Identification of Hub Genes and Pathways in Colitis-Associated Colon Cancer by Integrated Bioinformatic Analysis

Jie Yao
Jining Hospital of Traditional Chinese Medicine
Huang Liu
Surgery Teaching and Research Section Jining, Jining Medical University
Peng Wang
Affiliated Hospital of Jining Medical University, Jining Medical University
Wenzhi Sheng
Jining Medical University
Yongming Huang (asdf101f@126.com)
Affiliated Hospital of Jining Medical University, Jining Medical University

---

Research Article

Keywords: colitis-associated colon cancer, differentially expressed genes, bioinformatical analysis, Protein-protein interaction network, Functional enrichment analysis

Posted Date: May 4th, 2021

DOI: https://doi.org/10.21203/rs.3.rs-405409/v1

Abstract

Background

Colon cancer is the third leading cause of death in the world. According to the etiology, colon cancer can be divided into three categories: sporadic, hereditary, and colitis-associated colon cancer(CAC). In terms of clinical features, CAC patients have younger age of onset, more multiple lesions, more difficult to find under endoscope, stronger tumor invasiveness than sporadic colorectal cancer patients. Early detection of CAC by endoscopic monitoring is a challenge, and the incidence of septal colorectal cancer is still high. So it is indispensable to find the biomarkers to predict tumorigenesis of CAC.

Methods

The gene expression profiles of GSE43338 and GSE4904 were downloaded from Gene Expression Omnibus (GEO) database. The Limma package was used to standardize the data and obtain differentially expressed genes (DEGs). Protein-protein interaction (PPI) network was evaluated by STRING database. The Cytoscape software was used to establish PPI network and to perform modular analysis. Functional annotation analysis were performed by the DAVID. The transcription factor (TF) and miRNAs that regulate genes expression were analyzed by using the Network Analyst algorithm. Expression correlation analysis were conducted in GEPIA. The prognostic analysis of hub genes were conducted based on TCGA samples.

Results

A total of 275 DEGs including 103 up-regulated genes and 172 down-regulated genes were identified by bioinformatics analysis in CAC. IGF1, BMP4, SPP1, APOB, CCND1, CD44, PTGS2, CFTR, BMP2, KLF4, TLR2 were identified as hub DEGs, which were significantly enriched in PI3K-Akt signaling pathway, stem cell pluripotency regulation pathway, Focal adhesion-Hippo signal pathway and AMPK signal pathway respectively. The FOXC1 were crucial regulators for these hub gene according to the TF analysis. The Sankey diagram showed that both PI3K-AKT signal pathway and Focal adhesion were composed of up-regulated genes, such as SPP1, CD44, TLR2, CCND1, IGF1, which were related to core miRNAs, hsa-mir-16-5p, hsa-mir-129-2-3p and hsa-mir-37a-3p by research and prediction. Survival analysis showed that the differential expression of SPP1, CFRT and KLF4 was significantly related to the poor prognosis of CAC.

Conclusion

This study provided a novel way to illuminate the mechanism of colitis-associated colon cancer. The most of all, the hub genes and signaling pathways may contribute to the prevention, diagnosis and treatment of CAC.

Background

Colon cancer is the third leading cause of death in the world. According to etiology, colon cancer can be divided into three categories: sporadic, hereditary, and colitis-associated colon cancer (CAC). CAC is one of the major complications of inflammatory bowel disease (IBD). Compared with the age-and sex-matched general population, patients with IBD have a 2-fold increased risk of colon cancer (1). With the ever-rising incidence of IBD and duration extension, the population of CAC was also increasing. Epidemiological data shows that the incidence of CAC is 0.64%-0.87%, however, 8%-16% of IBD patients die of colon cancer (2–4). In terms of clinical features, CAC patients have younger age of onset, more multiple lesions, more difficult to find under endoscope, stronger tumor invasiveness and poorer prognosis than sporadic colorectal cancer patients (5). Early detection of CAC by endoscopic monitoring is a challenge, and the incidence of septal colorectal cancer is still high. The discovery of specific molecular markers of CAC has become the focus of attention.

