

IMCC: Quantitative Analysis of the Inter-module Connectivity for Bio-network

Pengqian Wang^{1,2}, Jun Liu², Yanan Yu², Yingying Zhang³, Bing Li^{2,4}, Dongfeng Li⁵,
Wenjuan Xu⁶, Qiong Liu², Zhong Wang^{2*}

*1 Institute of Chinese Materia Medica, China Academy of Chinese Medical Sciences, Beijing
100700, China*

*2 Institute of Basic Research in Clinical Medicine, China Academy of Chinese Medical Sciences,
Beijing 100700, China*

3 Dongzhimen Hospital, Beijing University of Chinese Medicine, Beijing 100700, China

*4 Institute of Information on Traditional Chinese Medicine, China Academy of Chinese Medical
Sciences, Beijing 100700, China*

5 School of Mathematical Sciences, Peking University, Beijing 100871, China

6 School of Life Sciences, Beijing University of Chinese Medicine, Beijing 100029, China

* Address correspondence to this author at the Institute of Basic Research in Clinical Medicine, China Academy of Chinese Medical Sciences, 18 Baixincang, Dongzhimennei, Beijing 100700, China; Fax +86-10-84046033; E-mail: zhouw@vip.sina.com

Abstract

Inter-module connectivity, which tend to connect different communities and maintain network architectural integrity, is contributing to functional coordination and information flow between modules in perturbations. Detecting the strength of inter-module connection is essential to characterize the reactive bio-systematical variation. A quantitative evaluation method for inter-module connections is needed. Here, based on the high-throughput microarray data from mouse, an evaluation approach (named as IMCC) for inter-module connectivity was developed. The IMCC model, which is an integration of direct and indirect inter-module connections, successfully excluded inter-module connections without statistical significance or below the cutoff value, and provided a more comprehensive landscape of inter-module relationships. We showed that the IMCC method reflected a more precise functional coordination between modules according to KEGG database, were validated by topological parameter. Application of IMCC in genome-scale stroke networks deciphered characteristic pathological “core-periphery” structure of modular map and functional coordination module pair.

Author summery

Inter-module connectivity, which tend to connect different communities and maintain network architectural integrity, is contributing to functional coordination and information flow between modules in perturbations. Moreover, modular rearrangements provide more efficient ways for phenotype alteration, inter-module connections have been considered to be “evolutionary interaction switches”. Such modular map rewiring can be used as a network biomarker to characterize the dynamics of drug responses. Detecting the strength of inter-module connection is essential to characterize the reactive bio-systematical variation response to disease or drug. We aim to construct a quantitative evaluation method for inter-module connections. Thus, this study has implications in systematical exploration detailed variation of inter-module pharmacological action mode of drugs.

1 Introduction

As accumulated data from high-throughput technologies delineate a holistic view of intra-cellular molecular network, the major challenge in the post-genomic era is deciphering how these entities in the cell work together to execute sophisticated functions [1-4]. The ongoing efforts have been making in decomposing a network into modules [5-9]. This “network dissection” methods may help to reveal modular structure organization in networks. However, inter-module connections, as the ‘backbone’ of cellular networks contributing to functional coordination and information flow between modules in most biological processes, are ever important [10-11]. Such connections tend to connect different communities and maintain network architectural integrity, according to Granovetter’s hypothesis [12-15]. Removal of inter-module connections may lead to collapsed network architecture as well as interrupted information propagation [13]. Moreover, modular rearrangements provide more efficient ways for phenotype alteration [16-17] than genetic variation or modular allostery [18-19], as inter-module edge is more transient and flexible compared with intra-module connections. Inter-module connections have been considered to be “evolutionary interaction switches” [4, 20-25], because functional innovations often emerge from the rewiring of conserved functional modules to adapt to environment, in the biological system, as one kind of complex adaptive system [26-28]. Such modular map rewiring can be used as a network biomarker to characterize the dynamics of drug responses [29], by identifying and evaluating the drug-conditional existence of collaborations between modules, as a result to reflect the dynamical fluctuations of molecular correlations [30-34]. Therefore, it is an extremely interesting and promising perspective to apply inter-module connectivity analysis to characterize the reactive bio-systematical variation to perturbation, especially the pharmacological mechanism of multiple-target compounds [35-39].

Recently, increasing attention has been paid to inter-module research. Several concepts, such as “bottleneck” [41], “connector” [42], “bridgeness” [42-44], and “fuzzy community” [44-48], have been proposed to investigate the node or module bridging one community to another, regulating the information flow between modules, maintaining dynamic rewiring in networks [41, 49]. The principle of inter-module correlations has been explored and suggested to constitute a form of functional coordination between modules in a pairwise fashion [10, 50]. A set of studies have also focused on the quantitative evaluation method for inter-module connections, such as counting the number of interactions or overlapping nodes [51] between modules or communities, which were commonly conducted to represent as the connections [17, 52-55]. Meanwhile, the significance of inter-module connections has also been calculated by performing hypergeometric test, Fisher’s test [10, 56], Wilcoxon rank-sum test [50], or optimization of the global score [56]. Studies on module coordination have shown that module cooperation should be established not only by direct interactions among modules but also through shared partners [57] or between-modules (or pathways) paths that consist of multiple proteins and interactions [58]. Although previous algorithms proved valuable, it was often ignored that the modules in biological network were of functions, and all of the evaluation methods for inter-module correlations should be oblige to biological functions. Therefore, it is pertinent to introduce biological function to measure reliability and validity of inter-module evaluations. Additionally, inter-module connectivity should be established not only considering direct inter-module connections bridged by shared nodes or interaction, but also through shared partners [57] or between-modular(pathway) path consisting of multiple protein and interactions [58]. To shed light on the inherent connections between modules and consequent inter-module coordination related

pathological process and pharmacological mechanism, it is pressed for developing an integrated and more precise evaluation method for quantitative analysis on inter-module connectivity.

In this paper, using genetic interactome and modules based on microarray of MCAO (middle cerebral artery obstruction) mice and WGCNA (weighted gene co-expression network analysis), we proposed a novel approach to evaluate inter-module connectivity, integrating quantitative analysis and statistical significance, named as inter-module connectivity coefficient (IMCC). Also, the biological similarity between modules from the KEGG database were employed to optimized the IMCC model. Then, the IMCC were compared with extant method on real network, based on the precision to reflect the inter-module coordination in KEGG database. And IMCC was also validated by topological parameter. Finally, we applied IMCC in genome-scale stroke networks to decipher characteristic inter-module connectivity related to pathological process.

