

Analysis of Co-Expression Module of Liver Metastasis Genes in Colorectal Cancer and Mining of Potential High-Expression Biomarkers

Xuehu Wang (✉ wangxuehu_tougao@163.com)

Hebei University <https://orcid.org/0000-0002-5845-7623>

Nie Li

Hebei University

Yun Wang

Hebei University

Xiaoping Yin

Hebei University

Yongchang Zheng

Chinese Academy of Medical Sciences and Peking Union Medical College

Research Article

Keywords: Colorectal cancer liver metastasis, Potential biomarker, Weighted gene co-expression network analysis, Gene expression analysis

Posted Date: May 19th, 2021

DOI: <https://doi.org/10.21203/rs.3.rs-516058/v1>

License:  This work is licensed under a Creative Commons Attribution 4.0 International License.

[Read Full License](#)

Abstract

Aims

The Hub genes highly related to the disease were found from the gene co-expression module, and the potential high expression genes were analyzed to predict the liver metastasis of colorectal cancer, so as to provide reference for subsequent targeted therapy.

Methods

In this study, we used the public data set of GEO database (GSE50760) to analyze the gene co-expression of liver metastasis of colon cancer, primary colon cancer and normal colon tissue (54 cases) and 50 cases of clinical cases. The functional annotations based on GO database are enriched, and the functional annotations of five gene modules are obtained through the enrichment of biological processes. Then the data mining is carried out to find the sub-networks with high adjacency in the gene co-expression network. At the same time, these sub-networks are annotated to find oncogenes related to liver metastasis of colorectal cancer.

Results

This experiment found that KRAS, APC, FBXW7, PIK3CA, TP53 were highly correlated with liver metastasis of colorectal cancer. Finally, two protein genes STAT1 and MAPK1 were found by MCODE, which may be highly correlated with liver metastasis of colorectal cancer. Two new genes with high expression proteins found in this experiment have potential cancer, which has not been reflected in previous studies.

Conclusion

According to clinical data, KRAS, APC, FBXW7, PIK3CA, TP53 are related to colorectal cancer liver metastasis, and the analysis of the data set shows that STAT1 and MAPK1 are not only related to colorectal cancer liver metastasis oncogene but also related to clinically obtained genes.

Introduction

Colorectal liver metastasis (CRLM) is one of the most common cancers in gastrointestinal diseases. According to global cancer data in 2018, colorectal cancer ranks third among the most easily diagnosed cancers, but it is one of the two leading causes of cancer-related death [1]. According to relevant data, more than 2.2 million new cases of colon cancer are expected to occur globally by 2030, including more than 1.1 million deaths [2,3]. According to some authoritative medical data, liver is the target organ that is most likely to produce blood metastasis in colorectal cancer [4,5]. Therefore, the treatment of liver metastasis in colorectal cancer is one of the key and difficult points in current medical technology. 16%~26% of colorectal cancer patients had liver metastases at the time of diagnosis, and 14%~24% of patients had liver metastases after primary radical resection [6-9]. Most patients with liver metastases

were unable to undergo radical resection. For patients with liver metastasis who have never been treated, their survival time is up to nine months, but for those who cannot be resected, their survival probability in five years is almost zero [10,11]. However, for patients with liver metastasis who can be completely resected [12,13], the survival probability in five years is about 30%-50% [14-16]. Therefore, in the study, the gene sequences of colorectal cancer can be detected by combining genomics, and the genes only related to liver metastasis of colorectal cancer can be found, which can reflect the state of the tumor based on molecular or process changes. This can be used to help the classification, screening, diagnosis, treatment and prognosis of cancer, so as to provide more convenient pathological analysis for doctors. The results of gene detection can be used to diagnose whether patients may suffer from liver metastasis of colorectal cancer, and thus provide certain treatment options for subsequent targeted therapy [17].

WGCNA (weighted gene co-expression network analysis) is a systematic biology method that describes the correlation patterns between genes in sequencing samples. It can construct weighted gene co-expression network based on gene expression and further divide it into co-expression modules [18]. Gene expression in the same co-expression module is similar and may have similar functions. Module carrier is the first major component of co-expression module, reflecting the overall expression of genes in co-expression module. This method can reveal the interaction mechanism of colorectal cancer related genes and identify potential biomarkers. WGCNA has been used to identify key genes or therapeutic targets in many areas, including mice, human brains and various cancers. Recently, there have been some studies on the identification of colorectal cancer co-expression gene modules and key genes by WGCNA [19].

The existing research on liver metastasis of colorectal cancer has the following progress. Mathews et al. conducted a multicenter retrospective analysis of 165 patients including non-liver metastasis, liver metastasis and simultaneous liver metastasis within one year. The average gray intensity, texture entropy and homogeneity (0.5-2.5) of the whole liver were analyzed, multivariate analysis was used to predict liver metastasis [20]. The results showed that there was only uniformity, so there was no texture for predicting early liver metastasis [21]. Acharya et al. studied the texture characteristics of liver CT in patients with liver metastasis of colorectal cancer. Through experimental analysis, it was found that the entropy values of liver metastasis group and normal colon tissue were different, and they were also different from the uniformity of extrahepatic diseases [22,23]. Yasuhiro Shoji MD et al. used immunohistochemical analysis to show that ZFP57 was overexpressed in breast cancer, gastric cancer, colorectal cancer and other cancers [24]. The experimental results showed that ZFP57 gene played a certain role in blood-borne liver metastasis of colorectal cancer [25]. Zhao H et al. found that exosomes from CD133 + cells of CircABCC1 can cause liver metastasis of colorectal cancer by quantitative reverse transcriptase reaction analysis [26]. ZPoon et al. developed a prediction model based on the retrospective study of colorectal cancer, lung cancer, ovarian cancer [27] and other cancers in the EHR system [28]. The experimental results can be used to predict disease risk of neutropenia in cancer patients. At the same time, other external data are used for experimental verification [29], providing reliable basis for patients' chemotherapy regimens [30].

At present, the clinical methods for the diagnosis of liver metastasis of colorectal cancer mainly include abdominal enhanced CT, serum CEA, CA19-9 examination, pathological staging [19,31,32], liver MRI examination and other equipment examination. Because there are certain differences in the genes of each person, doctors can analyze the gene expression of patients. If some genes have high expression similarities with oncogenes, the patient's condition can be preliminarily predicted, and it can provide reference for subsequent targeted therapy. Therefore, this paper proposes the method of gene expression analysis in the diagnosis of liver metastasis, mainly from the aspects of gene molecules and signaling pathways to analyze the interaction between gene proteins for the preliminary diagnosis of colorectal cancer.

The experiment correlated three types of genes, primary, liver metastasis, and normal colorectal cancer genes. Finding out genes that are only related to liver metastasis of colorectal cancer makes the research purpose more clear. After the genes are obtained, GO function enrichment analysis can know the molecular process of each gene module and what biological significance it has. It can be seen from the experimental results that the changes in the gene expression profile are consistent with the pathological changes, in the order of normal colon tissue->primary colon cancer->colon cancer liver metastasis. The discovery of gene expression profiling has a profound impact on pathological analysis and targeted therapy research. The results of these experiments can provide an effective theoretical basis for the study of the molecular mechanism of potentially highly expressed genes in liver metastasis of colorectal cancer.

Materials And Methods

Data

The public dataset of GEO database (GSE50760) was used to obtain the RNA-seq data of 18 groups of colon cancer liver metastasis, primary colon cancer and normal colon tissue samples and 50 clinical samples collected from the hospital. A total of 104 samples and 23505 genes were obtained. In order to reduce the computational complexity and improve the reliability of the results, the genes with the median absolute deviation of the first 75% of the standardized expression were selected, and the WGCNA package was used for gene co-expression clustering module analysis. The similarity of expression patterns of 17628 genes in 18 groups of colon cancer liver metastasis, primary colon cancer, normal colon tissue and hospital clinical samples (a total of 104 cases) was screened out, and 44 gene modules were obtained by clustering. Most of the genes were found in blue and blue-green modules.

