Identification of Molecular Subgroups in Liver Cirrhosis by Gene Expression Profiles

Ying-xue Zhang (✉ 1240617645@qq.com)
Beijing Hospital of Traditional Chinese Medicine https://orcid.org/0000-0003-0698-0532
Feng-xia Sun
Beijing Hospital of Traditional Chinese Medicine
Xiao-ling Li
Beijing Hospital of Traditional Chinese Medicine
Qing-hua Liu
Beijing University of Chinese Medicine School of Traditional Chinese Medicine
Zi-meng Chen
Beijing Hospital of Traditional Chinese Medicine
Yu-fei Guo
Beijing Hospital of Traditional Chinese Medicine

Research Article

Keywords: liver cirrhosis, gene expression profile, classification of subgroups, WGCNAmodule

Posted Date: May 10th, 2021

DOI: https://doi.org/10.21203/rs.3.rs-484882/v1

Version of Record: A version of this preprint was published at Hepatitis Monthly on January 19th, 2022.
See the published version at https://doi.org/10.5812/hepatmon.118535.

Identification of molecular subgroups in liver cirrhosis by gene expression profiles

Zhang Ying-xue[1] Sun Feng-xia[1] Li Xiao-ling[1] Liu Qing-hua[2] Chen Zi-meng[1] Guo Yu-fei[1]

(1 Department of Infections, Beijing Hospital of Traditional Chinese Medicine, Affiliated to Capital Medical University, Beijing, China 2 Beijing University of Chinese Medicine, Beijing, China)

Corresponding Author:

Sun Feng-xia

23 Art Gallery back street, Dongcheng District, Beijing,100010,China

Email address: sunfengxia1969@126.com

Abstract

Background: Cirrhosis is a common clinical chronic progressive liver disease and has become one of the main causes of death worldwide. The condition of liver cirrhosis is complex and there is also clinical heterogeneity. Identifying liver cirrhosis based on molecular characteristics has become a challenge.

Methods: To reveal the potential molecular characteristics of different types of cirrhosis, we divided 79 patients with cirrhosis into 4 subgroups based on gene expression profiles. These gene expression profiles were retrieved from the comprehensive gene expression database. In addition, these subgroups showed different expression patterns. To reveal the differences between subgroups, we used weighted gene co-expression analysis and identified six subgroup-specific gene co-expression analysis modules.

Results: The characteristics of WCGNA modules indicate that TGF - β signaling pathway,viral protein interaction with cytokines and cytokine receptors, including a variety of chemokines and inflammatory factors, are upregulated in subgroup I, indicating that subjects in subgroup I may show inflammatory characteristics; fatty acid metabolism, biosynthesis of cofactors, carbon metabolism and protein processing

pathway in endoplasmic reticulum were significantly enriched in subgroup II, which indicated that the subjects in subgroup II might have the characteristics of active metabolism; arrhythmogenic right ventricular cardiomyopathy and Neuroactive ligand−receptor interaction are significantly enriched in subgroup IV; we did not find a significant upregulation pathway in the third subgroup.

Conclusion: The subgroups classification of liver cirrhosis cases shows that patients from different subgroups may have unique gene expression patterns, which indicates that patients in each subgroup should receive more personalized treatment.

Keywords: liver cirrhosis, gene expression profile, classification of subgroups, WGCNA module

1 Introduction

Liver cirrhosis is a pathological stage characterized by diffuse liver fibrosis, pseudolobular formation, and blood vessel proliferation in and outside the liver as the characteristics of various chronic liver diseases. Approximately 2 million people die of liver disease each year, and 1 million people die from cirrhosis complicated. Liver cirrhosis is currently the 11th most common cause of death in the world, bringing a serious health and economic burden to the world[1,2]. Liver cirrhosis can cause serious life-threatening complications such as bleeding, liver failure, or hepatic encephalopathy. Thus far, there is still no clear consensus on the treatment of liver cirrhosis. The only way is to control symptoms and complications, in addition to slowing down the development of liver cirrhosis. If the liver is severely damaged, the only treatment option may be liver transplantation[3].Patients with early liver disease usually have no symptoms and may have reached the advanced stage of the disease when they are found. Chronic liver disease is characterized by progressive liver fibrosis, leading to cirrhosis, HCC, and liver failure. Cirrhosis can lead to irreversible damage, but the early stage of chronic liver disease can be reversed to a healthy state. Therefore, it is essential to find biomarkers that can identify the early stage of chronic liver disease for the prevention of severe liver damage. The discovery of biomarkers for liver cancer and liver cirrhosis has promoted the development of sequencing

technology, and high-throughput sequencing is one of the representative technological innovations in the biological field in recent decades[4].

With the development of microarray and high-throughput sequencing technology, it provides a good opportunity for further understanding of liver cirrhosis. Fan Y et al.[5] obtained the microarray data set from the GEO database, used EdgeR software to analyze the differential expression of circRNA between NASH cirrhosis samples and normal samples, and analyzed the functional enrichment and pathway enrichment of target genes, confirming that cyclic ribose abnormal regulation of nucleic acids in non-alcoholic steatohepatitis cirrhosis. Bioinformatics analysis shows that the abnormal expression of cyclic ribonucleic acid may be related to the occurrence and development of NASH liver cirrhosis. Chan KM et al.[6] used gene chip analysis to detect and compare the gene expression profiles of cirrhosis and non-cirrhosis, and found that compared with non-cirrhotic liver, a total of 213 important genes were differentially expressed in cirrhotic liver with more than two-fold changes. And further analyzed the signal pathways and main functions of the proteins encoded by these differentially expressed genes. The results showed that the liver with cirrhosis has a unique gene expression pattern related to inflammation, immune response, and cell growth, and may be related to cancer. Zhu M et al.[7] performed exome sequencing on the diseased liver samples of 82 patients and found that there were complex mutations in liver cirrhosis. And established an in vivo CRISPR screening method to study the functional effects of mutant genes. Consistent with the sequencing results, the detection of 147 genes once again revealed that the deletion of Pkd1, Kmt2d, and Arid1a promoted clonal expansion. Many of the recurrently mutated genes that have the greatest impact on regeneration are not often detected in liver cancer, and therefore do not seem to promote carcinogenesis. Gong Z et al.[8] used the hepatotoxic substance carbon tetrachloride to induce liver fibrosis in animal models, and used RNA sequencing to perform genome-wide identification of lncRNAs in fibrotic liver tissue compared with carbon tetrachloride untreated liver tissue , determined the gene expression pattern in liver fibrosis induced by carbon tetrachloride.