At present, with the rapid development of gene technology, gene microarray expression research and high-throughput sequencing technology are gradually mature. Bioinformatics technology is widely used in the study of the mechanism of various diseases. It not only enables us to understand the occurrence and development of diseases at the molecular level, but also provides a new and feasible method for finding specific biomarkers and therapeutic targets for diseases (6). Previously, the bioinformatics studies of CAC were mainly carried out on gene chips of ulcerative colitis and colon adenocarcinoma (7, 8), however, the gene of CAC was complex, which genetic correlation with ulcerative colitis is not so clear (9), and there are significant differences in genome-wide RNA profiles between sporadic CRC and CAC (10). Therefore, it is innovative and valuable to explore the abnormal signal pathways and important genes to research CAC. In this study, we download the original gene chip expression data set GSE43338 and GSE44904 from the publicly available Gene Expression Omnibus (GEO) database, and standardize the data to identify the differentially expressed gene (DEGs) between colitis-associated colon cancer and control tissues. In addition, this study provides a multi-level bioinformatics analysis strategy for differentially expressed genes, including modular analysis, functional enrichment analysis and screening of core genes by constructing protein-protein interaction network (PPI), construction of mulberry map of core genes, TFs and expression correlation analyses and miRNA network study by Network Analyst. The prognostic analysis of hub genes were conducted based on TCGA samples. This study may contribute to a better understanding of the mechanism of the occurrence and development of CAC. Above of all, highly specific hub genes and signaling pathways may contribute to the prevention, diagnosis and treatment of CAC.

Material And Methods

1. Acquisition and processing of gene expression set

The data sets of GSE44904 and GSE43338 gene expression profiles were downloaded from GEO database (Gene Expression Omnibus, https:// www.ncbi.nlm.nih.gov/geo). The platform of the dataset GSE44904 is GPL7202 (Agilent-014868 Whole Mouse Genome Microarray 4x44K G4122), which includes

AOM/DSS group (n=3), DSS group(n=3),AOM group(n=3)and Control group(n=3).The platform of data set GSE43338 was GPL339 ([MOE430A] Affymetrix Mouse Expression 430A Array). According to the needs of the study, CAC group(n=4), CAC control group(n=2) was selected. The R software limma package Version 4.0, (http://www.bioconductor.org/) (11) is used to calibrate the data, use the platform annotation file to annotate the probe, and remove the probe that does not match the gene (gene symbol). In addition, for multiple probes mapped to the same gene, the average value is calculated as the final expression value.

2. Screening and VENN analysis of DEGs

Compared with two or more groups of samples by the limma R package, the corrected $p < 0.05$and | log fold change (FC) | > 2 genes were considered to be DEGs. The up-regulated and down-regulated gene lists are saved as Excel files, and the TXT files of all gene lists sorted by logFC in each dataset are saved for subsequent analysis. The bioinformatics online tool (AlPuFu,www.aipufu.com) was used to analyze the data obtained by VENN. The up-regulated and down-regulated DEGs of the data set GSE44904 were screened by VENN, to screen the differential genes existing alone in the AOM/DSS group, and then the differential genes intersected with the up-regulated and down-regulated DEGs of the data set GSE43338 were used as the target DEGs, for follow-up analysis.

3. Construction of PPI protein interaction network and screening of important modules for DEGs

STRING (Search Tool for the Retrieval of Interacting Genes, http://strin gdb.org) is an online database that explores functional interactions between proteins encoded by differential genes and visualizes the PPI-protein interaction network of DEGs (12). We select the PPI relation pair with combined score > 0.4, eliminate the scattered PPI pairs, and map it to the network. Then, the PPI network diagram is constructed by using Cytoscape software, and the degree of connectivity is analyzed. Use the MCODE plugin in the Cytoscape software to filter the submodules based on the default parameters "Degree Cutoff = 2,"Node Score Cutoff = 0.2,"K-Core = 2" and "and" Max.Depth = 100 ". Hub gene was screened according to degrees score.