Construction of gene co-expression network

WGCNA is a method of studying gene set expression. Use WGCNA to construct a network in which genes are regarded as a point and the relationship between genes is regarded as a line [18]. By adding the Pearson coefficient to calculate the correlation based on gene expression, and then weighting to make the entire network close to the scale-free network distribution. The dynamic branch cutting method is used to divide the entire network into multiple cooperative expression modules [19]. The WGCNA package is used

to perform network construction at each stage of the above acquisition. The network construction steps mainly include correlation matrix calculation, soft threshold selection, adjacency matrix calculation, topology matrix calculation, dynamic branch cutting and module merging, and trait association analysis [18]. The adjacency matrix needs to be determined when constructing the network. is a symmetric $n \times n$ matrix with a value range of $[0,1]$. Its component represents the strength of the network connection at nodes i and j . To better compute the adjacency matrix, an intermediate variable (co-expression similarity) is introduced to represent the absolute value of the correlation coefficient between nodes i and j [18].

$$s_{ij} = |COR(x_i, x_j)| \quad (1)$$

When $i \neq j$ represent two different gene modules, the weighted co-expression network can also be characterized by improving the similarity between co-expression and power, as showed in Eq. (2)

$$a_{ij} = s_{ij}^\beta \quad (2)$$

Functional enrichment analysis of genes

Gene function enrichment analysis is to compare genes or genomes with function databases for overexpression analysis and function annotation. The commonly used databases mainly include the GO database and KEGG database [33]. The KEGG database contains information on various pathways of genes. The GO database mainly includes the functions of genes on three levels: BP (biological process), MF (molecular function) and CC (cell component) [34]. In this paper, the GO database is used to analyze the functional enrichment. Since we want to know whether the genes studied before have biological significance, the functional enrichment is used to analyze the BP (biological process) function in GO, and the R language is used. The ClusterProfiler package is used to analyze the paths and functions involved in the co-expression module.

Find and annotate adjacent subnets with high correlation

Core genes are genes that play an important role in the network and can represent the genetic characteristics of the modules to a certain extent. In the previous analysis, the purpose was to identify the modules that are significantly related to the clinical features to be studied. Since a characteristic gene (intrinsic gene) has been established for each module, it is only necessary to associate the characteristic gene with external characteristics, and then find the most important association, which is actually to calculate the correlation between the ME value of the module and the phenotype coefficient. GS and MM are used to screen core genes. GS describes the correlation between genes and clinical traits, reflecting the relationship between genes and traits [19]. MM describes the correlation between the gene and the module carrier, reflecting the core position of the gene in the module. Therefore, the GS and MM values of the genes (also expressed as K values) were calculated, and the condition for screening core genes was that $|GS| > 0.1$ and $|MM| > 0.8$ were simultaneously satisfied.

$$GS_i = -\log P_i \quad (3)$$

$$K_{cor,i}^{(q)} := cor(x_i, E^{(q)}) \quad (4)$$

P_i denotes the difference between genes, and $E(q)$ denotes the module characteristic gene of module q .

For trait association analysis, it is necessary to quantify the similarity between all genes on the array and each module to find important modules. Five gene modules are only related to colorectal cancer liver metastasis, and all genes are aggregated to obtain 2197 genes. Then, the sub-networks with high adjacency in the gene co-expression network are excavated. At the same time, the genes of the corresponding modules are more closely linked in the network by functional annotation of these sub-networks.

Tool

The data is collated and analyzed based on the R language (<https://www.r-project.org/>).

Results

Construction of weighted gene co-expression network

The weighted gene co-expression network analysis was performed on the clinical sample genes of liver metastasis patients obtained from GEO public data sets and hospitals to show the construction process of the network. First of all, the data needs to be processed simply. In order to make the calculation results more accurate, 17628 genes were obtained by selecting the genes with 75% of the median absolute deviation of the standardized expression, and the deletion values were detected. Second, select the threshold. The higher the square of the correlation coefficient, the closer the network is to the distribution without network scale. When , the degree of fit is the highest, so the corresponding soft threshold is selected. It can also be seen that the average connectivity of the network increases with the decrease of the soft threshold, and the soft threshold is selected to be as small as possible, and the final soft threshold is 5 (as determined in Fig. 1).

Thirdly, the weighted gene co-expression network is constructed by determining the soft threshold, and the co-expression modules are divided by dynamic cutting and module merging, as shown in Fig. 2. Finally, 44 gene modules are obtained. The upper part of the figure is the clustering tree of genes, and the following is the module based on similarity clustering. It can be seen that many genes exist in many blue and midnight blue modules.

In order to make the research results more accurate, the intrinsic genes were extracted from each gene module, used to calculate the adjacency between different gene modules, and the heat map was drawn to get Fig. 3. It can be seen from the figure that the 44 gene modules are divided into two parts, namely, the upper left corner and the lower right corner. Their internal adjacency is high, and there are few unrelated parts between gene modules, indicating that the correlation between gene modules is relatively good,

which is of research significance. According to the Eq.(1) and Eq.(3) of co-expression similarity, it can be seen that the greater the cor value of the gene module and the gene module and other gene modules, the two modules have better research significance, that is, the two modules are strongly related. The i and j in the formula respectively represent the module gene. When $i=j$, it represents the similarity between the gene modules. When $i \neq j$, it represents the similarity between two different gene modules.

Fourthly, the association between 44 gene modules and clinical traits was calculated. Since the public dataset and a small amount of clinical data were currently used, the total data were divided into three parts (liver metastasis of colon cancer, primary colon cancer, and normal colon tissue). The phenotypic data association diagram analysis was performed on the data of these three parts. The left column is colon cancer liver metastasis, the middle column is primary colon cancer, and the right column is normal colon tissue.

It can be seen from Fig. 4 that the positive correlation module of normal colon tissue is negatively correlated with the other two situations. There are many positive correlation common gene modules between liver metastasis and primary, and there are some positive correlation gene modules only related to liver metastasis. On the basis of this result, the genes related to colorectal cancer liver metastasis were selected for subsequent analysis.

GO functional enrichment analysis

Five gene modules that were only positively correlated with liver metastasis of colon cancer were selected and enriched based on the functional annotation of GO database as shown in Figure 5. In this figure, the abscissa represents the gene ratio, and the ordinate represents the biological process enriched by this type of gene. The larger the dot in the figure, the greater the number of genes enriched in the process. Specific enrichment results: In Fig. 5-(a), the genes of yellow-green module are mainly enriched in the regulation of lymphocyte activation, the regulation of cytokine production, the negative regulation of immune system process, T cell activation, neutrophil activation involved in immune response, neutrophil activation, regulation of cell activation, regulation of leukocyte activation and so on. In Fig. 5-(b), genes of brown modules are mainly enriched in small molecule catabolism, organic acid catabolism, carboxylic acid catabolism, cofactor catabolism, fatty acid catabolism, organic acid biosynthesis, carboxylic acid biosynthesis, steroid catabolism and so on. In Fig. 5-(c), the genes of orange4 modules are mainly enriched in transmembrane receptor protein serine and threonine kinase signaling pathways, ossification, negative regulation of transmembrane receptor protein serine / threonine kinase signaling pathway, negative regulation of cell growth factor stimulation, transforming growth factor beta receptor signaling pathway, ossification regulation and other biological processes. In Fig. 5-(d), the genes of the floral white module are mainly enriched in the main biological processes such as the regulation of actin filament tissue, extracellular structure tissue, regulation of tissue remodeling, renal epithelial development, renal tubular development and so on. In Fig. 5-(e), the genes of midnight blue module are mainly enriched in biological processes such as RNA splicing, ncRNA treatment, neurotransmitter transport, antigen processing through MHC class 1, presentation of exogenous peptides, and ectodermal cell differentiation.

