BMC Bioinformatics

RESEARCH

Open Access

DM-MOGA: a multi-objective optimization genetic algorithm for identifying disease modules of non-small cell lung cancer

Junliang Shang, Xuhui Zhu, Yan Sun*, Feng Li, Xiangzhen Kong and Jin-Xing Liu

*Correspondence:
sunyan225@126.com

School of Computer Science, Qufu Normal University, Rizhao 276826, China

Abstract

Background: Constructing molecular interaction networks from microarray data and then identifying disease module biomarkers can provide insight into the underlying pathogenic mechanisms of non-small cell lung cancer. A promising approach for identifying disease modules in the network is community detection.

Results: In order to identify disease modules from gene co-expression networks, a community detection method is proposed based on multi-objective optimization genetic algorithm with decomposition. The method is named DM-MOGA and possesses two highlights. First, the boundary correction strategy is designed for the modules obtained in the process of local module detection and pre-simplification. Second, during the evolution, we introduce Davies–Bouldin index and clustering coefficient as fitness functions which are improved and migrated to weighted networks. In order to identify modules that are more relevant to diseases, the above strategies are designed to consider the network topology of genes and the strength of connections with other genes at the same time. Experimental results of different gene expression datasets of non-small cell lung cancer demonstrate that the core modules obtained by DM-MOGA are more effective than those obtained by several other advanced module identification methods.

Conclusions: The proposed method identifies disease-relevant modules by optimizing two novel fitness functions to simultaneously consider the local topology of each gene and its connection strength with other genes. The association of the identified core modules with lung cancer has been confirmed by pathway and gene ontology enrichment analysis.

Keywords: Disease module identification, Biological network construction, Gene expression data, Genetic algorithm, Multi-objective optimization

Background

Lung cancer is the cancer with the highest mortality rate worldwide, and about 80% of cases are non-small cell lung cancer (NSCLC) that has a poor 5-year survival rate (average, 9–11 months) [1]. In recent years, research on molecular mechanisms of lung cancer has promoted the development of their corresponding targeted drugs which have

greatly improved the survival and prognosis of patients [2]. Meanwhile, more and more evidence has indicated that a group of genes related to a specific disease do not work in isolation, on the contrary, they usually interact with each other, thus gene co-expression networks (GCNs) have become a competitive model. Analysis on GCNs can help researchers identify disease modules. A disease module is considered as a subnetwork that contains most of the disease-related genes with compact topological connections and closely related functions, providing a system-level understanding of disease pathogenesis [3].

In recent years, many methods have been proposed for identifying disease modules. In general, these methods can be divided into four categories: local expansion, machine learning, mathematical programming and evolutionary algorithm (EA) [4]. DISeAse MOdule Detection algorithm (DIAMOnD) is a typical method based on local expansion which uses known disease-associated genes (seeds) to iteratively expand modules by evaluating the significance of the number of connections between seeds and other genes [5]. Although this method can identify disease modules, the coverage of disease-related genes may be low. The node and edge Prioritization-based Community Analysis is a knowledge-guided and network-based integration method to reveal functional modules in non-small cell lung cancer [6]. The protein–protein interaction network is prioritized by a random walk algorithm based on NSCLC seed genes and integrating edge weights, and then a "community network" is constructed in combination with Girvan-Newman and Label Propagation algorithms. MTGO is another method for functional module detection based on prior biological knowledge and topology information [7]. It directly utilizes gene ontology (GO) terms during module detection and labels each module with the most appropriate GO term, thereby simplifying the functional interpretation of modules. Molecular Complex Detection (MCODE) is a popular method without using prior knowledge for seeds [8]. It calculates the weight for each vertex according to the local neighborhood density. Nodes with the largest weight are selected as seeds, then the method traverses outwards and incorporates nodes with the weight higher than a given threshold into the module. SWItch Miner (SWIM) is a method to identify small modules containing key regulatory genes (switch genes) by introducing three topological attribute statistics for nodes [9]. SWIM was applied to a dataset from The Cancer Genome Atlas (TCGA) to characterize the etiology of interesting diseases. Identifying disease modules through machine learning methods is another efficient way. Wu and Stein used Markov Clustering (MCL) to cluster the weighted gene functional interaction network into a series of disease modules to respectively identify prognostic biomarkers of breast cancer and ovarian cancer [10]. PS-MCL (Parallel Shotgun Coarsened MCL) was proposed by Lim et al., a parallel community detection method that outperforms MCL in both runtime and the division quality [11]. PS-MCL adopts an effective coarsening scheme called shotgun coarsening (SC) to improve the module fragmentation problem of MCL, while providing a multi-core parallel algorithm for community detection to increase scalability. In addition, machine learning based methods are more efficient to identify disease modules from multi-omics data. A greedy decision forest is proposed to identify community structure from molecular interaction networks [12]. It obtains a high degree of interpretability by using shapley additive explanations. A strongly interconnected disease module identification method called SigMod is proposed by *Liu* et

al. based on mathematical programming [13]. It identifies disease modules by integrating the results of genome-wide association study (GWAS) and gene networks, as well as optimizing the binary quadratic objective function by a graph min-cut approach. EAs are popular and widely used in the field of disease module identification. Multi-objective evolutionary algorithm (MOEA) DiffCoMO identifies differential co-expression modules by maximizing the difference between module membership value of genes corresponding to two different infection stages [14]. A new method ModuleDiscoverer is proposed to identify regulatory modules from the protein–protein interaction network (PPIN) and gene expression data [15]. It uses a randomization heuristic-based approximation of community structure to discover modules according to the maximum clique enumeration problem.

In this research, we construct a GCN relying on the PPIN, then we develop a disease-related module identification method based on the multi-objective genetic algorithm, named DM-MOGA. This method is utilized to analyze the obtained network by optimizing two fitness functions which can evaluate the functional similarity and the density of the topological connection of modules, respectively. In addition, a boundary correction strategy is designed for local modules obtained by pre-simplification, to reconfirm the genes in the margin belonging to which module.

Methods

Network construction

Studies have confirmed that integrating gene expression data and PPIN helps people understand the complex multi-layered molecular structure of human diseases. Therefore, two gene expression datasets of NSCLC in the NCBI Gene Expression Omnibus (GEO) database are selected to construct GCNs with PPIN information being referred to, respectively. Detailed steps of data preprocessing and network construction are as follows. First of all, limma package in the R/Bioconductor software is utilized to identify differentially expressed genes (DEGs) whose t-statistics p value are adjusted by the Benjamini–Hochberg method [16]. Genes with the adjusted p value less than 0.05 are considered as DEGs. Only interactions between DEGs are used to construct GCNs.

To estimate the interaction intensity between DEGs, a new criterion called Gaussian Copula Mutual Information (GCMI) is introduced [17]. GCMI uses the concept of a statistical copula to provide the advantages of Gaussian parametric estimation for variables with any type of marginal distributions, and it is suitable for estimating MI between two continuous variables. At the same time, we use the sim_{Rel} score to calculate and compare functionally related products of a pair of DEGs which provides a similarity criterion for gene ontology (GO) terms of two gene products. The computation of sim_{Rel} has been implemented by the R package GOSemSim [18]. The definition of sim_{Rel} is as follows,

$$sim_{Rel}(c_1, c_2) = \max_{c \in S(c_1, c_2)} \left(\frac{2 \cdot \log p(c)}{\log p(c_1) + \log p(c_2)} \cdot (1 - p(c)) \right) \tag{1}$$

where $S(c_1, c_2)$ is the set of common ancestors of GO terms c_1 and c_2, $p(c)$ is the probability of c. It is utilized to calculate a fitness value in "Fitness functions" section.

After obtaining the correlation matrix between DEGs, it is compared with the PPIN. We downloaded the PPIN from the human protein reference database (HPRD) which

contains 39,240 interactions [19]. The original protein–protein interaction information is presented in Additional file 1. If a correlation does not exist in the PPIN, it will be modified to 0.

DM-MOGA framework

Due to the high complexity of identifying disease modules from a large-scale GCN, heuristic strategies are required to guide the search process. One of the most popular strategies is EA which are suitable for solving global optimization problems in the discrete search space [20]. EA is an optimization algorithm inspired by Darwin's principles of natural selection. Each solution is described as an individual in the population, and each individual is associated with one or more fitness functions optimized by natural selection process. In this paper, we propose a new method DM-MOGA based on MOEA and decomposition to identify modules which is regarded as biomarkers of NSCLC. The workflow of DM-MOGA is displayed in Fig. 1. In this section, we describe the framework of DM-MOGA, including pre-simplification with boundary correction, chromosome encoding and initialization scheme, operators of MOEA and the optimal solution selection strategy. After the evolution is completed, the result with the largest W' in the Pareto front is considered as the final solution that contains hundreds of modules. We only select the biggest module involving more biological information from the GCN.