4. Functional enrichment analysis of genes

The database used for annotation, visualization, and integrated discovery (DAVID, http://david d.ncifc rf.gov/)is an online tool that provides a comprehensive set of functional annotation methods for a range of genes or proteins provided by researchers(13, 14). The identified genes were analyzed by GO annotation and KEGG pathway enrichment analysis using DAVID tool. $P < 0.05$ was selected as the threshold of enrichment condition, and the TXT text of the above analysis results was downloaded for further analysis.

5. Analysis of transcriptional factors (TFs) and miRNAs of hub genes

NetworkAnalyst3.0 (https://www.networkanalyst.ca) is a comprehensive network visual analysis platform for gene expression analysis and meta- analysis(15). TarBase v8. Database was used for miRNAs

correlation analysis, and the JASPAR database on the platform was used to analyze the TFs related to hub gene.

6. Correlation analysis of TF and hub genes

Based on GEPIA (Gene Expression Profiling Interactive Analysis, http://gepia.cance r-pku.cn/), the correlation between hub gene and predicted core transcription factors was evaluated. The correlation analysis in GEPIA uses Pearson, Spearman and Kendall methods to perform pairwise gene expression correlation analysis for given TCGA and / or GTEx expression data (16).

7. Survival analysis of hub genes

The survival analysis of the identified hub gene was carried out by using the online software UALCAN (http://ualcan.path.uab.edu/index.html),which uses TCGA Level 3 RNA-seq and clinical data from 31 cancer types. UALCAN can estimate the effect of gene expression levels and clinicopathologic features on patient survival(17).

Results

1. Microarray data normalization and identification of DEGs

The chip expression data set GSE44904 and GSE43338 row standardization analysis, the results are shown in Fig. 1. The limma R package (adjusted $p < 0.05$ and | log fold change (fc) | > 2) was used to screen DEGs. Firstly, different groups in GSE44904 were compared, and then through venn analysis, a total of 1063 DEGs were screened out, including 503 up-regulated genes and 560 down-regulated genes (Fig. 2g). A total of 905 DEGs, including 496 up-regulated genes and 409 down-regulated genes, were screened from the data set GSE43338. Different volcanoes are shown in Fig. 2a, 2b, 2c, 2d. Heat map is drawn for the first 100 DEGs shown in Fig. 2e, 2f. According to the differentially expressed genes screened by the two data sets, venn analysis was performed again, and 275 overlapping genes were found, including 103 up-regulated genes and 172 down-regulated genes (Fig. 2h).

2. GO functional and KEGG pathway enrichment analysis

In order to study the functional annotation of selected DEGs, three kinds of enrichment analysis of GO in DAVID were used, including biological process (BP), molecular function (MF) and cellular component (CC). The results were considered statistically significant if $P < 0.05$, and the three parts of the GO results are shown in Fig. 3c. Biological processes mainly include: positive regulation of transcription from RNA polymerase II promoter, oxidation-reduction process, negative regulation of transcription from RNA polymerase II promoter, negative regulation of cell proliferation, positive regulation of transcription, DNA-templated, cell proliferation, transport, inflammatory response, negative regulation of transcription, DNA-templated, cell adhesion, etc. Cell components mainly include: extracellular space, plasma membrane, extracellular exosome, extracellular region, integral component of plasma membrane, endoplasmic reticulum membrane, Golgi apparatus, Endoplasmic reticulum and others. Molecular functions mainly

include: hormone activity, transporter activity, calcium ion binding, receptor binding, heparin binding and oxidoreductase activity. In order to further understand the pathway enrichment function of DEGS, we then constructed the KEGG. As shown in the Fig. 3e, the pathway is mainly enriched in ovarian steroidogenesis,fat digestion and absorption, metabolic pathways, vitamin digestion and absorption, signaling pathways regulating pluripotency of stem cells, arachidonic acid metabolism, foxO signaling pathway□aldosterone-regulated sodium reabsorption, bile secretion, PI3K-Akt signaling pathway, pathways in cancer, ether lipid metabolism, etc.