2 Materials and methods

Inter-module connectivity contributed to information propagation and functional coordination between modules. According to local conformations, connections between two modules composed of edges between nodes from distinct modules (direct-edge) or shared nodes from different modules (direct-node) were defined as direct inter-module connections (DIMC), and interactions mediated by genes that associate with both of the two modules were classified as indirect inter-module connections (IIMC).

In order to develop a more comprehensive, objective, and accurate method to measure the module-to-module relationships and reflect the biological coordination between modules, we firstly calculated two types of correlation parameters: SW for DIMC; CT and PS for IIMC; and then we screened these parameters using hypergeometric distribution or cutoff value and integrated the identified parameters; finally, we optimized the integration weight according to KEGG database (Figure.1).

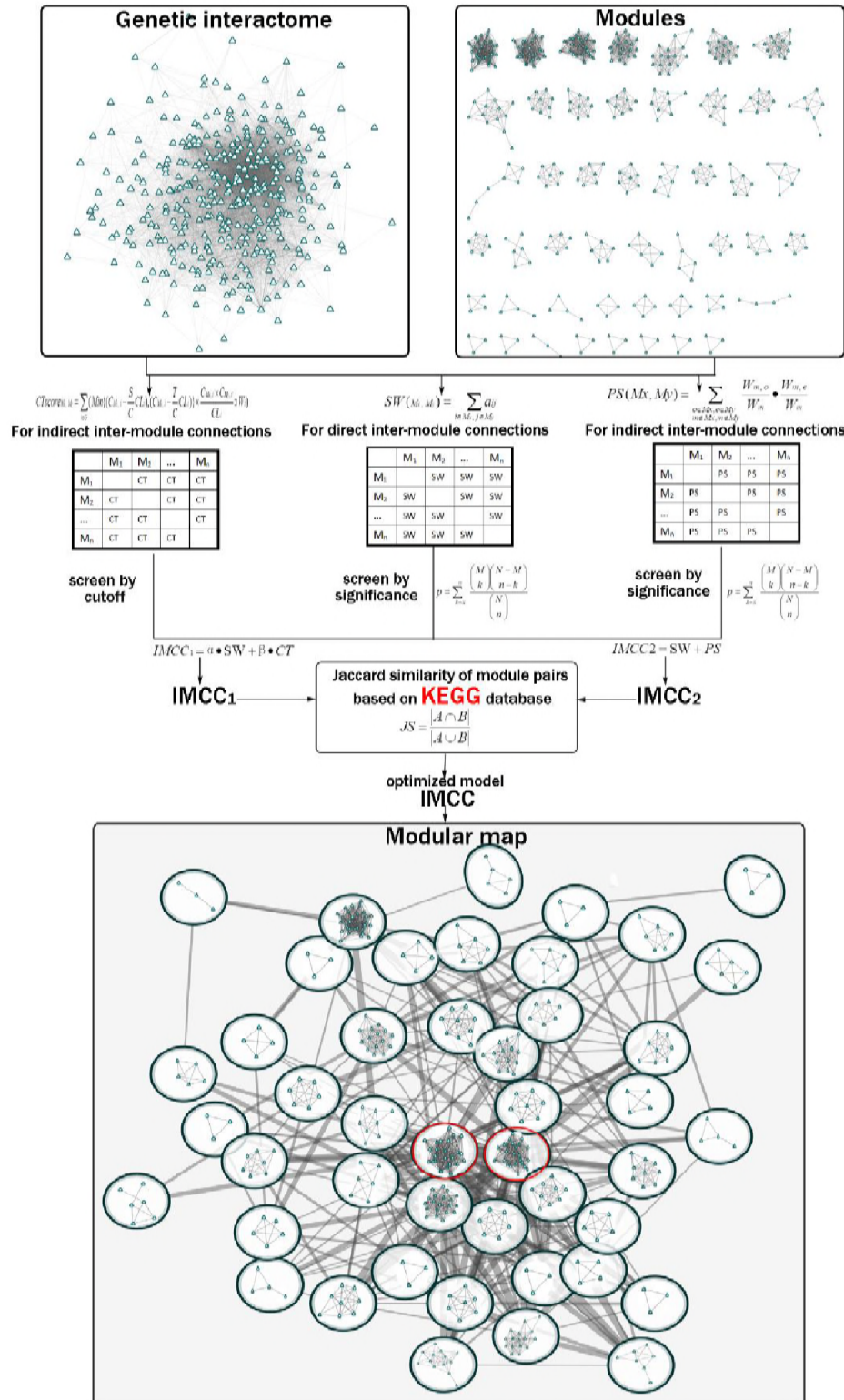


Figure 1. The procedure of computing IMCC model. Based on the genetic interactome and modules originated from microarray of MCAO mice, the parameters for direct or indirect inter-module connections were calculated, screened and integrated. Then the IMCC model were optimized according to KEGG database. The box at the bottom of the diagram was the modular map based on IMCC, in which vertices denoted modules and the thickness of edges linking pairs of modules was directly proportional to the corresponding IMCC. In this modular map, the red circled modules represented the “core” module pair (module-blue and module-brown).

2.1 Parameters calculations

For SW, we calculated the sum of weight of edges between pairs of modules:

$$SW(M_x, M_y) = \sum_{i \in M_x, j \in M_y} a_{ij} \quad (1)$$

where M_x and M_y denote any two modules connected by at least one edge, i and j are a gene in M_x and M_y , respectively, and a_{ij} is the weight of edge between gene i and j . Using this formula, we calculated the direct inter-module connections for any module pair possessing one or more edges.

To decide whether the inter-module direct connections were statistically significant, we used the P-value of the hypergeometric distribution [11], defined as

$$P = \sum_{k=x}^n \frac{\binom{M}{k} \binom{N-M}{n-k}}{\binom{N}{n}} \quad (2)$$

where x is observed inter-module connections, k and n are the numbers of inter-module connections and all possible edges between two modules, respectively; M and N represent the total numbers of inter-module connections and all combinational gene pairs between any two modules in a module to module network, respectively. In this paper, we set 0.05 as P-value threshold. Therefore, inter-module connections are supposed to be present, if P-value ≤ 0.05 ; and the SW is defined as a valid direct measurement.

For indirect inter-module connections, we introduced two parameters: path strength [58-59] and consistency score [57].