High Adjacent Subnetwork Data Mining

Network diagram is a commonly used visualization method in bioinformatics to show the correlation direction and correlation degree between different nodes. In enrichment analysis, network maps are often used to express the relationship between the function and the genes estimated to function [35]. For protein-protein interaction network, network subgraphs are often used to express the intensity of association, interaction and type of interaction between encoded genes. According to the above analysis, we can judge the importance of the gene node in the whole network by the number of connections or the strength of connections between one node and other related nodes [36], and then dig out the potential high-expression genes through network.

The 2197 genes in all the gene modules that were only positively correlated with liver metastasis of colon cancer were summarized. Since the correlation between genes in the fourth module was relatively low, it was impossible to make a high adjacency network, so the remaining 697 genes were selected. MCODE is used to mine high-adjacent sub-networks in gene co-expression networks (**Fig. 6**), and functional annotation of these sub-networks (**Table 1**) Only high-adjacent sub-networks are found in this paper. Because the high-adjacent value of MCODE5 is unstable, MCODE5 is not analyzed. The corresponding relationship between different color blocks and subnet numbers is shown in the lower right corner. The functional annotations of these sub-networks include class A/1 (rhodopsin-like receptors), peptide ligand binding receptors, G α signaling results, positive regulation of nitric oxide synthase biosynthesis, and DTX3L-PARP9-STAT1 complex, Cell response to gamma interferon, cell adhesion molecules, immune response activation signal transduction, immune response regulation signal pathway, activation of immune response, CD28 family co-stimulation, antigen processing and foreign peptide antigen MHC class I presentation, GRB2: SOS provides connections with integrin MAPK1 signaling and other metabolic processes. Among them, STAT1 and MAPK1 are highly related to cancer, and the others are related to metabolism and the micro-environment. STAT1, as an important protein connecting the signal transduction between cell membrane receptor and effector, plays a very important role. Studies have shown that this gene can play a role mainly by promoting cell apoptosis [37], negatively regulating cell cycle, inhibiting normal cell proliferation, inhibiting tumor angiogenesis, and weakening tumor migration and invasion [38].

As an important transmitter from the cell surface to the nucleus, MAPK1 gene plays an important role. It can not only regulate cell differentiation and growth, but also adapt to environmental stress, inflammation and other major cellular physiological processes. According to the mechanism of the signaling pathway, it can be understood that any protein functional problems in the signaling pathway can lead to serious diseases, and these diseases are generally associated with tumors [39].

From the previous analysis of the construction of gene co-expression network, we can find the gene modules related to liver metastasis, and then through experiments, we can get the genes that are only related to liver metastasis. At the same time, the biological function of these genes can be analyzed to know the timing of these genes. It has biological significance, so our research is meaningful. Through

gene co-expression analysis and mining of potential high-expression genes, we can perform auxiliary diagnosis and prognosis evaluation of liver metastasis of colorectal cancer, and formulate corresponding targeted treatment schemes for patients.

Table 1 Function enrichment of sub-networks identified by MCODE software

Sub-network number	Function Note Number	Description of function annotation	Log10(P)
MCODE_1	R-HSA-373076	Class A/1 (Rhodopsin-like receptors)	-27.0
MCODE_1	R-HSA-375276	Peptide ligand-binding receptors	-26.9
MCODE_1	R-HSA-418594	G alpha (i) signalling events	-25.8
MCODE_2	GO:0051770	positive regulation of nitric-oxide synthase biosynthetic process	-10.5
MCODE_2	CORUM:7385	DTX3L-PARP9-STAT1 complex	-10.2
MCODE_2	GO:0071346	cellular response to interferon-gamma	-10.2
MCODE_3	Ko05323	Rheumatoid arthritis	-9.6
MCODE_3	Hsa05323	Rheumatoid arthritis	-9.4
MCODE_3	ko04514	Cell adhesion molecules (CAMs)	-8.6
MCODE_4	GO:0002757	immune response-activating signal transduction	-3.6
MCODE_4	GO:0002764	immune response-regulating signaling pathway	-3.5
MCODE_4	GO:0002253	activation of immune response	-3.5
MCODE_5	R-HSA-389948	PD-1 signaling	-15.6
MCODE_5	R-HSA-388841	Costimulation by the CD28 family	-13.1
MCODE_5	Hsa04514	Cell adhesion molecules	-11.4
MCODE_7	GO:0002479	HALLMARK COAGULATION	-7.7
MCODE_7	GO:0042590	Complement cascade	-7.6
MCODE_7	R-HSA-1236974	ER-Phagosome pathway	-7.6

At the same time, we found that KRAS, APC, FBXW7, PIK3CA, TP53 were related to colorectal cancer liver metastasis from the clinical cases of our hospital. So we used these five genes and STAT1 and MAPK1 to do the adjacent self-network, and found that there is a correlation between each other, which has biological significance (**Fig. 7**).

For the genes associated with colorectal cancer liver metastasis (KRAS, NRAS, BRAF, PIK3CA) [40], a comparison of related genes in this experimental data set can be found (**Fig. 8**). The abscissa of each

subgraph represents eighteen samples, the ordinate represents the gene value, the blue circle represents the data of colorectal cancer liver metastasis, the red box represents the primary liver metastasis, and the green triangle represents the gene value of normal colon tissue. Among the four discovered gene proteins, the changes in transcriptome data of colorectal liver metastasis relative to primary and normal colon tissues are not obvious. The transcriptome data of MAPK1 and STAT1 changed significantly compared with the genes related to liver metastasis of colorectal cancer. Therefore, the MAPK1 and STAT1 obtained in this study have good research significance.

Conclusion

This paper proposes a new diagnosis of liver metastasis of colorectal cancer. Through the co-expression analysis of 104 cases in the public data set, 44 gene modules were first divided, and the threshold of the total gene co-expression network was determined to find correlation between the five gene modules with potential high expression. Secondly, the correlation analysis of three colon tissue genes in the data set was carried out to find the genes that were only positively correlated with colorectal liver metastasis, and the biological molecular significance was studied. The high correlation genes were mined by high neighbor network. STAT1 and MAPK1 were highly correlated with cancer, and the gene was correlated with the five cancer genes obtained by clinical analysis.

The expression analysis of colorectal cancer liver metastasis genes and identification of potential high expression biomarkers are currently not used in clinical. This technology can provide a good auxiliary diagnosis and prognosis evaluation for liver metastasis. According to the gene sequencing results of patients, the disease can be roughly diagnosed, and the diagnostic efficiency is improved.

Since the number of samples used in this paper is relatively small, there are still some shortcomings. In the next step, we will expand the data set, increase the number of samples, and improve the universality of the experiment. We can also combine machine learning, image processing and other analytical methods, and combine the characteristics of genomics and genomics to explore the relationship between imaging markers and molecular markers, and integrate genomics information into imaging methods to better promote the development and application of individualized medicine.

Declarations

Conflict of interest

The authors declare they have no conflicting financial interests.

Funding information

This work is supported in part by: Natural Science Foundation of Hebei Province, General Project, Grant/Award Number H2020201021; China Postdoctoral Fund, Grant/Award Number: 2018M631755; Hebei Province High-level Talent Funding Project-Post-doctoral Research Projects 5. Selective Funding,

Grant/Award Number: B2018003002; Hebei University improve comprehensive strength special funds in the Midwest, Grant/Award Number: 801260201011; Natural Science Foundation of China, Grant/Award Number: 61401308 and 61572063; Natural Science Foundation of Gansu Province, Grant/Award Number :18JR3RA029; Hebei provincial government funded Provincial Medical Talents Project in 2020 ,CT radiomics study on the correlation between colorectal cancer liver metastasis and microsatellite instability, 361007;Medical discipline cultivation project of Hebei University in 2020,CT radiomics study on the correlation between colorectal cancer liver metastasis and microsatellite instability,2020B05.The corresponding author of this paper is Xuehu Wang.