3. PPI network and modularization analysis of DEGs

The STRING online database is used to analyze the 275 intersecting DEGs, the PPI network is constructed as the Fig. 3a, and then the Cytoscape software is used to analyze the data. The degree score of DEGs was calculated, and the first 11 genes with the highest score were selected as hub gene (Fig. 4a), which were IGF1, BMP4, SPP1, APOB, CCND1, CD44, PTGS2, CFTR, BMP2, KLF4, TLR2. The detailed information of hub gene, including gene symbol, degree, full name and gene function, is shown in Table 1. Then, the MCODE plugin in Cytoscape software was used for modular analysis, and the sub-modules with high scores were selected with Score = 9. Module genes were SPP1, Tgoln2, ApoB, FSTL1, LAMB1, LAMC1, CHGB, BMP4, and CYR61 (Fig. 3b). David Databas was used to perform GO function and KEGG pathway enrichment analysis on all genes in the module. It was showed the GO function analysis results about submodule genes in Fig. 3d. BP mainly include extracellular matrix organization, cell adhesion, positive regulation of epithelial cell proliferation, positive regulation of cell migration. CP mainly includes extracellular region, extracellular space and extracellular exosome. MF mainly include Heparin binding, extracellular matrix binding and so on. KEGG pathway analysis showed that it was mainly enriched in ECM-receptor interaction, focal adhesion, PI3K-Akt signaling pathway, pathways in cancer, such as small cell lung cancer(Fig. 3f).

Table 1

PPI network was built, and then Cytoscape software was used to analyze the data.The DEGS degree score was calculated, and the top 11 with the highest score were selected as HUB genes, which were IGF1, BMP4, SPP1, ApoB, CCND1, CD44, PTGS2, CFTR, BMP2, KLF4, and TLR2.Detailed information about the HUB gene, including gene symbol, degree, full name, and gene function.(Table 1).

Gene symbols	Degree	Full name	Gene function
IGF1	24	Insulin Like Growth Factor 1	The protein is a member of a family of proteins involved in mediating growth and development
BMP4	23	Bone Morphogenetic Protein 4	The encoded protein may also be involved in the pathology of multiple cardiovascular diseases and human cancers
SPP1	22	Secreted Phosphoprotein 1	This protein is a cytokine that upregulates expression of interferon-gamma and interleukin-12
APOB	22	Apolipoprotein B	The protein affects plasma cholesterol and apolipoprotein levels in diseases
CCND1	20	Cyclin D1	This gene alters cell cycle progression, are observed frequently in a variety of human cancers
CD44	18	CD44 Molecule	This protein participates in a wide variety of cellular functions including lymphocyte activation, recirculation and homing, hematopoiesis, and tumor metastasis
PTGS2	18	Prostaglandin-Endoperoxide Synthase 2	This protein is responsible for the prostanoid biosynthesis involved in inflammation and mitogenesis
CFTR	16	CF Transmembrane Conductance Regulator	The encoded protein acts as a chloride channel, and controls ion and water secretion and absorption in epithelial tissues
BMP2	16	Bone Morphogenetic Protein 2	This protein plays a role in bone and cartilage development
KLF4	14	Kruppel Like Factor 4	This protein controls the G1-to-S transition of the cell cycle following DNA damage by mediating the tumor suppressor gene p53
TLR2	14	Toll Like Receptor 2	The protein regulates host inflammation and promotes apoptosis in response to bacterial lipoproteins.
Gene symbols	**Degree**	**Full name**	**Gene function**
IGF1	24	Insulin Like Growth Factor 1	The protein is a member of a family of proteins involved in mediating growth and development
BMP4	23	Bone Morphogenetic Protein 4	The encoded protein may also be involved in the pathology of multiple cardiovascular diseases and human cancers

