Article

A further study on the topological structure of tumor signaling pathways

WenJun Zhang, **XueBing Huang**
School of Life Sciences, Sun Yat-sen University, Guangzhou 510275, China
E-mail: zhwj@mail.sysu.edu.cn,wjzhang@iaees.org

Received 11 June 2021; *Accepted 16 November 2021*; *Published online 6 September 2022*; *Published 1 June 2023*

Abstract
Tumorigenesis is a multifactor and multistep process, of which the change of metabolic signaling pathways plays a key role. Most of the previous studies on tumor signaling pathways focused mainly on the metabolic process and chemical processes of some selected metabolites, the tumorigenesis induced by abnormal signaling from mutation of this metabolite or gene, and the chemical structure of ligands, receptors and signaling proteins. However their network biology is seldom studied. Based on the previous studies, the present study was conducted to further analyze the topological structure of tumor signaling pathways using Pajek and UCINET software. Some critical metabolites were found and sensitivity analyses for signaling pathways were conducted. Centrality and core skeleton analysis showed that the crucial metabolites of AKT signaling pathway are Akt-p and Akt; the crucial metabolites of JAK-STAT signaling pathway are JAKs and 23(STATs-P)2; the crucial metabolites of p53 signaling pathway are p53-P-P, Gene Expression, Ac-p53 and (Ac-p53-P)2; the crucial metabolites of Ras signaling pathway are Ras-GTP, Ras-GDP and MEKK1; the crucial metabolites of TNF signaling-pathway are MEKIKs-P-NIK-P and TRADD, and for VEGF signaling pathway, the crucial metabolites are PIP3 and ANGIOGENESIS. The performance of cascade model was poor in predicting topological properties of tumor signaling pathways.

Keywords tumor; signaling pathway; betweenness centrality; closeness centrality; degree; cascade model.

Network Pharmacology
ISSN 2415-1084
URL: http://www.iaees.org/publications/journals/np/online-version.asp
RSS: http://www.iaees.org/publications/journals/np/rss.xml
E-mail: networkpharmacology@iaees.org
Editor-in-Chief: WenJun Zhang
Publisher: International Academy of Ecology and Environmental Sciences

1 Introduction

The completion of the Human Genome Project indicated the arrival of the post-genome era. The focus of biological research has shifted from the local research on the function of individual genes or proteins in cells to various "omics" research that take all the genes, mRNAs, proteins and metabolites in the cell as the research objects, that is, holistic research. Various omics technologies such as metabolomics, genomics, proteomics, and transcriptomics have gradually pushed molecular biology into the era of systems biology, and also brought bioinformatics into the post-genomic informatics era. Since genes and proteins tend to affect the function of biological systems through network-like interactions in groups, it is necessary to analyze their interaction

networks to study biological functions (Zhang, 2012a, 2016b, 2018).

In the biological studies, the following four network types were generally defined and used, i.e., protein-protein interaction networks, metabolic networks, transcriptional regulatory networks and signal transduction networks. Networks (graphs) represent different interaction types between biological entities (such as transcription factors, genes, small chemical molecules, and proteins, etc.), nodes represent biological entities, and edges (links) between nodes represent their interactions. In addition, networks (graphs) can also be divided into directed graphs and undirected graphs according to the attributes of the links in the network, or divided into Bayesian networks and Boolean networks according to the representational meaning of the links in the network, or according to the distribution of node degrees in the network they are divided into homogeneous network and "scale-free" networks (Aittokallio and Schwikowski, 2006; Zhang, 2012a, 2018).

Generally speaking, we often carry out computational analysis of biological networks from the following three aspects: (1) Research on network statistics and topological properties, including node centrality, aggregation coefficient, betweenness centrality, node degree distribution, and the relationship between nodes (Zhang, 2018). In addition, robustness analysis of the network based on the shortest path and random removal of individual nodes, etc (Zhang, 2016d); (2) modularity analysis, which refers to identifying sub-graphs (sub-networks) of interconnected nodes with functionally or regionally specific connections in the network (Zhang, 2016c, 2018), such as a specific disease module or specific metabolic pathway module; (3) elemental analysis, which refers to the identification of small network modules that overlap with it when compared with a graph formed by random shuffling of nodes in the same network, often used for regulatory networks such as the graphical analysis of discrete biological processes. In addition, network graphs that characterize the interaction types of the same biological entity can be superimposed and compared to find their common components. For example, protein-protein interaction networks of different biological origins are superimposed to identify their possible evolutionary relationships (Pujol et al., 2010).

Numerous studies have proved that the position of a node in the topology of a biological network is related to its importance in intracellular functions (Zhang, 2012a, 2018). Changes in the external environment or internal conflicts acting on nodes with different topological properties will cause the network to exhibit different degrees of robustness or fragility (Albert et al., 2000; Jeong et al., 2001; Zhang, 2016d, 2018). Burst interference acts on random nodes in the network and has little impact on the network, because the failure of random selection mainly occurs in most non-critical nodes, and their absence generally does not destroy the overall characteristics of the network. At this time, the network shows robustness (Zhang, 2016d; Zhang and Feng, 2017). When crucial structures of the network are destroyed, the removal of a few crucial nodes will split the system into some small and isolated groups of nodes, and it is even possible to observe the phase transition of the system and the disintegration of the entire network, which is a manifestation of network vulnerability (Zhang, 2016b, 2018).

The research on crucial nodes of biological networks is mainly aimed at the evaluation of node centrality and the impact of removal on network structure and function. Specifically, identifying important nodes in a network has a series of computational metrics of topological properties. Two of the more important measures of topological properties are node degree distribution and aggregation coefficient (Zhang, 2018). The node degree distribution represents the different number of links connected by any node in the network, and the clustering coefficient represents the ratio of the number of adjacent nodes that a node actually connects to the maximum number of nodes it can connect to. The calculation criteria of these topological properties can effectively reflect the characteristics of the network structure (Zhang, 2018). For example, in the protein-protein interaction network, we found that the degree distribution of the network nodes conforms to the power-law type (Huang and Zhang, 2012), which belongs to the scale-free network, that is, in the network

most of the nodes are only connected to very few nodes, and very few nodes are highly connected. In addition, other common topological properties are degree centrality, which is used to measure the degree of a node in the path, and betweenness centrality, which is used to analyze the influence of the node on the flow of network information, and closeness centrality, which is used to measure the degree of deviation of a node from the center of the graph (Zhang, 2018; Zhang and Zhang, 2019; Xin and Zhang, 2020).

Metabolism is at the end of the regulation of life activities and is the chemical engine that drives life processes, generating energy to drive various cellular processes, degrading and synthesizing many different molecules. Metabolic network expresses all biochemical reactions in cells as a network, which reflects the interaction between all compounds involved in metabolic process and between all catalytic enzymes, and is an abstract expression of cell metabolism. The most important chemical feature of tumor cells is unrestricted rapid reproduction. In order to meet the needs of their rapid reproduction, tumor cells exhibit metabolic characteristics different from those of normal tissue cells.

The occurrence of tumor is a multifactor and multistep process. Among them, changes in cell signal transduction pathways play a key role (Rahman et al., 2013; Iqbal et al., 2014). Intracellular signaling pathways are closely related to the regulation of metabolic networks and affect the process of tumor metabolism (Ibrahim et al., 2011). Tumor signaling metabolism has become a hot research area today, and the research on tumor signaling pathways is relatively comprehensive. The tumor signaling pathways are mainly divided into several types: Akt, JAK-STAT, p53, Ras, TNF, and VEGF signaling pathways, etc (Kolch , 2002; Moustakas et alo., 2002; Katoh , 2005; Marrero, 2005; Stauffer et al., 2005; Ho et al., 2006).

These signal pathways are the main pathways of tumor signal metabolism, among which there are a variety of important ligands, receptors or signal proteins that affect each signal metabolism process, forming a complex network.

Existing tumor metabolism research mainly focuses on the metabolic pathway process and chemical process at a certain point of tumor signal metabolism, the process of inducing cancer caused by abnormal signal due to mutation, and the crucial ligands and receptors in the tumor signaling pathways and entire metabolic pathway.

Although there is no complete database on tumor signaling pathways, the two known websites on tumor signaling pathways provide reliable signaling pathway maps related to signaling pathways, which can be used for tumor signaling and metabolic network analysis. The original data sources are very reliable, and these metabolic pathways have been thoroughly studied, and the metabolic directions of metabolites and signals at each node are very clear.

Studies have shown that in the regulation of tumor metabolism, certain enzymes have the effect of affecting cell proliferation. And these nodes may be the most vulnerable control points in the metabolic network. Therefore, research on these nodes is crucial.

The tumor signaling pathways are an important research area, and there are few studies on this aspect (Huang and Zhang, 2012; Li and Zhang, 2013). Studying the tumor signaling pathways can help us better understand and utilize the metabolic process of tumor cells. Knowing the topological properties of the tumor metabolic network is essential to understand its dynamic behavior, biological function realization and characteristics. In present study, the relevant data of the above six signaling pathways were based, and Pajek software was used to conduct centrality analysis, network core skeleton analysis and cascade model analysis on tumor signaling pathways, so as to provide a useful and credible basis for future research on tumor signaling pathways. It is expected to provide valuable information for tumor diagnosis and treatment and tumor drug design from the perspective of network biology.

