

# Identify bilayer modules via pseudo-3D clustering: applications to miRNA-gene bilayer networks

Yungang Xu<sup>1,2,\*</sup>, Maozu Guo<sup>1,\*</sup>, Xiaoyan Liu<sup>1</sup>, Chunyu Wang<sup>1</sup>, Yang Liu<sup>1</sup> and Guojun Liu<sup>1</sup>

<sup>1</sup>School of Computer Science and Technology, Harbin Institute of Technology, Harbin 150001, China and <sup>2</sup>School of Life Science and Technology, Harbin Institute of Technology, Harbin 150001, China

Received March 18, 2015; Revised June 30, 2016; Accepted July 18, 2016

## ABSTRACT

**Module identification is a frequently used approach for mining local structures with more significance in global networks. Recently, a wide variety of bilayer networks are emerging to characterize the more complex biological processes. In the light of special topological properties of bilayer networks and the accompanying challenges, there is yet no effective method aiming at bilayer module identification to probe the modular organizations from the more inspiring bilayer networks. To this end, we proposed the pseudo-3D clustering algorithm, which starts from extracting initial non-hierarchically organized modules and then iteratively deciphers the hierarchical organization of modules according to a bottom-up strategy. Specifically, a modularity function for bilayer modules was proposed to facilitate the algorithm reporting the optimal partition that gives the most accurate characterization of the bilayer network. Simulation studies demonstrated its robustness and outperformance against alternative competing methods. Specific applications to both the soybean and human miRNA-gene bilayer networks demonstrated that the pseudo-3D clustering algorithm successfully identified the overlapping, hierarchically organized and highly cohesive bilayer modules. The analyses on topology, functional and human disease enrichment and the bilayer subnetwork involved in soybean fat biosynthesis provided both the theoretical and biological evidence supporting the effectiveness and robustness of pseudo-3D clustering algorithm.**

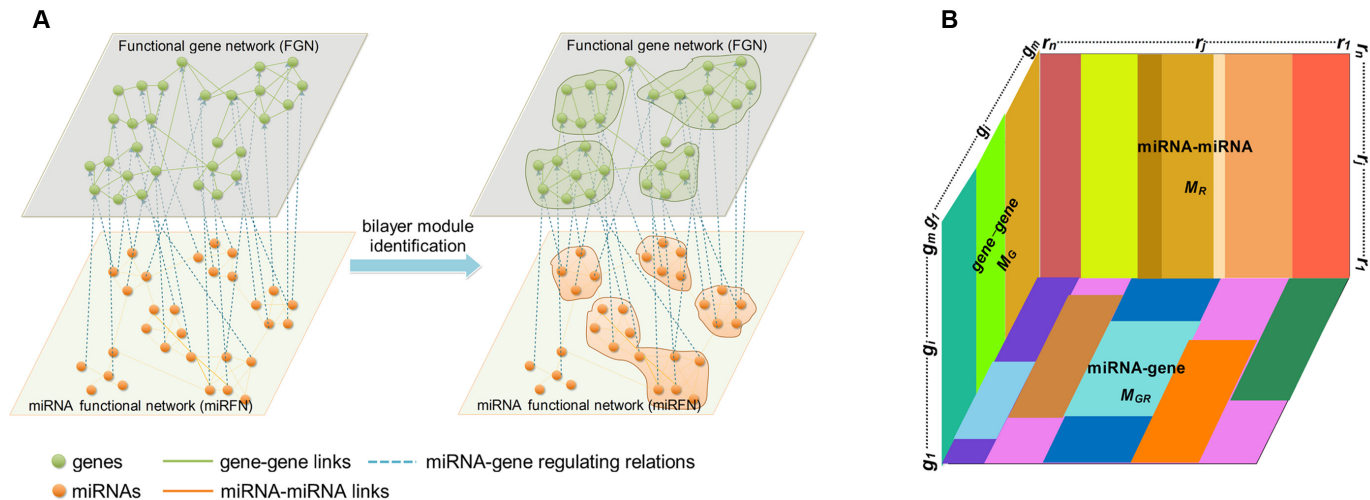
## INTRODUCTION

A biological network exhibits a modular organization. Module identification and analysis is one of the most frequently used approaches to exploring knowledge from complex biological networks. In parallel with the small world,

scale-free property and other basic local and/or global characteristics, the modular structure dependent on functional module is of great significance in understanding the organization and dynamics of network functions. Modularization is a ubiquitous phenomenon in various network systems (1). A functional module, composed of many types of interacting molecules, is a discrete local structure whose members have more internal links among themselves than external links with members of other modules (2). Normally, a module's function is separable from those of other modules (3), and many cellular functions are carried out by modules (2). Modular organization has been observed in metabolic (4), transcriptional regulation (5) and protein-protein interaction (PPI) (6) networks. Moreover, the exploration of modular structure has been proposed as a key factor in understanding the complexity of biological systems (7). In the past decade, researchers have proposed a wide variety of module identification or network decomposition methods, which can be broadly classified into six major categories: traditional clustering algorithms (8–19), network topological approaches (20–28), modularity optimization (29–32), seed expansion (33–36), matrix decomposition/factorization (37–41) and comparative network analysis (42–45). A detailed review of the above categories is beyond the scope of this paper and has already been presented by Chen *et al.* (46).

Nevertheless, there is yet no effective method for bilayer module identification from the increasingly booming bilayer networks, such as gene-disease bilayer network (47,48), gene-phenotype bilayer network (49), drug-target bilayer network (50), miRNA-disease bilayer network (51) and the miRNA-gene bilayer networks that we presented in this paper. A bilayer network often consists of three types of linkages between two kinds of nodes. Specifically, a miRNA-gene bilayer network contains two types of biomolecules (miRNAs and genes) and three types of interacting associations (miRNA-miRNA and gene-gene functional interactions, as well as miRNA-gene regulating relations) (Figure 1A). Existing module identification methods seem to have reported good results for traditional monolayer networks, but in the light of the special topological

\*To whom correspondence should be addressed. Tel: +86 451 86402407; Fax: +86 451 86221048; Email: maozuguo@hit.edu.cn  
Correspondence may also be addressed to Yungang Xu. Tel: +86 451 86402407; Fax: +86 451 86221048; Email: xuyungang@hit.edu.cn



**Figure 1.** Bilayer module identification and pseudo-3D clustering. (A) The topological view of a toy miRNA-gene bilayer network; (B) The pseudo-3D view of bilayer module identification using clustering approach.

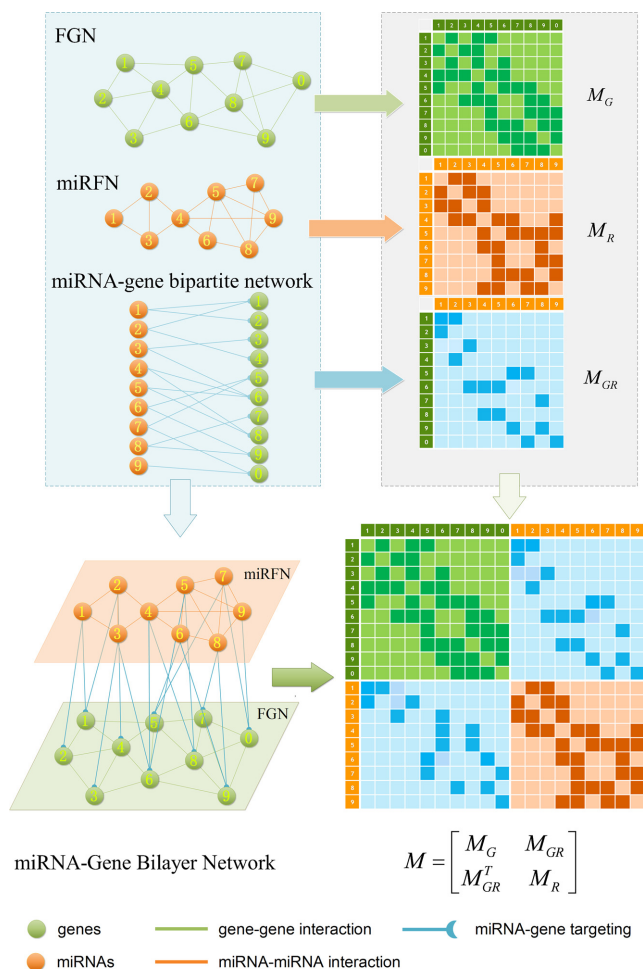
properties of bilayer networks, there are still many challenges that need to be overcome for bilayer module identification. Consequently, this work aims to present a method (or algorithm) to identify bilayer modules from a bilayer network, with specific applications to the miRNA-gene bilayer networks from both plant and human.