With the reduction of the cost of high-throughput technology and the development of bioinformatics, increasing researches will apply high-throughput technology and bioinformatics to reveal the pathogenesis of liver cirrhosis. However, most studies only focus on the difference between cirrhosis cases and normal controls, but rarely pay attention to the difference between cirrhosis cases. Due to the clinical heterogeneity of liver cirrhosis, it has become a challenge to identify liver cirrhosis based on transcriptome characteristics. In cancer research, tumor samples are usually divided into several subtypes based on gene expression patterns, which can reveal heterogeneity between tumors, predict clinical endpoints, and guide treatment[9]. To enhance our understanding of the molecular mechanism of liver cirrhosis, we divided cirrhosis cases into 4 subgroups based on gene expression profiles, and characterized the candidates by annotating the corresponding co-expression functional modules with the Kyoto Encyclopedia of Genes and Genomes (KEGG) pathway subgroup. The objective was to better characterize the differences between cirrhosis cases.

2 Materials and Methods

2.1 Data collection

2.1.1 Download data

Use R / Bioconductor package geoquery[10] to extract "Gene Expression Omnibus" (GEO) objects. First, go to the geo website（ https://www.ncbi.nlm.nih.gov/geo/ ） Enter the keyword "cirrhosis", select GEO Dataset Database, after entering the search results page, check the series option, and select the expression profiling by array. This study is Homo sapiens. The qualified gene chips were included in this study, and their platform files and sequence probe matrix files were downloaded, respectively.

2.1.2 GEO data annotation

The platform file downloaded from the GEO website is processed with Perl software to obtain a text with the row name of the gene name and the column name of the sample name, and name it the name of the corresponding gene chip as the data for subsequent research. Extract the information about the clinical features in the probe matrix file into the newly created EXCEL as the clinical data file for the research.

2.2 Removal of batch effect

Firstly, "limma" package and "sva" package in R / Bioconductor were used to merge the expression data, and batch correction was carried out[11]. When merging data, the mean value is taken for the data with multiple lines of one gene, only one line is reserved, and log2 is taken for the data with large value for conversion. When there is a difference between batch effects in the data, ComBat-seq can obtain better statistical ability and control the false positive rate compared with other available methods[12]. Therefore, we choose the combat method to eliminate the batch effect between platforms. Finally, the R / ggplot2 package is used for principal component analysis to evaluate whether the batch effect is removed.

2.3 Consensus clustering

First, we use the "limma" package and "Consensus Cluster Plus" package in the R/Bioconductor package to perform consensus clustering[13], and divide the included cases of cirrhosis into different subgroups. Clustering is performed using K-means algorithm with Spearman distance. Set the maximum cluster number to 10, and the final cluster number is determined by the consensus matrix and the cluster consensus score (>0.8).

2.4 extraction of specific upregulated genes in subtypes

By comparing the cases in a specific subgroup with those in other subgroups, the subgroup specific upregulated genes were identified. It should be noted that the Wilcoxon ' s - sum rank test was used to test the differential expression, the corrected p value threshold was < 0.05, and the absolute difference of means >0.2. For a given gene, the difference of means was calculated by subtracting the mean of expression in normal controls from that in case of the specified subgroup.

2.5 Gene set enrichment analysis

Gene Set Enrichment Analysis (GSEA) was used to observe whether the specific differential genes in each subgroup were also different from normal samples. Gene set enrichment analysis is implemented in GseaPrerank mode in GSEA 4.1.0. The gene set file (type-specific gene file) is composed of subgroup-specific upregulated genes. The gene list of each subgroup is sorted by the P value of T, which is calculated by

comparing the cirrhosis cases of each subgroup with the normal control group.

2.6 Weighted gene co-expression network analysis

Weighted Gene Co-Expression Analysis (WGCNA) was used to analyze the genotyping specific genes in subgroups. WGCNA has been proved to be an effective method to detect multiple coexpression modules, which can be used to find clustering (modules) of highly related genes[14,15]. Find the optimal power value through the power value scatter plot, and calculate the distance between genes. In addition, the average method and the dynamic method are used for hierarchical clustering analysis, the clustering diagram and the module classification of genes are respectively established, and similar modules are merged. We finally determined 6 functional modules. At the same time, we use the labeled heat map function option in the WGCNA package to draw the heat map, visualize the data, set the low expression of the gene as blue, the middle expression as white, and the high expression as red.

2.7 GO enrichment analysis and KEGG enrichment analysis

GO enrichment analysis and KEGG enrichment analysis were performed for each functional module of WGCNA, in which go was divided into three modules: biological process (BP), cellular component (CC), and molecular function (MF)[16]. KEGG is a reference knowledge base for biological interpretation of genome sequences and other high-throughput data. The genome of KEGG pathway is downloaded from MsigDB, and the gene species is selected as human[17]. In KEGG enrichment analysis, the *p*-value filter condition was set to < 0.05, and the corrected *p*-value filter condition was 1. The enrichment results were visualized and the bubble diagram was drawn.

3 Results

3.1 microarray data characteristics

This study included four independent microarray information, involving four independent clinical trials, all from the GEO database, GSE14323, GSE77627, GSE123932, GSE128726, a total of 127 samples (including 79 patients with liver cirrhosis and 48 healthy subjects). GSE123932 provides clinical information of

gender and age.

3.2 Removal of batch effect by cross-platform normalization

To eliminate the batch effect from different platforms and batches, we use ComBat method to eliminate the batch effect between data sets. A total of 11152 genes were detected by four microarrays.Before the removal of the batch effect, the samples were clustered into batches based on the top two principal components (PCs) of unnormalized expression values (Figure 1a). In contrast, the scatter plot is based on the standardized processing of principal component analysis, and the results show that the samples of four batches are mixed, indicating that the batch effect caused by different platforms is clearly removed (Figure 1b). The results show that the batch effect is successfully eliminated by cross-platform normalization.