In the light of the network, paths consisted of multiple vertexes and links between them [58]. To simplify the problem, we restricted the length of paths and only considered paths that consist of three nodes (outset (o), mediation (m) and end (e)) with two links.

$$PS(M_x, M_y) = \sum_{\substack{o \in M_x, e \in M_y \\ m \in M_x, m \in M_y}} \frac{W_{m,o}}{W_m} \cdot \frac{W_{m,e}}{W_m} \quad (3)$$

The path strength (PS) of a path is defined as the product of the weighted probabilities that mediation chooses outset and end. The weighted probability from m to o is the ratio of the weight between m and o ($W_{m,o}$) to the sum of the weights between m (W_m) and its first neighbors, the same as m to e .

Hypergeometric distribution was also used to screen the statistically significant PS. However, different from SW, in formula (2), x is observed nodes connecting a pair of module, k and n are the numbers of nodes connecting a pair of modules and all possible nodes connecting the two modules, respectively; M and N represent the total numbers of nodes connecting any pair of module and all possible nodes between any two modules in a module to module network, respectively. In this paper, we set 0.05 as P-value threshold.

We also employed consistency score to measure the inter-module connectivity as described in [57]

$$CTscore_{M_x, M_y} = \sum_{i \in G} (\text{Min}\{(C_{M_x, i} - \frac{S}{C} CL_i), (C_{M_y, i} - \frac{T}{C} CL_i)\}) \times \frac{C_{M_x, i} \times C_{M_y, i}}{CL_i} \times W_i \quad (4)$$

G is a gene set that consists of all genes in network, and C is the total number of genes in G . CL_i is the total number of links to gene i ; W_i is the weight of gene i in network. S and T are the numbers of

genes in modules M_x and M_y , respectively; $CM_{x,i}$ and $CM_{y,i}$ are the observed numbers of links connecting gene i and modules M_x and M_y , respectively. Formula (4) was used to compare the weights of genes correlated with a pair of modules with the weights of genes related to only one of the modules [57]. As the CT is a value after comparison with theoretical value, we set cutoff value (10) to screen out the valid CT.

We listed the parameters and screening procedures in Supplementary Table 1 and 2.

2.2 Measurement integration

To obtain a more accurate and objective relationship between modules, we merged DIMC and IIMC. Firstly, we set SW of the inter-module connections, whose P-value of hypergeometric distribution was equal to or less than 0.05, as weight of DIMC. In the process of $IMCC_1$ integration, SW and CT were two parameters of different dimensions, so it was adopted to correct the two parameters to the value of 0-1. We normalized the two measurements to be numbers between [0-1] by the follow formula:

$$f'_x = \frac{f_x - f_{\min}}{f_{\max} - f_{\min}} \quad (5)$$

The two parameters were weighted using weighted coefficient α and β for SW and CT, respectively.

And the two weighted parameters were summed up, as followed:

$$IMCC_1 = \alpha \bullet SW + \beta \bullet CT \quad (6)$$

We set $\alpha + \beta = 1$, and coefficient ratio $\rho = \alpha/\beta$. By adjusting ρ value, the effect of SW and CT on $IMCC_1$ would be altered. We calculated the $IMCC_1$ when $\rho = 1/10, 1/8, 1/4, 1/2, 1/1, 2/1, 4/1, 8/1, \text{ and } 10/1$, respectively.

In the integration of SW and PS, both of them belong to the same dimension, so we plused the two parameters without weighting, defined as:

$$IMCC_2 = SW + PS \quad (7)$$

The number of screened SW, CT, and PS were shown in figure 2A and 2B.

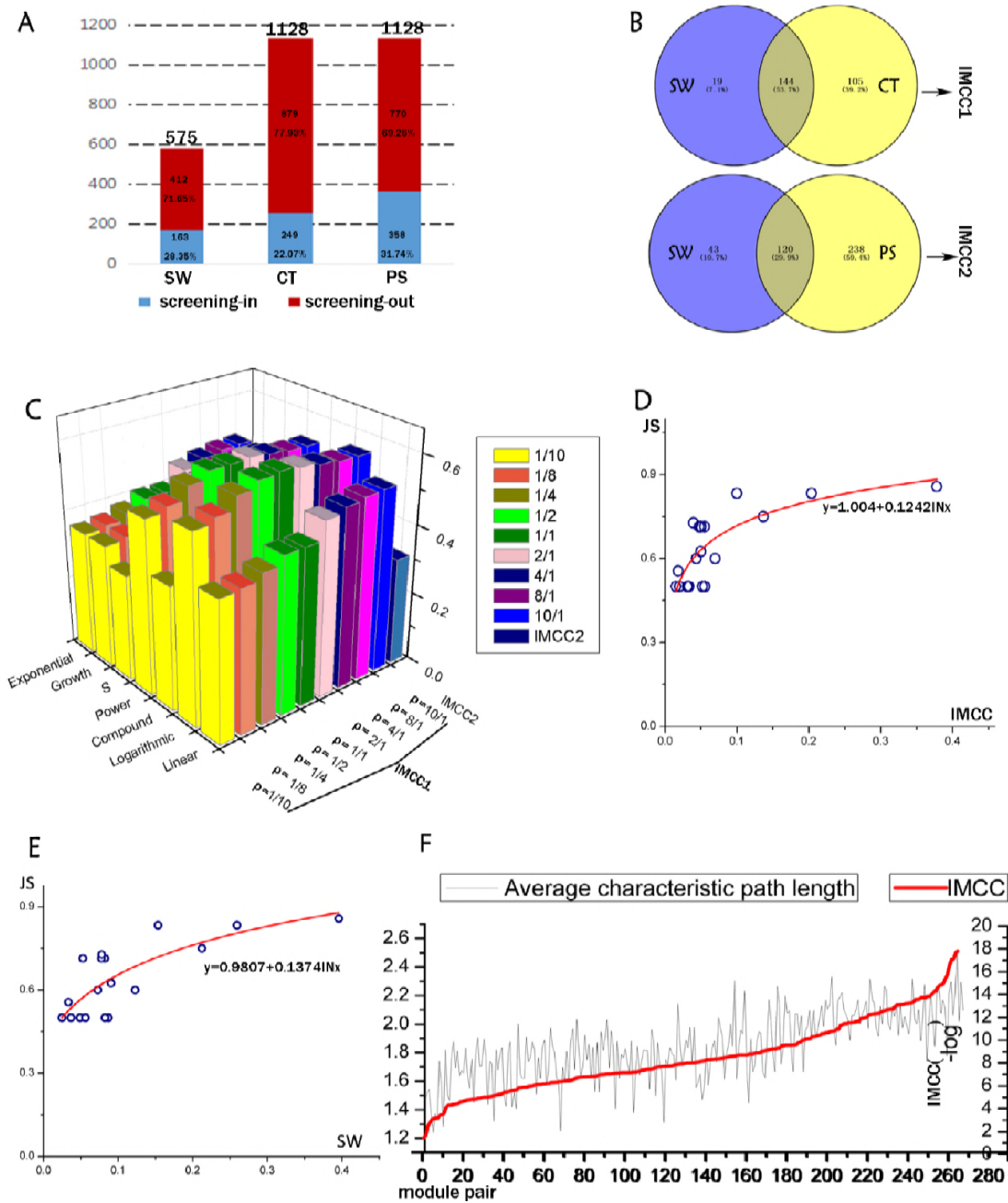


Figure 2. Quantitative evaluation of the IMCC method.