Local module pre-simplification and boundary correction

In order to improve the adaptability of DM-MOGA for large-scale biological networks, a pre-simplification strategy for local module (LM) is introduced from [21] and executed before the evolution. This strategy randomly selects one node a from the network, then a LM is defined as containing node a, its neighbor a_k with the largest degree, $a_k's$ neighbor a_{kk} that has the largest number of common neighbors with a_k, and all joint neighbors of a_k and a_{kk}. For neighbors of all nodes in the LM, those neighbors whose number of connections with LM is beyond half of its degree are also added to LM. Finally, a complete graph of order 3 in the LM is selected and simplified to a single node. Specifically, we will not consider the possibility that nodes in the LM do not belong to the same module during the evolution. Above operations are repeated to find another LM from the remaining nodes of the network until all nodes are assigned to a LM.

However, this strategy only considers the topology of the network without the weight of edges, and vertices of some edges with smaller weights do not necessarily belong to a LM. Therefore, we develop a module boundary correction strategy. In this strategy, the matrix *NodeTable* for all nodes is maintained, of which each row represents a node. The first column is the index of the node, the second column is the module to which the node belongs after the LM pre-simplification strategy, and the third column is the new module to which the node belongs after the boundary correction.

The constructed GCN is denoted as $G = (V, E)$, where $V = \{v_i | i = 1, 2, \ldots, N\}$ is the set of nodes, $E = \{e_i | i = 1, 2, \ldots, M\}$ is the set of connections between a pair of nodes. Firstly, calculate the weight $\{V_1^w, V_2^w, \ldots, V_i^w, \ldots V_N^w\}$ for each node in the network. Specifically, the most densely connected area in the module made up of node i and its immediate neighbors is defined as the highest k-core, and V_i^w is the product of the density and the minimum degree of the highest k-core. The density of the highest

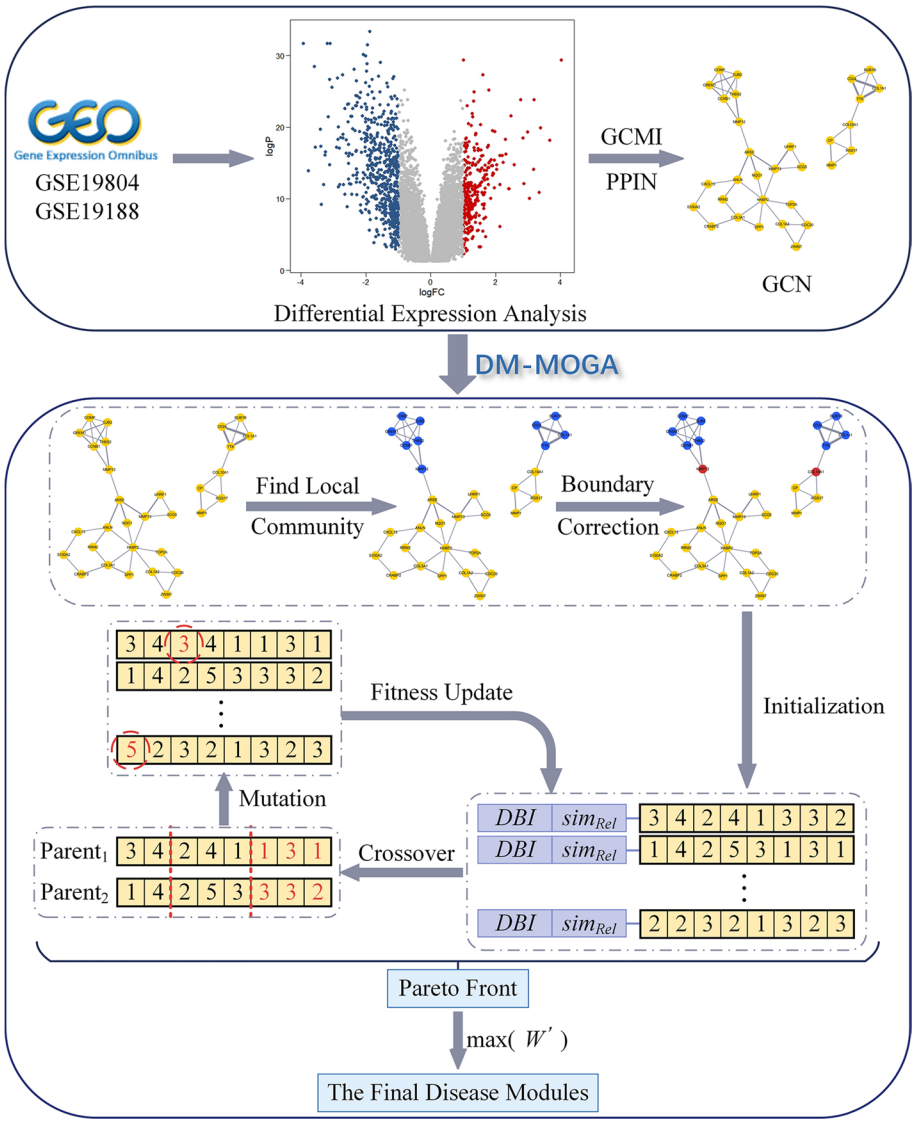

Fig. 1 The DM-MOGA workflow

k-core is $D^G = \frac{2E}{|V|(|V|-1)}$ [8]. Then, we get the attribute $\{V_1^a, V_2^a, \ldots, V_i^a, \ldots V_N^a\}$ for each node, where $V_i^a = \sum_j E_{ij},\ j \in LC_{NodeTable(i,2)}$. For each module, if node i satisfies $(V_i^w > 2)\&(V_i^w \geq \overline{V^w})$, node i is reserved in this module; otherwise, the third element of the corresponding row of node i in the *NodeTable* is set to 0, indicating that node i will be reassigned afterwards. Secondly, for each LM, the node with the largest sum of the weight of connecting edges is selected as the seed node, and neighbors of this seed are also recursively assigned to this module. For those nodes that still cannot be allocated to a LM, they are retained in the network independently.

Fitness functions

The proposed DM-MOGA identifies disease modules by minimizing the following two fitness functions. The first function is Davies–Bouldin Index (DBI) which should have

been a measure to evaluate the quality of clustering results [22]. The basic idea of DBI is to evaluate the distance between two clusters, considering that the distance between nodes belonging to different clusters should be as large as possible and the distance between nodes within a cluster should be as small as possible. With the similarity matrix obtained by calculating sim_{Rel} between genes, DBI is applied to assess the similarity of functions of genes belonging to the same disease module. The formula is as follows:

$$DBI = \frac{1}{C_{num}} \sum_{i=1}^{C_{num}} \max_{i:j \neq i} \frac{S_i + S_j}{dist\left(v_i, v_j\right)} \tag{2}$$

where C_{num} is the number of modules, $dist(\cdot)$ is the sim_{Rel} similarity between two nodes, $S_i = \frac{1}{|C_i|} \sum_{x_j \in C_i} dist\left(x_j, v_i\right)$ measures the extent of dispersion of module C_i, C_i represents the i-th module, x_j is the j-th node of C_i, and v_i represents the center of C_i.

The second function is a modified clustering coefficient (CC') suitable for weighted networks. In general, CC' quantifies the aggregation of one node and its neighbors, but it is modified to the sum of CC' of nodes to evaluate the overall module, and the specific formula is as follows. For a random module C_i, clustering coefficient is defined as:

$$CC_i' = \sum\nolimits_{x_j \in C_i} \frac{2 \sum_{x_l, x_k \in R_{x_j} \wedge l \neq k} E(x_l, x_k)}{\left|R_{x_j}\right|\left(\left|R_{x_j}\right| - 1\right)} \tag{3}$$

where R_{x_j} is the set of neighbors of node x_j. Since the result of MOEA is a group of modules, the second fitness function is set to the maximum of CC'.

The criterion W proposed by Zhao et al. is used to select a solution from the Pareto front obtained by MOEA as the final result [23]. The result with the largest W is considered as the final result. The original W is applied to extract a module from unweighted social networks. In order to put all modules into consideration and adapt to weighted GCNs, W is changed to the following form:

$$W' = \sum_{i=1}^{C_{num}} \left(\frac{O(C_i)}{|C_i|^2} - \frac{B(C_i)}{|C_i|\left|C_i'\right|} \right) \tag{4}$$

where C_i' is the complement of C_i, $O(C_i) = \sum\limits_{j,k \in C_i} E_{j,k}$, $B(C_i) = \sum\limits_{j \in C_i, k \in C_i'} E_{j,k}$.