Gene symbols	Degree	Full name	Gene function
SPP1	22	Secreted Phosphoprotein 1	This protein is a cytokine that upregulates expression of interferon-gamma and interleukin-12
APOB	22	Apolipoprotein B	The protein affects plasma cholesterol and apolipoprotein levels in diseases
CCND1	20	Cyclin D1	This gene alters cell cycle progression, are observed frequently in a variety of human cancers
CD44	18	CD44 Molecule	This protein participates in a wide variety of cellular functions including lymphocyte activation, recirculation and homing, hematopoiesis, and tumor metastasis
PTGS2	18	Prostaglandin-Endoperoxide Synthase 2	This protein is responsible for the prostanoid biosynthesis involved in inflammation and mitogenesis
CFTR	16	CF Transmembrane Conductance Regulator	The encoded protein acts as a chloride channel, and controls ion and water secretion and absorption in epithelial tissues
BMP2	16	Bone Morphogenetic Protein 2	This protein plays a role in bone and cartilage development
KLF4	14	Kruppel Like Factor 4	This protein controls the G1-to-S transition of the cell cycle following DNA damage by mediating the tumor suppressor gene p53
TLR2	14	Toll Like Receptor 2	The protein regulates host inflammation and promotes apoptosis in response to bacterial lipoproteins.

4. Analysis of KEGG pathway of Hub genes

In order to understand the pathway analysis enriched by the hub gene, the KEGG pathway analysis of the hub gene was constructed by DAVID, it was showed that the pathway is mainly enriched in signaling pathways regulating many biological functions in Fig. 4b, such as pluripotency of stem cells, pathways in cancer, proteoglycans in cancer, AMPK signaling pathway, PI3K-Akt signaling pathway, hippo signaling pathway, focal adhesion. Sankey diagram shows the distribution of hub genes in different signaling pathways (Fig. 4c): signaling pathways regulating pluripotency of stem cells (Enriched genes: IGF1, BMP4, BMP2, KLF4; $p = 0.0015$), pathways in cancer(enriched genes: BMP4, BMP2, CCND1, IGF1, PTGS2; $p = 0.0035$), proteoglycans in cancer(enriched genes: CCND1, IGF1, CD44, TLR2; $p = 0.0043$), AMPK signaling pathway(enriched genes: CCND1, IGF1, CFTR; $p = 0.0186$), PI3K-Akt signaling pathway(enriched genes: CCND1, SPP1, IGF1, TLR2; $p = 0.0196$), Hippo signaling pathway(enriched genes: BMP4, BMP2, CCND1; $p = 0.0273$), Focal adhesion(enriched genes: CCND1, SPP1, IGF1; $p = 0.0483$).

5. Analysis of transcription factor regulatory network of Hub genes

In order to identify the transcriptional regulation of hub gene and evaluate the effect of TF on hub gene expression, a TF-gene regulatory network was constructed based on the JASPAR database on Network Analyst platform. Figure 4d shows transcription factors that can regulate two or more genes. In addition to the hub gene, there are 46 transcription factors in the regulatory network, and 86 relationship pairs have been established. Among the predicted transcription factors, FOXC1 is considered to be the core TFs, that can regulate multiple genes: SPP1, IGF1, BMP4, TLR2, CD44, KLF4, CFTR. The correlation analysis was described in Fig. 5. The results showed that the expression of these up-regulated genes SPP1, IGF1, BMP4, TLR2, CD44, which were positively correlated with FOXC1, while the expression of down-regulated genes KLF4, CFTR, which was negatively correlated with FOXC1.

In order to further investigate some up-regulated genes, we performed gene-miRNA interactions with Tarbase V8.0 of NetworkAnalyst3.0, and these gene-miRNA could regulate more than three genes(Fig. 4e) showed a total of 5 genes, 259 miRNAs and 322 gene-miRNA pairs were registered in the network, and it was found that some miRNAs played an important role in regulating hub genesIt was predicted that hsa-miR-16-5p and hsa-miR-27a-3p could regulate CCND1, CD44, IGF1, SPP1 and hsa-miR-129-2-3p could regulate CCND1, CD44, SPP1, TLR2.