A The screening outcomes of the 3 inter-module connectivity parameters (SW, CT and PS). The height of the column represents the total amount of SW, CT and PS, respectively. Red and blue parts of the columns represent the screened out and the remaining parameters, respectively. B The overlapping condition of two parameters to be integrated as IMCC1 and IMCC2. C The R^2 of 7 fitting models for IMCC1 with different ρ and IMCC2. D The fitting curves of logarithmic model of JS and IMCC. E The fitting curves of logarithmic model of JS and SW. F The IMCC value against the average characteristic path length of module pairs.

2.3 Data sources

We used genetic interactomes and modules constructed based on the microarray datasets of MCAO mice and WGCNA, to illustrate the performance of IMCC method on the inter-module connection calculation.

2.4 Data analysis and weighting coefficient optimization

Considering the different characteristics in various types of networks, whether the emphasis of IMCC should be placed on DIMC or IIMC must be consistent with practical applications. In biological networks, the communication between certain modules is commonly mediated by component with important functions; for example, a gene might be a target regulated by two modules competitively. All the inter-module correlations summarized or predicted are presumed to contribute to biological functions. Therefore, it is imperative to introduce biological data to define the best weighting coefficient, in order to select the optimal IMCC. As a result, we employed the KEGG (Kyoto Encyclopedia of Genes and Genomes) database, according to which we calculated the jaccard coefficient of enriched pathways of each module pair.

$$JS = \frac{|A \cap B|}{|A \cup B|} \quad (8)$$

A and B is the category of enriched KEGG terms in module M_x and M_y, respectively. Therefore, A ∩ B represents the number of identical categories of KEGG terms between A and B; and A ∪ B represents the number of all the categories of KEGG terms in both A and B. For example, if A ∩ B = 6, and A ∪ B = 11, then JS will be 0.54545. We presumed that modules enriched with the same KEGG categories might form more dense connections than those with different KEGG categories. We benchmarked the IMCC₁ of different ρ values based on JS (Figure 2D, Supplementary Table 3), and the IMCC₁ scores were plotted versus the observed JS for each module pair. To obtain more precise results, we removed the outliers (Supplementary Table 3). Through linear and 6 nonlinear curve estimating, we compared the coefficient of determination (R²), and quantitatively identified the best ρ value and the most fitting model (Figure 2C). Our results suggested that the optimal ρ value was 1/1 and the most fitting model was logarithmic model with a R² of 0.616. Thus the final formula for IMCC could be simplified as

$$IMCC = SW + CT \quad (9)$$

We also plotted the IMCC₂ against JS, and compared R² of fitted curves of IMCC₁ and IMCC₂ to select the optimal integrative method.

2.5 Comparison and verification of inter-module average shortest path

For previous algorithms for inter-module were mainly based on the sum of weight of interactions (SW) between pairs of modules, we compared IMCC with SW by plotting the score versus the KEGG accuracy (JS calculated as formula 8), as shown in Figure 2D, E, Supplementary Table 3.

We also compared IMCC with inter-module average shortest path (IMASP), a topological parameter proposed to evaluate the distance of a module pair. IMCC was plotted versus the IMASP, and determination coefficient of curve fitting was also calculated (Figure 2F, Supplementary Table 4).

3 Results

In this paper, based on high-throughput microarray data from mouse, we integrated two types of inter-module connections, i.e. DIMC and IIMC, combined with statistical methods, and developed a novel quantitative algorithm termed as IMCC. Using IMCC, we drew more coarse grains from molecular network, and integrated the microscopic molecules into mesoscopic modules.

3.1 The IMCC model screens out noise

Biological networks, which were constructed based on a DNA microarray data set and a mathematical model [60], were considered to have high noise due to the false-positive levels inherent in the data set. Taking our experimental network as an example, there were 374 nodes in the co-expression network, whereas most (55.93%) of the shortest path length between any two nodes is gathering in 2, and the third most (13.41%) in 1 (Supplementary Table 5). This indicated that more than 60% of the node pairs could be connected directly or indirectly. Accordingly, in the modular map, the module-module connections also manifested a false-positive property. As for modules, all of the module pairs could be connected directly or indirectly to form a densely interacted modular map. In our data set, the number of direct inter-module connections (DIMCs) was 575, indicating that averagely each module in this map had 24 neighborhood. And the number of indirect inter-module connections (IIMCs) was up to 1128 in the modular map, indicating that any pair of modules was connected by IIMC. In face of such a large number of inter-module connections, how to remove noise interference to accurately screen out the real data about module connections?

Therefore, it seems imperative to screen out the random fluctuations of noise in inter-module connections. We introduced the hypergeometric distribution test (details are described in Methods), which calculates the probability that the specified target is selected from the whole population, so as to identify the valid value of SW (a parameter of DIMC) and PS (a parameter of IIMC) with significance. As the CT (another parameter of IIMC) value is drawn from the comparison with expectation of the whole network, we set the cutoff value of CT at 10 (an inter-module connection with a CT > 10 was considered valid). Comparison of the number of module-module interactions before and after screening revealed large differences. Among the 575 SW, we screened out 412 (71.65%) invalid SW and 163 (28.35%) valid SW remained, according to the P value of hypergeometric distribution. Meanwhile, out of the 1128 indirect module-module interactions, 249 (22.07%) valid CT and 358 (31.74%) valid PS remained after screening based on the cutoff value and hypergeometric distribution, respectively (Figure 2A). After the screening process, we successfully excluded a large number of inter-module connections without statistical significance or below the cutoff value.

3.2 The IMCC method provides a comprehensive landscape

The integration of parameters about these direct and indirect connections provides a more complete landscape of inter-module relationships. Two functional modules may bind to each other or target an identical molecule to constitute a competitive regulation, both of which are universally present in biological networks. There were 144 overlapping inter-module connections between SW and CT, and 19 (11.66% of the total SW) and 105 (42.17% of the total CT) specific (non-overlapping) inter-module connections in SW and CT, respectively. Between SW and PS, 120 overlapping inter-module connections were identified, and 43 (26.38% of the total SW) and 238 (66.48% of the total PS) specific inter-module connections were found in SW and PS, respectively (Figure 2B). These findings revealed that a certain number of module pairs were merely correlated by either direct or indirect connections. As a result, our integration model has a wider-range of coverage than solo-direct and solo-indirect

inter-module connections, which may provide a more comprehensive landscape of inter-module connections. Details of the integration process for SW and CT, as well as SW and PS were described in Supplementary Table 1 and 2.