2 Software and Data Sources

2.1 Introduction to UCINET software

UCINET integration software for network analysis includes NetDraw for one-dimensional and two-dimensional data analysis, as well as Mage, the software for three-dimensional display analysis under development, and Pajek's Free application software for large-scale network analysis. UCINET software can read text files, KrackPlot, Pajek, Negopy, VNA and other formats of data files. It can handle 32767 network nodes. Of course, from the practical point of view, when the number of nodes exceeds 5000 or so, some algorithms will run very slowly. The package has strong matrix analysis functions, such as matrix algebra and multivariate statistical analysis. Network analysis methods in the software include centrality analysis, subgroup analysis, role analysis, and permutation-based statistical analysis (Borgatti et al., 2011; Jiang et al., 2015; Zhang, 2012a-b, 2016a-b, 2018).

2.2 Pajek software

The Pajek software was written by Vladimir Batagelj and Andrej Mrvar and is freely available to users for non-commercial use. Pajek means spider in Slovenian, and the logo of the software is a spider, implying that it has the function of drawing a network (Kuang and Zhang, 2011; Jiang et al., 2015).

Pajek is a software for analyzing large and complex networks and is a powerful tool for studying various complex nonlinear networks that exist today. Pajek provides analysis and visualization manipulation tools to the following networks: Coauthoring Network, Chemical Organic Molecules, Protein Receptor Interaction Network, Genealogy, Internet, Citation Network, Communication Network (AIDS, News, Innovation), Data Mining (2-mode), etc. Pajek runs under Windows for analysis and visualization of large networks with thousands or even millions of nodes. It has the characteristics of fast calculation, visualization and abstraction.

2.3 Data sources

The original data of this study came from the two websites of providing tumor signaling pathways (Pathway Central, 2012):

http://www.sabiosciences.com/pathwaycentral.php

http://www.abcam.com/index.html?pageconfig=productmap&cl=2282

2.4 Data conversion

Tumor metabolism maps describe all metabolites and signaling pathways related to tumor metabolism, and these metabolic pathways are directional. Although the map is intuitive, it does not determine the order and associations of the various metabolites in the metabolism. For the convenience of research, firstly, according to the original data map, all the metabolites in the map are numbered, and each metabolite is a node (all entities, behaved as nodes in the pathways, are treated as metabolites in present study). Then in Excel, the starting node is listed as rows and ending nodes as columns. The association (i.e., interaction) between the two metabolites is recorded as 1, and non-association between the two metabolites is recorded as 0, and the metabolic map data is thus converted into matrix data. After the data conversion, open the Data/data editors/matrix editor in the UCINET software, import the Excel data, and then store it in the .##h format through the “Save as” in the MatrixEditor, and then select it through the Netdraw software File/Open/Ucinet dataset/network and open the file in ##h format that was just stored, and then save it as a file in .net, .clu, .vec and other formats through File/save data as/Pajek/net file. These documents constitute Pajek's basic analysis source documents.

3 Methods

3.1 Centrality analysis of tumor signaling pathways

The centrality of a metabolite can measure the relative importance of the metabolite in the network. Here we use three centrality measures: degree centrality, betweenness centrality and closeness centrality (Wasserman

and Faust, 1994; Zhang and Zhan, 2011; Shams and Khansari, 2014; Zhang, 2012a-b, 2016a-b, 2018, 2021; Zhang and Feng, 2017; Zhang and Zhang, 2019; Xin and Zhang, 2020, 2021; Yang and Zhang, 2022).

3.1.1 Degree Centrality (DC)

In the metabolic network, if a metabolite is directly related to many other metabolites, the metabolite plays a central role in metabolism. Central metabolic signals often have many associations with other signals, while peripheral metabolic signals often do not have such characteristics. Therefore, calculating the degree centrality of metabolites can measure the importance of metabolites. Degree centrality is the simplest and least informative measure, it only considers the number of other metabolites connected to metabolite *i*:

$$DC_i = DC_{in,i} + DC_{out,i}$$

where $D_{in,i}$ is the in-degree of metabolite *i*, $D_{out,i}$ is the out-degree of metabolite *i*. We use Net/Partitions/DC/All in Pajek to calculate DC_i.

3.1.2 Betweenness Centrality (BC)

If many metabolic pathways pass through a certain metabolite, it can be considered that this metabolite has a very important position, because it has the ability to control the connection of other metabolites, which is quite a bridge. A metabolite is said to have a high betweenness centrality if it is in the shortest path of many other metabolite pairs. Betweenness centrality can be obtained by calculating the probability that metabolite *i* appears on the shortest path of each pair of metabolites *j* and *k*. The standard BC is represented by:

$$BC_i = \frac{2 \times \sum_{j \le k} g_{jk}(i) / g_{jk}}{(N-1)(N-2)}$$

where $i \neq j \neq k$, g_{jk} is the shortest path between metabolites *j* and *k*, $g_{jk}(i)$ is the number of the shortest paths containing metabolite *i*, *N* is the total number of metabolites in the network.

3.1.3 Closeness Centrality (CC)

Degree centrality describes the local centrality of a metabolite, and measures the associations (i.e., interactions, connections) between the metabolite itself and other metabolites, regardless of whether it can control the associations between other metabolites. Betweenness centrality takes into account the ability of a metabolite to control associations between other metabolites, but not the degree to which it is controlled by other metabolites. Considering the degree to which a metabolite is controlled by other metabolites, if a metabolite in the network is less dependent on other metabolites in the process of interacting with other nodes, the metabolite has a higher centrality. Since metabolites at non-core positions need to pass other metabolites to transmit information, metabolites at core positions are less dependent on other metabolites when transmitting metabolic signals. Therefore, the closeness of this metabolite to other metabolites should be considered. The closer a metabolite is to other metabolites, the less dependent it can be on others. Closeness centrality can be obtained by calculating the shortest path from metabolite *i* to other metabolites:

$$CC_i = \frac{N-1}{\sum_{j=1}^{N} d_{ij}}$$

where $i \neq j$, d_{ij} is the length of the shortest path between metabolites *i* and *j*.

We use Net/Vector/Centrality/Betweenness and Net/Vector/Centrality/Closeness/All in Pajek to calculate BC_i and CC_i.

3.2 Core skeleton analysis of tumor signaling pathways

By comparing the topological structures of removing important metabolites, random metabolites, a large number of common metabolites and the complete network, it is further verified whether the metabolites with large centrality are crucial metabolites (Jiang and Zhang, 2015; Zhang and Feng, 2017). Referring to Kuang and Zhang (2011), the results of node analysis, link analysis and chain length analysis are selected for comparison. The important metabolites to be removed are the metabolites with large DC, BC, and CC values, and the random metabolites are obtained by the RANDBETWEEN function in MS Excel. In addition, in the presence of important metabolites and their interconnections, removing a large number of common metabolites (one third of the total number of common metabolites) is compared with other results to verify whether the removal of a large number of common metabolites will lead to the collapse of the network.

3.3 Cascade model

The original cascade model was proposed by Cohen and Newman in 1985 (Zhang, 2012a, Zhang, 2018). The model assumes that the adjacent matrix A of size $S*S$ (S is the number of metabolites) constructed from the metabolic network is a strict upper triangular matrix (that is, if $i \geq j$, then $a_{ij}=0$). In this case, the metabolic network is acyclic; and the species with sequence number 1 can only produce other metabolites, but cannot be obtained from other metabolites, and the metabolites with sequence number 2 can produce metabolites with sequence number 3 and above, but can also be obtained by metabolite 1 was obtained. Therefore, the metabolite with the serial number S can be produced from any other metabolite except itself (Zhang, 2012a; Zhang, 2016b, 2018; Zhang et al., 2014). Therefore, a strictly upper triangular matrix describes a strict metabolic hierarchy. Further, the model assumes that there is a positive real number c (in fact, $c=2CS^2/(S-1)$, where C is the connectivity), for $S \geq c$, the elements above the main diagonal in the upper triangular matrix obey the 0-1 distribution with parameter $p=c/S$, so that the relevant properties of the metabolic network can be deduced.

4 Results

Since the metabolites and metabolic processes of each signaling pathway are mainly expressed in the form of imagery data, they have been sorted and analyzed into the format required by the analysis software, and then imported into the software for analysis. The results are as follows.

4.1 Centrality analysis and core skeleton analysis

4.1.1 AKT signaling pathway

Use Pajek/Net/Partitions/DC/All,Net/Vector/Centrality/Betweenness, and Net/Vector/Centrality/Closeness/All to calculate the centrality values of AKT signaling pathway, the results are listed in Table 1.

Table 1 DC, BC and CC of metabolites in AKT signaling pathway.