Multi-objective optimization based on decomposition

The basic theory is multi-optimization problem (MOP) based on Pareto optimum which is to optimize a group of functions at the same time:

$$\min F(x) = \left(f_1(x), f_2(x), \ldots, f_k(x)\right)^T \tag{5}$$

where $x = [x_1, x_2, \ldots, x_N] \in \Omega$ and Ω is the feasible region. Then, the definition of dominance relationship is explained, that is, x_A dominates x_B (written as $x_A \succ x_B$, $x_A, x_B \in \Omega$) if and only if:

$$\forall i \in \left\{1, 2, \ldots, k\right\} f_i(x_A) \leq f_i(x_B) \wedge \exists j \in \left\{1, 2, \ldots, k\right\} f_j(x_A) < f_j(x_B) \tag{6}$$

If there is no vector $x \in \Omega$ such that $x \succ x^*$, x^* is called a non-dominated solution or Pareto-optimal solution.

MOEA/D-Net is a community detection method based on MOEA with decomposition. It decomposes a MOP into a number of scalar optimization subproblems and optimizes them simultaneously by population evolution. At each iteration, the population is made up of the best solution found for each subproblem since the beginning of evolution. In MOEA/D-Net, the popular Tchebycheff method is used to construct the aggregation function and therefore the scalar optimization subproblems are in the form:

$$\min g^{te}\left(x|\lambda_i, z^*\right) = \max_{j=1}^{2}\left\{\lambda_i^j\left|F_j(x) - z^*\right|\right\} \tag{7}$$

where $\lambda = \left\{\lambda_1, \lambda_2, \ldots, \lambda_{pop}\right\}$ is a series of weight vectors uniformly distributed on $\lambda_i^1 + \lambda_i^2 = 1$, $\lambda = \left\langle\lambda_i^1, \lambda_i^2\right\rangle \in [0,1]$, $i = \{1, 2, \ldots, pop\}$, pop is the population size, and $z^* = \left\langle z_1^*, z_2^*\right\rangle$ is the reference point in which each point z_j^* corresponds to the minimum value of a fitness function obtained from the population. For each target vector λ_i, calculate the Euclidean distance between all weight vectors and λ_i, and the neighborhood of λ_i, denoted as $Neib_i$, is made up of nm individuals with the smallest Euclidean distance to λ_i, where nm is a predefined parameter. For each non-dominated individual, there is a weight vector that makes it the optimal solution of Eq. 7, and each optimal solution of Eq. 7 is a Pareto-optimal solution of Eq. 5.

Initialization

Individuals are encoded and initialized based on the locus-based adjacency encoding schema which is popular in EA-based community detection algorithms [24]. In an individual, each element is initialized as a random neighbor index of its corresponding node or the index of the corresponding node itself, and then this element is recursively replaced by the index that most neighbors of this node share until the element is not changed. Another variable that needs to be initialized is the reference point z^*, and it is set to the minimum of two fitness functions in the initial population.

The main loop of DM-MOGA

DM-MOGA adopts the similar framework with MOEA/D-Net that is proposed by Gong et al. [25]. In this method, the following procedure is applied to evolve the population. Every individual $p_j\left(1 \le j \le pop\right)$ is used to perform the crossover and mutation operation with another randomly selected individual to generate a *child*. If the Tchebycheff value of the *child* is better than a neighbor in $Neib_j$, replace that neighbor with the *child* and update the reference point z^*. Specifically, we choose the two-point crossover to take advantage of protecting the effective connection between nodes. We randomly select two elements i and j (i.e., $1 \le i \le j \le N$), and elements in $[i, j]$ are exchanged between two parents in the population. After the crossover operation is finished, an individual p_j is randomly selected for mutation, on which the neighbor-based mutation is performed. According to the encoding strategy, the mutation operator is to randomly select an element e_l in the individual p_i and replace the neighbor index in it with the index of other neighbors of the node corresponding to e_l. DM-MOGA will continue to evolve until the maximum number of generations is reached.

Results and discussion

Datasets

In the experiments, two NSCLC expression microarray datasets obtained using the same platform (GPL570) were downloaded from the GEO database. The first dataset (ID: GSE19804) [26] is balanced and contains 60 disease and control samples, respectively. The second dataset (ID: GSE19188) [27] contains 91 disease samples and 65 control samples. In the above two datasets, each sample contains the expression data of 21,879 genes. After differential expression analysis, 7669 DEGs and 10,496 DEGs were respectively selected from GSE19804 and GSE19188. Detailed information of these two datasets is shown in Table 1.

Comparison with other methods

Ground-truth dataset

We integrated four kinds of lung cancer-related genes obtained from the MalaCards database as ground truth, including differentially expressed genes, genes related to lung cancer, genes contained in lung cancer related pathways, top affiliated genes of GO terms related to lung cancer [28].

Comparison methods

In the experiment, five methods were used to compare the performance with DM-MOGA, that is, a network reduction-based MOEA for community (module) detection (RMOEA), a disease module identification method SigMod, MCODE, and two classic module identification methods from the R package igraph, that is, Hierarchical Clustering [29] and Louvain [30].

To make a fair comparison, parameters in DM-MOGA and RMOEA were set to the same value, namely, the number of iterations $max\_gen = 100$, the population size $pop = 50$, the neighborhood size $nm = 40$, and the mutation rate was 0.1. In addition, to ensure that SigMod can search for modules smoothly, the parameter *maxjump* was set to 27, and we used the default value for parameters of other methods.

Classification performance of disease and healthy samples

Fivefold cross-validation was applied on the largest module to verify its effectiveness as a biomarker. The set of samples is randomly divided into five parts with the same size, one of which is selected as the test set each time, and the other four parts are used for training (the train set). Support vector machine (SVM) is used as the classifier, and the value of five criteria (Accuracy, Precision, Recall, F1 and AUC) of each experiment is taken as the cross-validation result. Since there is a random value during the five-fold cross-validation, for each identified module, we performed five-fold cross-validation for ten times independently, and the final result was the average value of each criterion in experiments. Figures 2 and 3 respectively display the classification performance of the disease module identified by six

Table 1 Details of two datasets

Datasets	Tumor samples	Normal samples	Genes
GSE19804	60	60	21879
GSE19188	91	65	21879

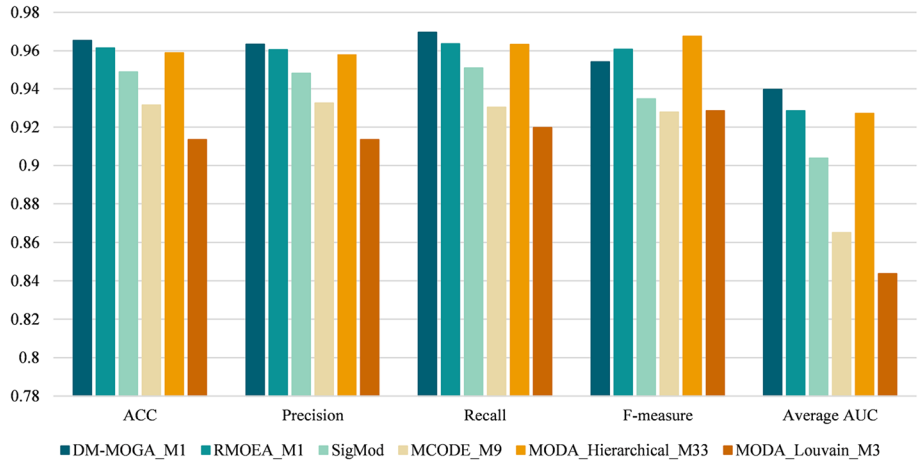

Fig. 2 The value of the five-fold cross-validated classification index of the optimal module in GSE19804