3.3 The IMCC method reflected inter-module functional coordination

When integrating these parameters, we compared and analyzed two integration models of IMCC (IMCC₁ and IMCC₂). All the inter-module correlations summarized or predicted are presumed to contribute to biological functions. We employed the KEGG to select the most fitting integration model. By plotting Jaccard Similarity (JS) coefficient of each module pair based on KEGG versus IMCC value, we calculated the precision of IMCC (Supplementary Table 3). Our results (Figure 2C and D) suggested that the optimal ρ value in IMCC₁ model was 1/1 and the most fitting model was logarithmic model with a R² of 0.616. In logarithmic model, the R² of different ρ values showed an obvious peak phenomenon: when $\rho=1/1$, the R² reached the peak with two sides sloping down to lower values; when $\rho=1/10$ or $\rho=10/1$, the minimum R² of each side were observed, respectively (Supplementary Figure 1). Therefore, it is proper to decide that the integrated parameter IMCC₁ is more consistent with the KEGG classification than any single index (SW or CT), which would provide more accurate evaluation of the relationship between modules. Using the same method, we plotted the IMCC₂ value against JS in KEGG, and found that there was no correlation between the two parameters and the fitting R² was much lower than IMCC₁ when $\rho=1/1$ (Figure 2C, Supplementary Figure 2). Therefore, we chose the IMCC₁ model as the final model with a ρ value of 1/1. The final formula can be simplified as follows:

$$IMCC = SW + CT \quad (10)$$

To some extent, the integrated method, which is based on the topology structure of networks, reflects the functional coordination of module pairs.

3.4 The IMCC method is more precise and validated by a topological parameter

In the comparison, IMCC showed increased performance of an integrated score relative to the SW. The fitting model of SW was $y = 0.1374\ln(x) + 0.9807$, with R² equal to 0.580, which is lower than 0.616 of IMCC (Figure 2E). Overall, the results indicated that IMCC achieved better performance on the weighted gene co-expression data. It also means that the results based on IMCC were more consistent with biological function than SW. After all, the IMCC provided a more precise evaluation for inter-module connections.

Unlike protein-protein interaction networks, gene co-expression networks are weighted networks. Thus, inter-module connections in such networks should not only include dichotomous edge (0 or 1), but also quantitative precision information. We introduced the SW, CT and PS to quantitatively analyze these inter-module connections, taking the weight and amount of the edge of molecular networks into consideration. We compared the correlation based on IMCC and a dichotomous topological parameter, inter-module average shortest path (also known as average characteristic path length). The results of the two assessments were generally consistent (Figure 2F), indicating that the two parameters had positive linear correlations with a R² of 0.495 (Supplementary Table 4, Supplementary Figure 3). Therefore, this result successfully validated the IMCC method.

3.5 Application of IMCC in genome-scale stroke networks deciphered pathological inter-module connectivity

We applied IMCC on genetic interactomes and modules based on microarray of MCAO mice to explore the characteristic inter-module connections related to stroke. By calculating IMCC between each pair of module, we constructed modular map, in which vertex and edge represent modules and IMCC (as shown in Figure 1). This may contribute in the following two aspects. First, we showed the architecture of the modular map in stroke condition. Second, we analyzed the enriched KEGG pathways of highly correlated gene modules to explore how the pathological pathway interactions contributed to stroke process.

As an overarching view, the modular map exhibited a characteristic “core-periphery” structure (as shown in Figure 1), where the periphery consisted of small well-defined communities, and the core comprised highly interconnected larger modules, which are harder to detect [61-64]. It is suggested that when a biological network is on stress, the inter-module connectivity would decrease [65], to promote this core-periphery structure [63], center of which is conserved and stable module, with crucial role for cell survival rather than development [64, 66], to adapt to the novel situation. In the stroke modular map, the most closely connected modules were a pair of modules (module-blue and module-brown), which can be regarded as “core” presenting characteristic pathological process for cell survival in stroke.

Thus, it is essential to enrich the KEGG pathways of this “core” to explore predominant pathological pathways in stroke. The module-blue was enriched with 12 KEGG pathways; the module-brown was enriched with 4 KEGG pathways ($p < 0.05$, shown in Table 1). Among these pathways, HTLV-I infection, which were suggested to associate with pathogenic mechanism of immune dysregulation in neuroinflammatory [67], was both enriched in module-blue and module-brown: 5 genes from module-blue and 5 genes from module-brown encoded proteins in this pathway (Figure 3). This indicated that the coordination between tightly connected module pair might be attributed to constituting an identical pathway of biological essence. Furthermore, pathways enriched by two modules showed cross-talk based on the KEGG background. For example, MAPK signaling pathway of module-brown participated in pathways in cancer, long-term potentiation, neurotrophin signaling pathway and foxO signaling pathway of module-blue. Therefore, pathway crosstalk might be responsible for the module coordination. Taken together, such topological inter-module connections turned out to constitute a form of functional coordination between modules. We speculated that this coordination is occurring typically from constituting an identical pathway or forming pathway crosstalk. This may also show the ability of our analysis to detect inter-module connectivity of biological function, which is consistent to prior knowledge and may strengthen the role of module coordination.

Table 1

Module	Term	P-Value	Genes
Module-blue	Pathways in cancer	0.001721	VEGFB, DVL3, TRAF2, ADCY1, BRAF, GNA11
	HTLV-I infection	0.003376	POLD3, DVL3, ADCY1, POLD1, IL2RG
	Long-term potentiation	0.010084	ADCY1, BRAF, CAMK4
	Aldosterone synthesis and secretion	0.016716	ADCY1, CAMK4, GNA11
	Insulin secretion	0.016716	ADCY1, GCK, GNA11
	Cholinergic synapse	0.027889	ADCY1, CAMK4, GNA11
	Neurotrophin signaling pathway	0.032131	PDPK1, BRAF, CAMK4
	Vascular smooth muscle contraction	0.034593	ADCY1, BRAF, GNA11
	FoxO signaling pathway	0.03816	PDPK1, BRAF, GADD45A
	Hepatitis C	0.039204	TRAF2, PDPK1, BRAF
	Insulin signaling pathway	0.041326	PDPK1, BRAF, GCK
Hippo signaling pathway	0.047383	DVL3, BBC3, BTRC	
Module-brown	Amyotrophic lateral sclerosis (ALS)	1.73E-04	BAX, GRIN2A, BCL2L1, SOD1
	MAPK signaling pathway	0.001913	RPS6KA4, JUND, MAP2K4, GNA12, HSPA1A
	HTLV-I infection	0.002702	BAX, POLD2, MAP2K4, SMAD3, BCL2L1
	Hepatitis B	0.040124	PTK2B, BAX, MAP2K4

Table 1 Enriched KEGG pathways of module-blue and module-brown.