3.3 Consensus clustering of Cirrhosis cases

Cluster analysis were performed using the expression file after batch effect correction and the sample information of the disease group. The 79 patients with cirrhosis were divided into subgroups (see section 2.3). According to the consistency scoring situation in the data statistics, the consistency cluster analysis of gene expression profiles is divided into 4 subgroups, among which the number of cases in subgroups I, II, III, and IV are 21, 30, 18, and 10, respectively. They have significantly different expression patterns. On the contrary, based on the consistency matrix, a high degree of similarity in gene expression patterns was observed in each subgroup (Figure 2a). Generally speaking, the higher the consistency score, the more stable the type. In the results of this study, when divided into 2, 3, or 4 groups, the cluster agreement score of each subgroup is higher than 0.8 (Figure 2b), which indicates that these classifications are more robust than other clusters. When considering more groups better, so 79 patients with liver cirrhosis were divided into 4 subgroups.

3.4 Identification of gene co-expression modules for each subgroup

To reveal the transcriptome differences between cirrhotic subgroups, WGCNA was performed under specific upregulated gene expression levels in each subgroup (see Section 2.6). Through the pairwise differential expression analysis between each two subgroups, 2033, 2134, 804, and 1892 genes that were specifically upregulated in

subgroups I, II, III, and IV were determined (the corrected threshold is $P<0.05$, The absolute difference of means >0.2). Differential expression analysis was performed by comparing the gene expression profile of each subgroup with the gene expression profile of the normal control group. The results of GSEA enrichment analysis showed that the specific upregulated genes in the subtypes were also significantly different when compared with normal samples (Figure 3a-d, FDR <0.05). It is worth noting that, compared with other subgroups, subgroup III has the least subgroup-specific upregulated genes, and the number of differentially expressed genes is relatively small by comparing the cases of each subgroup with the normal control group. (Table 1). Based on the expression levels of a total of 6863 specific upregulated genes in each subgroup, a gene heat map expression network was constructed, which identified 6 WGCNA modules. The relationship between WGCNA modules and corresponding subgroups is shown in Table 1. The gene of each WCGNA module is displayed by KEGG enrichment, and the TGF - β signaling pathway,viral protein interaction with cytokines and cytokine receptors, including a variety of chemokines and inflammatory factors is only enriched in the blue module significantly. At the same time, the PI3K/Akt signaling pathway is also significantly enriched in the blue module, and there are many upregulated genes. The genes in the blue module are upregulated in subgroup I, indicating that subjects in subgroup I may exhibit inflammatory characteristics ; fatty acid metabolism, biosynthesis of cofactors are only significantly enriched in the brown module, and carbon metabolism and protein processing pathways in the endoplasmic reticulum are only significantly enriched in the brown and green modules. The genes in the brown and green modules are in subgroup II, it is upregulated, indicating that subjects in subgroup II may show metabolic activity; arrhythmogenic right ventricular cardiomyopathy and neuroactive ligand−receptor interaction pathways are only significantly enriched in the turquoise module, and the genes in the turquoise module are upregulated in subgroup IV. We did not find a significant upregulation pathway in the third subgroup. These findings indicate that each subgroup has its own specific functional modules or pathways that can regulate the occurrence or progression of liver cirrhosis. (Figure 4a-b).

After GO enrichment analysis (Figure 4c-e), the biological process (BP):the blue module mainly involves the regulation of cell−cell adhesion,positive regulation of cell adhesion, positive regulation of cell−cell adhesion, blood coagulation and hemostasis; the brown module mainly involves small molecule catabolic process, cellular amino acid metabolic process, cellular ketone metabolic process, andresponse to endoplasmic reticulum stress; the green module mainly involves nucleotide−excision repair, respiratory electron transport chain, ATP synthesis coupled electron transport, mitochondrial ATP synthesis coupled electron transport, cellular respiration, etc; turquoise modules mainly involve axonogenesis, gland development, egulation of trans −synaptic signaling, modulation of chemical synaptic transmission,embryonic organ development, sensory organ morphogenesis,embryonic organ morphogenesis, etc.Cell components (CC): Blue modules were concentrated in focal adhesion, cell−substrate junctions, membrane rafts, collagen−containing extracellular matrix, membrane microdomain, and endocytosis vesicles; brown modules were concentrated in mitochondrial matrix, mitochondrial protein complex, mitochondrial inner membrane, focal adhesion, cell−substrate junction, collagen−containing extracellular matrix; green modules were concentrated in mitochondrial matrix, mitochondrial protein complex, mitochondrial inner membrane, organellar ribosome, respirasome, oxidoreductase complex. The turquoise modules were mainly concentrated in transmembrane transporter complex, transporter complex, synaptic membrane, neuron to neuron synapse, postsynaptic membrane, and ion cha, nnel complex.Molecular function (MF): Blue module mainly involves receptor ligand activity, signaling, receptor activator activity, DNA-binding, transcription factor binding, etc; the brown module mainly involves DNA−binding, transcription factor binding, oxidoreductase activity, acting on the CH−OH group of donors, NAD or NADP as acceptor, lyase activity, carbon−oxygen lyase activity, electron transfer activity; green module mainly involves It involves NADH dehydrogenase activity, NADH dehydrogenase (quinone) activity, NADH dehydrogenase (ubiquinone) activity; the turquoise module mainly involves signaling receptor activator activity, receptor ligand activity, metal ion transmembrane transporter activity, channel activity, neurotransmitter receptor activity,

etc.