Author contributions

The authors are grateful to the editors and the reviewers for their valuable comments and suggestions. The experimental processing and writing of this article were completed by Nie Li; Xuehu Wang, Yun Wang, and Xiaoping Yin provided academic guidance; Yongchang Zheng provided data collection and medical knowledge. All authors read and approved the final manuscript.

References

1. Freddie, Bray, Jacques, Ferlay, Isabelle, Soerjomataram et al. Global cancer statistics 2018: GLOBOCAN estimates of incidence and mortality worldwide for 36 cancers in 185 countries. *Ca A Cancer Journal for Clinicians*. 2018.
2. Arnold M, Sierra MS, Laversanne M, Soerjomataram I, Jemal A, Bray F. Global patterns and trends in colorectal cancer incidence and mortality. *Gut*. 2017;66(4):683-91.
3. Zhou XG, Huang XL, Liang SY, Tang SM, Wu SK, Huang TT et al. Identifying miRNA and gene modules of colon cancer associated with pathological stage by weighted gene co-expression network analysis. *Oncotargets & Therapy*. 2018;11:2815-30.
4. Tao J, Sun G. Application of deep learning based multi-fidelity surrogate model to robust aerodynamic design optimization. *Aerospace Science & Technology*. 2019;92(SEP):722-37.
5. Du M, Tao M, Hong J, Zhou D, Wang S. Application of Deep Learning Algorithm in Feature Mining and Rapid Identification of Colorectal Image. *IEEE Access*. 2020;PP(99):1-.
6. Qin Q, Wang K, Yang J, Xu H, Cao B, Yan W et al. Algorithms for immunochromatographic assay: review and impact on future application. *Analyst*. 2019;144.
7. A MMH, B KL, C MI, D AS, E LFB, B MAH. Dynamic multi-objective optimisation using deep reinforcement learning: benchmark, algorithm and an application to identify vulnerable zones based on water quality. *Engineering Applications of Artificial Intelligence*. 2019;86:107-35.
8. Koay E, Bajaj N, Elhammali A, Shiao J, Kopetz S. Clinical Associations of Treatment Failure After Radiation Therapy for Colorectal Liver Metastases. *International Journal of Radiation Oncologybiologyphysics*. 2020;108(2):E53-E4.

9. Brouwer NPM, Kruijssen DEWVD, Hugen N, Hingh IHJTD, Wilt JHWD. The Impact of Primary Tumor Location in Synchronous Metastatic Colorectal Cancer: Differences in Metastatic Sites and Survival. *Annals of Surgical Oncology*. 2019(5).
10. Abelson JS, Michelassi F, Sun T, Mao J, Milsom J, Samstein B et al. Simultaneous Resection for Synchronous Colorectal Liver Metastasis: the New Standard of Care? *Journal of Gastrointestinal Surgery*. 2017.
11. Jones T, Murphy A, Hussain LR, Guend H, Dunki-Jacobs EM, Lee DY. Trends and Outcomes of Synchronous Resection of Colorectal Metastasis in the Modern Era: Analysis of Targeted Hepatic NSQIP Database. *Journal of the American College of Surgeons*. 2017;225(4):S198.
12. Bao L, You B, Shi S, Shan Y, Zhang Q, Yue H et al. Metastasis-associated miR-23a from nasopharyngeal carcinoma-derived exosomes mediates angiogenesis by repressing a novel target gene TSGA10. *Oncogene*. 2018.
13. Arrazubi V, Mata E, Antelo ML, Tarifa A, Vera R. Circulating Tumor Cells in Patients Undergoing Resection of Colorectal Cancer Liver Metastases. Clinical Utility for Long-Term Outcome: A Prospective Trial. *Annals of Surgical Oncology*. 2019;26(12)
14. Karagiannis GS, Pastoriza JM, Wang Y, Harney AS, Oktay MH. Neoadjuvant chemotherapy induces breast cancer metastasis through a TMEM-mediated mechanism. *Science Translational Medicine* 2017;9(397).
15. Liu W, Sun Y, Zhang L, Xing BC. Negative surgical margin improved long-term survival of colorectal cancer liver metastases after hepatic resection: a systematic review and meta-analysis. *International Journal of Colorectal Disease*. 2015;30(10):1365-73.
16. A PS, B EWJ, A DP, A GL, A GW, C EB et al. Stereotactic radiofrequency ablation (SRFA) for recurrent colorectal liver metastases after hepatic resection. *European Journal of Surgical Oncology*. 2020;S0748-7983(20):30806-4.
17. Patel P, Alosaimi A, Bryant N, Kemp N, Parker D, Rancher E et al. Use of Doxorubicin for Radioiodine Targeted Therapy of Treatment Resistant Metastatic Breast Cancer, Hepatocellular Cancer, and Colon Cancer. *The Journal of Nuclear Medicine*. 2019;60:3025-3025.
18. Caroline Chung FC. Unique Glioma Requiring Unique Management. *International Journal of Radiation Oncology*Biophysics*. 2020;108(3):520-1.
19. Wang H, Liu J, Li J, Zang D, Song N. Identification of gene modules and hub genes in colon adenocarcinoma associated with pathological stage based on WGCNA analysis. *Cancer Genetics*. 2020;242.
20. Mathews SM, Chandra K, Barner KE. A novel application of deep learning for single-lead ECG classification. *Computers in Biology & Medicine*. 2018;99:53-62.
21. Poria S, Peng H, Hussain A, Howard N, Cambria E. Ensemble application of convolutional neural networks and multiple kernel learning for multimodal sentiment analysis. *Neurocomputing* 2017;261(oct.25):217-230.

22. Bychkov D, Turkki R, Haglund C, Linder N, Lundin J. Deep learning for tissue microarray image-based outcome prediction in patients with colorectal cancer. Paper presented at: Spie Medical Imaging: Digital Pathology, 2016.
23. Acharya UR, Fujita H, Shu LO, Hagiwara Y, Adam M. Application of Deep Convolutional Neural Network for Automated Detection of Myocardial Infarction Using ECG Signals. *Information Sciences*.2017;415.
24. Yasuhiro, Shoji, Hiroyuki, Takamura, Itasu, Ninomiya et al. The Embryonic Stem Cell-Specific Transcription Factor ZFP57 Promotes Liver Metastasis of Colorectal Cancer. *Journal of Surgical Research*. 2019.
25. Tada Y, Yamaguchi Y, Kinjo T, Song X, Akagi T, Takamura H et al. The stem cell transcription factor ZFP57 induces IGF2 expression to promote anchorage-independent growth in cancer cells. *Oncogene*. 2015;34(6).
26. Zhao H, Chen S, Fu QJJoCB. Exosomes from CD133 + cells carrying circBCC1 mediate cell stemness and metastasis in colorectal cancer. 2020;121(3).
27. Kumar A, Deep A, Gupta RK, Atam V, Mohindra S. Brain Microstructural Correlates of Cognitive Dysfunction in Clinically and Biochemically Normal Hepatitis C Virus Infection. *Journal of Clinical & Experimental Hepatology*. 2017;7(3).
28. Katz DH, Deo RC, Aguilar FG, Selvaraj S, Martinez EE, Beussink-Nelson L et al. Phenomapping for the Identification of Hypertensive Patients with the Myocardial Substrate for Heart Failure with Preserved Ejection Fraction. *Journal of Cardiovascular Translational Research*. 2017;10(3):275-284.
29. Montagnese S, Russo FP, Amodio P, Burra P, Gasbarrini A, Loguercio C et al. Hepatic encephalopathy 2018: A clinical practice guideline by the Italian Association for the Study of the Liver (AISF). *Digestive & Liver Disease*. 2018;51(2):190-205.
30. Wu Z, Jiang S, Zhou X, Wang Y, Liu QJCC. Application of image retrieval based on convolutional neural networks and Hu invariant moment algorithm in computer telecommunications. 2020;150:729-38.
31. Yasuyuki, Fukami, Yuji, Kaneoka, Atsuyuki, Maeda et al. Bilobar versus unilobar multiple colorectal liver metastases: a propensity score analysis of surgical outcomes and recurrence patterns. *Journal of Hepato Biliary Pancreatic Sciences*. 2017;24(3):153-160.
32. McCoy AM, Kemper AM, Boyce MK, Brown MP, Trumble TN. Differential gene expression analysis reveals pathways important in early post-traumatic osteoarthritis in an equine model. *Bmc Genomics*. 2020;21(1).
33. Yu G, Wang LG, Yan GR, He QY. DOSE: an R/Bioconductor package for disease ontology semantic and enrichment analysis. *Bioinformatics*. 2015(4):608-9.
34. Zhang B, Gaiteri C, Bodea L-G, Wang Z, McElwee J, Podtelezchnikov Alexei A et al. Integrated Systems Approach Identifies Genetic Nodes and Networks in Late-Onset Alzheimer's Disease. 2013;153:707-720.