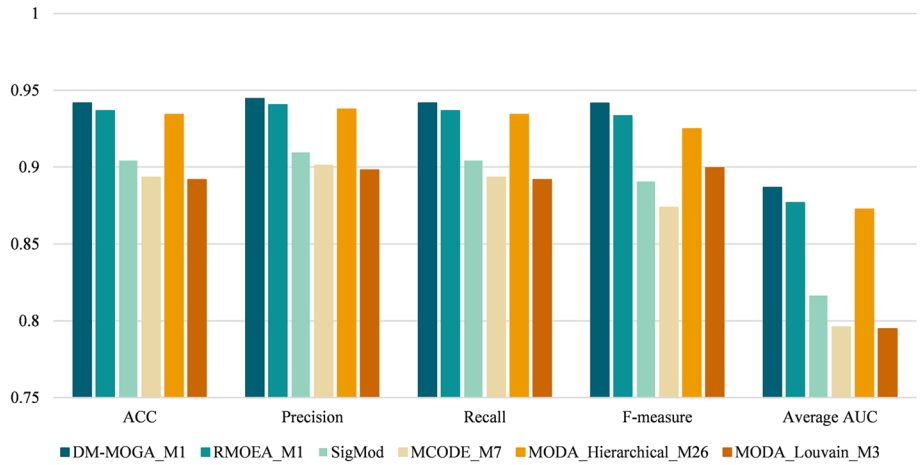

Fig. 3 The value of the five-fold cross-validated classification index of the optimal module in GSE19188

methods on two NSCLC gene expression datasets. The serial number of the disease module is marked after the abbreviation of comparison methods. According to the figures, module 1 (M1) in community detection results of DM-MOGA obtained the best classification performance on the GSE19804 dataset whose value of five criteria is better than other methods. As for the GSE19188 dataset, the classification performance of the module detected by DM-MOGA was also significant which obtained the maximum value on four of the five metrics. Therefore, the effectiveness of the module obtained from DM-MOGA is verified in guiding the classification of disease and control samples. Moreover, the basic framework of RMOEA is similar to that of DM-MOGA, and both have the ability to effectively guide sample classification. The difference between them is that RMOEA lacks the boundary correction strategy which may lead to unstable results in ten independent experiments. The reason for SigMod, hierarchical clustering and the Louvain algorithm that fails to provide reasonable results might be the same, that is, the default values of their key parameters are not applicable for GCNs. MCODE tends to obtain smaller modules because it has strict conditions for expanding nodes in modules.

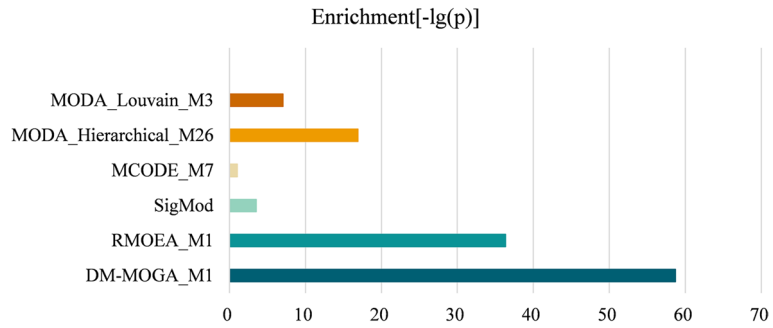

Fig. 4 Enrichment of the module discovered by different methods from GSE19804

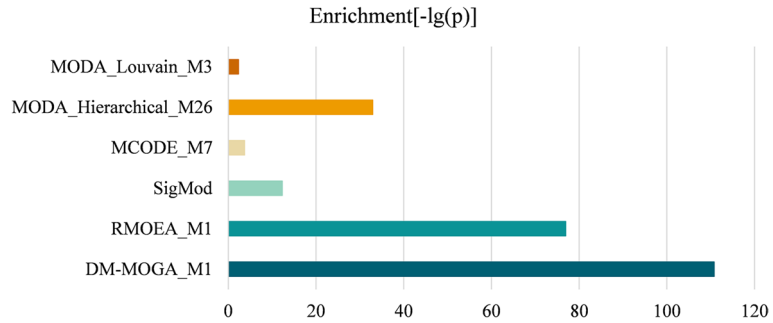

Fig. 5 Enrichment of the module discovered by different methods from GSE19188

Enrichment analysis for the disease module

The quality of the identified module was further quantitatively studied by gene set enrichment analysis [31]. A hypergeometric test is used to estimate the enrichment of the identified module with reference to the ground truth set, where the statistical significance of the enrichment is defined as follows [32]:

$$p = \sum_{k=N_{gm}}^{N_g} \left[\binom{N_g}{k} \binom{N - N_g}{N_m - k} \Big/ \binom{N}{N_m} \right] \tag{8}$$

where N_g is the number of genes in the ground-truth dataset, N_m is the number of genes in the identified module, N_{gm} is the number of genes that belong to both the ground truth set and the identified module. A smaller p value indicates a more significant enrichment of genes in the identified module. Figures 4 and 5 display the $-\log(p)$ value of the module obtained by comparison methods running on GSE19804 and GSE19188, respectively. It can be observed that the proposed method obtains significantly better p value than the other methods on both two GCNs.

Effectiveness of GCMI

In order to study whether the gene–gene interaction metric can affect the efficiency of modules identified by DM-MOGA, we employed the most commonly used Pearson Correlation Coefficient (PCC) to reconstruct GCNs for comparison. Except for the

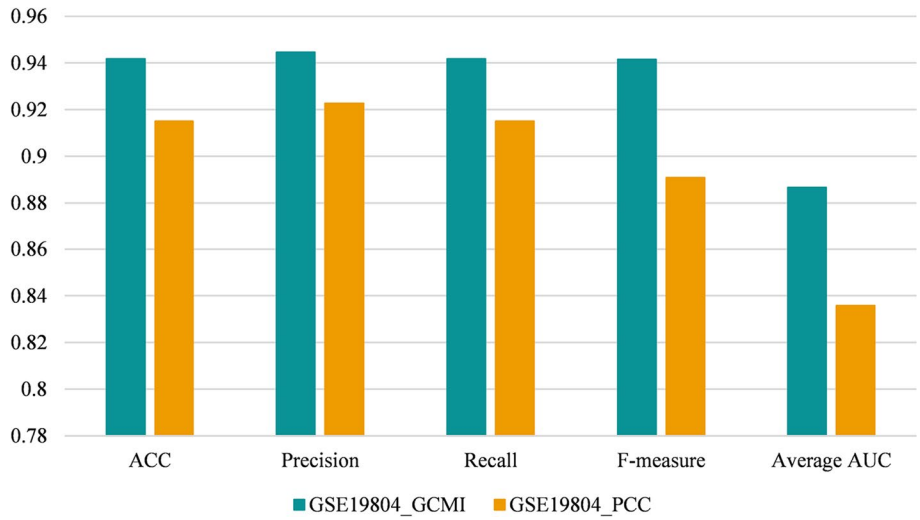

Fig. 6 The classification performance of the module identified based on different interaction metrics from GSE19804

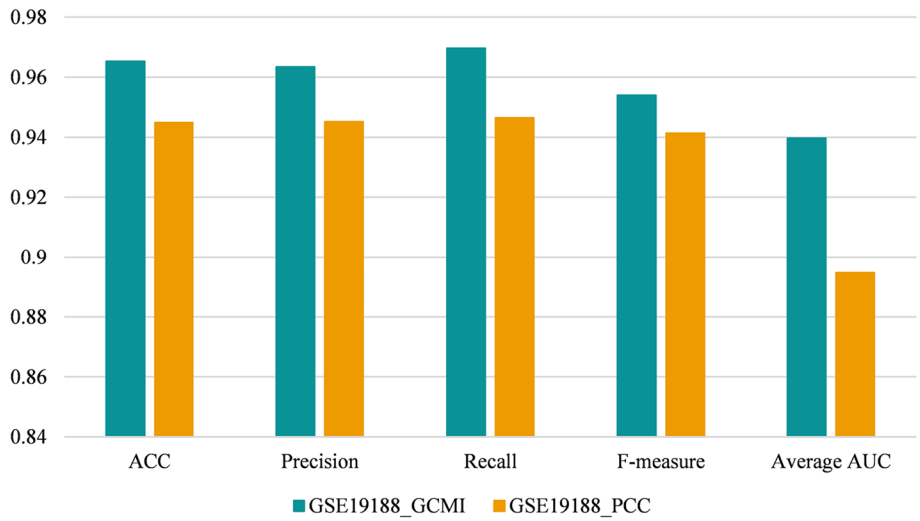

Fig. 7 The classification performance of the module identified based on different interaction metrics from GSE19188

interaction criterion, other steps of network construction remain unchanged. Then the proposed method was applied to detect modules on the new networks. Figures 6 and 7 respectively show the classification performance of the module identified by DM-MOGA based on different criteria on the two datasets. It can be observed that GCMI we choose to calculate edge weights can improve the biological significance of the disease module.