We constructed one bilayer network each of plant and human, respectively, by combining network topological information based on previously presented gene-gene interactions and miRNA-miRNA interactions, as well as targeting information introduced by miRNAs binding their targets (52–55). Clustering is commonly accepted as a powerful approach to partition the data items into a list of disjoint groups, such that the similarities within each group are maximized and those between different groups are minimized. For such a bilayer network, module identification means simultaneously clustering two kinds of objects and three types of links (Figure 1A). Without loss of generality, the miRNA-gene bilayer network can be mathematically expressed as three adjacency matrices:  $M_G$  for gene-gene links,  $M_R$  for miRNA-miRNA links and  $M_{GR}$  for miRNA-gene regulating relations. From the traditional clustering viewpoint, the bilayer module identification consists of two traditional one-way clustering processes ( $M_G$  and  $M_R$ ) and one two-way clustering (biclustering) process ( $M_{GR}$ ), which can be organized as the pseudo 3-dimensional (pseudo-3D) view as shown in Figure 1B. Based on this concept, we proposed a method (or algorithm) called pseudo-3D clustering to achieve bilayer module identification, because it is neither a simple combination of traditional one- and two-way clustering methods, nor a real 3D clustering (or triclustering) method used in gene expression data (56).

In addition, the research reported in this paper has its roots in works which study the identification of miRNA-gene (or miRNA-mRNA) regulatory modules. Various methods were proposed to discover miRNA-gene regulatory modules (also named co-modules). However, the miRNA-gene regulatory module identification is different

from the miRNA-gene bilayer module detection referred to herein; the former is either based only on the predicted miRNA-gene regulatory relations (57,58), or based only on the (anti-) expression correlations between miRNAs and genes (59–61), or at best incorporates the gene-gene relationships inferred from protein-protein interactions (62). However, the latter aims to respectively clustering the miRNAs and genes in two connecting layers based on three types of relations (Figure 1A). Zhang *et al.* (62) and Pio *et al.* (58) have given the detailed reviews of existing methods. Furthermore, specifically for a biological application, the bilayer module identification method has to consider a number of other important issues, of which the most critical is that the resulting bilayer modules should meet the following conditions:

1. Module overlap, since genes and miRNAs can be involved in multiple biological processes for the pleiotropy. Generally speaking, module overlaps show that nodes or links may belong to two or more modules. Several common clustering algorithms, including CFinder (12), MCL (63), MCODE (15), DetMod (11), ClusterONE (64) and MINE (65) permit overlaps between the modules. The pseudo-3D clustering algorithm will identify the overlaps either on gene layer or miRNA layer.
2. Hierarchical organization, a fundamental characteristic of many complex networks, implies that small groups of nodes organize in a hierarchical manner into increasingly large groups (66,67). A module at a higher level should contain multiple modules of lower levels. Some algorithms, like hierarchical clustering, can destruct this kind of organization by tuning a cutoff. Hierarchical modules allow studying the intermolecular interactions of different granularity. Bilayer modules should also be hierarchically organized.
3. High cohesiveness, which means the dense connections within the module, but only sparse connections between different modules. Specific to miRNA-gene bilayer modules, high cohesiveness means that the genes



**Figure 2.** The construction and mathematical representation of the miRNA-gene bilayer network.

and miRNAs in the same bilayer module should be closely linked in respective layer and show strong regulation from miRNA layer to gene layer.

Taking into account all above considerations, we propose a method, named pseudo-3D clustering algorithm that provides a solution to the issues raised up by the specific module identification task for the mushrooming bilayer biological networks. The pseudo-3D clustering algorithm is demonstrated to be able to identify the overlapping, hierarchically organized and highly cohesive bilayer modules from, but not limited to, the miRNA-gene bilayer networks and further reveal their implications to specific biological process, such as soybean fatty acid synthesis.

## MATERIALS AND METHODS

### The mathematical representation of bilayer network and necessary definitions

To be algorithm-friendly, some useful definitions are necessary. Figure 2 provides the workflow for construction and mathematical representation of a bilayer network, taking the soybean miRNA-gene bilayer network for instance.

In previous studies, we have reconstructed four functional gene networks (FGN) and four miRNA functional networks (miRFN) of soybean (*Glycine max*), respectively (52,53). Integrating the predicted regulating relations between miRNAs and their target genes, we can easily consolidate a FGN and a miRFN to form a bilayer network such as shown in Figure 1A. For simplicity, we only integrate the most inclusive FGN and miRFN (i.e. FGN-INT and miRFN-INT) without considering all other possible combinations between four FGNs and four miRFNs. The brief statistics of the final bilayer network and the component networks are shown in Supplementary Table S1.

Let  $G = (U, D)$  and  $R = (V, E)$  be the FGN and miRFN, where  $U = \{g_1, \dots, g_i, \dots, g_m\}$  and  $V = \{r_1, \dots, r_j, \dots, r_n\}$  are node sets representing  $m$  genes and  $n$  miRNAs, respectively;  $D$  and  $E$  denote edges corresponding to respective gene-gene and miRNA-miRNA interactions. The topology structures can be mathematically captured by their corresponding adjacency matrices  $M_G$  and  $M_R$ :

$$\begin{aligned} M_G^{m \times m} &= \{G_{ik} = FS(g_i, g_k) \mid i, k \in [1, m]\} \\ M_R^{n \times n} &= \{R_{jl} = FS(r_j, r_l) \mid j, l \in [1, n]\} \end{aligned} \quad (1)$$

where  $FS(a, b)$  is the functional similarity of the genes (or miRNAs)  $a$  and  $b$ .

Let  $T = (W, F)$  be the miRNA-gene bipartite network, where  $W = U \cup V$  is the union of genes in FGN and miRNAs in miRFN;  $F$  denotes the edge set that with each edge having exact one end vertex in  $U$  and the other end vertex in  $V$ , corresponding to the regulating relations between miRNAs and their target genes. Homoplasticly, the topology structure can be mathematically captured by its corresponding adjacency matrix  $M_{GR}$ :

$$M_{GR}^{m \times n} = \{T_{ij} \mid i \in [1, m], j \in [1, n]\}, \quad (2)$$

where  $T_{ij}$  is the regulatory strength value for a given miRNA-gene targeting relation ( $r_j - g_i$ ), which was defined in our previous work (52).