35. Jiang H, Zhan F, Wang C, Qiu J, Zeng X. A Robust Algorithm Based on Link Label Propagation for Identifying Functional Modules from Protein-protein Interaction Networks. *IEEE/ACM Transactions on Computational Biology & Bioinformatics*. 2020;PP(99).
36. Lim JY, Lee JH, Yun DH, Lee YM, Kim DK. Inhibitory effects of nodakenin on inflammation and cell death in lipopolysaccharide-induced liver injury mice - *ScienceDirect. Phytomedicine*. 2020;81:153411.
37. Dilber IDY, Aysun Z, Gnbe ZR, Cemre CS, Burcu A, Mustafa AJJoP et al. Antimigratory effect of pyrazole derivatives through the induction of STAT1 phosphorylation in A549 cancer cells. 2021.
38. Zhang J, Wang F, Liu F, Xu GJTAIMO. Predicting STAT1 as a prognostic marker in patients with solid cancer. 2020;12.
39. A ESPA, F AJDSP, D LBDS, F ELDS, F FPM, C WCR et al. 22 β -hydroxytingenone induces apoptosis and suppresses invasiveness of melanoma cells by inhibiting MMP-9 activity and MAPK signaling- *ScienceDirect. Journal of Ethnopharmacology*. 2020; 267:113605.
40. Luisa, Foltran, Giovanna, De, Maglio, Nicoletta et al. Prognostic role of KRAS, NRAS, BRAF and PIK3CA mutations in advanced colorectal cancer. *Future Oncology*. 2015;14(1):292.

Figures

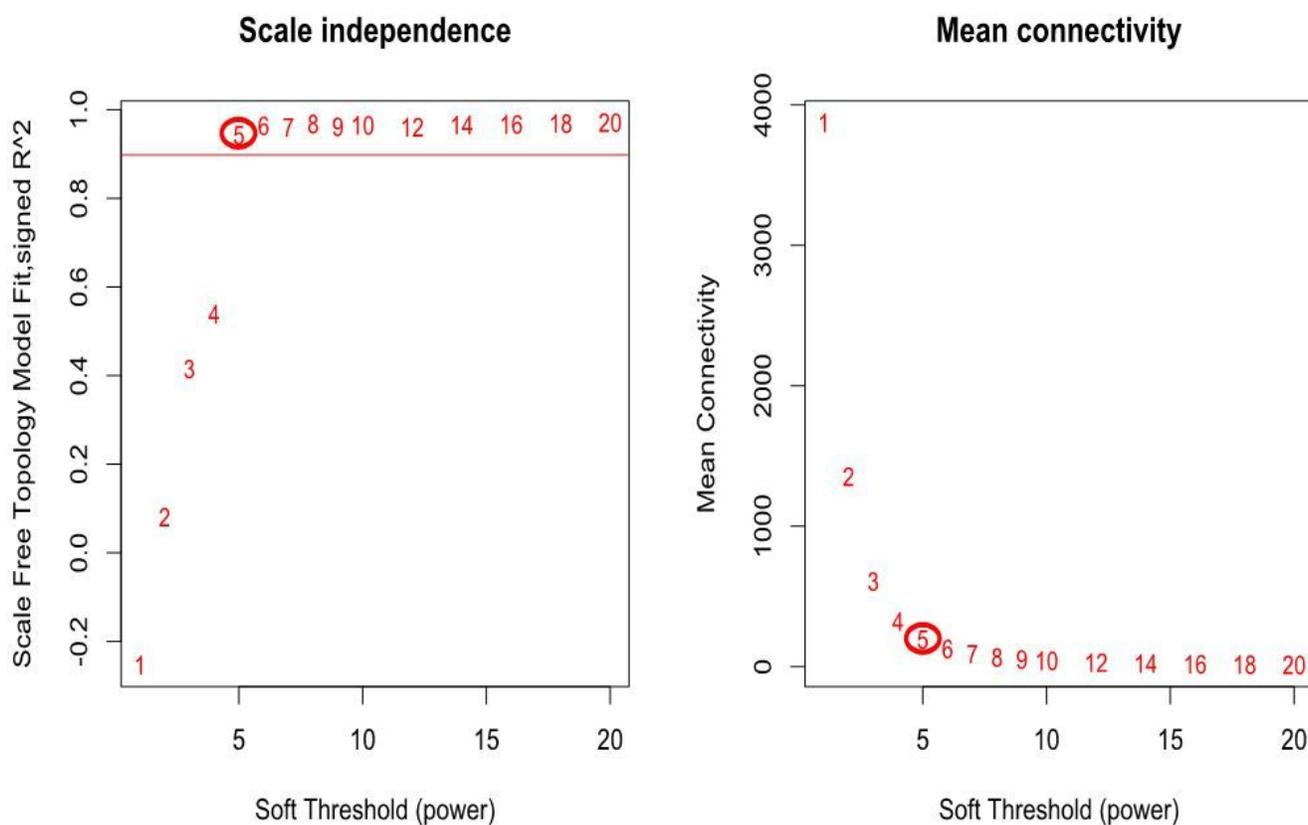


Figure 1

Determination of soft threshold.