Identification of modules associated with lung cancer

Pathways

Pathway enrichment analysis was implemented by the KOBAS v3.0 web server, in which four datasets are considered in the analysis, including the KEGG pathway [33], BioCyc,

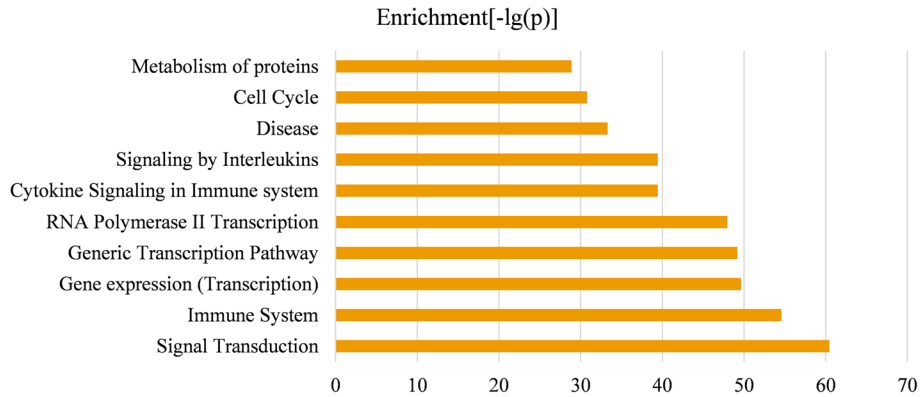

Fig. 8 Top 10 significantly enriched pathway terms associated with genes in the identified module of GSE19804

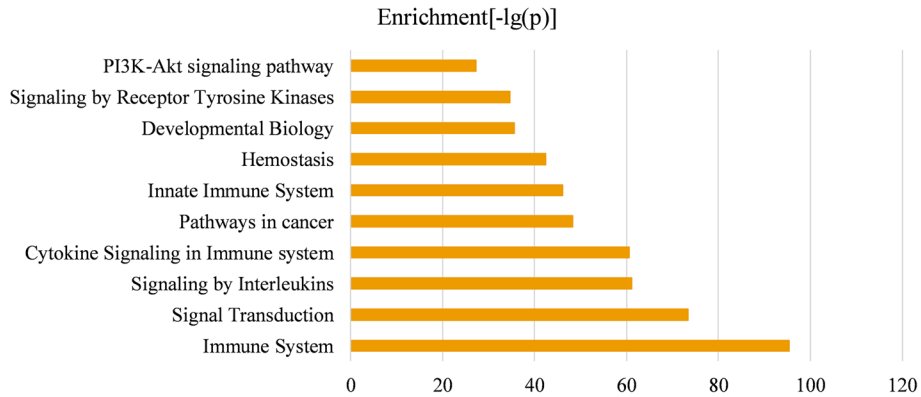

Fig. 9 Top 10 significantly enriched pathway terms associated with genes in the identified module of GSE19188

PANTHER, and Reactome [34, 35]. Pathways that are considered to be significant only if the Benjamini–Hochberg adjusted p value ≤ 0.05. Pathway enrichment analysis was performed on the disease modules of GSE19804 and GSE19188 independently, and ten pathways with the smallest p values $(-\log(p))$ were respectively displayed in Figs. 8 and 9. Moreover, in Additional file 2, we provided the enrichment significance of these pathways in the two disease modules. Except for the Pathways in cancer pathway that is from the KEGG pathway database, the other pathways in Figs. 8 and 9 are all from the Reactome pathway database [33, 34]. Each pathway was proved to be connected with lung cancer more or less. To be specific, recent studies have found that the mutation frequency of the PTEN locus in lung cancer is high, and there is a strong correlation between loss of PTEN function and positive expression (p value < 0.05) of EGFR, TGF-α and P-AKT signal transduction pathway (adjusted p value $= 4.39\text{E}{-61}$) in the development of NSCLC [36]. Besides, there are evidence demonstrating that other signaling pathways enriched in the two disease modules are associated with lung cancer, for instance, the PI3K/Akt signaling pathway (adjusted p value $= 5.87\text{E}{-28}$) inhibiting the metastasis of A549 cell line from lung adenocarcinoma, the receptor tyrosine kinases (RTKs) (adjusted p value $= 2.1\text{E}{-35}$) participating in the signal transmission across

the plasma membrane, signaling by interleukin (adjusted p value = 4.29E−40) and cell cycle (adjusted p value = 1.97E−31) [37–40]. In Figs. 8 and 9, all pathways related to the immune system were confirmed to participate in the progression of NSCLC, including the immune system (adjust p-value = 3.23E-55), cytokine signaling in immune system (adjusted p value = 4.23E−40) and innate immune systems (adjusted p value = 6.65E−47) [41–43] RNA Polymerase II Transcription pathway participates gene expression (transcription) pathway (adjusted p value = 2.72E−50) and generic transcription pathway (adjusted p value = 7.61E−50). By inhibiting RNA polymerase II-dependent transcription (adjust p value = 1.29E−48), cell growth in the malignant cell line A549 can be effectively inhibited [44, 45]. Studies have found that metabolism of protein (adjusted p value = 1.53E−29) is related to cancer cachexia. In cancer cachexia, overall protein synthesis is decreasing that is directly proportional to tumor growth [46]. Some researchers believe that cancer is a developmental biology (adjusted p value = 2.48E−36) problem. They found that embryos and cancer have a number of common cellular and molecular features [47]. It is known that hemostatic biomarkers (adjusted p value = 3.31E−43) can affect the survival and venous thromboembolism (VTE) occurrence in lung cancer patients [48].

GO terms

GO enrichment analysis was implemented by the R package clusterProfiler [49]. In Figs. 10 and 11, the top ten GO terms that were most significantly enriched in the modules identified from the two datasets are respectively shown. The modules obtained in GSE19804 and GSE19188 were both significantly enriched in regulation of protein serine/threonine kinase activity (GO:0071900, adjusted p value = 3.87E−44). Under EGF-stimulated conditions, it is revealed that proteins interacting with B-Raf are enriched in

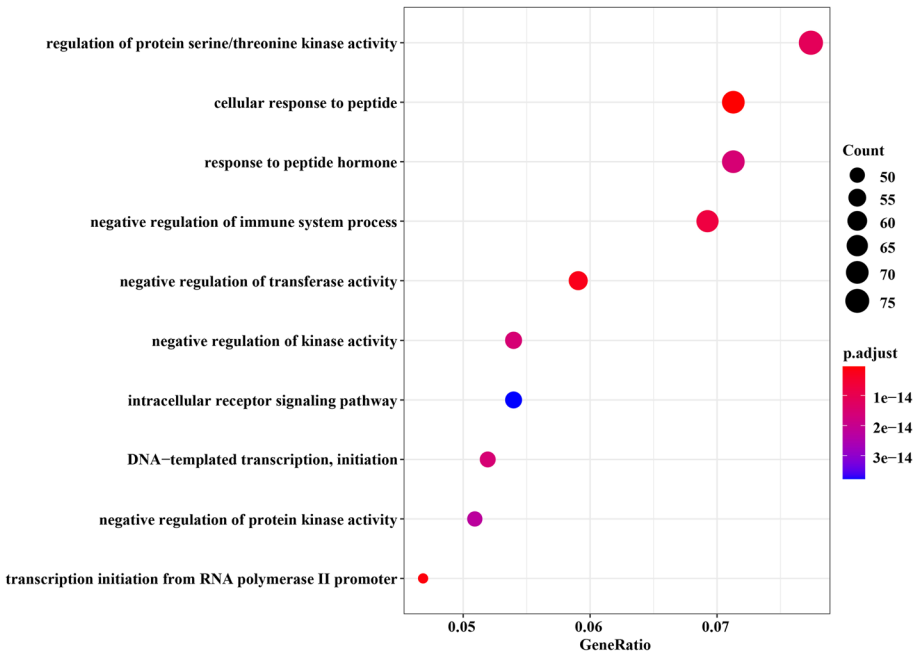

Fig. 10 Top 10 significantly enriched GO terms associated with genes in the identified module of GSE19804