4 Discussion

In this study, we analyzed the gene expression profiles of liver cirrhosis cases and normal controls from 4 independent GEO data sets. Eliminates the batch effect of different platforms or batches. In addition, we successfully divided 79 patients with liver cirrhosis into 4 subgroups based on gene expression profiles. Transcriptome classification revealed subgroup-specific functional modules or pathways in further analysis. A significant link was observed between clinical characteristics and transcriptome classification. Consistent clustering based on large sample size and high cluster consensus score (>0.8) shows that our transcriptome classification is robust. The motivation of this research is the transcriptome classification of cancer, which can identify subgroups based on their gene expression profiles or other omics data. The current molecular subtypes guide the diagnosis and clinical treatment of many cancer types[18]. For example, gene expression profiles are of great significance to our understanding of breast cancer biology. The five molecular subtypes of breast cancer show significant differences in incidence, risk factors, prognosis，and treatment sensitivity[19]. To study the genetic changes of molecular subtypes of bladder cancer, Choi W et al.[20]used the complete TCGA RNA-seq data set and developed three different classifiers to classify bladder cancer into molecular subtypes.It was found that changes involving RB1 and NFE2L2 were enriched in basal cell carcinoma, while changes involving FGFR3 and KDM6A were enriched in intraluminal tumors,the results further strengthen the conclusion that the molecular subtypes of bladder cancer are different disease entities with specific genetic changes, and have a direct impact on the clinical management of bladder cancer patients.In addition to cancer research, non-cancerous diseases, such as to reveal the potential molecular characteristics of different types of coronary heart disease, Peng XY et al.[21]divided 352 patients with coronary heart disease into three subgroups based on gene expression profiles. Transcriptome classification shows that cases from different subgroups may have unique gene expression patterns. Chang Y et al.[22]analyzed the

genome-wide blood gene expression of 229 former smokers in the ECLIPSE study and determined a new, clinically relevant molecular subtype of COPD. Although these studies have certain limitations and confounding factors, these studies have indeed improved our understanding of the link between molecular mechanisms and disease development.

Similar to cancer, complex diseases like liver cirrhosis exhibit clinical heterogeneity. Unlike previous studies that only studied the gene expression profiles of cirrhosis cases, or compared them with the gene expression profiles of normal controls, we further divided the cirrhosis cases into subgroups. Compared with previous studies, the subgroups are specific sexual function modules and pathways that not only confirm the regulatory pathways related to cirrhosis, but also link specific pathways with specific subgroups.For example, it is well known that the TGF-β signaling pathway plays an important role in the regulation of different cellular processes, including proliferation, differentiation, migration or cell death, which are critical to the homeostasis of tissues and organs. As far as the liver is concerned, TGF-β signaling is involved in all stages of disease progression, from initial liver damage to inflammation and fibrosis, to cirrhosis and cancer. TGF-β plays a key role in cell growth and apoptosis of hepatocytes and promotes liver differentiation during embryogenesis and physiological liver regeneration. However, due to chronic liver injury, high levels of TGF-β cause stellate cells to activate myofibroblasts and a large number of hepatocytes to die, promoting liver fibrosis and cirrhosis[23].In our study, the TGF-β signaling pathway is most significantly enriched in the first subcomponent, including the significant upregulation of SMAD7, which is one of I-SMAD and acts as an antagonist of TGF-β signaling. An important negative regulator of effective TGF-β and bone morphogenetic protein (BMP) signaling pathways. Studies in mice have shown that SMAD7 has a protective effect in the fibrosis of various organs by inhibiting pro-fibrotic TGF-β signal transduction[24].Overexpression of SMAD7 has been shown to antagonize TGF-β-mediated fibrosis, carcinogenesis and inflammation[25]. Viral protein interactions with cytokine and cytokine receptors are also most significantly enriched in the first subcomponent type, including a variety of

chemokines and inflammatory factors, such as chemokines CCL2, CX3CL1, etc, bile duct epithelial cells (BEC) Damage to the small bile ducts in primary biliary cirrhosis (PBC) shows the characteristics of aging. Senescent BECs showed upregulation of a variety of chemokines and chemotactic activities. The expression of CCL2 and CX3CL1 in PBC senescent BECs increased, which may promote the infiltration of corresponding cells expressing CCR2 and CX3CR1, and further aggravate the PBC bile duct Inflammation of the lesion [26,27].At the same time, PI3K / Akt signaling pathway is also significantly enriched in subgroup I, and there are many upregulated genes. After PI3K / Akt signaling pathway regulates the activation of downstream signaling molecules, it plays a key role in regulating the immune response and the release of inflammatory factors in vivo and in vitro[28].Combined with the results of GO functional enrichment analysis, the first subcomponent mainly involves the regulation of cell−cell adhesion,positive regulation of cell adhesion, positive regulation of cell−cell adhesion, blood coagulation and hemostasis and other biological processes. The combination of cell and extracellular matrix and cell-cell adhesion depend on intercellular adhesion molecules. Cell adhesion molecules participate in various interactions. They can broadly support fibrosis formation by releasing TGFβ, promoting fibrosis, or mediating leukocyte migration. Some candidate cell adhesion molecules have been evaluated as targets for preventing and/or reversing liver inflammation and fibrosis[29].Anticoagulant mechanisms and damages the fibrinolytic system. Inflammatory mediators can increase platelet count, platelet reactivity, downregulate the natural anticoagulation mechanism, activate the coagulation system, promote the spread of coagulation response, and damage fibrinolysis. Similarly, coagulation can increase the inflammatory response by releasing mediators from platelets and activating cells, thereby promoting the interaction between cells, thereby increasing the inflammatory response[30]. In summary, it is indicated that subjects in subgroup I may exhibit inflammatory features.Upregulation of metabolism was observed in subgroup II, including fatty acid metabolism, biosynthesis of cofactors， carbon metabolism and protein processing pathway in the endoplasmic reticulum, combined with the results of GO functional

enrichment analysis, subgroup II mainly involves the small molecule catabolic process, cellular amino acid metabolic process, cellular ketone metabolic process, andresponse of endoplasmic reticulum stress，nucleotide−excision repair, respiratory electron transport chain and other biological processes, indicating that subjects in subgroup II may exhibit metabolic activity feature. Systems biology studies have shown that carbohydrates, amino acids, and lipid-related pathways have changed in liver fibrosis and cirrhosis, and further studies are needed to verify the metabolic characteristics and determine the therapeutic targets of liver fibrosis[31]. Arrhythmogenic right ventricular cardiomyopathy and neuroactive ligand−receptor interaction pathways are only significantly enriched in subgroup IV. Studies have shown that neuroactive ligand−receptor interaction pathways are closely related to the occurrence and development, invasion, and metastasis of liver cancer. A variety of traditional Chinese medicine and Chinese patent medicine exert their curative effects by influencing this signaling pathway[32-34]，but the relationship with liver cirrhosis needs further the study. We did not find a significant upregulation pathway in subgroup III, which may be related to the small sample size. These results further prove that transcriptome classification represents the development stage and intrinsic biological characteristics of liver cirrhosis.