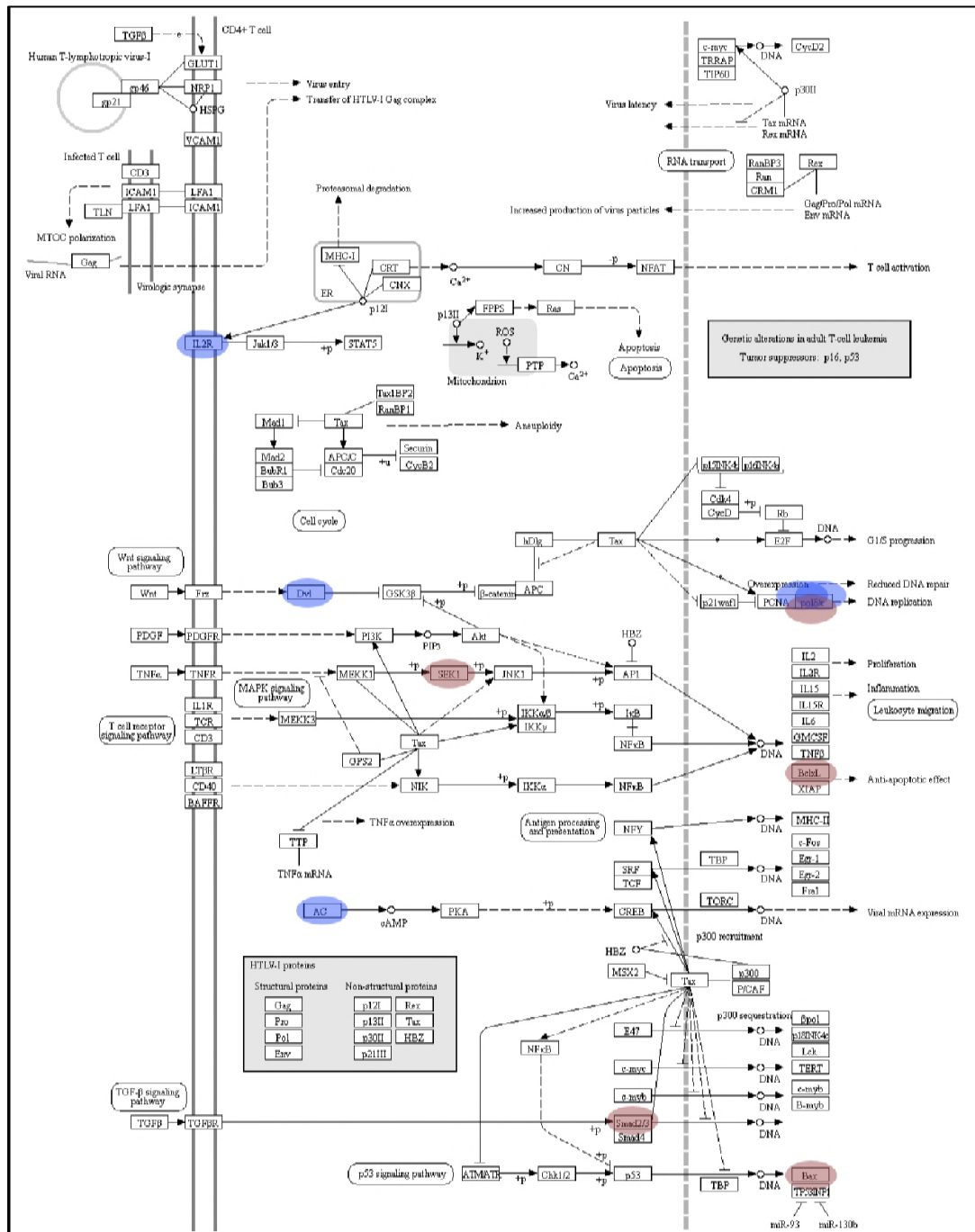


Figure 3. Proteins encoded by genes from module-blue (labelled by blue) and module-brown (labelled by brown) in HTLV-1 infection based on KEGG database.

4 Discussion

In this paper, we described an integrative, quantitative approach (IMCC) to quantify inter-module connections. There were several unique features in our methodology: i) Using the statistical screen, the IMCC method presented a novel and powerful tool to filter out the random fluctuations of noise from significant inter-module connections. ii) By integrating the analysis of direct and indirect inter-module connections, we provide a more comprehensive approach for quantification. iii) The biological functions were introduced to evaluate the reliability of inter-module connections. By fitting the IMCC score to the JS of the KEGG category, IMCC is considered a more precise tool for inter-module topology structure analysis reflecting the functional coordination of module pairs, in the comparison with extant parameter. Taken together, this novel model showed good performance to reveal functional interactions among modules.

This capability can be extremely useful for bio-network. In the application in genome-scale stroke networks, we deciphered characteristic pathological “core-periphery” structure and functional coordination module pair. Thus, this method can be attempted to apply in case-control or drug perturbation network to explore detailed variation of inter-module connectivity related to pathological process or pharmacological action. This inter-module connectivity of disease- or drug-conditional existence can be served as network biomarkers to characterize the dynamic response to perturbation. For its flexibility and impact on phenotype alteration, inter-module connection can also be served as targets to design specific drugs. This concept may bring about reinterpreting the connotation of pharmacology. After all, various desirable properties of this algorithm discussed in this work will facilitate the inter-module analysis to be applicable to biological complex systems for disease and drug discovery.

Supporting information

S1 Text. Supplementary Figures and tables for model and validation. (DOC)

Acknowledgments

This work was supported by National Science and Technology Major Projects “Major New Drug Development Project”. [grant number 2017ZX09301-059].

The authors declare that they have no competing interests.

