic. However, the node biomarkers assume those single molecules are independent which is not the case in practice. There are complex functional relationships among molecules, which should be taken into account when detecting biomarkers. The edge biomarkers

take a step further to use the functional relationship between a pair of molecules as biomarkers. However, the pairwise edge biomarkers still treat the molecules functionally independent to some extent. The network biomarkers consider the functional context of molecules and identify sub-networks as biomarkers which are more biologically reasonable, but it is not practical to use them as biomarkers in clinic considering the number of molecules as well as the complicated functional relationship in the network biomarkers. Table 2 summarizes the characteristics of the 3 classes of biomarkers, and some typical tools for identification of the network-based biomarkers are listed in Table 3.

When detecting biomarkers based on molecular networks, one important issue is to quantify the biomarkers, especially for edge and network biomarkers. For node biomarkers, the expression changes of themselves or their downstream nodes are generally used. For edge or

Table 2. The characters, advantages and disadvantages of 3 kinds of network-based biomarkers.

	Node biomarker	Edge biomarker	Network biomarker
Utilization of molecular network	Network information is used to evaluate the impact of individual nodes	The network information is used to detect the rewired interactions	The network is used as the context to detect the dysfunctional modules in the network
Advantages	More simple and applicable	The function interactions between pairs of nodes, i.e., edges, are considered	The functional interactions among nodes are considered and have more predictive power
Disadvantages	The functional relationships among nodes is not considered	Edges are usually considered independently	Difficult to interpret and cannot be used directly in clinics

Table 3. Tools for the identification of the network-based biomarkers.

Tool	Biomarker type	Description	Availability
NetworkAnalyst	Node biomarker, network biomarker	Analysis of degree and betweenness of nodes, detection of modules and shortest paths, and functional enrichment analysis	http://www.networkanalyst.ca/
DGscore	Node biomarker	Detection of driver genes with a score for each sample	http://fafner.meb.ki.se/biostatwiki/driver-genes/
OncoIMPACT	Node biomarker	Quantify the impact of candidate driver mutations with genes connected to them	http://sourceforge.net/projects/oncoimpact
Network-based stratification (NBS)	Node biomarker	The mutation profiles are smoothed over the network and the subtypes are detected	http://idekerlab.ucsd.edu/software/NBS/
stSVM	Node biomarker	Both mRNA and miRNA profiles were smoothed, which are further used for classification	The R package netClass
DRAGEN	Edge biomarker	The identification of significant differentially regulated gene sets between different phenotypes	http://bioinfo.au.tsinghua.edu.cn/dragen/
ProMISe	Edge biomarker	The detection of miRNA–mRNA interaction signatures for diagnosis	http://www.bioconductor.org/packages/release/bioc/html/Roleswitch.html
miR_Path	Network biomarker	The cancer-related miRNAs are detected	http://comp-sysbio.org/miR_Path/
HyperModules	Network biomarker	The module biomarkers with frequently mutated genes are detected	http://www.baderlab.org/Sofware/HyperModules
FERAL	Network biomarker	Multiple operators are used to quantify the sets of connected genes as biomarker	http://homepage.tudelft.nl/53a60/resources/FERAL/FERAL.zip

network biomarkers, they can be quantified as the average or summation of those values of the nodes within them, where the biomarkers work like mega-nodes [8,19,27]. Other complicated approaches, such as linear regression model, SVM, and the DNB method, have also been proposed to quantify the biomarkers [13,20,34,44]. However, it is still a challenging task to choose an appropriate way to quantify the biomarkers. Recently, Allahyar and de Ridder have proposed to use Sparse Group Lasso (SGL) to select the most appropriate operators for quantifying network modules, which may be used to quantify the network biomarkers in the future [38].

ACKNOWLEDGEMENTS

COMPLIANCE WITH ETHICS GUIDELINES

The authors Guanghui Zhu, Xing-Ming Zhao, and Jun Wu declare that they have no conflict of interests.

This article does not contain any studies with human or animal subjects performed by any of the authors.

REFERENCES

1. Akbani, R., Ng, P. K., Werner, H. M., Shahmoradgoli, M., Zhang, F., Ju, Z., Liu, W., Yang, J. Y., Yoshihara, K., Li, J., *et al.* (2014) A pan-cancer proteomic perspective on The Cancer Genome Atlas. Nat. Commun., 5, 3887

2. Bell, D., Berchuck, A., Birrer, M., Chien, J., Cramer, D. W., Dao, F., Dhir, R., DiSaia, P., Gabra, H., Glenn, P., *et al.* (2011) Integrated genomic analyses of ovarian carcinoma. Nature, 474, 609–615