In addition, although our findings suggest that cases from different subgroups may have different expression patterns, they also need further prospective validation in the population. Because only gse123932 provided clinical information of gender and age, and the sample size was small, the correlation between clinical characteristics and WGCNA module could not be carried out, similar to transcriptome typing in cancer, future liver cirrhosis research should also introduce multi omics data to reveal more accurate molecular subgroups of liver cirrhosis. In conclusion, inspired by the cancer subgroup study, we applied a similar strategy to reveal the molecular subgroup of liver cirrhosis. Current studies have shown that cirrhosis cases from different subgroups may have their own unique gene expression patterns, which suggests that patients in each subgroup should receive more personalized treatment.

Declarations

Funding

The work was supported by Special Research Project of Capital Health Development (NO.2018-2-2173)

Conflicts of interest/Competing interests

The author(s) declare(s) that there is no conflict of interest regarding the publication of this paper.

Availability of data and material

The data and material used to support the findings of this study are included within the article.

Code availability

Not applicable.

Authors' contributions

YXZ conceived this research and contributed to data collation, original draft preparation, and draft review. FXS contributed to the writing-review and editing and Supervision. XLL and QHL performed the image analyses and analyzed the data. ZMC and YFG contributed to the writing-original draft preparation. All authors designed the experiments, wrote the manuscript, and revised the paper. All authors read and approved the final version of the manuscript.

References

[1] Xu Xiaoyuan, Ding Huiguo, Li Wengang, Xu Jinghang, Han Ying, Jia Jidong, Wei Lai, Duan Zhongping, Linghu Enqiang, Zhuang Hui. Guidelines for diagnosis and treatment of liver cirrhosis[J]. Journal of Practical Hepatology, 2019, 22(06): 770-786. doi: 10.3969/j.issn.1672-5069.2019.06.004.

[2] Asrani Sumeet K,Devarbhavi Harshad,Eaton John,et al.Burden of liver diseases in the world[J].

Journal of hepatology,2019,70(1):151-171. doi: 10.1016/j.jhep.2018.09.014.

[3] Lee NY, Suk KT. The Role of the Gut Microbiome in Liver Cirrhosis Treatment. Int J Mol Sci. 2020 Dec 28;22(1):199. doi: 10.3390/ijms22010199.

[4] Oh S, Jo Y, Jung S, Yoon S, Yoo KH. From genome sequencing to the discovery of potential biomarkers in liver disease. BMB Rep. 2020 Jun;53(6):299-310. doi: 10.5483/BMBRep.2020.53.6.074.

[5] Fan Y, Zheng Y, Wang J, Zhao T, Liang T. Differential expression and bioinformatic analysis of circRNA in nonalcoholic steatohepatitis cirrhosis. Int J Clin Exp Pathol. 2020 Nov 1;13(11):2820-2830.

[6] Chan KM, Wu TH, Wu TJ, Chou HS, Yu MC, Lee WC. Bioinformatics microarray analysis and identification of gene expression profiles associated with cirrhotic liver. Kaohsiung J Med Sci. 2016 Apr;32(4):165-76. doi: 10.1016/j.kjms.2016.03.008.

[7] Zhu M, Lu T, Jia Y, Luo X, Gopal P, Li L, Odewole M, Renteria V, Singal AG, Jang Y, Ge K, Wang SC, Sorouri M, Parekh JR, MacConmara MP, Yopp AC, Wang T, Zhu H. Somatic Mutations Increase Hepatic Clonal Fitness and Regeneration in Chronic Liver Disease. Cell. 2019 Apr 18;177(3):608-621. doi: 10.1016/j.cell.2019.03.026.

[8] Gong Z, Tang J, Xiang T, Lin J, Deng C, Peng Y, Zheng J, Hu G. Genome- wide identification of long noncoding RNAs in CCl4- induced liver fibrosis via RNA sequencing. Mol Med Rep. 2018 Jul;18(1):299-307. doi: 10.3892/mmr.2018.8986.

[9] Giordano TJ. The cancer genome atlas research network: a sight to behold. Endocr Pathol. 2014 Dec;25(4):362-5. doi: 10.1007/s12022-014-9345-4.

[10] Davis S, Meltzer PS. GEOquery: a bridge between the Gene Expression Omnibus (GEO) and BioConductor. Bioinformatics. 2007 Jul 15;23(14):1846-7. doi: 10.1093/bioinformatics/btm254.

[11] Leek JT, Johnson WE, Parker HS, Jaffe AE, Storey JD. The sva package for removing batch effects and other unwanted variation in high-throughput experiments. Bioinformatics. 2012 Mar 15;28(6):882-3. doi: 10.1093/bioinformatics/bts034.

[12] Zhang Y, Parmigiani G, Johnson WE. ComBat-seq: batch effect adjustment for RNA-seq count data. NAR Genom Bioinform. 2020 Sep;2(3):lqaa078. doi: 10.1093/nargab/lqaa078.

[13] Wilkerson MD, Hayes DN. ConsensusClusterPlus: a class discovery tool with confidence assessments and item tracking. Bioinformatics. 2010 Jun 15;26(12):1572-3. doi: 10.1093/bioinformatics/btq170.

[14] Langfelder P, Horvath S. WGCNA: an R package for weighted correlation network analysis. BMC Bioinformatics. 2008 Dec 29;9:559. doi: 10.1186/1471-2105-9-559.

[15] Yin L, Cai Z, Zhu B, Xu C. Identification of Key Pathways and Genes in the Dynamic Progression of HCC Based on WGCNA. Genes (Basel). 2018 Feb 14;9(2):92. doi: 10.3390/genes9020092.

[16] [12]Du Z, Zhou X, Ling Y, Zhang Z, Su Z. agriGO: a GO analysis toolkit for the agricultural community. Nucleic Acids Res. 2010 Jul;38(Web Server issue):W64-70. doi: 10.1093/nar/gkq310.

[17] Kanehisa M, Sato Y, Furumichi M, Morishima K, Tanabe M. New approach for understanding genome variations in KEGG. Nucleic Acids Res. 2019 Jan 8;47(D1):D590-D595.

[18] Collisson EA, Bailey P, Chang DK, Biankin AV. Molecular subtypes of pancreatic cancer. Nat Rev Gastroenterol Hepatol. 2019 Apr;16(4):207-220. doi: 10.1038/s41575-019-0109-y.