References

1. Jia J, Zhu F, Ma X, Cao Z, Li Y, Chen YZ. Mechanisms of drug combinations: interaction and network perspectives. *Nat Rev Drug Discov.* 2016; **8**:111-28.
2. Wang Z, Wang YY. Navigating personalized medicine dependent on modular flexibility. *Trends Mol Med.* 2013; **19**:393-395.
3. Wang Z, Liu J, Yu YN, Wang YY. Modular pharmacology: the next paradigm in drug discovery. *Expert Opin Drug Discov.* 2012; **7**:667-77.
4. Kim I, Lee H, Han SK, Kim S. Linear Motif-Mediated Interactions Have Contributed to the Evolution of Modularity in Complex Protein Interaction Networks. *PLoS Comput Biol.* 2014; **10**:e1003881.
5. Hartwell LH, Hopfield JJ, Leibler S, Murray AW. From molecular to modular cell biology. *Nature.* 1999; **2**:c47–c52.
6. Ravasz E, Somera AL, Mongru DA, Oltyai ZN, Barabasi AL. Hierarchical organization of modularity in metabolic networks. *Science.* 2002; **297**:1551-1555.
7. Maraziotis IA, Dimitrakopoulou K, Bezerianos A. An in silico method for detecting overlapping functional modules from composite biological networks. *BMC Syst Biol.* 2008; **2**:93.
8. Rhrissorrakrai K, Gunsalus KC. MINE: module identification in networks. *BMC Bioinformatics.* 2011; **12**:192.
9. Nepusz T, Yu H, Paccanaro A. Detecting overlapping protein complexes in protein-protein interaction networks. *Nat Methods.* 2012; **9**:471-472.
10. Valente AX, Cusick ME. Yeast Protein Interactome topology provides framework for coordinated-functionality. *Nucleic Acids Res.* 2006; **34**:2812-9.
11. Lin CY, Lin YW, Yu SW, Lo YS, Yang JM. MoNetfamily: a web server to infer homologous modules and module-module interaction networks in vertebrates. *Nucleic Acids Res.* 2012; **40**:263-70.
12. Granovetter MS. *The Strength of Weak Ties.* Social Science Electronic Publishing, New York. 1973.
13. Onnela JP, Saramäki J, Hyvönen J, Szabó G, Lazer D, Kaski K, et al. Structure and tie strengths in mobile communication networks. *Proc Natl Acad Sci USA.* 2009; **104**:7332-6.
14. Jo HH, Pan RK, Kaski K. Emergence of bursts and communities in evolving weighted networks. *Plos One,* 2011; **6**:e22687.

15. Ma X, Gao L. Discovering protein complexes in protein interaction networks via exploring the weak ties effect. *BMC Syst Biol.* 2012; **6**:1-15.
16. Meda SA., Ruaño G, Windemuth A, O'Neil K, Berwise C, Dunn SM, et al. Multivariate analysis reveals genetic associations of the resting default mode network in psychotic bipolar disorder and schizophrenia. *Proc Natl Acad Sci USA.* 2014; **111**:2066-75.
17. Amar D, Safer H, Shamir R. Dissection of regulatory networks that are altered in disease via differential co-expression. *PLoS Comput Biol.* 2013; **9**:e1002955.
18. Parter M, Kashtan N, Alon U. Facilitated variation: how evolution learns from past environments to generalize to new environments. *PLoS Comput Biol.* 2008; **4**: e1000206.
19. Roguev A, Bandyopadhyay S, Zofall M, Zhang K, Fischer T, Collins SR, et al. Conservation and rewiring of functional modules revealed by an epistasis map in fission yeast. *Science*, 2008; **322**:405-410.
20. Neduva V, Russell RB. Linear motifs: evolutionary interaction switches. *FEBS Lett.* 2005; **579**:3342-3345.
21. Akiva E, Friedlander G, Itzhaki Z, Margalit H. A dynamic view of domain-motif interactions. *PLoS Comput Biol.* 2012; **8**:e1002341.
22. Van Roey K, Dinkel H, Weatheritt RJ, Gibson TJ, Davey NE. The switches ELM resource: a compendium of conditional regulatory interaction interfaces. *Sci Signal.* 2013; **6**:rs7.
23. Mosca R, Pache RA, Aloy P. The role of structural disorder in the rewiring of protein interactions through evolution. *Mol Cell Proteomics.* 2012; **11**:M111.
24. Kim J, Kim I, Yang JS, Shin YE, Hwang J, Park S, et al. Rewiring of PDZ domain-ligand interaction network contributed to eukaryotic evolution. *PLoS Genet.* 2012; **8**, e1002510.
25. Sun MG, Sikora M, Costanzo M, Boone C, Kim PM. Network evolution: rewiring and signatures of conservation in signaling. *PLoS Comput Biol.* 2012; **8**:e1002411.
26. Gerhart J, Kirschner M. The theory of facilitated variation. *Proc Natl Acad Sci USA.* 2007; **104**:8582-8589.
27. King N, Westbrook MJ, Young SL, Kuo A, Abedin M, Chapman J, et al. The genome of the choanoflagellate *Monosiga brevicollis* and the origin of metazoans. *Nature*, 2008; **451**:783–788.
28. Parter M, Kashtan N, Alon U. Facilitated variation: how evolution learns from past environments to generalize to new environments. *PLoS Comput Biol.* 2008; **4**:e1000206.

29. Zeng T, Wang DC, Wang X, Xu F, Chen L. Prediction of dynamical drug sensitivity and resistance by module network rewiring-analysis based on transcriptional profiling. *Drug Resist Updat*. 2014; **17**: 64-76.
30. Yu X, Li G, Chen L. Prediction and early diagnosis of complex diseases by edge-network. *Bioinformatics*. 2014; **30**:852-9.
31. Zeng T, Sun SY, Wang Y, Zhu H, Chen L. Network biomarkers reveal dysfunctional gene regulations during disease progression. *FEBS J*. 2013; **280**:5682-5695.
32. Li M, Zeng T, Liu R, Chen L. Detecting tissue-specific early warning signals for complex diseases based on dynamical network biomarkers: Study of type 2 diabetes by cross-tissue analysis. *Brief Bioinform*. 2014; **15**:229-243.
33. Breitkreutz D, Tuszynski JA. Molecular signaling network complexity is correlated with cancer patient survivability. *Proc Natl Acad Sci USA*. 2012; **109**:9209-12.
34. Sim SC, Ingelman-Sundberg M. Pharmacogenomic biomarkers: new tools in current and future drug therapy. *Trends Pharmacol Sci.*, 2011; **32**:72-81.
35. Ding F, Zhang Q, Ung CO, Wang Y, Han Y, Hu Y, et al. An Analysis of Chemical Ingredients Network of Chinese Herbal Formulae for the Treatment of Coronary Heart Disease. *Plos One*. 2015; **10**:e0116441.
36. Wang Z, Liu J, Cheng Y, Wang Y. Fangjiomics: in search of effective and safe combination therapies. *J Clin Pharmacol*. 2011; **51**:1132-51.
37. Duan DD, Wang Z, Zhang BL, Wang YY. Fangjiomics: revealing adaptive omics pharmacological mechanisms of the myriad combination therapies to achieve personalized medicine. *Acta Pharmacol Sin*. 2015; **36**:651-3.
38. Liu J, Wang Z. Diverse array-designed modes of combination therapies in Fangjiomics. *Acta Pharmacol Sin*. 2015; **36**:680-8.
39. Wang Z, Duan DD, Wang YY. Editorial: Combination Therapy of Vascular Diseases and Fangjiomics: When West Meets East in the Era of Phenomics. *Curr Vasc Pharmacol*. 2015; **13**: 420-2.
40. Yu H, Kim PM, Sprecher E, Trifonov V, Gerstein M. The importance of bottlenecks in protein networks: correlation with gene essentiality and expression dynamics. *PLoS Comput Biol*. 2007; **3**: e59.
41. Zhu F, Wang W, Di Z, Fan Y. Identifying and characterizing key nodes among communities based