ID	Metabolite	DC	BC	CC
1	GABA(A)R	1	0.000000	0.341176
2	CPCR	1	0.000000	0.195286
3	Ras	3	0.026013	0.241667
4	RTK	1	0.000000	0.195286
5	PI3Ky	1	0.000000	0.210909
6	PIP3	6	0.101028	0.266055
7	PDK-1	2	0.108590	0.327684
8	PI3K-GAB1-GAB2	2	0.038113	0.310160

9	GAB2	1	0.000000	0.300518
10	Akt	7	0.192377	0.426471
11	IRS1-PI3K	2	0.004537	0.327684
12	PIP2	1	0.000000	0.210909
13	PTEN	1	0.000000	0.210909
14	CTMP	1	0.000000	0.300518
15	PI3K	3	0.038113	0.215613
16	ILK	1	0.000000	0.300518
17	JAK1	1	0.000000	0.177914
18	BCAP	2	0.013007	0.179012
19	SYK	1	0.000000	0.152231
20	Akt-p	31	0.217786	0.513274
21	Caspase9-P	1	0.000000	0.341176
22	PDE3B-P	1	0.000000	0.341176
23	TSC2-TSC1	2	0.025408	0.358025
24	mTOR	3	0.019964	0.272300
25	p70S6K	1	0.000000	0.214815
26	4EBP1	2	0.006957	0.216418
27	eIF4E	1	0.000000	0.178462
28	Raf1	1	0.000000	0.341176
29	XIAP-Ser87-P	1	0.000000	0.341176
30	BAD-P	2	0.006325	0.345238
31	BAD-P-(14-3-3)	1	0.000000	0.257778
32	Chk1	1	0.000000	0.341176
33	P21CIP1-P	1	0.000000	0.341176
34	P27KIP1-P-(14-3-3)	1	0.000000	0.341176
35	FKHR-P-(14-3-3)	2	0.000000	0.345238
36	FKHR-Death Genes	1	0.000000	0.257778
37	CREB-P	2	0.006352	0.345238
38	CREB-P-Survival Genes	1	0.000000	0.257778
39	MDM2-P	2	0.006352	0.345238
40	MDM2-P-p53-Ub	1	0.000000	0.257778
41	GSK3	3	0.012704	0.349398
42	Glycogen Synthase	1	0.000000	0.260090
43	CyclinD	1	0.000000	0.260090
44	JIP1	1	0.000000	0.341176
45	ASK1-P	1	0.000000	0.341176
46	eNOS-P	1	0.000000	0.341176
47	AR-P	1	0.000000	0.341176
48	Ataxin-(14-3-3)	1	0.000000	0.341176
49	Htt-P	1	0.000000	0.341176
50	YAP-(14-3-3)	1	0.000000	0.341176

51	P47Phox	1	0.000000	0.341176
52	PRAS40-(14-3-3)	1	0.000000	0.341176
53	WNK1-P	1	0.000000	0.341176
54	IKKs-P	1	0.000000	0.341176
55	PFK1-P-PFK2-P	1	0.000000	0.341176
56	GLUT4	1	0.000000	0.341176
57	DNA-PK	1	0.000000	0.341176
58	PP2A	1	0.000000	0.341176
59	CDC37-HSP90	1	0.000000	0.341176

It can be found from Table 1 that some metabolites have a very large degree, including the three metabolites with degrees 6, 7, and 31. Most of the metabolites have degrees between 1 and 3, accounting for 94.9% of the total metabolites, indicating that a few metabolites have a high degree, and most metabolites have a low degree. Among them, the metabolites with the degree of 1 are the most popular, indicating that the topology of the AKT signaling pathway is relatively concentrated, the chain structure is less, and the central metabolites are restricted by the surrounding metabolites. This is consistent with the results of Huang and Zhang (2012), i.e., the AKT network type is a scale-free complex network, and the degree distribution conforms to the power-law distribution (Zhang and Li, 2016), that is, a few metabolites in the network have high degrees, and most metabolites have low degrees (Huang and Zhang, 2012).

Metabolites with higher degrees tend to be critical components of the metabolic process. We use degree centrality, betweenness centrality and closeness centrality to evaluate the relative importance of metabolites. The larger the centrality value, the more important the metabolite is. It can be seen from Table 1 that the top five metabolites with DC, BC and CC values are 20, 10, 6, 15, 3; 20, 10, 7, 6, 8; 20, 10, 41, 37, 39. The number of metabolites with DC, BC, and CC values all ranking in the top five is 20 and 10. From this, it can be speculated that Akt-p and Akt are important metabolites in the AKT network. According to the results of Li and Zhang (2013), FKHR-P-(14-3-3), IKKs-P, Akt, Akt-P, mTOR, TSC2-TSC1, PDK-1, PIP3, IRS1-PI3K, PI3K are important metabolites of the AKT metabolic network because these ten metabolites have the largest k value (k=2). Akt-p and Akt are important metabolites and the results of Li and Zhang (2013) are the same, and other results are different mainly due to the different degrees of accuracy caused by different measures. Selecting k cores to measure the importance of metabolites, the results are not detailed enough. For example, the k values of TSC2-TSC1, Akt-P, FKHR-P-(14-3-3), IKKs-P, and mTOR are all 2, but their importance in the network is different. Akt-p is the center in the network, with more adjacent metabolites, has a greater ability to bear and transmit information and a faster information transmission speed. Therefore, the importance of metabolites cannot be measured simply by the k value. The results obtained by experiment are more accurate than those obtained by the k value In terms of topological properties, Akt-p is the most critical metabolite of the AKT network, and Akt is the second most important metabolite.

Akt-p and Akt are important metabolites in the AKT network and also verify some of the existing research results of the Akt signaling pathway. Huang et al. (2008) believed that Akt signaling pathway-related tumors, PI3K/Akt and its related genes can be used as targets for gene therapy. The reason why the importance of the results of PI3K metabolite centrality analysis is not reflected here may be that PI3K is located in the initial part of the Akt signaling network, activates Akt by generating the second messenger PIP3, and is at the edge of the network, so the betweenness centrality (BC=0.038113) and closeness centrality (CC=0.215613) are smaller, but the degree DC=3, k=2, and the importance is also relatively high.

The following will verify the results obtained by the centrality analysis by removing some metabolites and comparing with the complete network. The results above are verified by comparing the topological structures of removing important metabolites, random metabolites, a large number of common metabolites and the complete network. Referring to Kuang and Zhang (2011), the results obtained by three measures of metabolite analysis, link analysis and chain length analysis are selected for comparison. The important metabolites to be removed are 20 and 10 whose DC, BC, and CC values are in the top five. The two random metabolites obtained by the RANDBETWEEN function in MS Excel are: 28 and 35. In total of 17 (one-third of the total number of common metabolites) random common metabolites are removed including 15, 25, 32, 3, 29, 14, 39, 47, 28, 27, 43, 13, 4, 18, 11, 58, 17. The comparison results are shown in Table 2.

Table 2 Comparison of the topological structure of metabolic networks after removing important metabolites, random metabolites and a large number of common metabolites and that without removing metabolites.

		Removing two important metabolites S=57	Removing two random metabolites S=57	Removing 1/3 common metabolites S=42	Unremoved S=59
Metabolite analysis	Averaged degree	0.7719	1.9649	1.8095	2
	Isolated metabolites	24	1	3	0
Link analysis	Total number of links	22	56	38	59
	Maximum degree	6	29	26	31
	Link density	0.386	0.9825	0.9048	1
	Connectivity	0.0068	0.0172	0.0215	0.01
Chain length analysis	Maximum chain length	ID 19：4	ID 19：10	ID 5 and 12：7	ID 19：10

It can be found from Table 2 that, compared with the analysis results without removing important metabolites, the Akt metabolic network with two important metabolites, Akt-p and Akt removed:

(1) There are 24 isolated metabolites, which is much larger than the value of removing two random metabolites and 17 common metabolites. That is, after removing the important metabolites, Akt-p and Akt, a large number of metabolite connections in the network are broken, a large number of metabolites fail, and the network characteristics of the metabolic network no longer exist, and the network can only eventually collapse.

(2) The total number of links, the maximum degree, the connection density and the connectivity are all reduced, and the degree of reduction is greater than removing two random metabolites.

(3) The maximum chain length changes, but the maximum chain length of the network without random metabolites remains unchanged.

It can be seen that after removing the two important metabolites, Akt-p and Ak, the topology of the Akt metabolic network has changed greatly, which verifies the results obtained by the above centrality analysis. In

addition, after removing 17 common metabolites, although only 3 isolated metabolites appeared, the total number of links, the maximum degree, the link density and the connectivity all decreased, the maximum chain length also changed, and the network topology also changed greatly. It can be seen that in the presence of important metabolites and their interconnections, the removal of a large number of common metabolites will also lead to the collapse of the network.

In conclusion, Akt-p and Akt constitute the core skeleton of the AKT metabolic network, which is crucial for the maintenance of the AKT metabolic network.

4.1.2 JAK-STAT signaling pathway

Centrality values of JAK-STAT signaling pathway are calculated and listed in Table 3.

Table 3 DC, BC and CC of metabolites in JAK-STAT signaling pathway.