[19] Prat A, Pineda E, Adamo B, Galván P, Fernández A, Gaba L, Díez M, Viladot M, Arance A, Muñoz M. Clinical implications of the intrinsic molecular subtypes of breast cancer. Breast. 2015 Nov;24 Suppl 2:S26-35. doi: 10.1016/j.breast.2015.07.008.

[20] Choi W, Ochoa A, McConkey DJ, Aine M, Höglund M, Kim WY, Real FX, Kiltie AE, Milsom I, Dyrskjøt L, Lerner SP. Genetic Alterations in the Molecular Subtypes of Bladder Cancer: Illustration in the Cancer Genome Atlas Dataset. Eur Urol. 2017 Sep;72(3):354-365. doi: 10.1016/j.eururo.2017.03.010.

[21] Peng XY, Wang Y, Hu H, Zhang XJ, Li Q. Identification of the molecular subgroups in coronary artery disease by gene expression profiles. J Cell Physiol. 2019 Feb 25. doi: 10.1002/jcp.28324. Epub ahead of print.

[22] Chang Y, Glass K, Liu YY, Silverman EK, Crapo JD, Tal-Singer R, Bowler R, Dy J, Cho M, Castaldi P. COPD subtypes identified by network-based clustering of blood gene expression. Genomics. 2016 Mar;107(2-3):51-58. doi: 10.1016/j.ygeno.2016.01.004.

[23] Fabregat I, Moreno-Càceres J, Sánchez A, Dooley S, Dewidar B, Giannelli G, Ten Dijke P; IT-LIVER Consortium. TGF-β signalling and liver disease. FEBS J. 2016 Jun;283(12):2219-32. doi: 10.1111/febs.13665.

[24] de Ceuninck van Capelle C, Spit M, Ten Dijke P. Current perspectives on inhibitory SMAD7 in health and disease. Crit Rev Biochem Mol Biol. 2020 Dec;55(6):691-715. doi: 10.1080/10409238.2020.1828260.

[25] Yan X, Liu Z, Chen Y. Regulation of TGF-beta signaling by Smad7. Acta Biochim Biophys Sin (Shanghai). 2009 Apr;41(4):263-72. doi: 10.1093/abbs/gmp018.

[26] Sasaki M, Miyakoshi M, Sato Y, Nakanuma Y. Modulation of the microenvironment by senescent biliary epithelial cells may be involved in the pathogenesis of primary biliary cirrhosis. J Hepatol. 2010 Aug;53(2):318-25. doi: 10.1016/j.jhep.2010.03.008.

[27] Sasaki M, Miyakoshi M, Sato Y, Nakanuma Y. Chemokine-chemokine receptor CCL2-CCR2 and CX3CL1-CX3CR1 axis may play a role in the aggravated inflammation in primary biliary cirrhosis. Dig Dis Sci. 2014 Feb;59(2):358-64. doi: 10.1007/s10620-013-2920-6.

[28] Zhou T, Shi J, Li X. [Role of PI3K/Akt signaling pathway in the innate immune of sepsis]. Zhonghua Wei Zhong Bing Ji Jiu Yi Xue. 2018 Nov;30(11):1091-1094. Chinese. doi: 10.3760/cma.j.issn.2095-4352.2018.011.016.

[29] Hintermann E, Christen U. The Many Roles of Cell Adhesion Molecules in Hepatic Fibrosis. Cells. 2019 Nov 24;8(12):1503. doi: 10.3390/cells8121503.

[30] Esmon CT. The interactions between inflammation and coagulation. Br J Haematol. 2005 Nov;131(4):417-30. doi: 10.1111/j.1365-2141.2005.05753.x.

[31] Chang ML, Yang SS. Metabolic Signature of Hepatic Fibrosis: From Individual Pathways to Systems Biology. Cells. 2019 Nov 12;8(11):1423. doi: 10.3390/cells8111423.

[32] He Rensong. Screening and analysis of lncRNA related differential expression in liver cancer[D]. Fujian Medical University, 2019. doi: 10.27020/d.cnki.gfjyu.2019.000645.

[33] Wang Minggang, Chen Qi, Gao Xiang, Ye Zhihua, Xiao Mingzhong, Li Hanmin. Bioinformatics analysis of Diwuyanggan Capsules in the prevention and treatment of liver cancer based on expression profile chips[J]. Journal of Integrated Traditional Chinese and Western Medicine Liver Diseases, 2019, 29(03): 244 -247.

[34] Zhang Zexin, Li Bingbing, Qin Xianfeng, Li Jing, Zhong Chong, Lin Lizhu. Based on network pharmacology, research on the mechanism of action of tricornu-zedoary turmeric on the treatment of hepatocellular carcinoma[J]. Journal of Shandong University of Traditional Chinese Medicine, 2021, 45(01) :81-89. doi: 10.16294/j.cnki.1007-659x.2021.01.014.

Table 1:

The number of differentially expressed genes in each subgroup was obtained by comparing the control group and the case group and the weighted gene co-expression analysis module.

Subtypes	The specific genes were compared with the normal group	Specific genes compared with other subgroups	Specific upregulated genes in subgroup	modular
I	6021	4848	2033	blue
II	3387	3169	2134	brown、green
III	3929	2274	804	
IV	6404	6407	1892	turquoise