6. Survival analysis of Hub genes in colon cancer

Considering CAC as an etiological classification of colon cancer, we use colon cancer data from TCGA database to analyze the survival of hub gene. The results were showed by Fig. 6. Survival analysis data contains information of high or low expression of target gene and the correlation between hub gene and colon cancer. Among the eleven hub genes, three hub genes were found to be related to the survival of colon cancer patients, including SPP1 ($p = 0.019$), CFTR ($p = 0.031$) and KLF4 ($p = 0.048$).

Discussion

Not all patients with inflammatory bowel disease can develop into CAC, so comparing the differentially expressed genes in the CAC model and the differentially expressed genes in the IBD model may enable us to understand the pathogenesis of colon cancer based on the Colitis-Associated Cancer. In this study, the data of GEO database (GSE44904 and GSE43338) were standardized, then different groups of GSE44904 data sets were analyzed, and the differential genes in AOM/DSS group, AOM group and DSS group were screened. Then through venn analysis, the differential genes expressed alone in colitis associated colon cancer (AOM/DSS) were screened. Through intersection analysis with gene microarray data of CAC animal model in GSE43338 data set, a total of 275 specific genes (including 103 up-regulated genes and 172 down-regulated genes) were found to be expressed in CAC. The selected DEGs were analyzed for functional enrichment, including GO analysis and KEGG signal analysis. It was found that some biological processes and functions were associated with colitis associated colon cancer to varying degrees, such as, regulation of transcription from RNA polymerase II promoter, reduction process, cell proliferation, inflammatory response, cell adhesion, extracellular space, plasma membrane, extracellular exosome, transporter activity, calcium ion binding, receptor binding. Furthermore, the modular analysis of

the differential genes shows that the enrichment results of the genes in the submodules with the highest scores also verify the importance of these biological processes and functions to a certain extent. In the analysis of KEGG pathway, we found that a large number of differential genes are enriched in Metabolicpathways, which is consistent with the published studies (18). LuY and WangJ found that there were many metabolic pathway changes in colon cancer induced by AOM/DSS through metabonomics analysis (19). Our study also found that Fat digestion and absorption, Ovarian steroidogenesis, Vitamin digestion and absorption, Arachidonic acid metabolism and Ether lipid metabolism and other metabolic pathways are closely related to the occurrence and development of colitis-associated colon cancer.

However, what interests us is that in the signaling pathway of DEGs enrichment, in addition to the metabolic pathway, a large number of differential genes are enriched in Pathways in cancer, Signaling pathways regulating pluripotency of stem cells, PI3K-Akt signaling pathway, FoxO signaling pathway. Then, through the sub module analysis, we selected the genes in the module with the highest score, and analyzed the KEGG pathway of these genes. We found that the pathway is similar to the Pathways enriched by DEGs, such as Pathways in cancer, PI3K-Akt signaling pathway and Focal adhesion pathway. The results suggest that these pathways and their genes play a key role in the occurrence and development of CAC. Focal adhesion is the contact point between cells and the surrounding environment, which can drive cell migration. This signal pathway plays an important role in wound healing and tumor metastasis. It has been found that the low expression of miR-4728-3p in ulcerative colitis-associated colorectal cancer can act on the CAV1, THBS2, and COL1A2 gene on focal adhesion signaling, which is related to the pathogenesis of the tumor (20). Li J and Wang D found that activation of adhesion macular kinase prevented the development of ulcerative colitis and colitis associated tumors (21). PPI network analysis was further performed on DEGs. According to the degree score value, we identified the differential genes with the highest score and significance as hub genes, which were BMP4, SPP1, APOB, CCND1, CD44, PTGS2, CFTR, BMP2, KLF4, TLR2 and IGF1. By analyzing the KEGG pathway of hub genes, it was found that these genes were not only enriched in Pathways in cancer, Signaling pathways regulating pluripotency of stem cells, PI3K-Akt signaling pathway, Focal adhesion, but also enriched in Hippo signal pathway and AMPK signal pathway. It is emphasized that these genes and their enriched pathways are closely related to the occurrence and development of CAC. Pluripotency is characteristic of stem cells, and a small number of cells in tumors have the ability to renew themselves and produce heterogeneous tumor masses (22). P53 can inhibit pluripotency of tumor stem cells and thus mediate cancer stem cell function. In a preclinical animal model of colitis associated colorectal cancer, targeted knockout of stem cell specific P53 was found to significantly increase tumor size and incidence (23). Josse C also found that PI3K/Akt is the main pathway affected by AOM/DSS model through miRNA chip experiment (24). This is consistent with our research. In inflammatory infiltrated human colon tissue, the PI3K/Akt pathway is activated, which mediates the progression of colitis and CAC through a positive feedback loop that maintains the recruitment of inflammatory cells (25).