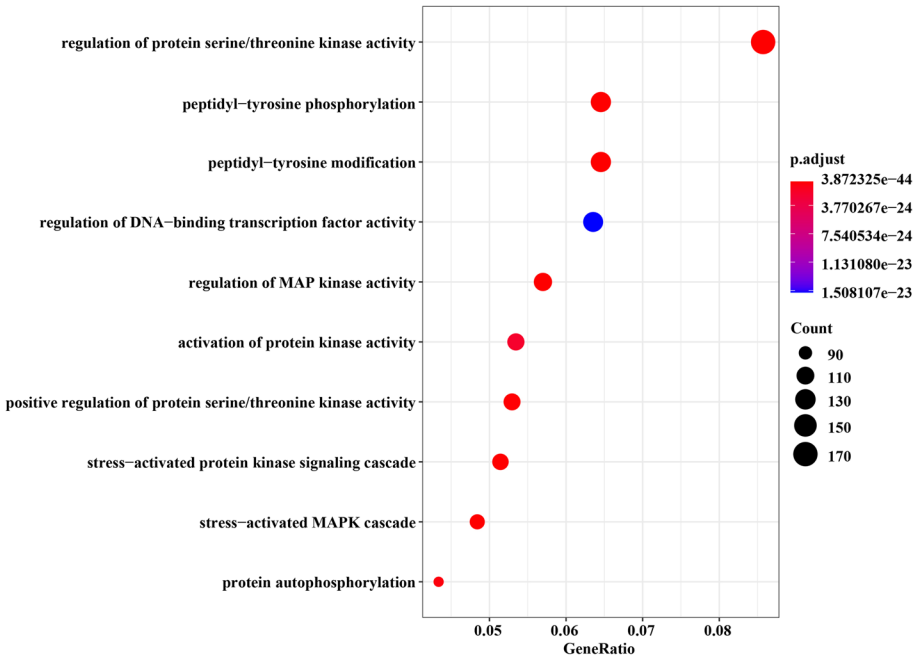

Fig. 11 Top 10 significantly enriched GO terms associated with genes in the identified module of GSE19188

regulation of protein serine/threonine kinase activity, and several interacting partners of B-Raf are enriched in NSCLC [50].

Besides, the core module from GSE19804 was mainly enriched in GO:1901653, GO:0043434, GO:0002683, GO:0051348, GO:0033673, GO:0030522, GO:0006352, GO:0006469, GO:0006367, corresponding to cellular response to peptide, response to peptide hormone, negative regulation of immune system process, negative regulation of transferase activity, negative regulation of kinase activity, intracellular receptor signaling pathway, DNA-templated transcription, initiation, negative regulation of protein kinase activity, transcription initiation from RNA polymerase II promoter. Several studies have confirmed that these GO terms are related to the pathogenesis and development of NSCLC. For instance, studies have found that licorice is a potential NSCLC treatment drug, and the targets of licorice are enriched in the biological process of cell response to peptide (GO:1901653, adjusted p value = 3.16E−16) which is an important way to stimulate the acquired immune system [51]. DHX36 plays a role in lung cancer cells by regulating signaling pathways such as response to peptide hormone (GO:0043434, adjusted p value = 1.57E−14) [52]. Interleukin-34 (IL-34) is significantly enriched in "negative regulation of immune system processes" (GO:0002683, adjusted p value = 7.18E−15) which is highly expressed in primary lung cancer tissues and associated with poor prognosis [53]. Exposure to cigarette smoke (CS) can cause injury to the epithelial cells of the respiratory tract and is considered to be one of the pathogenic factors of lung cancer. DEGs were extracted by comparing BEAS-2B cells (a he human bronchial epithelial cell line) before exposure to CS with after, and they are significantly enriched in the negative regulation of transferase activity (GO:0051348, adjusted p value = 2.21E−15) [54]. In addition, negative regulation of kinase activity (GO:0033673, adjusted p value = 1.57E−14), intracellular receptor signaling pathway (GO:0030522, adjusted p value = 3.78E−14),

DNA-templated transcription, initiation (GO:0006352, adjusted p value = 1.57E−14), negative regulation of protein kinase activity (GO:0006469, adjusted p value = 2.25E−14) and transcription initiation from RNA polymerase II promoter (GO:0006367, adjusted p value = 1.10E−15) are the most enriched terms by GO enrichment analysis on key lung cancer-related gene sets that have been reported in other studies [55–59].

In Fig. 11, the identified module was mainly enriched in GO:0018108, GO:0018212, GO:0051090, GO:0043405, GO:0032147, GO:0071902, GO:0031098, GO:0051403, GO:0046777, corresponding to peptidyl-tyrosine phosphorylation, peptidyl-tyrosine modification, regulation of DNA-binding transcription factor activity, regulation of MAP kinase activity, activation of protein kinase activity, positive regulation of protein serine/threonine kinase activity, stress-activated protein kinase signaling cascade, stress-activated MAPK cascade, protein autophosphorylation. These GO terms have a certain correlation with lung cancer. Somatic variants that can be detected in the matched lymph node metastases but not in the primary lung cancer, are termed as LME-SMs genes. They are enriched in GO terms, for instance, peptidyl-tyrosine phosphorylation (GO:0018108, adjusted p value = 6.70E−33) and peptidyl-tyrosine modification (GO:0018212, adjusted p value = 1.16E−32) [60]. In [61], there are 35 genes that have been reported to be related to lung cancer. They are mainly related to GO terms in biological pathways, such as regulation of DNA-binding transcription factor activity (GO:0051090, adjusted p value = 1.51E−23), positive regulation of DNA-binding transcription factor activity, etc. In [62], it is found that cPLA2 is over-expressed in NSCLC cells transformed by oncogenic Ras, and cPLA2 is a well-known substrate of MAP kinase and closely related to the regulation of MAP kinase activity (GO:0043405, adjusted p value = 8.96E−32). Stem cell factor (SCF) and its receptor c-kit proto-oncogene are co-expressed in at least 70% of small cell lung cancer tumors and tumor-derived cell lines. The binding of SCF to c-Kit leads to receptor dimerization and activation of protein kinase activity (GO:0032147, adjusted p value = 1.62E−24) [63]. In other studies, the rest GO terms are also the most enriched terms in the results of GO enrichment analysis on key lung cancer-related gene sets [64–66].

Conclusions

In this work, to identify disease modules in GCNs, a multi-objective optimization method DM-MOGA is proposed based on the MOEA framework with decomposition. In DM-MOGA, the first step is to respectively construct the GCN on two NSCLC gene expression datasets, in which GCMI between all genes is calculated and considered as edge weights, and then the edges are filtered by referring to the prior knowledge of the PPI network. Secondly, DM-MOGA is separately executed on two GCNs that searches for disease modules by simultaneously optimizing two novel fitness functions, DBI and CC'. After the evolution is finished, the Pareto-optimal solution with the largest W' is selected as the final result.

To examine the validity of disease modules obtained through the above process, a series of experiments performed. First of all, DM-MOGA was compared with several other module identification methods from the following aspects, specifically, the classification effect of disease and control samples guided by modules, the enrichment of

modules in the disease-related gene set, and the validity of the edge weight criterion. Then, the correlation between modules and lung cancer was verified by pathway and GO term enrichment analysis. Experiments proved that the biological meaning of key modules obtained by DM-MOGA was more significant.

The proposed method possesses two main advantages. First, two fitness functions that have never been used for module identification problems are introduced which effectively improve the accuracy of the module in guiding patient classification. Second, the boundary correction strategy is designed for local modules, so that nodes with high correlation strength and low degree can be incorporated into the module. However, there are still some works in this field that can be further studied. On the one hand, it is necessary to develop fitness functions that are more suitable for disease module identification; on the other hand, studying the improvement strategies of EAs can further improve search efficiency.

Abbreviations

NSCLC	Non-small cell lung cancer
EA	Evolutionary algorithm
DIAMOnD	DISeAse MOdule Detection algorithm
MCODE	Molecular complex detection
MCL	Markov clustering
GWAS	Genome-Wide Association Study
MOEA	Multi-objective evolutionary algorithm
PPIN	Protein–protein interaction network
GEO	Gene expression omnibus
DEGs	Differentially expressed genes
GCMI	Gaussian copula mutual information
GO	Gene ontology
HPRD	Human Protein Reference Database
LM	Local module
DBI	Davies–Bouldin Index
CC	Clustering coefficient
MOP	Multi-optimization problem
SVM	Support vector machine
PCC	Pearson correlation coefficient
RTKs	Receptor tyrosine kinases
VTE	Venous thromboembolism
IL-34	Interleukin-34
CS	Cigarette smoke
LN	Lymph node
SCF	Stem cell factor

Supplementary Information

The online version contains supplementary material available at https://doi.org/10.1186/s12859-023-05136-z.

Additional file 1: The protein–protein interaction information downloaded from HPRD.