Therefore, we define the bilayer miRNA-gene network  $M = (W, E_G, E_R, E_T)$  by integrating the three networks  $G = (U, D)$ ,  $R = (V, E)$  and  $T = (W, F)$ , which share the same set of nodes  $W$ . The bilayer network  $M$  has three types of interactions, where  $E_G = D$  represents the gene-gene interactions within the FGN,  $E_R = E$  represents the miRNA-miRNA interactions within miRFN and  $E_T = F$  the miRNA-gene targeting information. The topology structure of this bilayer network can be mathematically captured by the block adjacency matrix  $M$ :

$$M^{N \times N} = \begin{bmatrix} M_G & M_{GR} \\ M_{GR}^T & M_R \end{bmatrix}, \quad (3)$$

where  $N = m + n$ . As an example shown in Supplementary Table S1,  $m = 5263$  and  $n = 472$  for soybean bilayer network. In the later actual use, the matrix  $M$  is also considered to consist of  $m$  row vectors for genes and  $n$  row vectors for miRNAs, namely

$$M = [\vec{g}_1, \dots, \vec{g}_i, \dots, \vec{g}_m, \vec{r}_1, \dots, \vec{r}_j, \dots, \vec{r}_n]^T, \quad (4)$$

where each gene (or miRNA) in the bilayer network is represented as a row vector in a  $(m+n)$ -dimension space, i.e.

$$\begin{aligned} \vec{g}_i &= [G_{i1}, \dots, G_{ik}, \dots, G_{im}, T_{i1}, \dots, T_{il}, \dots, T_{in}] \\ \vec{r}_i &= [T_{i1}, \dots, T_{ik}, \dots, T_{im}, R_{i1}, \dots, R_{il}, \dots, R_{in}] \end{aligned} \quad (5)$$

Let  $C$  denotes a bilayer module, which is defined in this research as a bilayer sub-network that consists of two objects (genes and miRNAs) and three inter-object relations. Subsequently, a bilayer module containing  $p$  ( $\leq m$ ) genes and  $q$  ( $\leq n$ ) miRNAs can also be represented by a sub-matrix with  $p + q$  row vectors in the  $(m+n)$ -dimension space. To be noted that, in order to facilitate the description of the pseudo-3D algorithm, hereafter a bilayer module  $C$  refers specifically to a collection of genes and miRNAs (also known as a bicluster), and does not include the relations between them. Nevertheless, it will not hamper the identification of bilayer modules defined hereinbefore, since each one of these biclusters will be easily reverted to a previously defined module with bilayer topology by decoding the relations from the matrix  $M$ . Based on this specific definition of a bilayer module  $C$ , we use  $C_g = U \cap C$  to denote the gene set of the module, and  $C_r = V \cap C$  to denote the miRNA set.

### The problem description of bilayer module identification in miRNA–gene bilayer network

As aforementioned, our aim in this work is to identify the bilayer network modules with the characteristics of overlap, hierarchy and high cohesiveness. Based on the mathematical representations and definitions, the bilayer module identification problem raised in this research can be described as follows:

**Given:**

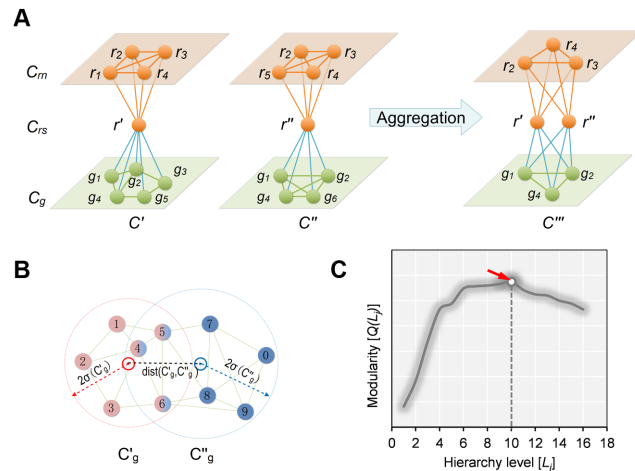
- the set of genes  $U$  and the set of miRNAs  $V$ ;
- the block adjacency matrix  $M^{N \times N}$ ;
- a cohesiveness function  $q(C, M) \rightarrow \mathbb{R}$ ,  $C$  is a bilayer module;
- a cohesiveness threshold  $\alpha$  for  $q(C, M)$ ;
- a modularity function  $Q(L_j) \rightarrow \mathbb{R}$ ,  $L_j$  is the  $j$ -th level bilayer modules.

**Find:** a list of bilayer modules  $L_j$ , for each level  $j = 1, \dots, k$  such that:

- for each list  $L_j$ ,  $j = 2, \dots, k$ ,  $\forall C' \in L_j \exists C'' \in L_{j-1}$  satisfies  $C'' \subseteq C'$  (hierarchy);
- for each list  $L_j$ ,  $j = 1, \dots, k$ ,  $\exists C', C'' \in L_j$  satisfies  $C'_g \cap C''_g \neq \emptyset$  or  $C'_r \cap C''_r \neq \emptyset$  (overlapping);
- $\forall C' \in L_j$ ,  $q(C, M) > \alpha$  (cohesiveness);
- the optimal level of bilayer modules  $L_j$  according to  $Q(L_j)$ .

### Pseudo-3D clustering algorithm

The pseudo-3D clustering starts from extracting initial non-hierarchically organized modules and then iteratively decipher the hierarchical organization of modules according to a bottom-up strategy. Based on the above descriptions, we



**Figure 3.** The schematic view of the biclique aggregation and module merging. (A) The initial bicliques and the process of biclique aggregation (miRNA-to-gene direction); (B) An example of the object distribution of two candidate gene clusters for merging. (C) A toy example of the modularity curve and the corresponding optimal hierarchy level (red arrow).

here describe the pseudo-3D clustering algorithm in high-level as Supplementary Figure S1 and following subsections. Supplementary Method M1 provides the full idea and detailed descriptions of the algorithm.

**Initial module identification.** The pseudo-3D clustering algorithm starts from extraction of a set of initial non-hierarchically organized bilayer modules, namely the lowest-level module partition ( $L_1$ ). Each initial module is a primary bicluster aggregated from bicliques (Figure 3A, left), which are subgraphs extracted from the miRNA–gene bipartite network in two directions, i.e. miRNA-to-gene and gene-to-miRNA. The algorithm starts the identification by extracting a set of initial bicliques and then iteratively aggregates two bicliques into a new one (Figure 3A). The iteration will stop when there are no candidates for aggregation. The resulting initial modules derived from both in miRNA-to-gene and gene-to-miRNA direction are combined to form  $L_1$  level modules, by simply removing modules that appear more than once and those that are a subset of others. The bidirectional biclique extraction and aggregation process not only cluster the miRNA–gene regulating relations, but also the interactions of miRNA–miRNA and gene–gene, so that it realize the simultaneous clustering of two kinds of objects (genes and miRNA) and their three types of interrelationships, namely the original idea of a pseudo-3D clustering algorithm.

**Iteration of overlap detection and module merging.** Overlap detection means to determine whether some objects (miRNAs or genes) belonging to a module appear in another module of the same level. However, not all pairs of overlapping bilayer modules will be merged subsequently. Module merging processes merge the overlapping modules when the spheres in gene (or miRNA) space of two merging candidates are close enough with their distance less than a heuristic threshold (Figure 3B). At each iteration, several pairs of modules can be merged, and an additional level of the hi-

erarchy may or may not be added depending on whether merging is performed. Obviously, a module could be the merging candidate with more than one other module. In order to obtain a result independent of the order in which pairs of modules are considered, merging is actually performed after finding out all merging candidates and guarantees to maximize the cohesiveness of the resulting modules. The iteration will stop when neither overlaps nor merges are performed in the last iteration.