- on electrical-circuit networks. Plos One. 2014; **9**:e97021.
42. Tamás N, Andrea P, László N, Bazsó F. Fuzzy communities and the concept of bridgeness in complex networks. Phys Rev E Stat Nonlin Soft Matter Phys. 2008; **31**:119-136.
 43. Liu C, Li J, Zhao Y. Exploring hierarchical and overlapping modular structure in the yeast protein interaction network. BMC Genomics, 2010; **11**:1-12.
 44. Kovács IA, Palotai R, Szalay MS, Csermely P. Community Landscapes: An Integrative Approach to Determine Overlapping Network Module Hierarchy, Identify Key Nodes and Predict Network Dynamics. Plos One, 2010; **5**:e12528.
 45. Wang W, Liu D, Liu X, Pan L. Fuzzy overlapping community detection based on local random walk and multidimensional scaling. Physica A. 2013; **392**:6578-6586.
 46. Zhang S, Wang RS, Zhang XS. Uncovering fuzzy community structure in complex networks. Phys Rev E. 2007; **76**:70-80.
 47. Reichardt J, Bornholdt S. Detecting fuzzy community structures in complex networks with a Potts model. Phys Rev Letters. 2004; **93**:218701.
 48. Cui L, Qin L, Lu N. A fast overlapping community detection algorithm with self-correcting ability. Scientific World J. 2014; **4**:281-316.
 49. Missiuro PV, Liu K, Zou L, Ross BC, Zhao G, Liu JS, et al. Information flow analysis of interactome networks. PLoS Comput Biol. 2009; **5**:1000350.
 50. Ulitsky I, Shlomi T, Kupiec M, Shamir R. From E-MAPs to module maps: dissecting quantitative genetic interactions using physical interactions. Mol Syst Biol. 2008; **4**:173-178.
 51. Yu Q, Li GH, Huang JF. MOfinder: a novel algorithm for detecting overlapping modules from protein-protein interaction network. J Biomed Biotechnol. 2012;**2012**, 103702.
 52. Yang SK, Wang YC, Chao CC, Chuang YJ, Lan CY, Chen BS. Dynamic cross-talk analysis among TNF-R, TLR-4 and IL-1R signalings in TNF alpha-induced inflammatory responses. BMC Med Genomics. 2009; **3**:1-19.
 53. Collins SR, Schuldiner M, Krogan NJ, Weissman JS. A strategy for extracting and analyzing large-scale quantitative epistatic interaction data. Genome Biol. 2006; **7**:1-14.
 54. Schuldiner M, Collins SR, Thompson NJ, Denic V, Bhamidipati A, Punna T, et al. Exploration of the Function and Organization of the Yeast Early Secretory Pathway through an Epistatic Miniarray Profile. Cell. 2005; **123**:507-19.

55. Bandyopadhyay S, Mehta M, Kuo D, Sung MK, Chuang R, Jaehnig E J, et al. Rewiring of genetic networks in response to DNA damage. *Science*. 2010; **330**:1385-1389.
56. David A, Ron S. Constructing module maps for integrated analysis of heterogeneous biological networks. *Nucleic Acids Res*. 2014; **42**:4208-19.
57. Hsu JT, Peng CH, Hsieh WP, Lan CY, Tang CY. A novel method to identify cooperative functional modules: study of module coordination in the *Saccharomyces cerevisiae* cell cycle. *BMC Bioinformatics*. 2011; **12**:1-22.
58. Kelder T, Eijssen L, Kleemann R, van Erk M, Kooistra T, Evelo C. Exploring pathway interactions in insulin resistant mouse liver. *BMC Syst Biol*. 2011; **5**:1814-1817.
59. Cho YR, Zhang A. Identification of functional hubs and modules by converting interactome networks into hierarchical ordering of proteins. *BMC Bioinformatics*. 2010; **11**:1-10.
60. Hatfield GW, Hung SP. Differential analysis of DNA microarray gene expression data. *Mol Microbiol*. 2003; **47**:871-877.
61. Lancichinetti A, Kivela M, Saramaki J, Fortunato S. Characterizing the Community Structure of Complex Networks. *Plos One*. 2010; **5**:e11976-e11976.
62. Bollobás B. *Modern Graph Theory*. Springer, Netherlands. 1998.
63. Verma T, Russmann F, Arajo NA, Nagler J, Herrmann HJ. Emergence of core-peripheries in networks. *Nat Commun*. 2016; **7**:36-41.
64. Han H W, Ohn JH, Moon J, Kim JH. Yin and Yang of disease genes and death genes between reciprocally scale-free biological networks. *Nucleic Acids Res*. 2013; **41**:9209-17.
65. Mihalik Á, Csermely P. Heat Shock Partially Dissociates the Overlapping Modules of the Yeast Protein-Protein Interaction Network: A Systems Level Model of Adaptation. *PLoS Comput Biol*. 2011; **7**:e1002187.
66. Kim D, Kim MS, Cho KH. The core regulation module of stress-responsive regulatory networks in yeast. *Nucleic Acids Res*. 2012; **40**:8793-802.
67. Natsumi A, Tomoo S, Naoko Y, Hitoshi A, Atae U, Steven J, et al. Human t-lymphotropic virus type 1 (htlv-1) and regulatory t cells in htlv-1-associated neuroinflammatory disease. *Viruses*. 2011; **3**:1532-1548.