ID	Metabolite	DC	BC	CC
1	JAK1	3	0.000000	0.185841
2	JAK2	5	0.000000	0.254545
3	TYK2	2	0.000000	0.141414
4	JAKs	10	0.132985	0.308824
5	SHP1-STATIP	2	0.000000	0.247059
6	Growth Hormones Receptor	2	0.000000	0.235955
7	IFNyR1	2	0.000000	0.185022
8	IFNyR2	2	0.000000	0.214286
9	IFNAR1	2	0.000000	0.185022
10	IFNAR2	2	0.000000	0.141414
11	Cytokines Receptor	2	0.000000	0.238636
12	STAT5-P	3	0.015099	0.278146
13	STAT3-P	2	0.005226	0.276316
14	STAT1-P	7	0.005807	0.224599
15	STAT2-P	3	0.001161	0.163424
16	STATs-P	3	0.008130	0.276316
17	SH28	1	0.000000	0.237288
18	STAT-P	2	0.000000	0.274510
19	SHP2-SOS-GRB2	2	0.013937	0.176471
20	Ras	2	0.023810	0.166667
21	RTK	2	0.002904	0.205882
22	(STAT5-P)2	2	0.023229	0.304348
23	(STATs-P)2	9	0.147793	0.336000
24	(STAT3-P)2	2	0.013937	0.302158
25	(STAT1-P)2	1	0.000000	0.184211
26	IRF9	1	0.000000	0.160920
27	STAT1-P-STAT2-P-IRF9	3	0.000000	0.190909
28	(SUMO)3-(STATs)2-PIAS-Ubsc9	2	0.000000	0.254545
29	KPNA1-RAN	3	0.008130	0.262500
30	(STATs-P)2-Cofactors-CTFS-P	3	0.126887	0.269231

31	STAM-P	2	0.017131	0.244186
32	SOCS	3	0.150407	0.251497
33	PI3K	2	0.016260	0.240000
34	Akt	1	0.000000	0.194444
35	c-Myc	2	0.010163	0.206897
36	Gene Expression	5	0.157375	0.240000
37	Raf	2	0.032520	0.185022
38	MEK	2	0.040070	0.217617
39	ERKs	2	0.046458	0.264151
40	JAK-(Ub)3	3	0.016260	0.245614
41	Proteasome	1	0.000000	0.198113
42	GAS	1	0.000000	0.194444
43	ISRE	1	0.000000	0.194444

Table 3 demonstrates that some metabolites have a large degree, among which some metabolites have degrees 7, 9, and 10 respectively. Most of the metabolites have degrees between 1 and 5, accounting for 93% of all metabolites, indicating that a few metabolites have a high degree, and most metabolites have a low degree. Among them, the metabolites with degree 2 are the most, indicating that the topology of the JAK-STAT signaling pathway is sparse and there are many chain structures. This is consistent with the findings of Huang and Zhang (2012) that the JAK-STAT signal network is a scale-free network, and the degree distribution conforms to a power-law distribution, that is, a few metabolites in the network have high degrees, and most metabolites have low degrees (Huang and Zhang, 2012; Zhang and Li, 2016).

Metabolites with higher degrees tend to be critical components of the metabolic process. From the centrality values in Table 3, it can be known that the top five metabolite numbers with DC, BC, and CC values are 4, 23, 14, 2, 36; 36, 32, 23, 4, 30; 23, 4, 22, 24, 12. The metabolites with DC, BC, and CC values all ranking in the top five are 4 and 23. It is speculated that JAKs and (STATs-P)2 are important metabolites in the JAK-STAT network, indicating that JAKs and (STATs-P)2 play an important role in the JAK-STAT network. According to the results of Li and Zhang (2013), in the JAK-STAT metabolic pathway, only IFNyR2, IFNAR1, SH28, Ras, PI3K, Akt, (STAT3-P)2, ISRE, and Proteasome have a k value equal to 1, and the k values of other metabolites are all equal to 2. This shows that various metabolites are closely correlated in this pathway (Li and Zhang, 2013). The JAK-STAT network has many looped structures. JAKs and (STATs-P)2 are hub metabolites connecting sub-loops, so JAKs and (STATs-P)2 are very important in the network. This is consistent with previous research findings that the constitutive activation of the JAK-STAT signaling pathway, especially the abnormal activation of STAT3, is closely related to the occurrence, development, invasion and metastasis of liver cancer. Overexpression of SOCS protein inhibits the activation of JAKs and the activity of STATs, thereby inducing apoptosis. On the contrary, the deletion of SOCS protein can lead to the overexpression of STAT3, which leads to the occurrence of malignant tumors, such as liver cancer (Darnell , 2005; Croker et al., 2008; Morales et al., 2010).

The results above are verified by comparing the topological structures of removing important metabolites, random metabolites and the complete network. Referring to the Kuang and Zhang (2011), the results obtained from three measures of metabolite analysis, link analysis and chain length analysis are selected for comparison. The important metabolites to be removed are metabolites 4 and 23 in the top five with DC, BC, and CC values.

The two random metabolites obtained by the RANDBETWEEN function in MS Excel are metabolites 20 and 11. The comparison results are listed in Table 4.

Table 4 Comparison of topological structure of metabolic networks after removing important metabolites and random metabolites and that without removing metabolites.

		Removing two important metabolites *S*=41	Removing two random metabolites *S*=41	Unremoved *S*=43
Metabolite analysis	Averaged degree	1.9024	2.5853	2.6512
	Isolated metabolites	1	1	0
Link analysis	Total number of links	39	53	57
	Maximum degree	7	10	10
	Link density	0.9512	1.2927	1.3256
	Connectivity	0.0232	0.0315	0.0308
Chain length analysis	Maximum Chain length	ID 5、28：6	ID 2、6、37：9	ID 21：12

It can be found from Table 4 that compared with the results without removing important metabolites, the JAK-STAT metabolic network with two important metabolites, JAKs and (STATs-P)2 removed:

(1) The average degree is 1.9024, which is greatly changed compared with the average degree of the complete JAK-STAT metabolic network. Although the number of isolated metabolites that appear after removing important metabolites and random metabolites is 1, the metabolic network maintains relative integrity after removing random metabolites, and after removing important metabolites, the network is divided into four independent metabolite clusters, and the network characteristics of the metabolic network no longer exist, and the network can only eventually collapse.

(2) The total number of links, the maximum degree, the connection density and the connectivity are all reduced, and the degree of reduction is greater than removing two random metabolites.

(3) Maximum chain length: The maximum chain length has changed and cannot be compared effectively.

It can be seen that after removing the two important metabolites, JAKs and (STATs-P)2, the topology of the JAK-STAT metabolic network has changed greatly, which verifies the results obtained by the above centrality analysis.

4.1.3 P53 signaling pathway

Centrality values of P53 signaling pathway (Fig. 1) are calculated and listed in Table 5.

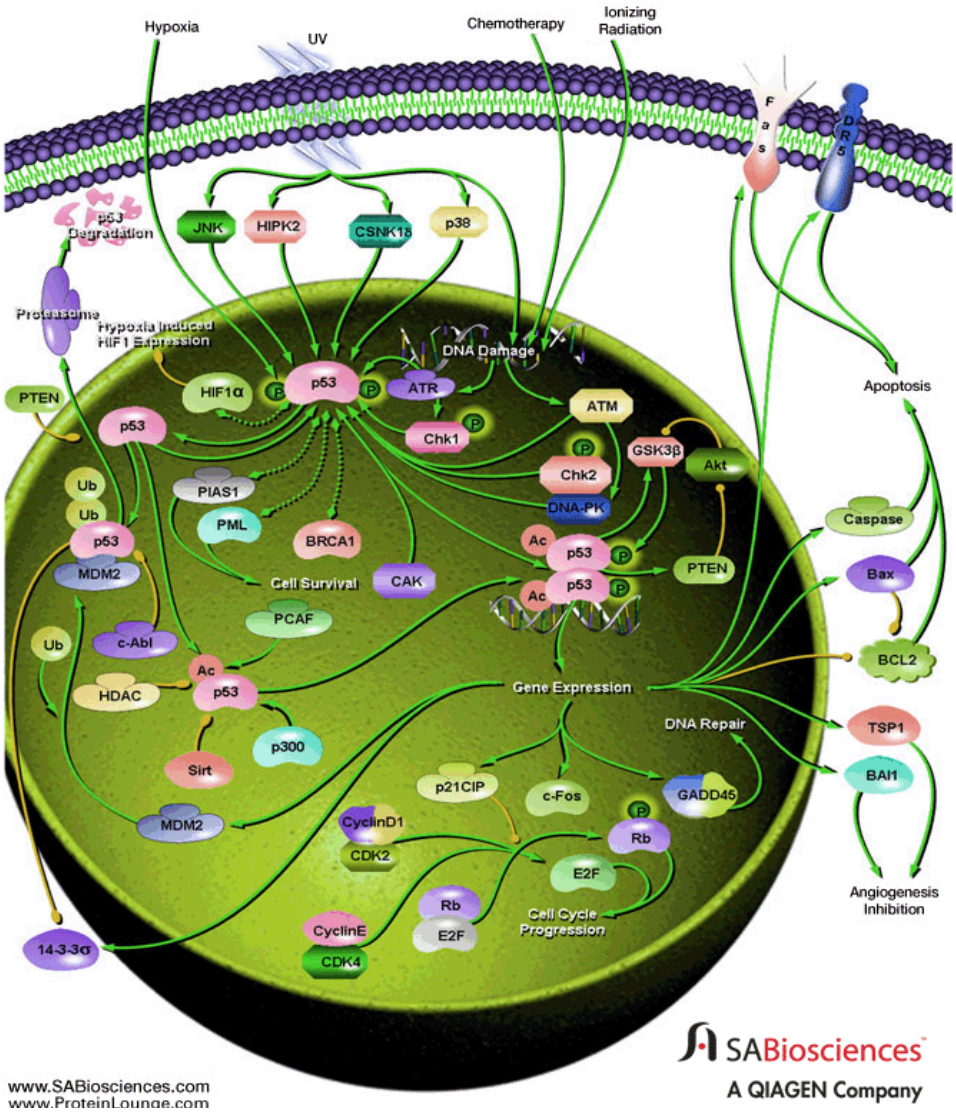

Fig. 1 p53 signaling pathway (Pathway Central, 2012).