In inflammation-related tumor models, inhibiting IGF1 signaling can reduce the number and size of colon tumors in wild-type mice (26).The results of IgF-1R knockout can activate the LKB1/AMPK pathway and play a protective role in colitis and CAC (27). Chen G found that Extract of Ilex rotunda can affect the

abnormal expression of YAP gene in the Hippo pathway in experimental CAC mouse model, and has anti-inflammatory and preventive effects on CAC (28). Yes-associatedprotein1(YAP1) is a transcriptional coactivator in the Hippo signaling pathway. PGE2 signal can increase the expression and transcriptional activity of YAP1, and YAP1 further activates PTGS2 and PTGER4, which in turn can activate PGE2. This positive feedback loop plays an important role in colon regeneration and promotes the canceration of colitis-related cancer and colon cancer (29). In the study of azoxomeane/dextran sodium sulfate (AOM/DSS) induced colitis associated colon cancer in mice (30). Boswellia serrata (BS) resin extract changed intestinal flora composition and alleviated tumor growth by mediating Akt/GSK3/Cyclin D1 signaling pathway. In a mouse model of colitis associated colorectal cancer, Mattaveewongt demonstrated that the size and number of tumors can be reduced by increasing AMPK activity, inhibiting NF-κB-regulated inflammatory response, and decreasing CCND1 expression (31).

Furthermore, other hub genes are significantly related to the development of colitis-associated colon cancer. For example, abnormal BMPs protein is a common feature of cancer. In colon mucosa, BMP pathway overlaps with many molecular pathways of colon cancer (32). George S and Karagiannis found that the inhibition of BMP pathway is an early event of inflammation-driven colon tumor in mice (33). TLR2 is highly expressed in tumor tissues of CRC patients. Gene knockout and knockdown of TLR2 can inhibit the proliferation of inflammation-related colorectal cancer and sporadic colorectal cancer (34). SPP1 is an important inflammatory mediator. It is up-regulated in inflammation-related intestinal tumors and mediates the progression of colon cancer (35). Yang VW and Liu Y found that the deletion of KLF4 causes genetic instability, which in turn leads to the progress of CAC (36). The mutation of APOB gene in CRC associated with ulcerative colitis was found by whole exon sequencing, and there was significant difference between ulcerative colitis associated CRC and scattered CRC (37). CD44 is an adhesion and anti-apoptotic molecule that is highly expressed in colon cancer (38). However, in a comparative study, CD44 expression was found to be lower in ulcerative colitis associated dysplasia and cancers than in sporadic colonic tumors (39).