3. Verhaak, R. G., Tamayo, P., Yang, J. Y., Hubbard, D., Zhang, H., Creighton, C. J., Fereday, S., Lawrence, M., Carter, S. L., Mermel, C. H., *et al.* (2013) Prognostically relevant gene signatures of high-grade serous ovarian carcinoma. J. Clin. Invest., 123, 517–525

4. Wu, G. and Stein, L. (2012) A network module-based method for identifying cancer prognostic signatures. Genome Biol., 13, R112

5. Zhang, W., Zang, J., Jing, X., Sun, Z., Yan, W., Yang, D., Shen, B. and Guo, F. (2014) Identification of candidate miRNA biomarkers from miRNA regulatory network with application to prostate cancer. J. Transl. Med., 12, 66

6. Li, Y., Vongsangnak, W., Chen, L. and Shen, B.(2014) Integrative analysis reveals disease-associated genes and biomarkers for prostate cancer progression. BMC Med. Genomics, 7, S3

7. Santiago, J. A. and Potashkin, J. A. (2015) Network-based metaanalysis identifies HNF4A and PTBP1 as longitudinally dynamic biomarkers for Parkinson's disease. Proc. Natl. Acad. Sci. USA, 112, 2257–2262

8. Li, Y., Xu, J., Chen, H., Bai, J., Li, S., Zhao, Z., Shao, T., Jiang, T., Ren, H., Kang, C., *et al.* (2013) Comprehensive analysis of the functional microRNA-mRNA regulatory network identifies miRNA signatures associated with glioma malignant progression. Nucleic Acids Res., 41, e203

9. Ozgür, A., Vu, T., Erkan, G. and Radev, D. R. (2008) Identifying gene-disease associations using centrality on a literature mined gene-interaction network. Bioinformatics, 24, i277–i285

10. Bertrand, D., Chng, K. R., Sherbaf, F. G., Kiesel, A., Chia, B. K. H., Sia, Y. Y., Huang, S. K., Hoon, D. S. B., Liu, E. T., Hillmer, A., *et al.* (2015) Patient-specific driver gene prediction and risk assessment through integrated network analysis of cancer omics profiles. Nucleic Acids Res., 43, e44

11. Suo, C., Hrydziuszko, O., Lee, D., Pramana, S., Saputra, D., Joshi, H., Calza, S. and Pawitan, Y. (2015) Integration of somatic mutation, expression and functional data reveals potential driver genes predictive of breast cancer survival. Bioinformatics, 31, 2607–2613

12. Gao, C., Dang, X., Chen, Y. and Wilkins, D. (2009) Graph ranking for exploratory gene data analysis. BMC Bioinformatics, 10, S19

13. Cun, Y. and Fröhlich, H. (2013) Network and data integration for biomarker signature discovery via network smoothed T-statistics. PLoS One, 8, e73074

14. Hofree, M., Shen, J. P., Carter, H., Gross, A. and Ideker, T. (2013) Network-based stratification of tumor mutations. Nat. Methods, 10, 1108–1115

15. Qin, G. M. Li, R. Y. and Zhao X. M. (2016) Identifying disease associated miRNAs based on protein domains. IEEE/ACM Trans. Comput. Biol. Bioinform

16. Zhang, W., Zeng, T. and Chen, L. (2014) EdgeMarker: identifying differentially correlated molecule pairs as edge-biomarkers. J. Theor. Biol., 362, 35–43

17. Liu, X., Liu, Z. P., Zhao, X. M. and Chen, L. (2012) Identifying disease genes and module biomarkers by differential interactions. J. Am. Med. Inform. Assoc., 19, 241–248

18. Ben-Hamo, R., Gidoni, M. and Efroni, S. (2014) PhenoNet: identification of key networks associated with disease phenotype. Bioinformatics, 30, 2399–2405

19. Ma, S., Jiang, T. and Jiang, R. (2015) Differential regulation enrichment analysis via the integration of transcriptional regulatory network and gene expression data. Bioinformatics, 31, 563–571

20. Li, Y., Liang, C., Wong, K. C., Jin, K.and Zhang, Z. (2014) Inferring probabilistic miRNA-mRNA interaction signatures in cancers: a role-switch approach. Nucleic Acids Res., 42, e76

21. Yu, X., Li, G. and Chen, L. (2014) Prediction and early diagnosis of complex diseases by edge-network. Bioinformatics, 30, 852–859

22. Liu, K. Q., Liu, Z. P., Hao, J. K., Chen, L. and Zhao, X. M. (2012) Identifying dysregulated pathways in cancers from pathway interaction networks. BMC Bioinformatics, 13, 126

23. Wang, Y., Chen, J., Li, Q., Wang, H., Liu, G., Jing, Q. and Shen, B. (2011) Identifying novel prostate cancer associated pathways based on integrative microarray data analysis. Comput. Biol. Chem., 35, 151–158