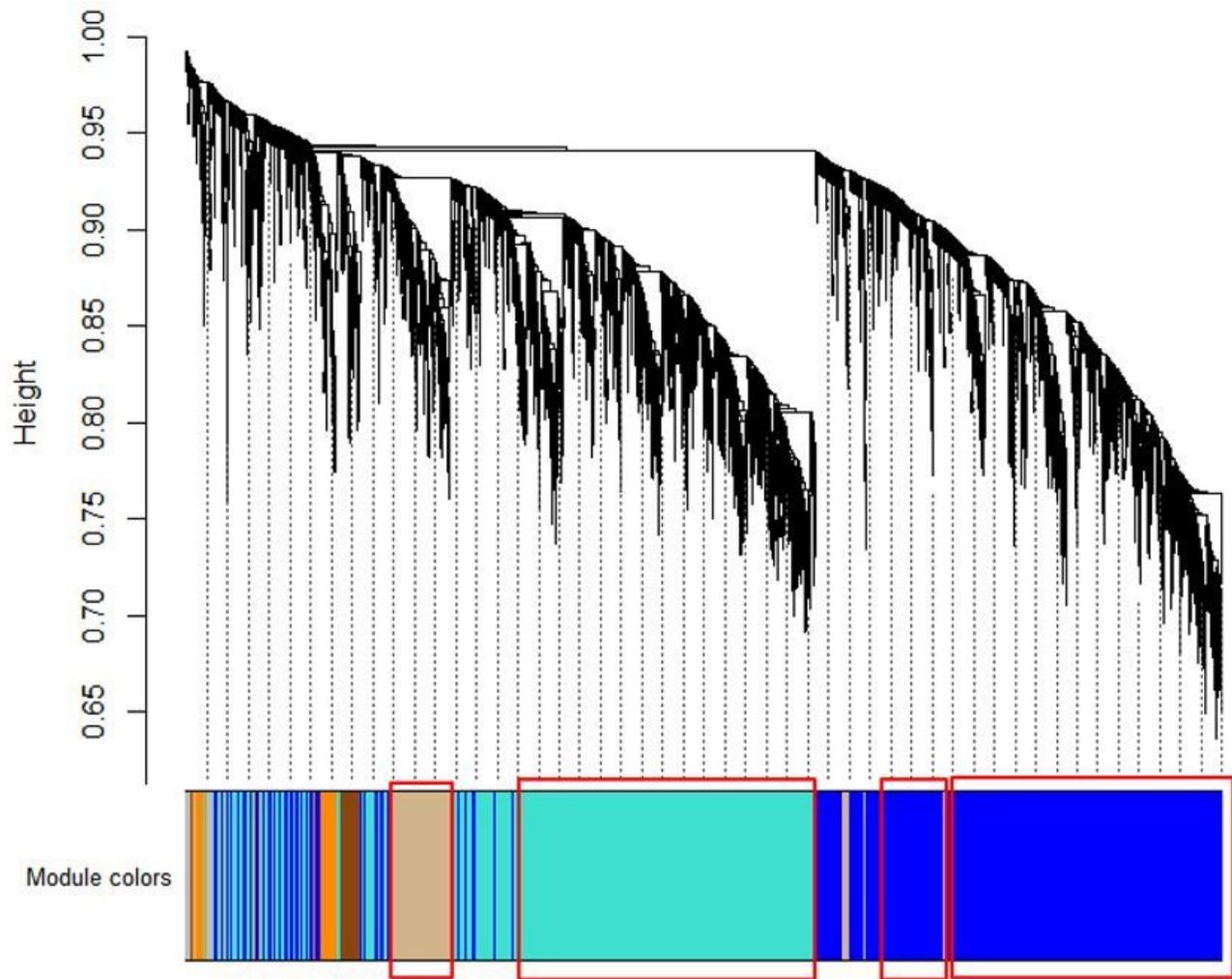


Figure 2

Gene module clustering results

Eigengene adjacency heatmap

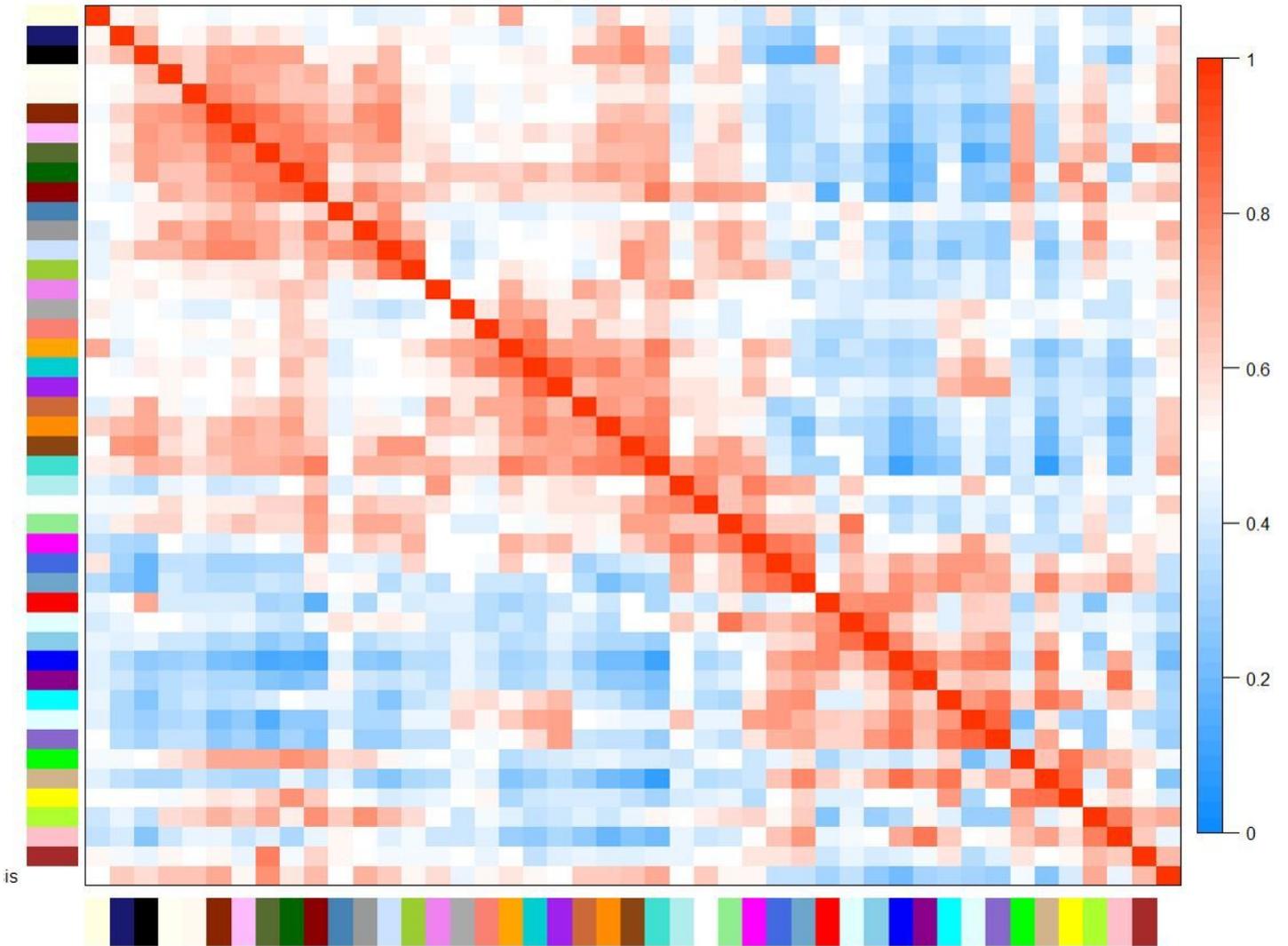


Figure 3

Intrinsic gene adjacent heat map of 44 gene modules, the correlation between left upper corner and right lower corner is relatively high, and there are few unrelated gene modules in the map.