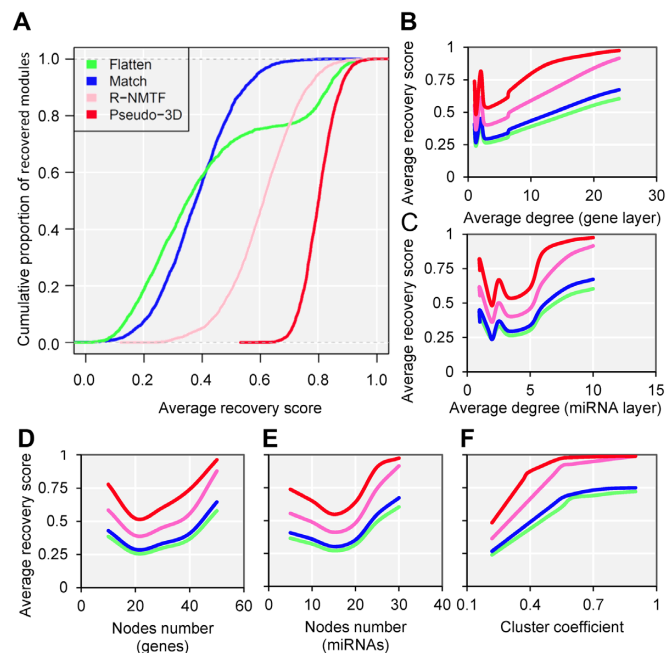
**Determination of the optimal module partition.** The iteration of overlap detection and merging will produce the partitions of a bilayer network in different levels. We need to further determine the optimal level at which the partition gives a more accurate characterization of the modular organization of the bilayer network. We here, based on the modularity function proposed by Newman (30) specifically for evaluation of a monolayer network partition, give a modularity function for the bilayer network partition. According to the modularity value, the optimal partition is specified as the level that produces the maximum modularity value across all hierarchy levels (Figure 3C).

**Time complexity.** The time complexity of the pseudo-3D algorithm depends on the time complexity of each single step. The initial module identification occupies a complexity of  $O(m^3 + n^3)$  in two directions. One iteration of overlap detection, merging and computing the  $Q(L_j)$  value occupies  $O(s^2 * (avg\_gene^2 + avg\_mirna^2))$ , where  $s$  denotes the number of modules. So, for  $k$  levels of partitions with maximum  $s$  modules at each level, the pseudo-3D algorithm will occupy a total complexity of  $O(m^3 + n^3 + k * 2 * s^2 * (avg\_gene^2 + avg\_mirna^2))$ ; it will be reduced to  $O(m^3)$  in the most cases that  $m$  (the number of genes in the bilayer network) is significantly larger than the other parameters. A detailed analysis of time complexity is provided in Supplementary Methods M1.

## RESULTS AND DISCUSSION

### Simulation and performance evaluation

To explore the advantages of pseudo-3D clustering algorithm in terms of graph theory before the specific biological applications, we performed the comparison with other network clustering strategies using a stimulated bilayer network (Supplementary Methods, M2). One well-defined bilayer miRNA–gene network with extremely modular structure was simulated (Supplementary Figure S4A,B and Table S1) and three clustering strategies were used for comparison: (i) clustering the gene network and miRNA network separately and match them together via miRNA–gene links ('Match'); (ii) Combining the three networks as one network and doing clustering ('Flatten'); (iii) clustering the bilayer network using R-NMTF (68), a method was developed to co-cluster phenotypes and genes, which are also organized as a bilayer network. The performance was evaluated by the extent, to what a clustering method can recover the maintained bilayer modules from a randomly perturbed bilayer network. Based on 1000 perturbations, the pseudo-3D clustering outperforms all three other strategies significantly ( $P$ -value  $< 2.2e-16$ , Kolmogorov–Smirnov test), fol-



**Figure 4.** Pseudo-3D clustering algorithm outperforms other methods based on simulation. (A) The cumulative proportion of recovered modules for each methods based on a simulated bilayer network. (B–F) The performance of each method regarding different node numbers, cluster densities and degree distributions. The (B and C) average degrees and (F) cluster coefficients are of the whole bilayer networks, while the (D and E) node numbers are of the individual bilayer modules.

lowed by R-NMTF (Figure 4A and Supplementary Figure S4C), and the Flatten gets the worst performance. This indicates that, for a bilayer network, the connections between two layers are important to get better module detection and that simultaneously clustering on three types of edges between two types of nodes will reinforce the clustering performance. The superior performance than another bilayer-like module detection method, R-NMTF, indicates the outstanding role of pseudo-3D clustering algorithm specifically for module identification of bilayer networks.

To investigate how the network density and other relevant topological properties affect the clustering performances and further verify the robustness of pseudo-3D clustering algorithm with respect to the node numbers, cluster densities and degree distributions, we generated 10 simulated miRNA–gene bilayer networks; each bilayer network is composed of 100 well-defined bilayer modules and possess the combination of different node numbers, cluster coefficients and degree distributions (Supplementary Method M2, Supplementary Table S2). Simulation results were provided as boxplots in Supplementary Table S2, shown as the average recovery scores of all four clustering strategies varying coordinately with the nodes number, degree distribution and cluster coefficient, with the same trend that pseudo-3D outperforms other three methods (Figure 4B–F). Specifically, the average degree of the network will facilitate the clustering when the average degree is more than 5 for gene layer and more than 3 for miRNA layer, which indicates that a higher density will produce better performance for all methods (Figure 4B and C). Nevertheless, the absolute node

number (either genes or miRNAs) does not contribute consistently to the performance of clustering (Figure 4D and E). Additionally, we found a significantly consistent correlation between the cluster coefficients and recovery scores, indicating that cluster coefficient is the most important determining factor for clustering. This is in line with the expectation that a well modularized network has a larger cluster coefficient and is easier to be partitioned into clusters (modules). Moreover, extremely high cluster coefficient reduces the difference in performance between pseudo-3D and R-NMTF rather than between these two bilayer clustering methods and the other two conventional monolayer clustering methods (Flatten and Match), indicating that considering bilayer structure will improve the clustering performance. Taken together, the simulation studies based on either one benchmark or a series of bilayer networks of the combination of different cluster densities and degree distributions, demonstrated outperformance and robustness of the pseudo-3D clustering method.

### The bilayer module identification of the soybean miRNA–gene network

We then applied pseudo-3D clustering to the soybean miRNA–gene network. In the phase of initial module identification, we obtained 472 initial bicliques for ‘miRNA-to-gene’ direction and 5263 for ‘gene-to-miRNA’ direction. After aggregation, all bicliques were consolidated into 300 and 2823 initial modules, respectively. A pruning process was performed to incorporate these two sets of initial modules and generate the first level partition of the miRNA–gene bilayer network, i.e.  $L_1$ , which contains 1497 modules. The modules at  $L_1$  have 83 genes and 6 miRNAs in average, with the average clustering coefficient (0.7363, see Supplementary Table S3) is much higher than the global bilayer network (0.696, see Supplementary Table S1).

Taking  $L_1$  as the input of the iteration for overlap detection and merging, we got another 8 upper levels of module partitions, i.e.  $L_2 \sim L_9$ . In other words, the soybean miRNA–gene bilayer network was partitioned into modules that hierarchically organized at 9 different levels. The compositions and main topological properties of the 9 levels of bilayer modules are given in Supplementary Table S3. It shows that, with the increasing of module level from  $L_1$  to  $L_9$ , the average module size increases, the number of modules decreases and the average cohesiveness diminishes, which are consistent with the theoretical characterizations of merging modules into larger ones. However, the average clustering coefficient of the modules at different levels have no significant increasing or decreasing trend (with the variance of 0.000346), indicating that all modules identified by pseudo-3D clustering algorithm are valid. In addition, the average clustering coefficients of modules at all levels (0.7507) are larger than that of the global miRNA–gene bilayer network (0.696, Supplementary Table S1), further illustrating the effectiveness of module partition at each level.

Based on the aforementioned definition, the maximum modularity corresponds to the best module partition. Therefore,  $L_6$  (in bold font, Supplementary Table S3) is the optimal level of module partitions. In the subsequent sections,  $L_6$  are used to perform deep analyses and discussions

on the module topologies, functional enrichments and biological experimental evidences, aiming at getting the insight into the regulatory characteristics between the soybean miRNAs and their target genes at module level and the further evaluation of the effectiveness of the pseudo-3D clustering algorithm.