24. Zhao, X. M., Liu, K. Q., Zhu, G., He, F., Duval, B., Richer, J. M., Huang, D. S., Jiang, C. J., Hao, J. K. and Chen, L. (2015) Identifying cancer-related microRNAs based on gene expression data. Bioinformatics, 31, 1226–1234

25. Bader, G. D. and Hogue, C. W. (2003) An automated method for finding molecular complexes in large protein interaction networks. BMC Bioinformatics, 4, 2

26. Nepusz, T., Yu, H. and Paccanaro, A. (2012) Detecting overlapping protein complexes in protein-protein interaction networks. Nat. Methods, 9, 471–472

27. Zhang, X., Gao, L., Liu, Z. P. and Chen, L. (2015) Identifying module biomarker in type 2 diabetes mellitus by discriminative area of functional activity. BMC Bioinformatics, 16, 92

28. Chuang, H. Y., Lee, E., Liu, Y. T., Lee, D. and Ideker, T. (2007) Network-based classification of breast cancer metastasis. Mol. Syst. Biol., 3, 140

29. Zeng, T., Wang, D. C., Wang, X., Xu, F. and Chen, L. (2014) Prediction of dynamical drug sensitivity and resistance by module network rewiring-analysis based on transcriptional profiling. Drug Resist. Updat., 17, 64–76

30. Zeng, T., Zhang, W., Yu, X., Liu, X., Li, M. and Chen, L. (2015) Big-data-based edge biomarkers: study on dynamical drug sensitivity and resistance in individuals. Brief. Bioinform., 17, 576–592

31. Leung, A., Bader, G. D. and Reimand, J. (2014) HyperModules: identifying clinically and phenotypically significant network modules with disease mutations for biomarker discovery. Bioinformatics, 30, 2230–2232

32. Kim, Y. A., Cho, D. Y., Dao, P. and Przytycka, T. M. (2015) MEMCover: integrated analysis of mutual exclusivity and functional network reveals dysregulated pathways across multiple cancer types. Bioinformatics, 31, i284–i292

33. Kim, Y. A., Salari, R., Wuchty, S. and Przytycka, T. M. (2013) Module cover — a new approach to genotype-phenotype studies, In Proceedings of the Pacific Symposium, Biocomputing, 135–146, Singapore: World Scientific

34. Chen, L., Liu, R., Liu, Z. P., Li, M.and Aihara, K. (2012) Detecting early-warning signals for sudden deterioration of complex diseases by dynamical network biomarkers. Sci. Rep., 2, 342

35. Liu, R., Li, M., Liu, Z. P., Wu, J. , Chen, L. and Aihara, K. (2012) Identifying critical transitions and their leading biomolecular networks in complex diseases. Sci. Rep., 2, 813

36. Li, Y., Jin, S., Lei, L., Pan, Z. and Zou, X. (2015) Deciphering deterioration mechanisms of complex diseases based on the construction of dynamic networks and systems analysis. Sci. Rep., 5, 9283

37. Zeng, T., Zhang, C. C., Zhang, W., Liu, R., Liu, J. and Chen, L. (2014) Deciphering early development of complex diseases by progressive module network. Methods, 67, 334–343

38. Allahyar, A. and de Ridder, J. (2015) FERAL: network-based classifier with application to breast cancer outcome prediction. Bioinformatics, 31, i311–i319

39. de Gramont, A., Watson, S., Ellis, L. M., Rodón, J. , Tabernero, J., de Gramont, A. and Hamilton, S. R. (2015) Pragmatic issues in biomarker evaluation for targeted therapies in cancer. Nat. Rev. Clin. Oncol., 12, 197–212

40. Qin, G. and Zhao, X. M. (2014) A survey on computational approaches to identifying disease biomarkers based on molecular networks. J. Theor. Biol., 362, 9–16

41. Liu, R., Wang, X., Aihara, K. and Chen, L. (2014) Early diagnosis of complex diseases by molecular biomarkers, network biomarkers, and dynamical network biomarkers. Med. Res. Rev., 34, 455–478

42. Zeng, T., Zhang, W., Yu, X., Liu, X., Li, M., Liu, R. and Chen, L. (2014) Edge biomarkers for classification and prediction of phenotypes. Sci. China Life Sci., 57, 1103–1114

43. Zeng, T., Sun, S. Y., Wang, Y., Zhu, H. and Chen, L. (2013) Network biomarkers reveal dysfunctional gene regulations during disease progression. FEBS J., 280, 5682–5695

44. Liu, X., Liu, R., Zhao, X. M. and Chen, L. (2013) Detecting early-warning signals of type 1 diabetes and its leading biomolecular networks by dynamical network biomarkers. BMC Med Genomics, 6, S8