Figures

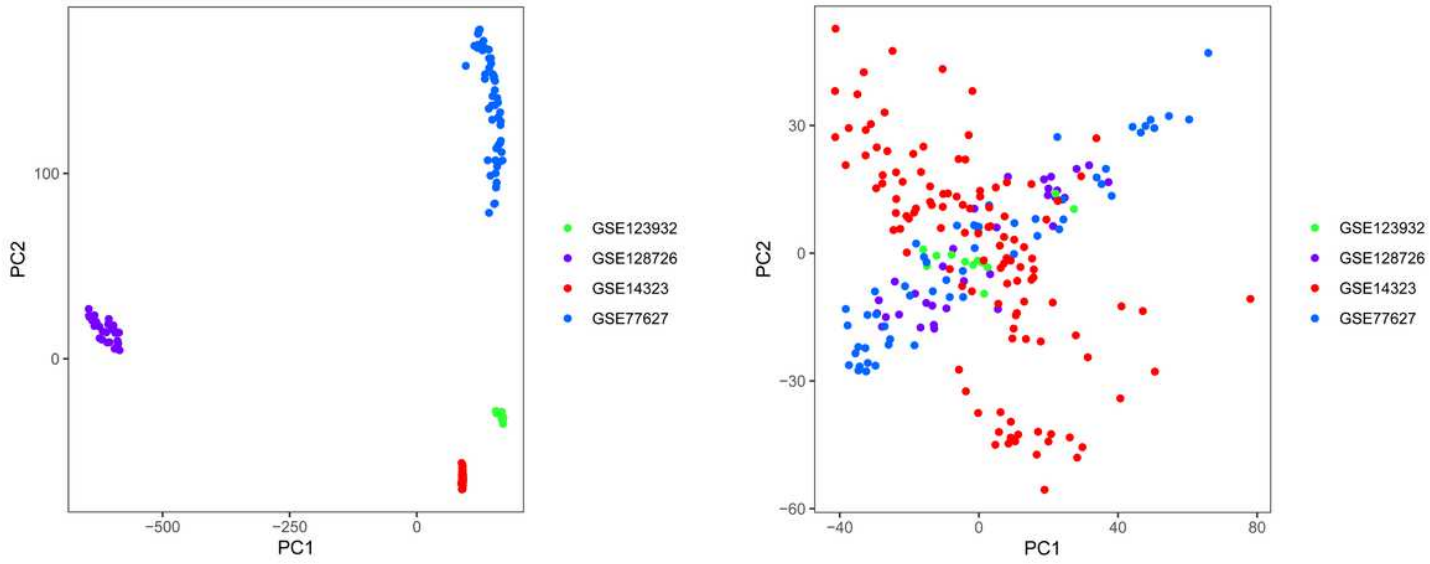

Figure 1

Removal of batch effect by cross-platform normalization

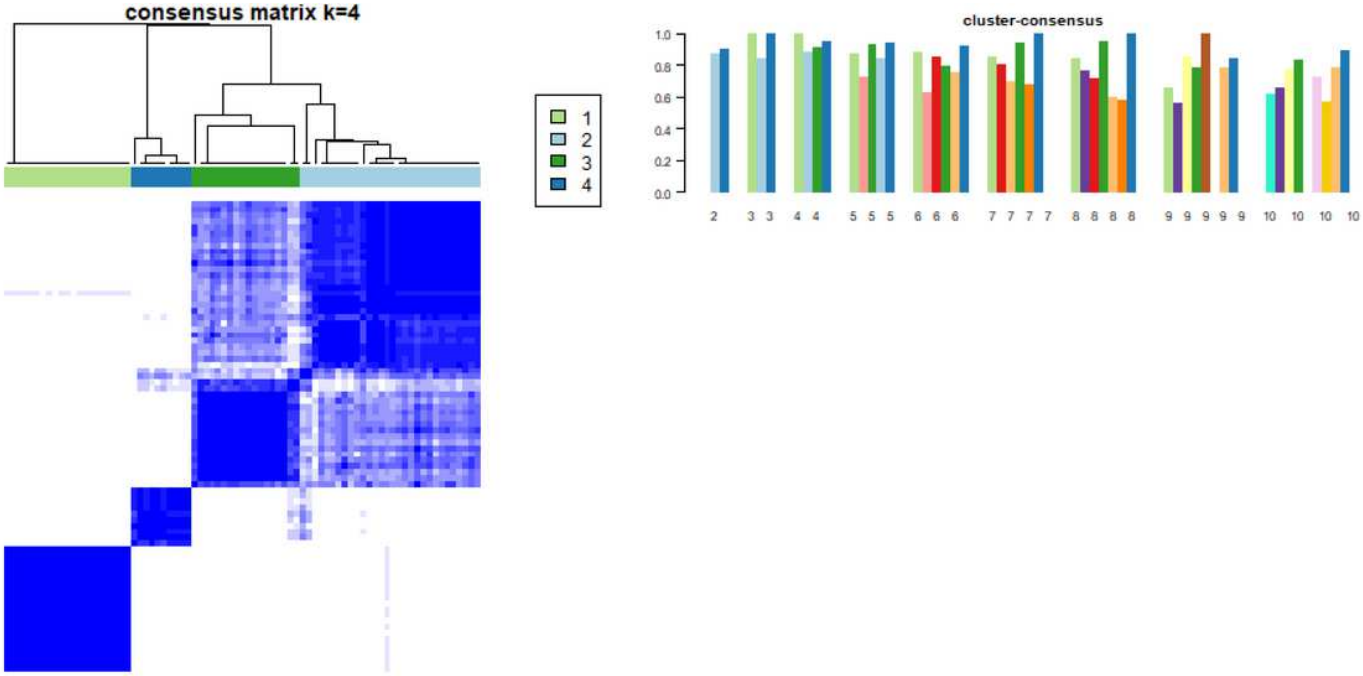

Figure 2

Consensus clustering of Cirrhosis cases