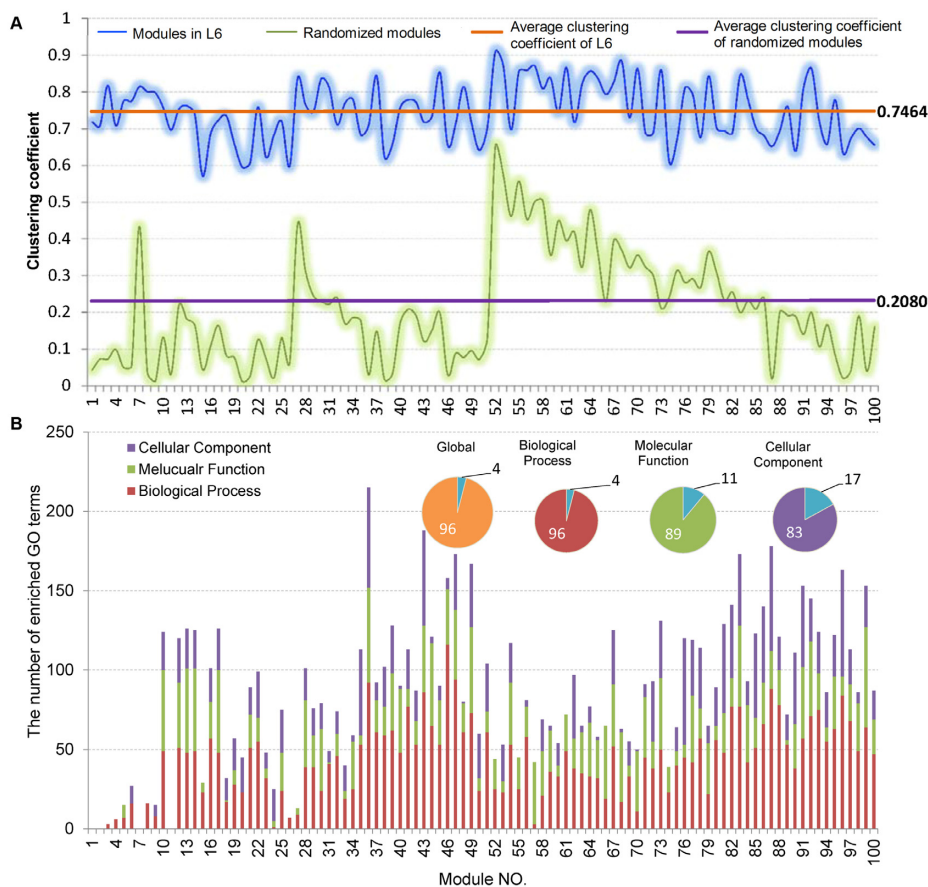
### The topological analysis of the soybean bilayer modules

At the 6th level ( $L_6$ ), the soybean miRNA–gene bilayer network is partitioned into 100 overlapping bilayer modules.  $L_6$  has an average of 333 genes and 72 miRNAs per module, with the average cohesiveness of 0.457 and the average clustering coefficient of 0.7464. Supplementary file 2 (sheet 1) provides the summary statistics of the 100 modules at  $L_6$ . In average, each miRNA targets 9.3 intra-module genes (i.e.  $avg\_gene = 9.3$ ), and each gene is targeted by 1.5 intra-module miRNAs (i.e.  $avg\_mirna = 1.5$ ).

Firstly, we analyzed the variation of the clustering coefficients of 100 modules at  $L_6$ . Shown in Figure 5A and Supplementary File 2 (sheet 1), the most of clustering coefficients vary insignificantly from 0.7 to 0.8 with the average of 0.7464 and the variance of 0.006. To compare, we built a set of 100 randomized modules maintaining as same node numbers and degree distributions for each module as those at  $L_6$  by edge perturbations. In contrast, the clustering coefficients of the 100 randomized modules vary significantly from 0.01 to 0.65 with a lower average of 0.2080 and a larger variance of 0.022. There was additional significant difference between the two sets of modules regarding the clustering coefficients ( $P$ -value =  $1.2e-81$  by ANOVA). The above results show that the nodes (genes or miRNAs) of the modules at  $L_6$  are connected more closely than those of the randomized ones, which implies the modular characterization of the genes and/or miRNAs in functionality.

Secondly, based on the conclusion of our previous research (52) that the miRNAs in the same family or in the same cluster share a larger functional similarities than randomly selected miRNAs, we investigated whether the miRNAs in the same family or in the same cluster tend to be clustered in the same module. For the 81 miRNA families, 76 families (94%) have more than half of their member miRNAs been clustered in same modules, of which 50 families (62%) are completely clustered in same modules. For the 50 families, half of them are completely clustered in more than one module, with the maximum of 15 modules. These suggest that the miRNAs in the same family tend to participate in forming the same functional module and have pleiotropy.

For the 59 miRNA clusters, 50 clusters (85%) have more than half of their members been clustered in same modules, of which 39 clusters (66%) are completely clustered in same modules. This suggests a similar conclusion to miRNA family that the miRNAs located in the same cluster tend to be a same functional module. However, we also found a significant difference between them. That is only three of 59 miRNA clusters are completely clustered in more than one module, wherein only one cluster is completely in maximum 3 modules, another two are completely in 2 modules. It suggests that the miRNAs in the same cluster do not exhibit characterization of pleiotropy as the miRNA in the same family. We speculated on the reason for the difference is



**Figure 5.** Topological and functional analysis on the optimally partitioned bilayer modules. (A) The clustering coefficient distribution of the 100 modules at  $L_6$ . Horizontal lines indicate the average cluster coefficients. (B) The statistics for GO functional enrichment of the 100 modules at  $L_6$ . Inset pies show the percentage of enriched (color filled) and unenriched (blue filled) modules.

that miRNA regulates its target genes by base-paired binding the UTR (untranslated region) of the mRNA, so that the pleiotropy of a set of miRNAs depends more on whether they share the same seed sequences rather than the locations in the genome, because a miRNA family always shares a same seed sequence.

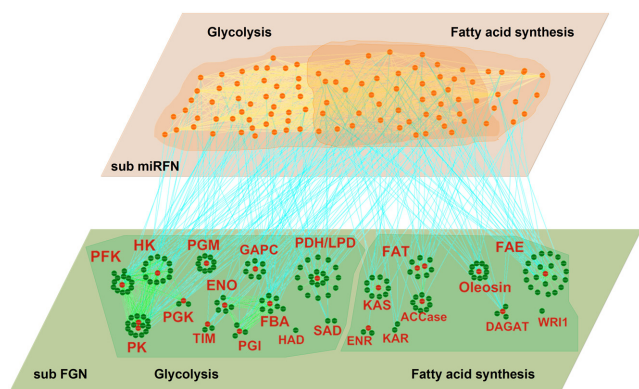
### Functional enrichment analysis and a case study in soybean fatty acid synthesis

Enrichment analysis is a computational method that widely used to determine whether an *a priori* defined set of genes shows statistically significant, concordant differences between two biological states (e.g. Gene Ontology) (69). To estimate whether the genes are significantly enriched in a module extracted by pseudo-3D clustering algorithm, we performed functional enrichment analysis on 100 bilayer modules at  $L_6$  based on Gene ontology (GO) annotation. Except four modules (i.e. #1, #2, #7 and #11), the rest 96 modules are significantly enriched in three aspects of GO, i.e. biological process (BP), molecular function (MF) and cellular component (CC) (see Figure 5B). In details, 96 modules are enriched in BP, 89 modules in MF and 83 modules in CC, covering an average of 91% genes of each module (#34 module has the maximum coverage of 99.435%). The online database SoyFN (70) allows the re-

trieval of detailed enrichment results for all modules (<http://nclab.hit.edu.cn/SoyFN/SoyModule.php>).