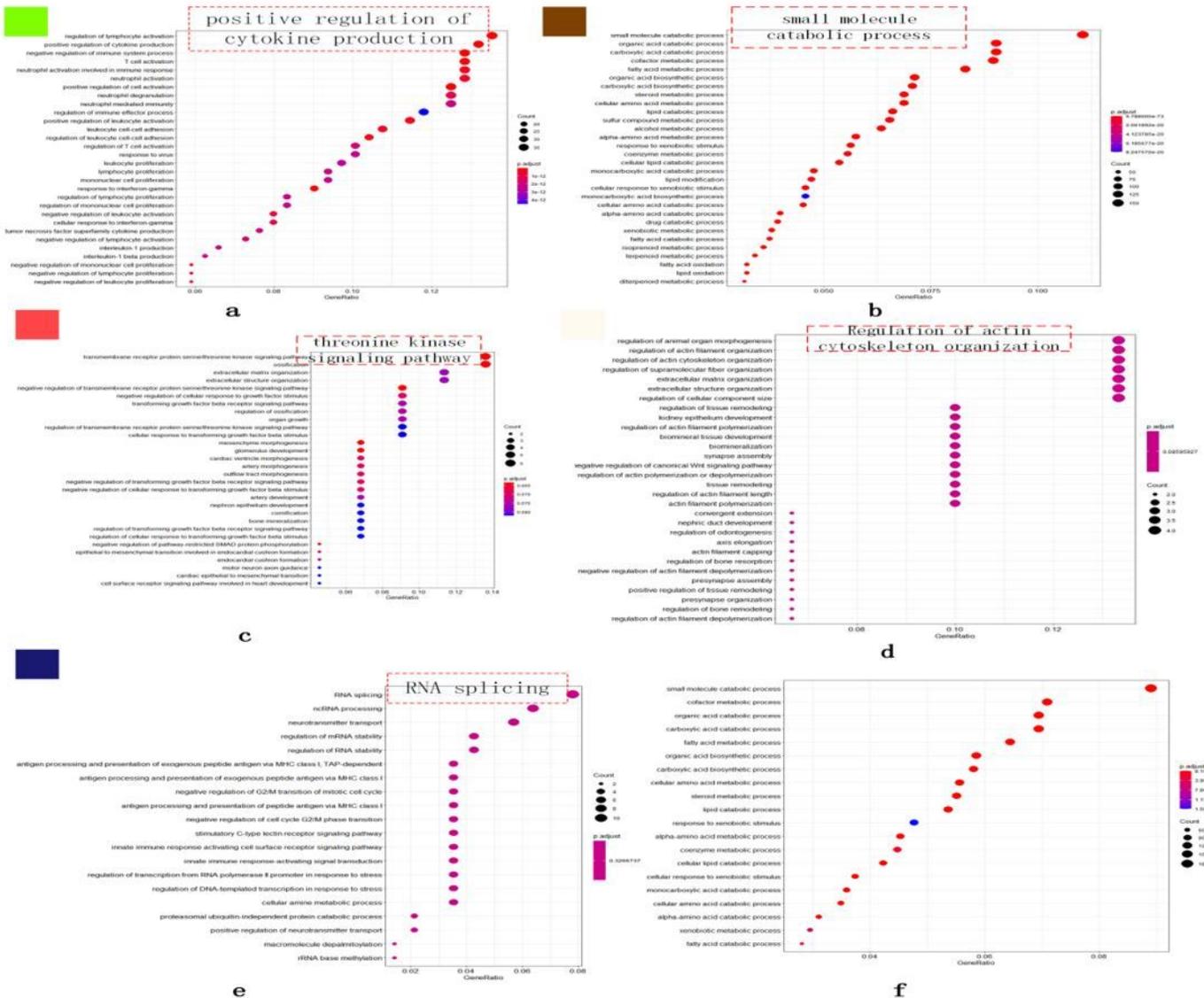


Figure 5

Function enrichment analysis. (a) The genes of yellow-green modules are enriched in the main biological processes such as regulating lymphocyte activation and regulating the production of cytokines. (b) The genes of brown modules are mainly enriched in the main biological processes such as small molecule catabolism and organic acid catabolism. (c) The genes of orange-red 4 modules are mainly enriched in the main biological processes such as transmembrane receptor protein serine and threonine kinase signaling pathway. (d) The genes of floral white modules are mainly enriched in the main biological processes such as the regulation of actin filament tissue and extracellular structure tissue. (e) The genes of midnight blue modules are mainly enriched in the main biological processes such as RNA splicing, ncRNA treatment and neurotransmitter transport. (f) The total five gene modules are mainly enriched in the biological processes such as small molecule catabolism and cofactor metabolism.