Additional file 2: The union of the 10 most significantly enriched pathways in the two disease modules and the genes overlapping between these modules.

Acknowledgements
Not applicable.

Author contributions
XZ wrote the main manuscript text. XZ and YS designed the project and acquired the data. FL and XK interpreted the results. JS, XZ and YS drafted the manuscript. After that, JXL revised and refined the manuscript substantively. All authors reviewed and approved the final manuscript.

Funding
This work was funded by the National Science Foundation of China to J.S. (61972226), F.L. (61902216) and J-X.L. (61872220).

Availability of data and materials
The code for DM-MOGA is available at https://github.com/LyanMelrose/DM-MOGA.git. The PPIN is available from HPRD (http://www.hprd.org; see also Additional file 1). Information of genes associated with NSCLC is compiled from the MalaCards database (https://www.malacards.org/). Gene expression dataset GSE19804 and GSE19188 are available from the GEO database (GSE19804: https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE19804; GSE19188: https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE19188). All tools used in this work are available from the respective links (the KOBAS v3.0 web server: http://kobas.cbi.pku.edu.cn/kobas3, R package limma: http://bioconductor.org/packages/release/bioc/html/limma.html; R package GOSemSim: http://bioconductor.org/packages/release/bioc/html/GOSemSim.html; R package clusterProfiler: http://bioconductor.org/packages/release/bioc/html/clusterProfiler.html; and R package igraph: https://igraph.org/r/).

Declarations

Ethics approval and consent to participate
Not applicable.

Consent for publication
Not applicable.

Competing interests
The authors declare that they have no competing interests.

Received: 23 April 2022 Accepted: 4 January 2023
Published online: 09 January 2023

References

1. Andonegui-Elguera SD, Zamora-Fuentes JM, Espinal-Enríquez J, Hernández-Lemus E. Loss of long distance co-expression in lung cancer. Front Genet. 2021;12:625741.
2. Yousefi M, Bahrami T, Salmaninejad A, Nosrati R, Ghaffari P, Ghaffari SH. Lung cancer-associated brain metastasis: molecular mechanisms and therapeutic options. Cell Oncol. 2017;40(5):419–41.
3. Mahapatra S, Mandal B, Swarnkar T. Biological networks integration based on dense module identification for gene prioritization from microarray data. Gene Reports. 2018;12:276–88.
4. Tian Y, Su X, Su Y, Zhang X. EMODMI: a multi-objective optimization based method to identify disease modules. IEEE Trans Emerg Top Comput Intell. 2020;5(4):570–82.
5. Sharma A, Menche J, Huang CC, Ort T, Zhou X, Kitsak M, et al. A disease module in the interactome explains disease heterogeneity, drug response and captures novel pathways and genes in asthma. Hum Mol Genet. 2015;24(11):3005–20.
6. Wang F, Han S, Yang J, Yan W, Hu G. Knowledge-guided, "community network" analysis reveals the functional modules and candidate targets in non-small-cell lung cancer. Cells. 2021;10(2):402.
7. Vella D, Marini S, Vitali F, Di Silvestre D, Mauri G, Bellazzi R. MTGO: PPI network analysis via topological and functional module identification. Sci Rep. 2018;8(1):1–13.
8. Bader GD, Hogue CW. An automated method for finding molecular complexes in large protein interaction networks. BMC Bioinform. 2003;4(1):1–27.
9. Paci P, Colombo T, Fiscon G, Gurtner A, Pavesi G, Farina L. SWIM: a computational tool to unveiling crucial nodes in complex biological networks. Sci Rep. 2017;7(1):1–16.
10. Wu G, Stein L. A network module-based method for identifying cancer prognostic signatures. Genome Biol. 2012;13(12):1–14.
11. Lim Y, Yu I, Seo D, Kang U, Sael L. PS-MCL: parallel shotgun coarsened Markov clustering of protein interaction networks. BMC Bioinform. 2019;20(13):1–12.
12. Pfeifer B, Baniecki H, Saranti A, Biecek P, Holzinger A. Multi-omics disease module detection with an explainable greedy decision forest. Sci Rep. 2022;12(1):1–15.
13. Liu Y, Brossard M, Roqueiro D, Margaritte-Jeannin P, Sarnowski C, Bouzigon E, et al. SigMod: an exact and efficient method to identify a strongly interconnected disease-associated module in a gene network. Bioinformatics. 2017;33(10):1536–44.
14. Ray S, Maulik U. Identifying differentially coexpressed module during HIV disease progression: a multiobjective approach. Sci Rep. 2017;7(1):1–13.
15. Vlaic S, Conrad T, Tokarski-Schnelle C, Gustafsson M, Dahmen U, Guthke R, et al. Modulediscoverer: identification of regulatory modules in protein-protein interaction networks. Sci Rep. 2018;8(1):1–11.
16. Kumar SU, Saleem A, Kumar DT, Preethi VA, Younes S, Zayed H, et al. A systemic approach to explore the mechanisms of drug resistance and altered signaling cascades in extensively drug-resistant tuberculosis. In: Donev R, Karabencheva-Christova T, editors., et al., Advances in protein chemistry and structural biology. Amsterdam: Elsevier; 2021. p. 343–64.
17. Ince RA, Giordano BL, Kayser C, Rousselet GA, Gross J, Schyns PG. A statistical framework for neuroimaging data analysis based on mutual information estimated via a Gaussian copula. Hum Brain Mapp. 2017;38(3):1541–73.
18. Yu G, Li F, Qin Y, Bo X, Wu Y, Wang S. GOSemSim: an R package for measuring semantic similarity among GO terms and gene products. Bioinformatics. 2010;26(7):976–8.
19. Keshava Prasad T, Goel R, Kandasamy K, Keerthikumar S, Kumar S, Mathivanan S, et al. Human protein reference database—2009 update. Nucleic Acids Res. 2009. https://doi.org/10.1093/nar/gkn892.