Soybean is the largest oilseed crop produced and consumed worldwide, accounting for 56% of the world oilseed production (SoyStats 2014, <http://soystats.com>). In this subsection, therefore, we take soybean oil (fat) synthesis as an example to investigate whether and/or how the bilayer modules produced by pseudo-3D clustering algorithm can characterize a real biological process. In order to get the genes and miRNAs involved in fat synthesis, we searched in four public databases and consulted seven literatures (Supplementary Methods, M3). As a result, we got 177 genes that encode 24 key enzymes and 102 miRNA types (see Supplementary Methods M3 for 'miRNA type' definition), which are also represented as Supplementary Figure S5 according to the three stages of fat synthesis.

We firstly analyzed the distribution of 177 genes encoding 23 key enzymes (except MCMT, EC:2.3.1.39) in the 100 bilayer modules at  $L_6$ . The results are provided as a matching matrix shown in Supplementary Figure S6 and show that the enzyme encoding genes of glycolysis and fatty acid synthesis are significantly concentrated in the different functional modules, indicating that the genes in the modules partitioned by pseudo-3D clustering algorithm have a higher correlation in function. Exceptionally, the genes that en-



**Figure 6.** A graphic view of the miRNA–gene bilayer subnetwork for soybean fat synthesis. All abbreviations of the enzyme names are given in Supplementary Figure S5.

code the enzymes involved in fatty acid modification and TAG assembly (the third stage of the fat synthesis) are not matched to any modules. We think the possible reasons are: (i) compared with such basic metabolic processes as glycolysis, the research of this process was not thorough enough, accumulating less functional annotation data of the related genes; (ii) there are relatively less enzyme encoding genes involved in this process. We also found that different modules can gather the same enzyme encoding gene (set), which is the evidence that the functional modules identified by pseudo-3D clustering algorithm are overlapping.

Secondly, using the 177 genes and 102 miRNA types, we extracted a subnetwork regarding fat synthesis from the global soybean miRNA–gene bilayer network. As shown in Figure 6, the subnetwork contains 380 edges among 175 genes (gene–gene), 553 edges among 102 miRNA types (miRNA–miRNA) and 242 edges between 102 miRNAs and 68 their target genes (miRNA–gene). Supplementary file 3 provides the detailed information of the sub network, which can be recreated by imported into Cytoscape 3.3.0 (71). Supplementary Table S4 provides the detailed statistics of the global topological properties of this subnetwork.

As apparent from Supplementary Table S4, in the miRNA–gene bilayer subnetwork for soybean fat synthesis, there are relatively sparse connections among genes with smaller clustering coefficient (0.333) and network density (0.025), but dense connections among miRNAs with larger clustering coefficient (0.528) and density (0.107). This indicates that the genes, as the enzyme encoding units, are relatively more homogeneous in function, while the miRNAs, as the post-transcriptional regulators, exhibit multifunctional and synergistic characteristics. Figure 6 shows that genes encoding the same enzyme (in red font) themselves constitute a spokewise module with one or two genes as the hub (red nodes). The hub genes play more critical roles in encoding corresponding enzymes and should be the preferential candidates for biological verification. Also the spokewise topology results in the small clustering coefficient and density. The spokewise modules are not completely independent. The modules encoding the same type of enzymes, such as the modules PFK, HK, PGK and PK,

are connected to form a complete module network (the left of lower layer, Figure 6).

### Application to human miRNA–gene bilayer network reveals the consistent robustness of pseudo-3D clustering

We further employed a more informative human bilayer network to better illustrate the usefulness and biological implications of pseudo-3D clustering algorithm. The human bilayer network was constructed based on the public available data sources in the context of disease (54,55,72) (Supplementary Methods M4) that contains much more biological information than that from an unpopular plant, soybean. Pseudo-3D clustering algorithm clustered this human bilayer network at 14 different hierarchies, with the 14th level ( $L_{14}$ ) being the optimal partition based on the modularity measure (Supplementary Table S5). At  $L_{14}$ , the human miRNA–gene bilayer network is partitioned into 37 overlapping bilayer modules.  $L_{14}$  has an average of 696 genes and 158 miRNAs per module, with the average cohesiveness of 0.4404 and the average clustering coefficient of 0.4343. Supplementary File 2 (sheet 2) provides the summary statistics of the 37 modules at  $L_{14}$ . In average, each miRNA targets 8.9 intra-module genes (i.e.  $avg\_gene = 8.9$ ), and each gene is targeted by 2.0 intra-module miRNAs (i.e.  $avg\_mirna = 2.0$ ).

To further illustrate how the pseudo-3D clustering algorithm groups the human bilayer network consistently with the functional categories, we first performed GO enrichment analysis (biological process) on the gene set of each module, separately, to investigate the enriched functions. The top ten enriched GO terms were merged to compare the functional similarity of pair-wise modules (Supplementary Methods M4 and Supplementary File 4). Then the hierarchical clustering was performed based on functional enrichment and pseudo-3D partitions to compare their consistency. We found that all modules were completely clustered into the same three clusters based on these two types of independent information (Supplementary Figure S7). A quantitative score, named consensus score, was defined to indicate this consistency (Supplementary Methods M4). The results shows that pseudo-3D clustering algorithm can produce a bilayer network partition with a significantly high consistency with the functional enrichment (consensus score = 1.0,  $P$ -value <  $2.2e-16$ ).

Additionally, enrichment analysis of human diseases on all 37 modules shows additional evidence indicating the robustness of pseudo-3D algorithm (Supplementary Methods M4). We found that only a small part of 37 modules significantly enriched in a sub group of diseases, therein most of these disease-related modules are extremely enriched in one or a limited diseases, such as module 31 (m31) is significantly related to breast cancer (Supplementary Figure S8A). Hierarchical clustering based on disease enrichment also shows a significantly high consistency with the pseudo-3D partitioned hierarchies (consensus score = 1.0,  $P$ -value <  $2.2e-16$ ) (Supplementary Figure S8B).

To sum up, the pseudo-3D clustering algorithm is demonstrated to successfully cluster the bilayer networks from both the plant and human into hierarchically organized and overlapping bilayer modules with high topological cohe-



siveness. Furthermore, functional enrichment analyses, as well as the biological evidence derived from database retrieval and literature collection, proved its excellent performance in identifying bilayer modules with functional consistency from a bilayer network. Additionally, the pseudo-3D clustering algorithm has been implemented as a runnable JAR file for public download and application to other types of bilayer networks. The manual and JAR file can be accessed at <http://nclab.hit.edu.cn/SoyFN/SoyModule.php>.

## SUPPLEMENTARY DATA

Supplementary Data are available at NAR Online.

## ACKNOWLEDGEMENTS

The authors thank the editors and reviewers for their valuable comments and suggestions.

## FUNDING

National Natural Science Foundation of China [61571163, 61532014, 61402132, 91335112 and 61271346]. Funding for open access charge: National Natural Science Foundation of China [61571163].

*Conflict of interest statement.* None declared.