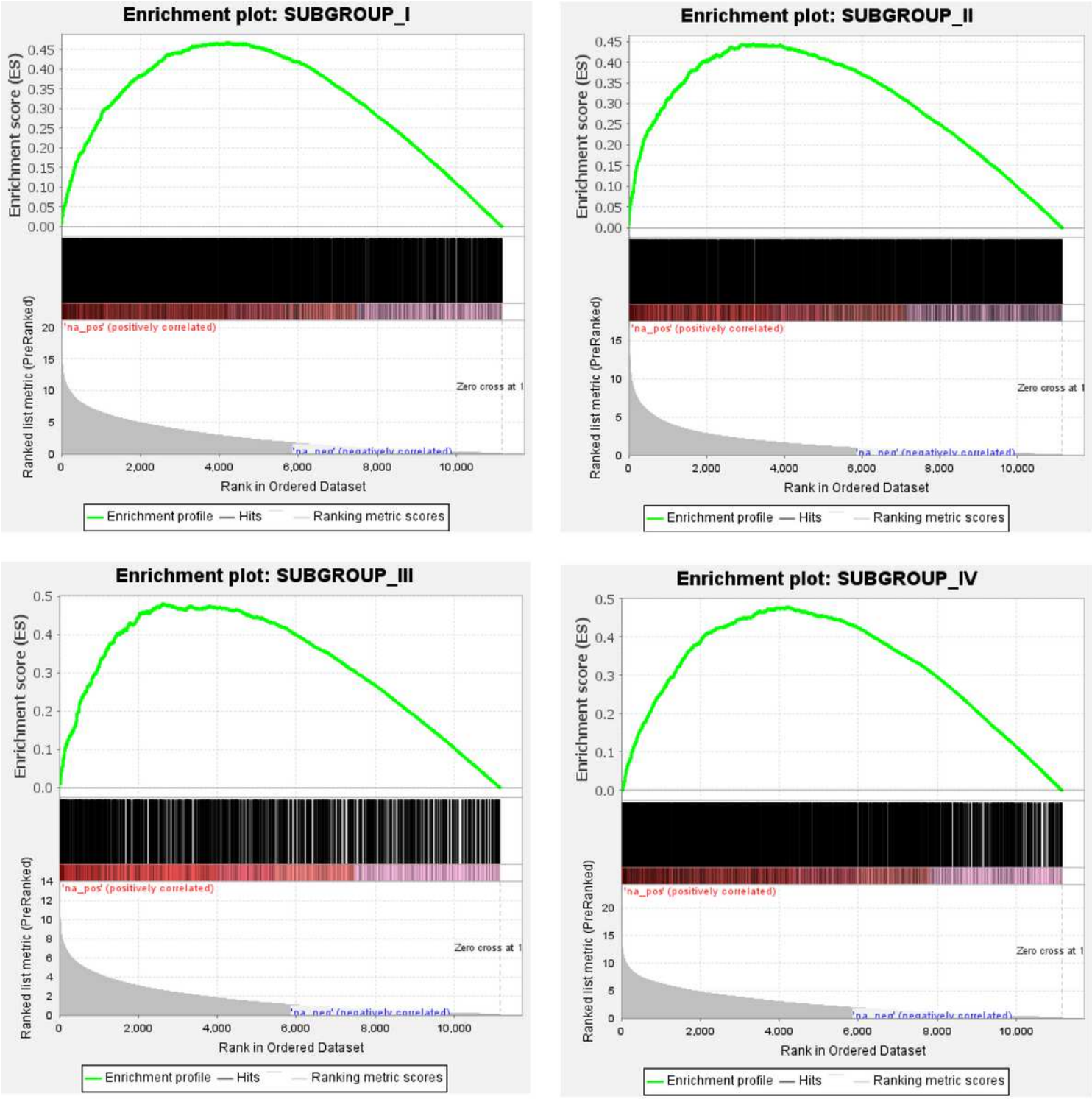

Figure 3

Identification of gene co‐expression modules for each subgroup

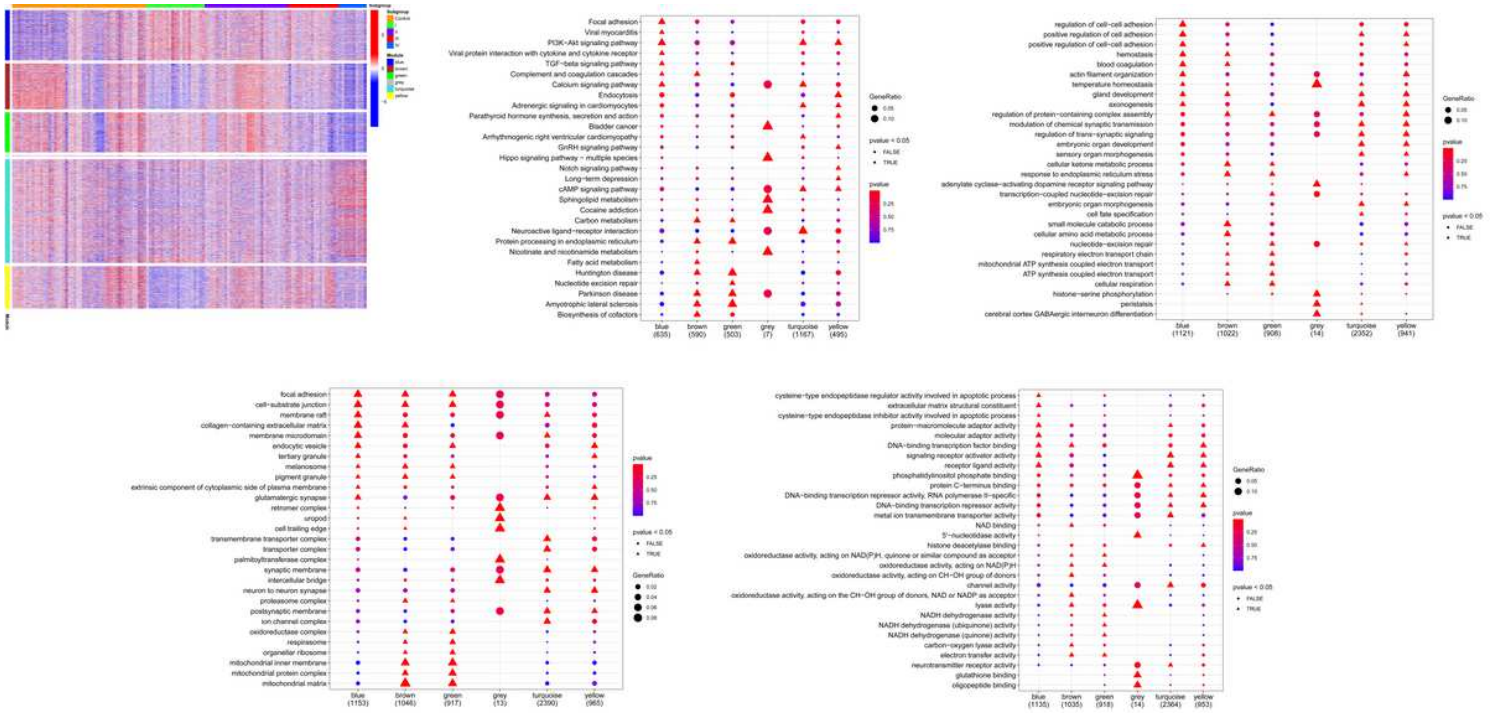

Figure 4

Identification of gene co‐expression modules for each subgroup