Table 5 DC, BC and CC of metabolites in P53 signaling pathway.

ID	Metabolite	DC	BC	CC
1	Hypoxia	1	0.000000	0.294798
2	UV	5	0.000000	0.250000
3	Chemotherapy	1	0.000000	0.202381
4	Ionizing Radiation	1	0.000000	0.202381
5	JNK	2	0.002647	0.303571
6	HIPK2	2	0.002647	0.303571
7	CSNK1	2	0.002647	0.303571
8	p38	2	0.002647	0.303571
9	p53-P-P	22	0.197059	0.414634
10	HIFα	2	0.000000	0.294798

11	proceasome	2	0.000000	0.258883
12	PTEN	3	0.018039	0.320755
13	p53	4	0.023529	0.337748
14	Ub-Ub-p53-MDM2	6	0.011373	0.329032
15	Ub	1	0.000000	0.248780
16	c-Abl	1	0.000000	0.248780
17	HDAC	1	0.000000	0.251232
18	14-3-3θ	2	0.000000	0.301775
19	PIAS1	2	0.000000	0.294798
20	PML	2	0.000000	0.294798
21	BRCA1	2	0.000000	0.294798
22	CAK	1	0.000000	0.294798
23	PCAF	1	0.000000	0.251232
24	Ac-p53	6	0.034706	0.333333
25	Sirt	1	0.000000	0.251232
26	p300	1	0.000000	0.251232
27	MDM2	2	0.004314	0.301775
28	DNA damage	5	0.025882	0.252475
29	ATR	3	0.017451	0.309091
30	Chk1-P	2	0.000000	0.301775
31	ATM	3	0.017451	0.309091
32	Chk2-P	1	0.000000	0.294798
33	DNA-PK	2	0.000000	0.301775
34	(Ac-p53-P)2	6	0.196471	0.428571
35	GSK3β	3	0.013333	0.305389
36	Akt	2	0.006667	0.246377
37	p21CIP	2	0.012157	0.301775
38	c-Fos	1	0.000000	0.284916
39	GADD45	1	0.000000	0.284916
40	CyclinD1-CDK2	1	0.000000	0.166124
41	CyclinE-CDK4	1	0.000000	0.166124
42	Rb-E2F	1	0.000000	0.195402
43	E2F	3	0.000784	0.198444
44	Rb-P	3	0.000000	0.241706
45	Fas	1	0.000000	0.284916
46	DR5	1	0.000000	0.284916
47	Caspase	1	0.000000	0.284916
48	Bax	2	0.000000	0.286517
49	BCL2	2	0.000000	0.286517
50	TSP1	1	0.000000	0.284916
51	BAI1	1	0.000000	0.284916
52	Gene Expression	13	0.152157	0.395349

From Table 5 we can find that some metabolites have a large degree, among which two metabolites have degrees of 13 and 22, and most of the metabolites have degrees between 1 and 6, accounting for 96.2% of the total metabolites in p53 signaling pathway. It indicates that a few metabolites in the network have a high degree, and most metabolites have a low degree. This is consistent with the results of Huang and Zhang (2012), the network type of p53 is a scale-free complex network, and the degree distribution follows the power-law distribution (Huang and Zhang, 2012; Zhang and Li, 2016).

It can be seen from Table 5 that the top five metabolites with DC, BC and CC values are 9, 52, 14, 24, 34; 9, 34, 52, 24, 28; 34, 9, 52, 13, 24. The metabolites with DC, BC, and CC values all ranking in the top five are 9, 52, 24, and 34. It is speculated that metabolites 9, 52, 24 and 34 are important metabolites in the P53 signaling pathway, indicating that p53-P-P, Gene Expression, Ac-p53, and (Ac-p53-P)2 play important roles in the p53 metabolic network. According to the results of Li and Zhang (2013), The *k* value of UV, ATM, DNA damage, ART, JNK, Chk1-P, HIPK2, CSNK1, p38, PTEN, proceasome, MDM2, 14-3-3θ, DNA-PK, Akt, GSK3β, Bax, BCL2, Ub-Ub-p53-MDM2, Gene Expression, (Ac-p53-P)2, and p53-P-P are 2, and these metabolites are more important in the network, and the *k*=1 for remaining metabolites. Both Gene Expression, (Ac-p53-P)2, and p53-P-P are the crucial metabolites of the network using *k* value analysis and centrality analysis, but Li and Zhang (2013) believes that the *k* value of Ac-p53 is 1, so it doesn't seem to be important. The centrality analysis shows that Ac-p53 has larger DC, BC and CC values (all top five), and is closely related to (Ac-p53-P)2 and p53-P-P, so it can be considered that this metabolite is also very critical. This is also consistent with the results of Blattner (2002) that the level of p53 increases after being stimulated by external stress, and as a transcription factor, it is activated by phosphorylation (by generating (Ac-p53-P)2, p53-P-P, Ac-p53). The expression of downstream genes can prevent the malignant transformation of cells and the occurrence of tumors by initiating cell cycle arrest and apoptosis. After mutation of p53, the tumor suppressor function of wild-type p53 is lost, and the function similar to that of an oncogene is obtained. Therefore, the most studied gene therapy using p53 is wild-type p53 gene replacement therapy (Blattner et al., 2002)).

The results above are verified by comparing the topological structures of removing important metabolites, random metabolites, a large number of common metabolites and the complete network. The important metabolites to be removed are 9, 52, 24, and 34 whose DC, BC, and CC values are in the top five. The 4 random metabolites obtained by the RANDBETWEEN function in MS Excel are metabolites 45, 11, 29, and 33. The comparison results are listed in Table 6.

Table 6 Comparison of topological structure of metabolic networks after removing important metabolites and random metabolites and that without removing metabolites.

		Removing four important metabolites *S*=48	Removing four random metabolites *S*=48	Unremoved *S*=52
Metabolite analysis	Averaged degree	1.0833	2.5833	2.6923
	Isolated metabolites	18	0	0
Link analysis	Total number of links	26	62	70
	Maximum degree	6	20	22
	Link density	0.5417	1.2917	1.3462

	Connectivity	0.0113	0.0269	0.0259
Chain length analysis	Maximum chain length	ID 2、3、4：3	ID 3、4：7	ID 3、4：7

It can be seen from Table 6 that compared with the results without removing the important metabolites, the p53 metabolic network with the four important metabolites of p53-P-P, Gene Expression, Ac-p53 and (Ac-p53-P)2 removed:

(1) There are 18 isolated metabolites, and the average degree is 1.0833, which is more variable than the complete metabolic network. However, there is no isolation after removing random metabolites, and the average degree is 2.5833, which is not much different from the complete metabolic network. That is, after removing the important metabolites 9, 52, 24, and 34, a large number of metabolite connections in the network are broken, and a large number of metabolites fail, so the network characteristics of the metabolic network no longer exist, and the network can only eventually collapse.

(2) The total number of links, the maximum degree, the connection density and connectivity are all reduced, and the degree of reduction is greater than that of removing random metabolites.

(3) The maximum chain length varies greatly.

It can be seen that after the removal of the four important metabolites, p53-P-P, Gene Expression, Ac-p53, and (Ac-p53-P)2, the topology of the p53 metabolic network has undergone great changes, which verifies that the above centrality results obtained.

4.1.4 Ras signaling pathway

The centrality values of Ras signaling pathway are calculated and listed in Table 7.

Table 7 DC, BC and CC of metabolites in Ras signaling pathway.