20. Muraro D, Simmons A. An integrative analysis of gene expression and molecular interaction data to identify dysregulated sub-networks in inflammatory bowel disease. BMC Bioinform. 2016;17(1):1–11.
21. Zhang X, Zhou K, Pan H, Zhang L, Zeng X, Jin Y. A network reduction-based multiobjective evolutionary algorithm for community detection in large-scale complex networks. IEEE Trans Cybern. 2018;50(2):703–16.
22. Karo IMK, MaulanaAdhinugraha K, Huda AF. A cluster validity for spatial clustering based on Davies Bouldin index and polygon dissimilarity function. In: 2017 second international conference on informatics and computing (ICIC); 2017. p. 1–6.
23. Zhao Y, Levina E, Zhu J. Community extraction for social networks. Proc Natl Acad Sci. 2011;108(18):7321–6.
24. Shahabi Sani N, Manthouri M, Farivar F. A multi-objective ant colony optimization algorithm for community detection in complex networks. J Ambient Intell Humaniz Comput. 2020;11(1):5–21.
25. Gong M, Ma L, Zhang Q, Jiao L. Community detection in networks by using multiobjective evolutionary algorithm with decomposition. Phys A Stat Mech Appl. 2012;391(15):4050–60.
26. Lu T-P, Tsai M-H, Lee J-M, Hsu C-P, Chen P-C, Lin C-W, et al. Identification of a novel biomarker, sema5a, for non–small cell lung carcinoma in nonsmoking women. Cancer Epidemiol Prev Biomark. 2010. https://doi.org/10.1158/1055-9965.
27. Hou J, Aerts J, Den Hamer B, Van Ijcken W, Den Bakker M, Riegman P, et al. Gene expression-based classification of non-small cell lung carcinomas and survival prediction. PLoS ONE. 2010. https://doi.org/10.1371/journal.pone.0010312.
28. Rappaport N, Twik M, Plaschkes I, Nudel R, Iny Stein T, Levitt J, et al. MalaCards: an amalgamated human disease compendium with diverse clinical and genetic annotation and structured search. Nucleic Acids Res. 2016. https://doi.org/10.1093/nar/gkw1012.
29. Zhang B, Horvath S. A general framework for weighted gene co-expression network analysis. Stat Appl Genet Mol Biol. 2005;4(1):17.
30. Blondel VD, Guillaume J-L, Lambiotte R, Lefebvre E. Fast unfolding of communities in large networks. J Stat Mech Theory Exp. 2008;2008(10):P10008.
31. Subramanian A, Tamayo P, Mootha VK, Mukherjee S, Ebert BL, Gillette MA, et al. Gene set enrichment analysis: a knowledge-based approach for interpreting genome-wide expression profiles. Proc Natl Acad Sci. 2005;102(43):15545–50.
32. Liu X, Liu Z-P, Zhao X-M, Chen L. Identifying disease genes and module biomarkers by differential interactions. J Am Med Inform Assoc. 2012;19(2):241–8.
33. Kanehisa M. Toward understanding the origin and evolution of cellular organisms. Protein Sci. 2019;28(11):1947–51.
34. Gillespie M, Jassal B, Stephan R, Milacic M, Rothfels K, Senff-Ribeiro A, et al. The reactome pathway knowledgebase. Nucleic Acids Res. 2022;50(D1):D687–92.
35. KOBAS-i. http://kobas.cbi.pku.edu.cn/kobas3. Accessed 23 Aug 2021.
36. Yun F, Jia Y, Li X, Yuan L, Sun Q, Yu H, et al. Clinicopathological significance of PTEN and PI3K/AKT signal transduction pathway in non-small cell lung cancer. Int J Clin Exp Pathol. 2013;6(10):2112.
37. Pan B, Shen J, Cao J, Zhou Y, Shang L, Jin S, et al. Interleukin-17 promotes angiogenesis by stimulating VEGF production of cancer cells via the STAT3/GIV signaling pathway in non-small-cell lung cancer. Sci Rep. 2015;5(1):1–13.
38. Singhal S, Vachani A, Antin-Ozerkis D, Kaiser LR, Albelda SM. Prognostic implications of cell cycle, apoptosis, and angiogenesis biomarkers in non–small cell lung cancer: a review. Clin Cancer Res. 2005;11(11):3974–86.
39. Lee Y-C, Lin H-H, Hsu C-H, Wang C-J, Chiang T-A, Chen J-H. Inhibitory effects of andrographolide on migration and invasion in human non-small cell lung cancer A549 cells via down-regulation of PI3K/Akt signaling pathway. Eur J Pharmacol. 2010;632(1–3):23–32.
40. Pisick E, Jagadeesh S, Salgia R. Receptor tyrosine kinases and inhibitors in lung cancer. Sci World J. 2004;4:589–604.
41. Carbone DP, Gandara DR, Antonia SJ, Zielinski C, Paz-Ares L. Non-small-cell lung cancer: role of the immune system and potential for immunotherapy. J Thorac Oncol. 2015;10(7):974–84.
42. Rusek AM, Abba M, Eljaszewicz A, Moniuszko M, Niklinski J, Allgayer H. MicroRNA modulators of epigenetic regulation, the tumor microenvironment and the immune system in lung cancer. Mol Cancer. 2015;14(1):1–10.
43. Chen J-B, Kong X-F, Qian W, Mu F, Lu T-Y, Lu Y-Y, et al. Two weeks of hydrogen inhalation can significantly reverse adaptive and innate immune system senescence patients with advanced non-small cell lung cancer: a self-controlled study. Med Gas Res. 2020;10(4):149.
44. MacCallum DE, Melville J, Frame S, Watt K, Anderson S, Gianella-Borradori A, et al. Seliciclib (CYC202, R-Roscovitine) induces cell death in multiple myeloma cells by inhibition of RNA polymerase II–dependent transcription and down-regulation of Mcl-1. Can Res. 2005;65(12):5399–407.
45. Koparal AT, Zeytinoglu M. Effects of carvacrol on a human non-small cell lung cancer (NSCLC) cell line, A549. Cytotechnology. 2003;43(1):149–54.
46. Tijerina AJ. The biochemical basis of metabolism in cancer cachexia. Dimens Crit Care Nurs. 2004;23(6):237–43.
47. Aiello NM, Stanger BZ. Echoes of the embryo: using the developmental biology toolkit to study cancer. Disease Models Mech. 2016;9(2):105–14.
48. Reitter EM, Kaider A, Ay C, Quehenberger P, Marosi C, Zielinski C, et al. Longitudinal analysis of hemostasis biomarkers in cancer patients during antitumor treatment. J Thromb Haemost. 2016;14(2):294–305.
49. Yu G, Wang L-G, Han Y, He Q-Y. clusterProfiler: an R package for comparing biological themes among gene clusters. Omics J Integr Biol. 2012;16(5):284–7.
50. Zhang X-Y, Guo H, Huang Y, Hao P-Q, Yang Y, Liu Y, et al. Comparative interactome analysis reveals distinct and overlapping properties of Raf family kinases. Biochem Biophys Res Commun. 2019;514(4):1217–23.
51. Zhu J, Huang R, Yang R, Xiao Y, Yan J, Zheng C, et al. Licorice extract inhibits growth of non-small cell lung cancer by down-regulating CDK4-Cyclin D1 complex and increasing CD8+ T cell infiltration. Cancer Cell Int. 2021;21(1):1–18.
52. Cui Y, Li Z, Cao J, Lane J, Birkin E, Dong X, et al. The G4 resolvase DHX36 possesses a prognosis significance and exerts tumour suppressing function through multiple causal regulations in non-small cell lung cancer. Front Oncol. 2021;11:655757.

53. Baghdadi M, Wada H, Nakanishi S, Abe H, Han N, Putra WE, et al. Chemotherapy-induced IL34 enhances immunosuppression by tumor-associated macrophages and mediates survival of chemoresistant lung cancer cells. Can Res. 2016;76(20):6030–42.
54. Chen H, Chen X, Shen Y, Yin X, Liu F, Liu L, et al. Signaling pathway perturbation analysis for assessment of biological impact of cigarette smoke on lung cells. Sci Rep. 2021;11(1):1–15.
55. Xie ZC, Tang RX, Gao X, Xie QN, Lin JY, Chen G, et al. A meta-analysis and bioinformatics exploration of the diagnostic value and molecular mechanism of miR-193a-5p in lung cancer. Oncol Lett. 2018;16(4):4114–28.
56. Gao G, Yao Z, Shen J, Liu Y. Identification of Key miRNAs in the treatment of Dabrafenib-resistant melanoma. Biomed Res Int. 2021;2021:5524486.
57. Shahid M, Azfaralariff A, Law D, Najm AA, Sanusi SA, Lim SJ, et al. Comprehensive computational target fishing approach to identify Xanthorrhizol putative targets. Sci Rep. 2021;11(1):1–11.
58. Sarmadi VH, Ahmadloo S, Boroojerdi MH, John CM, Al-Graitte SJR, Lawal H, et al. Human mesenchymal stem cells-mediated transcriptomic regulation of leukemic cells in delivering anti-tumorigenic effects. Cell Transplant. 2020;29:0963689719885077.
59. Jin Q, Lu J, Gao R, Xu J, Pan X, Wang L. Systematically deciphering the pharmacological mechanism of fructus aurantii via network pharmacology. Evid Based Complement Altern Med. 2021;2021:6236135.
60. Chen Y, Mao B, Peng X, Zhou Y, Xia K, Guo H, et al. A comparative study of genetic profiles of key oncogenesis-related genes between primary lesions and matched lymph nodes metastasis in lung cancer. J Cancer. 2019;10(7):1642.
61. Wang Y, Wang P, Liu M, Zhang X, Si Q, Yang T, et al. Identification of tumor-associated antigens of lung cancer: SEREX combined with bioinformatics analysis. J Immunol Methods. 2021;492: 112991.
62. Boonstra J, Verkleij AJ. Regulation of enzyme activity in vivo is determined by its localization. Adv Enzyme Regul. 2004;44:61–73.
63. Krystal GW, Hines SJ, Organ CP. Autocrine growth of small cell lung cancer mediated by coexpression of c-kit and stem cell factor. Can Res. 1996;56(2):370–6.
64. Casarrubios M, Cruz-Bermúdez A, Nadal E, Insa A, Campelo MdRG, Lázaro M, et al. Pretreatment tissue TCR repertoire evenness is associated with complete pathologic response in patients with NSCLC receiving neoadjuvant chemoimmunotherapy. Clin Cancer Res. 2021;27(21):5878–90.
65. Zhuang Z, Chen Q, Huang C, Wen J, Huang H, Liu Z. A comprehensive network pharmacology-based strategy to investigate multiple mechanisms of HeChan tablet on lung cancer. Evid Based Complement Altern Med. 2020;2020:7658342.
66. Eathiraj S, Palma R, Volckova E, Hirschi M, France DS, Ashwell MA, et al. Discovery of a novel mode of protein kinase inhibition characterized by the mechanism of inhibition of human mesenchymal-epithelial transition factor (c-Met) protein autophosphorylation by ARQ 197. J Biol Chem. 2011;286(23):20666–76.

Publisher's Note