## REFERENCES

- Lorenz, D.M., Jeng, A. and Deem, M.W. (2011) The emergence of modularity in biological systems. *Phys. Life Rev.*, **8**, 129–160.
- Hartwell, L.H., Hopfield, J.J., Leibler, S. and Murray, A.W. (1999) From molecular to modular cell biology. *Nature*, **402**, C47–C52.
- Tornow, S. and Mewes, H.W. (2003) Functional modules by relating protein interaction networks and gene expression. *Nucleic Acids Res.*, **31**, 6283–6289.
- Ravasz, E., Somera, A.L., Mongru, D.A., Oltvai, Z.N. and Barabasi, A.L. (2002) Hierarchical organization of modularity in metabolic networks. *Science*, **297**, 1551–1555.
- Ihmels, J., Friedlander, G., Bergmann, S., Sarig, O., Ziv, Y. and Barkai, N. (2002) Revealing modular organization in the yeast transcriptional network. *Nat. Genet.*, **31**, 370–377.
- Han, J.D.J., Bertin, N., Hao, T., Goldberg, D.S., Berriz, G.F., Zhang, L.V., Dupuy, D., Walhout, A.J.M., Cusick, M.E., Roth, F.P. et al. (2004) Evidence for dynamically organized modularity in the yeast protein-protein interaction network. *Nature*, **430**, 88–93.
- Csermely, P., Korsmaros, T., Kiss, H.J.M., London, G. and Nussinov, R. (2013) Structure and dynamics of molecular networks: A novel paradigm of drug discovery A comprehensive review. *Pharmacol. Ther.*, **138**, 333–408.
- Zhang, Y.J., Xuan, J.H., de los Reyes, B.G., Clarke, R. and Ressom, H.W. (2009) Reverse engineering module networks by PSO-RNN hybrid modeling. *BMC Genomics*, **10**(Suppl. 1), S15.
- Segal, E., Shapira, M., Regev, A., Pe'er, D., Botstein, D., Koller, D. and Friedman, N. (2003) Module networks: identifying regulatory modules and their condition-specific regulators from gene expression data. *Nat. Genet.*, **34**, 166–176.
- Jaimovich, A., Rinott, R., Schuldiner, M., Margalit, H. and Friedman, N. (2010) Modularity and directionality in genetic interaction maps. *Bioinformatics*, **26**, i228–i236.
- Maraziotis, I.A., Dimitrakopoulou, K. and Bezerianos, A. (2008) An in silico method for detecting overlapping functional modules from composite biological networks. *BMC Syst. Biol.*, **2**, 93.
- Palla, G., Derenyi, I., Farkas, I. and Vicsek, T. (2005) Uncovering the overlapping community structure of complex networks in nature and society. *Nature*, **435**, 814–818.
- Guruharsha, K.G., Rual, J.F., Zhai, B., Mintseris, J., Vaidya, P., Vaidya, N., Beekman, C., Wong, C., Rhee, D.Y., Cenaj, O. et al. (2011) A protein complex network of drosophila melanogaster. *Cell*, **147**, 690–703.
- Brohee, S. and van Helden, J. (2006) Evaluation of clustering algorithms for protein-protein interaction networks. *BMC Bioinformatics*, **7**, 488.
- Bader, G.D. and Hogue, C.W. (2003) An automated method for finding molecular complexes in large protein interaction networks. *BMC Bioinformatics*, **4**, 2.
- Pu, S.Y., Ronen, K., Vlasblom, J., Greenblatt, J. and Wodak, S.J. (2008) Local coherence in genetic interaction patterns reveals prevalent functional versatility. *Bioinformatics*, **24**, 2376–2383.
- Langfelder, P. and Horvath, S. (2008) WGCNA: an R package for weighted correlation network analysis. *BMC Bioinformatics*, **9**, 559.
- Segal, E., Friedman, N., Koller, D. and Regev, A. (2004) A module map showing conditional activity of expression modules in cancer. *Nat. Genet.*, **36**, 1090–1098.
- Jeon, J., Jeong, J.H., Baek, J.H., Koo, H.J., Park, W.H., Yang, J.S., Yu, M.H., Kim, S. and Pak, Y.K. (2011) Network clustering revealed the systemic alterations of mitochondrial protein expression. *PLoS Comput. Biol.*, **7**, e1002093.
- Georgii, E., Dietmann, S., Uno, T., Pagel, P. and Tsuda, K. (2009) Enumeration of condition-dependent dense modules in protein interaction networks. *Bioinformatics*, **25**, 933–940.
- Zhao, S.W. and Li, S. (2012) A co-module approach for elucidating drug-disease associations and revealing their molecular basis. *Bioinformatics*, **28**, 955–961.
- Kovacs, I.A., Palotai, R., Szalay, M.S. and Csermely, P. (2010) Community Landscapes: An integrative approach to determine overlapping network module hierarchy, identify key nodes and predict network dynamics. *Plos One*, **5**, e12528.
- Freyre-Gonzalez, J.A., Alonso-Pavon, J.A., Trevino-Quintanilla, L.G. and Collado-Vides, J. (2008) Functional architecture of *Escherichia coli*: new insights provided by a natural decomposition approach. *Genome Biol.*, **9**, R154.
- Saez-Rodriguez, J., Gayer, S., Ginkel, M. and Gilles, E.D. (2008) Automatic decomposition of kinetic models of signaling networks minimizing the retroactivity among modules. *Bioinformatics*, **24**, 1213–1219.
- Ma, H.W., Zhao, X.M., Yuan, Y.J. and Zeng, A.P. (2004) Decomposition of metabolic network into functional modules based on the global connectivity structure of reaction graph. *Bioinformatics*, **20**, 1870–1876.
- Yoon, J., Si, Y.G., Nolan, R. and Lee, K. (2007) Modular decomposition of metabolic reaction networks based on flux analysis and pathway projection. *Bioinformatics*, **23**, 2433–2440.
- Chen, J.C. and Yuan, B. (2006) Detecting functional modules in the yeast protein-protein interaction network. *Bioinformatics*, **22**, 2283–2290.
- Luo, F., Yang, Y.F., Chen, C.F., Chang, R., Zhou, J.Z. and Scheuermann, R.H. (2007) Modular organization of protein interaction networks. *Bioinformatics*, **23**, 207–214.
- Ziv, E., Middendorf, M. and Wiggins, C.H. (2005) Information-theoretic approach to network modularity. *Phys. Rev. E Stat. Nonlin. Soft Matter Phys.*, **71**, 046117.
- Newman, M.E.J. (2006) Modularity and community structure in networks. *Proc. Natl. Acad. Sci. U.S.A.*, **103**, 8577–8582.
- Nacher, J.C. and Schwartz, J.M. (2012) Modularity in protein complex and drug interactions reveals new polypharmacological properties. *Plos One*, **7**, e30028.
- Chen, Z.J., He, Y., Rosa, P., Germann, J. and Evans, A.C. (2008) Revealing modular architecture of human brain structural networks by using cortical thickness from MRI. *Cereb. Cortex*, **18**, 2374–2381.
- Jia, P.L., Zheng, S.Y., Long, J.R., Zheng, W. and Zhao, Z.M. (2011) dmGWAS: dense module searching for genome-wide association studies in protein-protein interaction networks. *Bioinformatics*, **27**, 95–102.
- Ulitsky, I. and Shamir, R. (2007) Identification of functional modules using network topology and high-throughput data. *BMC Syst. Biol.*, **1**, 8.
- Ulitsky, I. and Shamir, R. (2009) Identifying functional modules using expression profiles and confidence-scored protein interactions. *Bioinformatics*, **25**, 1158–1164.