ID	Metabolite	DC	BC	CC
1	Integrins	1	0.000000	0.196532
2	Rap1A-GTP	2	0.008913	0.242857
3	PLC-∑	2	0.016043	0.311927
4	Ras-GDP	5	0.055258	0.311927
5	Ras-GTP	12	0.000000	0.425000
6	GRB2	1	0.000000	0.239437
7	GAP	2	0.000000	0.306306
8	GEF	2	0.000000	0.306306
9	PMA	1	0.000000	0.239437
10	CD-GECII	3	0.053476	0.311927
11	TCR	1	0.000000	0.239437
12	Lck	1	0.000000	0.300885
13	RalGDS	2	0.040107	0.330097
14	Ral	3	0.035651	0.265625
15	PLD	1	0.000000	0.211180
16	RalBP1	2	0.019608	0.216561

17	CDC42	2	0.010695	0.180851
18	Actin Cytoskeleton	1	0.000000	0.153846
19	Raf-P	2	0.032086	0.323810
20	MEKs-P	2	0.026738	0.257576
21	ERKs-P	2	0.019608	0.211180
22	ERKs	2	0.010695	0.177083
23	Elk1	1	0.000000	0.151111
24	PI3K	2	0.024064	0.317757
25	Rac	4	0.037433	0.267717
26	PAKs	2	0.021390	0.272000
27	MEKK1	3	0.062389	0.343434
28	JNKK	2	0.053476	0.274194
29	JNK	3	0.042781	0.225166
30	c-Jun-c-Fun	1	0.000000	0.184783
31	ATF2	2	0.015152	0.186813
32	Gene Expression	1	0.000000	0.158140
33	p120-GAP	2	0.008021	0.311927
34	p190-B	2	0.008021	0.251852
35	Rho	1	0.000000	0.212500

Table 7 shows that three metabolite have degrees 4, 5, and 12 respectively. Most of the metabolites have degrees between 1 and 3, accounting for 91.4% of the total metabolites, among them the metabolites with degrees 1 and 2 are the most, indicating that the topology of the Ras signal metabolism network is sparse and there are many chain structures. The network type of Ras is a scale-free complex network, and the degree distribution conforms to the power-law distribution (Huang and Zhang, 2012).

It can be seen from Table 7 that the top five metabolites with DC, BC and CC values are 5, 4, 25, 27, 10; 27, 4, 10, 28, 29; 5, 27, 13, 19, 24. Among them, the DC, BC, and CC values are all ranked in the top five metabolites are 5, 4, and 27. From this, it can be speculated that 5, 4, and 27 are important metabolites in the Ras network, indicating that Ras-GTP, Ras-GDP, and MEKK1 play important roles in the Ras network. According to the results of Li and Zhang (2013), in the Ras metabolic pathway, only four metabolites, Actin Cytoskeleton, PMA, TCR, and Rho, have *k* values equal to 1, and the *k* values of other metabolites are equal to 2. The metabolism of 35 metabolites in the Ras metabolic pathway is closely related and the process is complex (Li and Zhang, 2013). The *k* value cannot further discriminate the relative importance of these 31 metabolites with *k* value of 2. By comparing the three centrality values, the three most important metabolites can be found. Regardless of the *k* value or the centrality analysis, it can be concluded that Ras-GTP, Ras-GDP and MEKK1 are the crucial metabolites in the Ras network. Because Ras-GTP is the pivot metabolite of many loop-forming structures and chain-like structures, Ras-GDP and MEKK1 are located at the overlapping points of loop-like and chain-like structures. Ras proteins include an active GTP-binding conformation and an inactive GDP-binding conformation, which can be interconverted under certain conditions to form the Ras cycle. About 30% of human tumors have point mutations in the Ras gene. The point-mutated Ras protein loses the GTPase activity, preventing the hydrolysis of the active form of the Ras protein by the GTPase-activating protein, resulting in the existence of the Ras protein in the active bound form (Bos, 1989).

The results above are verified by comparing the topological structures of removing important metabolites, random metabolites, a large number of common metabolites and the complete network. The important metabolites to be removed are Ras-GTP, Ras-GDP, and MEKK1 whose DC, BC, and CC values are ranked in the top five. The three random metabolites obtained by the RANDBETWEEN function in MS Excel are metabolites 12, 35, and 25. The comparison results are listed in Table 8.

Table 8 Comparison of topological structures of metabolic networks after removing important metabolites and random metabolites and that without removing metabolites.

		Removing three important metabolites *S*=32	Removing three random metabolites *S*=32	Unremoved *S*=35
Metabolite Analysis	Averaged degree	1.375	2.125	2.2286
	Isolated metabolites	4	0	0
Link analysis	Total number of links	22	34	39
	Maximum degree	3	11	12
	Link Density	0.6875	1.0625	1.1142
	Connectivity	0.0215	0.0332	0.0318
Chain length Analysis	Maximum chain length	ID 13、19：4	ID 6、8、9、11：7	ID 33：8

It can be found from Table 8 that compared with the results without removing important metabolites, the Ras metabolic network with three important metabolites, Ras-GTP, Ras-GDP and MEKK1 removed:

(1) The average degree is 1.375, which varies greatly compared with the complete metabolic network. Although only 4 isolated metabolites appeared, the metabolic network was broken into six independent modules (i.e., clusters), the overall characteristics of the original metabolic network no longer existed, and the network collapsed.

(2) The total number of links, the maximum degree, the connection density and the connectivity are all reduced, and the degree of reduction is greater than that of removing random metabolites.

(3) The maximum chain length varies greatly.

It can be seen that after removing the three important metabolites, Ras-GTP, Ras-GDP, and MEKK1, the topology of the Ras metabolic network has changed greatly, which verifies the results obtained by the above centrality analysis.

4.1.5 TNF signaling pathway

The centrality values of TNF signaling pathway are calculated and listed in Table 9.

Table 9 DC, BC and CC of metabolites in TNF signaling pathway.

ID	Metabolite	DC	BC	CC
1	FADD	1	0.000000	0.205479
2	RAIDD	1	0.000000	0.211268
3	RIP	2	0.000000	0.283019
4	TRADD	4	0.011494	0.315789
5	TRAF2	2	0.000000	0.277778
6	SODD	2	0.000000	0.250000
7	Caspase8	4	0.006897	0.256410
8	Caspase2	5	0.006897	0.265487
9	Caspase1	4	0.003448	0.260870
10	Caspase3	4	0.000000	0.283019
11	Caspase6	3	0.000000	0.227273
12	Caspase7	3	0.000000	0.227273
13	BID	2	0.004598	0.208333
14	Caspase9	4	0.006897	0.236220
15	tBID	1	0.000000	0.173410
16	Cytoc	1	0.000000	0.163934
17	CytoC- Caspase 9- APAF1	2	0.004598	0.194805
18	MEKIKs-P-NIK-P	6	0.036782	0.337079
19	ERKs-P	2	0.008621	0.291262
20	p38-P	2	0.008621	0.291262
21	IKKs-P	2	0.011494	0.277778
22	(NF-kB)-IkBs	2	0.008046	0.247934
23	EIk1	2	0.004023	0.260870
24	ATFs	2	0.004023	0.260870
25	NF-kB	2	0.002299	0.227273
26	Ceramides	1	0.000000	0.120000
27	TAK1	2	0.004598	0.135747
28	JNKK1-P	2	0.006897	0.154639
29	JNK1-P	2	0.006897	0.177515
30	(c-Jun)-(c-Fos)	2	0.004598	0.205479
31	Gene Expression	4	0.000000	0.240000

Table 9 indicates that two metabolites have degrees 5 and 6, and most metabolites have degrees between 1 and 4, which is 93.5% of the total metabolites, among them the metabolites with degree 2 are the most, indicating that the topology of the TNF signaling metabolism network is sparse and there are many chain structures. Network type of TNF is a scale-free random network, and the degree distribution does not conform to the binomial distribution and Poisson distribution (Huang and Zhang, 2012).

It can be seen from Table 9 that the top five metabolites with DC, BC and CC values are 18, 8, 4, 7, 10; 18, 4, 21, 19, 20; 18, 4, 19, 20, 10. Among them, the metabolites whose DC, BC, and CC values are ranked in the top five are metabolites 18 and 4. It can be speculated that MEKIKs-P-NIK-P and TRADD are important

metabolites in TNF network. According to the results of Li and Zhang (2013), there are Caspase3, Caspase2, Caspase6, Caspase7, Caspase1, and Caspase9 in the TNF metabolic pathway with the *k* value equal to 3, which are crucial parts of the network, and other metabolites have the *k* values equal to 2 or 1 (Li and Zhang, 2013). However, the centrality analysis showed that MEKIKs-P-NIK-P and TRADD are the important metabolites of TNF network. The large *k* values of Caspase3, Caspase2, Caspase6, Caspase7, Caspase1, and Caspase are mainly because these six metabolites are connected to each other into a small module, but this small module is not located in the center of the network, while MEKIKs-P-NIK-P is more important because it is the pivot metabolite for different modules. Topological property analysis showed that MEKIKs-P-NIK-P may play an important role in the mutual mediation and restriction of the three signaling pathways of JNK, NF-kB and Caspase.

The results above are verified by comparing the topological structures of removing important metabolites, random metabolites, a large number of common metabolites and the complete network. The important metabolites to be removed are 18 and 4 in the top five with DC, BC, and CC values. The two random metabolites obtained by the RANDBETWEEN function in MS Excel are metabolites 3 and 14. The comparison results are shown in Table 10.

Table 10 shows that compared with the results without removing important metabolites, the TNF metabolic network with two important metabolites MEKIKs-P-NIK-P and TRADD removed:

(1) The average degree is reduced. Although there are no isolated metabolites, the metabolic network is divided into three independent small clusters after removing important metabolites. The integrity of the metabolic network is destroyed, and the network characteristics no longer exist.

(2) The total number of links, the maximum degree, the connection density and the connectivity are all reduced, and the degree of reduction is greater than that of removing random metabolites.

(3) The maximum chain length of ID 26 is the same as that of the original network, but the chain lengths of other metabolites vary greatly.