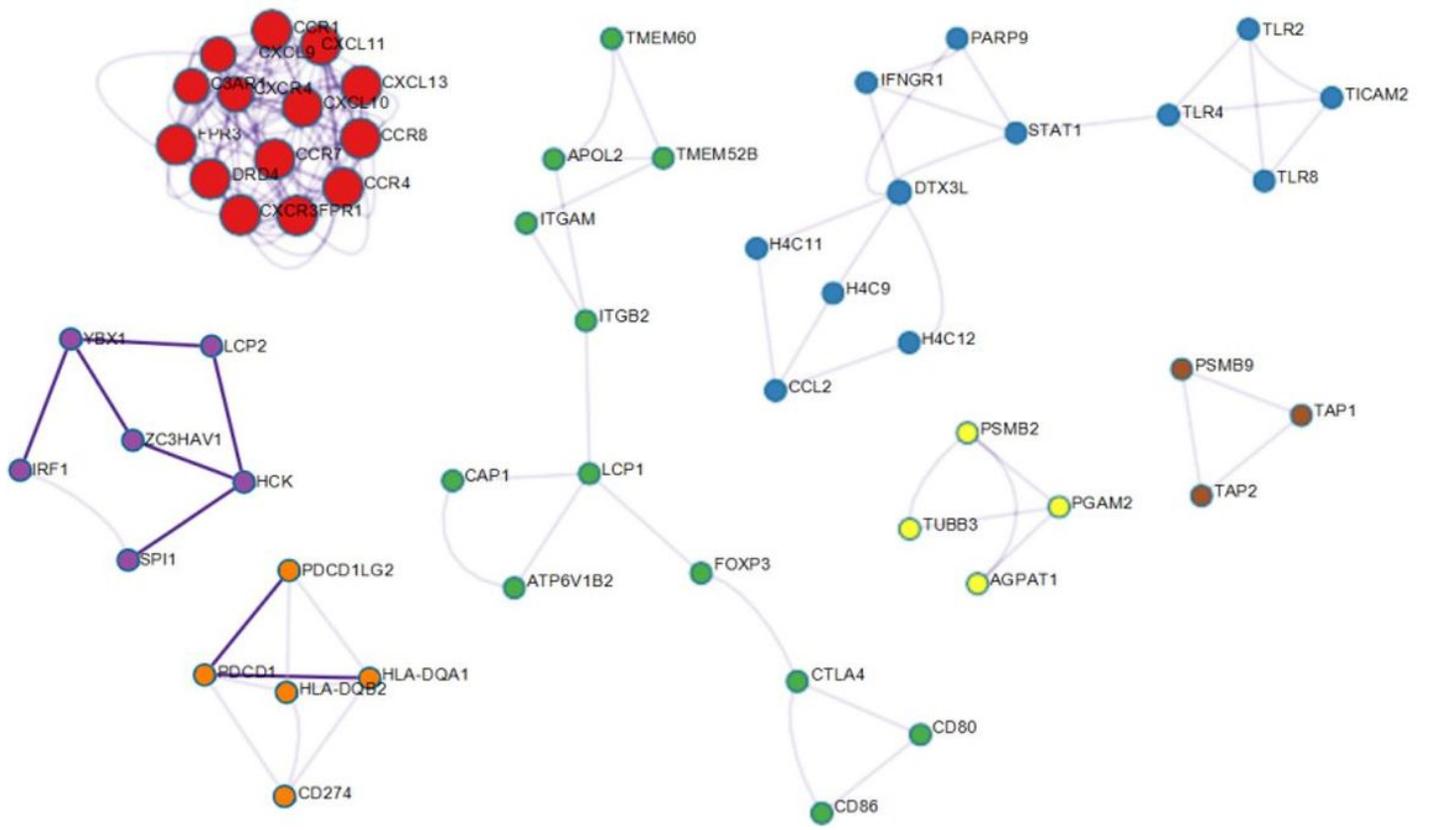


Figure 6

shows the high adjacency subnets excavated by MCODE

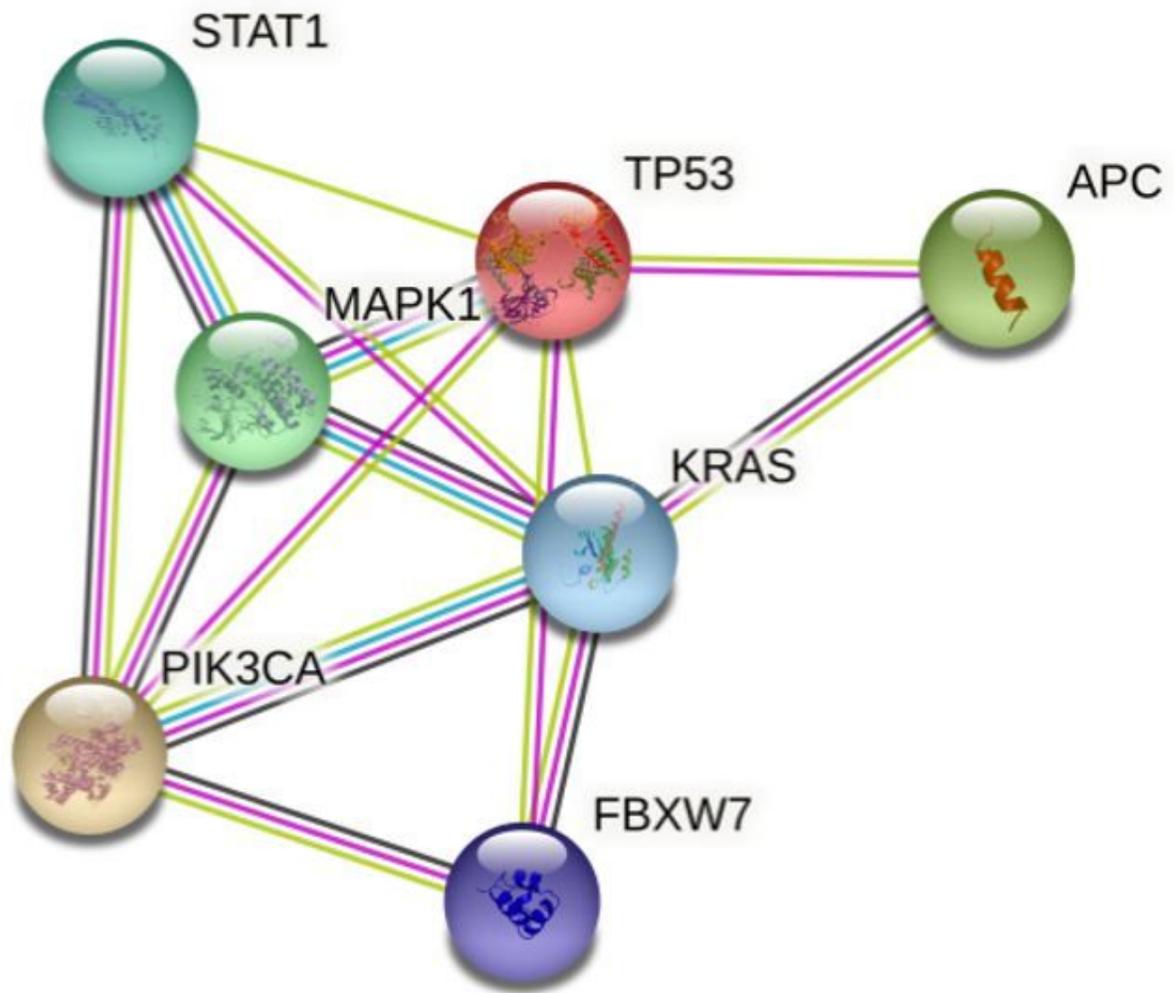


Figure 7

KRAS, APC, FBXW7, PIK3CA, TP53 and the mined protein interaction network

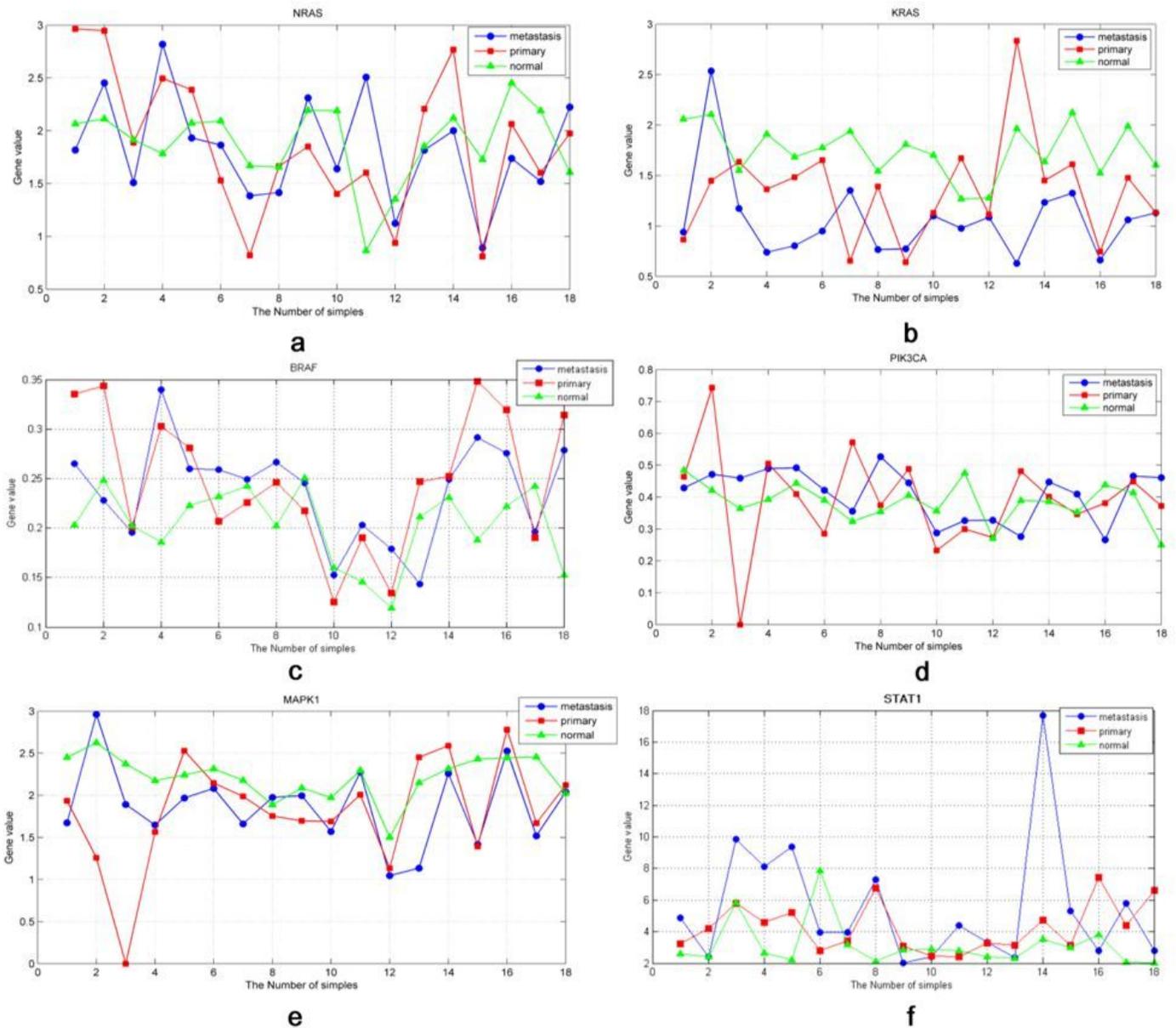


Figure 8

Comparison of six gene transcriptome data. (a) is the gene value of each sample of NRAS gene in three diseases; (b) is the gene value of each sample of KRAS gene in three diseases; (c) is the gene value of each sample of BRAF gene in three diseases; (d) is the gene value of each sample of PIK3CA gene in three diseases; (e) is the gene value of each sample of MAPK1 gene in three diseases; (f) is the gene value of each sample of STAT1 gene in three diseases.