36. Hou, L., Wang, L., Qian, M.P., Li, D., Tang, C., Zhu, Y.P., Deng, M.H. and Li, F.T. (2011) Modular analysis of the probabilistic genetic interaction network. *Bioinformatics*, **27**, 853–859.
37. Ciriello, G., Cerami, E., Sander, C. and Schultz, N. (2012) Mutual exclusivity analysis identifies oncogenic network modules. *Genome Res.*, **22**, 398–406.
38. Yang, X.R., Zhou, Y., Jin, R. and Chan, C. (2009) Reconstruct modular phenotype-specific gene networks by knowledge-driven matrix factorization. *Bioinformatics*, **25**, 2236–2243.
39. Li, H., Sun, Y. and Zhan, M. (2007) The discovery of transcriptional modules by a two-stage matrix decomposition approach. *Bioinformatics*, **23**, 473–479.
40. Zhang, W.S., Edwards, A., Fan, W., Zhu, D.X. and Zhang, K. (2010) svdPPCS: an effective singular value decomposition-based method for conserved and divergent co-expression gene module identification. *BMC Bioinformatics*, **11**, 338.
41. Zou, Q., Li, J., Hong, Q., Lin, Z., Shi, H., Wu, Y. and Ju, Y. (2015) Prediction of microRNA-disease associations based on social network analysis methods. *BioMed Res. Int.*, **2015**, 810514.
42. Kelley, R. and Ideker, T. (2005) Systematic interpretation of genetic interactions using protein networks. *Nat. Biotechnol.*, **23**, 561–566.
43. Pinter, R. Y., Rokhlenko, O., Yeager-Lotem, E. and Ziv-Ukelson, M. (2005) Alignment of metabolic pathways. *Bioinformatics*, **21**, 3401–3408.
44. Sharan, R., Suthram, S., Kelley, R.M., Kuhn, T., McCuine, S., Uetz, P., Sittler, T., Karp, R.M. and Ideker, T. (2005) Conserved patterns of protein interaction in multiple species. *Proc. Natl. Acad. Sci. U.S.A.*, **102**, 1974–1979.
45. Ali, W. and Deane, C.M. (2009) Functionally guided alignment of protein interaction networks for module detection. *Bioinformatics*, **25**, 3166–3173.
46. Chen, Y.Y., Wang, Z. and Wang, Y.Y. (2014) Spatiotemporal positioning of multipotent modules in diverse biological networks. *Cell Mol. Life Sci.*, **71**, 2605–2624.
47. Bauer-Mehren, A., Bundschuh, M., Rautschka, M., Mayer, M.A., Sanz, F. and Furlong, L.I. (2011) Gene-Disease Network Analysis reveals functional modules in mendelian, complex and environmental diseases. *Plos One*, **6**, e20284.
48. Zeng, X., Liao, Y., Liu, Y. and Zou, Q. (2016) Prediction and validation of disease genes using HeteSim Scores. *IEEE/ACM Trans. Comput. Biol. Bioinform.*, **99**, 1.
49. Dhami, P., Saffrey, P., Bruce, A.W., Dillon, S.C., Chiang, K., Bonhoure, N., Koch, C.M., Bye, J., James, K., Foad, N.S. *et al.* (2010) Complex exon-intron marking by histone modifications is not determined solely by nucleosome distribution. *Plos One*, **5**, e12339.
50. Yildirim, M.A., Goh, K.I., Cusick, M.E., Barabasi, A.L. and Vidal, M. (2007) Drug-target network. *Nat. Biotechnol.*, **25**, 1119–1126.
51. Zou, Q., Li, J., Song, L., Zeng, X. and Wang, G. (2016) Similarity computation strategies in the microRNA-disease network: A Survey. *Brief. Funct. Genomic.*, **15**, 55–64.
52. Xu, Y., Guo, M., Liu, X., Wang, C. and Liu, Y. (2013) Inferring the soybean (*Glycine max*) microRNA functional network based on target gene network. *Bioinformatics*, **30**, 94–103.
53. Xu, Y.G., Guo, M.Z., Zou, Q., Liu, X.Y., Wang, C.Y. and Liu, Y. (2014) System-level insights into the cellular interactome of a non-model organism: inferring, modelling and analysing functional gene network of soybean (*Glycine max*). *Plos One*, **9**, e113907.
54. Zaman, N., Li, L., Jaramillo, M.L., Sun, Z., Tibiche, C., Banville, M., Collins, C., Trifiro, M., Paliouras, M., Nantel, A. *et al.* (2013) Signaling network assessment of mutations and copy number variations predict breast cancer subtype-specific drug targets. *Cell Rep.*, **5**, 216–223.
55. Bhajun, R., Guyon, L., Pitaval, A., Sulpice, E., Combe, S., Obeid, P., Haguët, V., Ghorbel, I., Lajaunie, C. and Gidrol, X. (2015) A statistically inferred microRNA network identifies breast cancer target miR-940 as an actin cytoskeleton regulator. *Sci. Rep.*, **5**, 8336.
56. Mahanta, P., Ahmed, H., Bhattacharyya, D. and Kalita, J.K. (2011) *Emerging Trends and Applications in Computer Science (NCETACS), 2011 2nd National Conference on*. IEEE, pp. 1–6.
57. Yoon, S.R. and De Micheli, G. (2005) Prediction of regulatory modules comprising microRNAs and target genes. *Bioinformatics*, **21**, 93–100.
58. Pio, G., Ceci, M., D’Elia, D., Loglisci, C. and Malerba, D. (2013) A novel biclustering algorithm for the discovery of meaningful biological correlations between microRNAs and their target genes. *BMC Bioinformatics*, **14**(Suppl. 7), S8.
59. Peng, X.X., Li, Y., Walters, K.A., Rosenzweig, E.R., Lederer, S.L., Aicher, L.D., Proll, S. and Katze, M.G. (2009) Computational identification of hepatitis C virus associated microRNA-mRNA regulatory modules in human livers. *BMC Genomics*, **10**, 373.
60. Tran, D.H., Satou, K. and Ho, T.B. (2008) Finding microRNA regulatory modules in human genome using rule induction. *BMC Bioinformatics*, **9**(Suppl. 12), S5.
61. Joung, J.G., Hwang, K.B., Nam, J.W., Kim, S.J. and Zhang, B.T. (2007) Discovery of microRNA-mRNA modules via population-based probabilistic learning. *Bioinformatics*, **23**, 1141–1147.
62. Zhang, S.H., Li, Q.J., Liu, J. and Zhou, X.J. (2011) A novel computational framework for simultaneous integration of multiple types of genomic data to identify microRNA-gene regulatory modules. *Bioinformatics*, **27**, 1401–1409.
63. Enright, A.J., Van Dongen, S. and Ouzounis, C.A. (2002) An efficient algorithm for large-scale detection of protein families. *Nucleic Acids Res.*, **30**, 1575–1584.
64. Nepusz, T., Yu, H.Y. and Paccanaro, A. (2012) Detecting overlapping protein complexes in protein-protein interaction networks. *Nat. Methods*, **9**, 471–481.
65. Rhrissorakkrai, K. and Gunsalus, K.C. (2011) MINE: Module Identification in Networks. *BMC Bioinformatics*, **12**, 192.
66. Ravasz, E. and Barabasi, A.L. (2003) Hierarchical organization in complex networks. *Phys. Rev. E Stat. Nonlin. Soft Matter Phys.*, **67**, 026112.
67. Barabasi, A.L., Ravasz, E. and Oltvai, Z. (2003) Hierarchical organization of modularity in complex networks. *Lect. Notes Phys.*, **625**, 46–65.
68. Hwang, T., Atluri, G., Xie, M., Dey, S., Hong, C., Kumar, V. and Kuang, R. (2012) Co-clustering phenome-genome for phenotype classification and disease gene discovery. *Nucleic Acids Res.*, **40**, e146.
69. Subramanian, A., Tamayo, P., Mootha, V.K., Mukherjee, S., Ebert, B.L., Gillette, M.A., Paulovich, A., Pomeroy, S.L., Golub, T.R., Lander, E.S. *et al.* (2005) Gene set enrichment analysis: A knowledge-based approach for interpreting genome-wide expression profiles. *Proc. Natl. Acad. Sci. U.S.A.*, **102**, 15545–15550.
70. Xu, Y.G., Guo, M.Z., Liu, X.Y., Wang, C.Y. and Liu, Y. (2014) SoyFN: a knowledge database of soybean functional networks. *Database (Oxford)*, **2014**, bau019.
71. Smoot, M.E., Ono, K., Ruscheinski, J., Wang, P.-L. and Ideker, T. (2011) Cytoscape 2.8: new features for data integration and network visualization. *Bioinformatics*, **27**, 431–432.
72. Ru, Y., Kechris, K.J., Tabakoff, B., Hoffman, P., Radcliffe, R.A., Bowler, R., Mahaffey, S., Rossi, S., Calin, G.A., Bemis, L. *et al.* (2014) The multiMiR R package and database: integration of microRNA-target interactions along with their disease and drug associations. *Nucleic Acids Res.*, **42**, e133.