It can be seen that after removing the two important metabolites, MEKIKs-P-NIK-P and TRADD, the topology of the TNF metabolic network changed greatly, which verified the results obtained by the above centrality analysis.

Table 10 Comparison of topological structure of metabolic networks after removing important metabolites and random metabolites and that without removing metabolites.

		Removing two important metabolites *S*=29	Removing two random metabolites *S*=29	Unremoved *S*=31
Metabolite analysis	Averaged degree	2.0689	2.2759	2.5161
	Isolated metabolites	0	0	0
Link analysis	Total number of links	30	33	39
	Maximum degree	5	5	6
	Link density	1.0345	1.2379	1.2581
	Connectivity	0.0357	0.0393	0.0406
Chain length analysis	Maximum chain length	ID 26：5	ID 6、26：5	ID 6、26：5

4.1.6 VEGF signaling-pathway

The centrality values of VEGF signaling pathway are calculated and listed in Table 11.

Table 11 DC, BC and CC of metabolites in VEGF signaling pathway.

ID	Metabolite	DC	BC	CC
1	VEGFR2	3	0.000000	0.237762
2	PIP3	3	0.016711	0.306306
3	PIP2	3	0.000000	0.280992
4	PI3K-P	3	0.008021	0.314815
5	Src	2	0.000000	0.269841
6	PLCy-P	2	0.000000	0.215190
7	IP3	3	0.006462	0.244604
8	DAG	3	0.010695	0.248175
9	GRB2-SHC-SOS	2	0.007130	0.219355
10	FAK-Paxillin	3	0.004456	0.259542
11	MKK3/6	2	0.003565	0.201183
12	Akt/PKB	4	0.019385	0.267717
13	P	2	0.000000	0.206061
14	BAD-P	3	0.003936	0.253731
15	Caspase9-P	3	0.003936	0.253731
16	eNOS-HSP90	2	0.005273	0.250000
17	p38	2	0.006239	0.203593
18	MAPKAPK2/3	2	0.007130	0.223684
19	HSP27	2	0.006239	0.259542
20	Focal Adhesion Turnover	2	0.004456	0.255639
21	Cell Migration	2	0.002674	0.283333
22	Ras	2	0.013369	0.220779
23	Raf1	3	0.037433	0.242857
24	MEK1/2	2	0.036542	0.246377
25	ERK1/2	3	0.033868	0.269841
26	PKC	2	0.016043	0.232877
27	Ca++	2	0.007353	0.232877
28	Prostagiandin Production	3	0.015374	0.263566
29	cPLA	2	0.014260	0.232877
30	Cell Suvival	3	0.004308	0.280992
31	NO production	2	0.001708	0.272000
32	Actin Reorganization	3	0.005348	0.317757
33	Gene Expression &Cell Proliferation	2	0.006239	0.288136
34	Vascular Cell Permeability	2	0.004679	0.283333
35	ANGIOGENESIS	6	0.000000	0.343434

Table 11 demonstrates that two metabolites have degrees 4 and 6, and most metabolites have degrees between 2 and 3, which is 94.3% of the total metabolites, among them the metabolites with degree 2 are the most, indicating that the topology of the VEGF signaling metabolism network is sparse and the chain structure is more. Network type of VEGF is a scale-free random network, and the degree distribution does not conform to the binomial distribution and Poisson distribution, but conforms to the power-law distribution (Huang and Zhang, 2012).

Table 11 shows that the top five metabolites with DC, BC and CC values are 35, 12, 2, 23, 32; 23, 24, 25, 12, 2; 35, 32, 4, 2, 33. Among them, the metabolite whose DC, BC, and CC values are all ranked in the top five is 2; the highest value of DC and CC is metabolite 35. It can be speculated that metabolites 2 and 35 are important metabolites in the VEGF network, indicating that PIP3 and ANGIOGENESIS (angiogenesis) may play an important role in the VEGF network. According to the results of Li and Zhang (2013), the *k* value of all metabolites in the VEGF metabolic pathway is equal to 4. These metabolites constitute a core network (Li and Zhang, 2013). The VEGF metabolic network is a loop-shaped closed structure, and the *k* values are all the same. Therefore, in terms of DC, BC, and CC values, PIP3 is the most critical metabolite in the VEGF network, and ANGIOGENESIS is the second key metabolite.

The results above are also verified by comparing the topological structures of removing important metabolites, random metabolites, a large number of common metabolites and the complete network. The important metabolites to be removed are the top 2 with DC, BC, and CC values, and the 35 with the largest DC and CC. The 2 random metabolites obtained by the RANDBETWEEN function in MS Excel are metabolites 11, 1. The comparison results are shown in Table 12.

It can be seen from Table 12 that compared with the results without removing important metabolites, the VEGF metabolic network with two important metabolites, ANGIOGENESIS and PIP3 removed:

(1) The average degree is reduced. Although there are no isolated metabolites, the metabolic network is divided into two independent small clusters after removing important metabolites. The integrity of the metabolic network is destroyed, and the network characteristics no longer exist.

(2) The total number of links, the maximum degree, the connection density and the connectivity are all reduced, and the degree of reduction is greater than that of removing random metabolites.

(3) The maximum chain length of metabolite 1 is the same as the original network, but the chain lengths of other metabolites vary greatly.

After removing the two important metabolites, PIP3 and ANGIOGENESIS, the topology of the VEGF metabolic network changed greatly, which verified the results of the above centrality analysis.

Table 12 Comparison of topological structure of metabolic networks after removing important metabolites and random metabolites and that without removing metabolites.

		Removing two important metabolites *S*=33	Removing two random metabolites *S*=33	Unremoved *S*=35
Metabolite analysis	Averaged degree	2.1818	2.4848	2.5714
	Isolated metabolites	0	0	0
Link analysis	Total number of links	36	41	45
	Maximum degree	3	6	6
	Link density	1.0909	1.2424	1.2857

	Connectivity	0.0331	0.0376	0.0367
Chain length analysis	Maximum chain length	ID 1：8	ID 8、9：7	ID 1：8

4.2 Cascade model analysis

From Table 13, we can find that among the property values predicted by the cascade model for Akt, p53, Ras, TNF and VEGF signaling pathways, there is one that is significantly different from the true value (error greater than 1), and the other six predicted values are not different from the true value. The seven values of JAK-STAT signaling pathway are not much different from the true values. Therefore, the cascade model can better predict the properties of the JAK-STAT signaling pathway, but cannot well predict the properties of Akt, p53, Ras, TNF, VEGF and other signaling pathways. Because the relative error is very large, the reason may be that the assumptions such as "elements in the matrix obey the 0-1 distribution of $p=c/S$" in the cascade model do not hold in these five tumor signaling pathways.

Table 13 Prediction of network properties of tumor signaling pathways using cascade model*.

Tumor signaling pathways	Proportion type	U	M	D	UM	UD	MM	MD
Akt	True	0.2203	0.2712	0.5085	0.2034	0.0170	0.3220	0.4576
	Predicted	0.3799	0.2402	0.3799	0.2716	0.2989	0.1579	0.2716
	Rel. error	0.1155	0.0035	0.0325	0.0229	4.6898	0.0836	0.0756
JAK-STAT	True	0.1860	0.7209	0.0903	0.2105	0.0526	0.6667	0.0702
	Predicted	0.3251	0.3498	0.3251	0.2810	0.2123	0.2257	0.2810
	Rel. error	0.1039	0.1910	0.5789	0.0236	0.4845	0.2916	0.6333
p53	True	0.2885	0.5000	0.2115	0.2571	0.0143	0.6715	0.0571
	Predicted	0.3223	0.3554	0.3223	0.2806	0.2078	0.2310	0.2806
	Rel. error	0.0040	0.0418	0.0580	0.0021	2.6200	0.2889	0.8740
Ras	True	0.1143	0.6857	0.2000	0.1026	0	0.7180	0.1795
	Predicted	0.3594	0.2812	0.3594	0.2775	0.2652	0.1798	0.2775
	Rel. error	0.5257	0.2386	0.1271	0.2985	*	0.4034	0.0535
TNF	True	0.2258	0.6129	0.1613	0.2820	0	0.4359	0.2821
	Predicted	0.3350	0.3300	0.3350	0.2813	0.2281	0.2093	0.2813
	Rel. error	0.0528	0.1305	0.1870	0	*	0.1178	0
VEGF	True	0.1429	0.8286	0.0285	0.2	0	0.6667	0.1333
	Predicted	0.3309	0.3382	0.3309	0.2813	0.2214	0.2160	0.2813
	Rel. error	0.2474	0.2901	3.1983	0.0330	*	0.3046	0.1641

*: U, M, D represent the ratio of upstream, midstream and downstream metabolites to the total metabolites in the signaling pathways, respectively; UM, UD, MM, MD represent the ratio of upstream-midstream, upstream-downstream, midstream-midstream, and midstream-downstream links, respectively to the total number of links in the signaling pathways, repectively. The calculation of the relative error is: $(\text{true} - \text{predicted})^2/\text{true}$, and the value * is infinity.

5 Conclusion

Akt, p53, Ras, TNF, JAK-STAT and VEGF are scale-free networks, and their degree distribution is the power-law distribution (Zhang and Li, 2016). It indicates that a few metabolites have a high degree, and most metabolites have a low degree. Higher metabolites with high degrees tend to be some critical metabolites in the metabolic process.

Centrality and core skeleton analysis show that the crucial metabolites of AKT signaling pathway are Akt-p and Akt; the crucial metabolites of JAK-STAT signaling pathway are JAKs and 23(STATs-P)2; the crucial metabolites of p53 signaling pathway are p53-P-P, Gene Expression, Ac-p53 and (Ac-p53-P)2; the crucial metabolites of Ras signaling pathway are Ras-GTP, Ras-GDP and MEKK1; the crucial metabolites of TNF signaling-pathway are MEKIKs-P-NIK- P and TRADD, and for VEGF signaling pathway, the crucial metabolites are PIP3 and ANGIOGENESIS. When these important metabolites are removed from the pathways, a large number of metabolite connections in the network are broken, that is, a large number of metabolites fail, the network characteristics of the tumor metabolism network no longer exist, and the network collapses. However, when a certain number of common metabolites in the network are randomly removed, such consequences will not occur. The tumor metabolism network still has a high level of integrity in structure. This is consistent with the existing research, that is, the scale-free network has an amazing ability to withstand random external attacks, and is prone to collapse when important metabolites are attacked. The study of the topological structure of these six tumor metabolic networks is helpful to find the crucial factors in the network as molecular targets for drug screening.

In terms of models, the cascade model is very inaccurate in simulating tumor signaling pathways such as Akt, p53, Ras, TNF, and VEGF, etc., because the relative error of the simulation is large, so we should build models with higher simulation accuracy. In this study, we have achieved a preliminary understanding of the network structure and related characteristics of six tumor signal pathways, which is helpful to discover and summarize the further correlations and rules.

Acknowledgment

We are thankful to the support of The National Key Research and Development Program of China (2017YFD0201204), and Discovery and Crucial Node Analysis of Important Biological and Social Networks (2015.6-2020.6), from Yangling Institute of Modern Agricultural Standardization. The two authors are all the first authors of the article.

References

Aittokallio T, Schwikowski B. 2006. Graph-based methods for analysing networks in cell biology. Briefings in Bioinformatics, 7: 243-255

Albert R, Jeong H, Barabasi AL. 2000. Error and attack: tolerance of complex network. Nature, 406: 378-382

Blattner C, Hey T, Meek DW, Lane DP. 2002. Hypophosphorylation of Mdm2 augments p53 stability. Molecular and Cell Biology, 22(17): 6170-6182

Borgatti SP, Everett MG, Freeman LC. 2011. Software for Social Network Analysis. 5-9, Analytic Technologies, USA

Bos JL. 1989. Ras oncogenes in human cancer a review. Cancer Research, 49(17): 4682-4689

Cohen JE, Newman CM. 1985. A stochastic-theory of community food web Models and aggregated data. Proceedings of the Royal Society of London B, Biological Science, 224: 421-448

Croker BA, Kiu H, Nicholson SE. 2008. SOCS regulation of the JAK/STAT signaling pathway. Seminars in

Cell and Developmental Biology, 19: 414-422

Darnell JE. 2005. Validating Stat3 in cancer therapy. Nature Medicine, 11: 595-596

Ho CC, Siu WY, Lau A, Chan WM, Arooz T, Poon RY. 2006. Stalled replication induces p53 accumulation through distinct mechanisms from DNA damage checkpoint pathways. Cancer Research, 66(4): 2233-2241

Huang JQ, Zhang WJ. 2012. Analysis on degree distribution of tumor signaling networks. Network Biology, 2(3): 95-109

Huang XL, Cui GH, Zhou KY. 2008. Recent progress in the relationship between PI3K-Akt signal pathway and cancer cell apoptosis. Cancer, 3: 331-336

Ibrahim SS, Eldeeb MAR, Rady MAH. 2011. The role of protein interaction domains in the human cancer network. Network Biology, 1(1): 59-71

Iqbal S, Ejaz H, Nawaz MS, et al. 2014. Meta-analysis of cancer transcriptomes: A new approach to uncover molecular pathological events in different cancer tissues. Network Biology, 4(1): 1-20

Jeong H, Mason SP, Barabasi AL, Oltvai ZN. 2001. Lethality and centrality in protein networks. Nature, 499: 41-42

Jiang LQ, Zhang WJ, Li X. 2015. Some topological properties of arthropod food webs in paddy fields of South China. Network Biology, 5(3): 95-112

Jiang LQ, Zhang WJ. 2015. Effects of parasitism on robustness of food webs. Selforganizology, 2(2): 21-34

Kolch W. 2002. Ras/Raf signalling and emerging pharmacotherapeutic targets. Expert Opinion on Pharmacotherapy, 3(6): 709-718

Katoh M. 2005. WNT/PCP signaling pathway and human cancer. Oncology Reports, 14(6): 1583-1588

Kuang WP, Zhang WJ. 2011. Some effects of parasitism on food web structure: a topological analysis. Network Biology, 1(3-4): 171-185

Li JR, Zhang WJ. 2013. Identification of crucial metabolites/reactions in tumor signaling networks. Network Biology, 3(4): 121-132

Marrero MB. 2005. Introduction to JAK/STAT signaling and the vasculature. Vascular Pharmacology, 43(5): 307-309

Morales JK, Falanga YT, Depcrynski A, Fernando J, Ryan JJ. 2010. Mast cell homeostasis and the JAK-STAT pathway. Genes and Immunity, 11: 599-608

Moustakas A, Pardali K, Gaal A, Heldin CH. 2002. Mechanisms of TGF-beta signaling in regulation of cell growth and differentiation. Immunology Letters, 82(1-2): 85-91

Pathway Central. 2012. http://www.sabiosciences.com/pathwaycentral.php. SABiosciences, QIAGEN, USA

Pujol A, Mosca R, Farres J, et al. 2010. Unveiling the role of network and systems biology in drug discovery. Trends in Pharmacological Sciences, 31: 115-123

Rahman KMT, Md. Islam F, Banik RS, et al. 2013. Changes in protein interaction networks between normal and cancer conditions: Total chaos or ordered disorder? Network Biology, 3(1): 15-28

Shams B, Khansari M. 2014. Using network properties to evaluate targeted immunization algorithms. Network Biology, 4(3): 74-94

Stauffer F, Holzer P, Garcia-Echeverria C. 2005. Blocking the PI3K/PKB pathway in tumor cells. Current Medicinal Chemistry - Anti-Cancer Agents, 5(5): 449-462

Wasserman S, Faust K. 1994. Social Network Analysis: Methods and Applications. Cambridge University Press, USA

Xin SH, Zhang WJ. 2020. Construction and analysis of the protein-protein interaction network for the olfactory system of the silkworm *Bombyx mori*. Archives of Insect Biochemistry and Physiology, 105(3): e21737

Xin SH, Zhang WJ. 2021. Construction and analysis of the protein-protein interaction network for the detoxification enzymes of the silkworm, *Bombyx mori*. Archives of Insect Biochemistry and Physiology, 108(4): e21850

Yang S, Zhang WJ. 2022. Systematic analysis of olfactory protein-protein interactions network of fruitfly, *Drosophila melanogaster*. Archives of Insect Biochemistry and Physiology, 110(2): e21882

Zhang GL, Zhang WJ. 2019. Protein-protein interaction network analysis of insecticide resistance molecular mechanism in *Drosophila melanogaster*. Archives of Insect Biochemistry and Physiology, 100(1): e21523

Zhang WJ. 2012a. Computational Ecology: Graphs, Networks and Agent-based Modeling. World Scientific, Singapore

Zhang WJ. 2012b. Several mathematical methods for identifying crucial nodes in networks. Network Biology, 2(4): 121-126

Zhang WJ. 2016a. Screening node attributes that significantly influence node centrality in the network. Selforganizology, 3(3): 75-86

Zhang WJ. 2016b. Selforganizology: The Science of Self-Organization. World Scientific, Singapore

Zhang WJ. 2016c. A method for identifying hierarchical sub-networks / modules and weighting network links based on their similarity in sub-network / module affiliation. Network Pharmacology, 1(2): 54-65

Zhang WJ. 2016d. Network robustness: Implication, formulization and exploitation. Network Biology, 6(4): 75-85

Zhang WJ, Li X. 2016. Generate networks with power-law and exponential-law distributed degrees: with applications in link prediction of tumor pathways. Network Pharmacology, 1(1): 15-35

Zhang WJ. 2018. Fundamentals of Network Biology. World Scientific Europe, London, UK

Zhang WJ. 2021. Construction and analysis of the word network based on the Random Reading Frame (RRF) method. Network Biology, 11(3): 154-193

Zhang WJ, Feng YT. 2017. Metabolic pathway of non-alcoholic fatty liver disease: Network properties and robustness. Network Pharmacology, 2(1): 1-12

Zhang WJ, Jiang LQ, Chen WJ. 2014. Effect of parasitism on food webs: Topological analysis and goodness test of cascade model. Network Biology, 4(4): 170-178

Zhang WJ, Zhan CY. 2011. An algorithm for calculation of degree distribution and detection of network type: with application in food webs. Network Biology, 1(3-4): 159-170