Finally, the results of TFs gene pretest analysis showed that FOXC1, FOXL1, NFKB1, STAT3, JUN, E2F1, CREB1, GATA2 was significantly related to the regulatory network of hub gene. Recent studies have emphasized the important role of transcription factor nuclear factor kappa B (NF-κB) and signal transducer and activator of transcription 3 (STAT3) in the progression of inflammation-associated cancer (40 ~ 41). Meanwhile, transcription factors Jun (42), E2F1 (43) and GATA2 (44) have been reported to be closely related to the occurrence and development of colitis associated tumors. FoxC1, as a core transcription factor, interacts most closely with hub genes. FoxC1 belongs to the fork head box (FOX) transcription factor family. A large number of studies have confirmed that at least 14 proteins in FOX transcription factor family are closely related to the pathogenesis of colorectal cancer (45). At present, as a new cancer marker and therapeutic target, the regulatory role of FOXC1 in many types of cancer has been widely studied (46). The lack of research in CAC is worthy of our continuous attention.

To sum up, based on GSE44904 and GSE43338 data sets, bioinformatics analysis identified 275 DEGs in CAC, including 103 up-regulated genes and 172 down-regulated genes. IGF1, BMP4, SPP1, APOB, CCND1,

CD44, PTGS2, CFTR, BMP2, KLF4 and TLR2 are hub proteins, which are mainly related to PI3K-Akt signaling pathway, Focal adhesion, Hippo signaling pathway, AMPK signaling pathway and stem cell pluripotency regulation pathway. A study on TF-genes regulatory network of hub gene shows that the predicted transcription factors: FOXC1, FOXL1, NFkb1, STAT3, JUN, E2F1, CREB1 and GATA2 can regulate many genes, among which FOXC1 is the core transcription factor, which has the closest interaction with hub gene. According to the distribution of hub gene in different signal pathways, it was found that both PI3K-AKT signal pathway and Focal adhesion were composed of up-regulated genes, which were SPP1, CD44, TLR2, CCND1, IGF1, respectively. These up-regulated genes are mainly polysaccharides in cancer, and their related network studies predict core miRNAs, such as hsa-mir-16-5p, hsa-mir-27a-3p, etc., which provides more evidence for elucidating the mechanism of CAC progress. Survival analysis showed that the differential expression of SPP1, CFRT and KLF4 was significantly related to the poor prognosis of CAC. This study suggests that the hub gene may play an important role in the pathogenesis of CAC, and may serve as a potential biomarker for the diagnosis and treatment of CAC in the future.

Abbreviations

CAC: Colitis-associated colon cancer; GEO: Gene expression omnibus; DEGs: Differentially expressed genes; PPI: Protein-protein interaction; BP: Biological processes; CC: Cell component; MF: Molecular function; TF: Transcription factor ; IGF1: Insulin like growth factor 1; BMP4: Bone morphogenetic protein 4; SPP1: Secreted phosphoprotein 1; APOB: Apolipoprotein B; CCND1: Cyclin D1; CD44: CD44 molecule; PTGS2: Prostaglandin-endoperoxide synthase 2; CFTR: CF transmembrane conductance regulator; BMP2: Bone morphogenetic protein 2; KLF4: Kruppel like factor 4; TLR2:Toll like receptor 2;

Declarations

Acknowledgements and Authors' contributions

Sincerely thank the GEO and TCGA platforms and the authors who uploaded the original data. In addition, Thanks to the publisher for supporting this article.This article was done in collaboration with all the following authors. HYM and SWZ determined the research theme and formulated the main research plan.YJ and LH analyzed the data, and wrote the manuscript. WP helped collect data and references. All authors read and approved the final manuscript.

Funding

Not applicable.

Availability of data and materials

All data generated or analysed during this study are included in this published article

Ethics approval and consent to participate

Not applicable.

Consent for publication

Not applicable.

Competing interests

All authors declare no conflict of interest in this study.

References

1. Lutgens MW, van Oijen MG, van der Heijden GJ, Vleggaar FP, Siersema PD, Oldenburg B: Declining risk of colorectal cancer in inflammatory bowel disease: an updated meta-analysis of population-based cohort studies. Inflamm Bowel Dis 2013, 19(4):789-799.
2. Chu TPC, Moran GW, Card TR: The Pattern of Underlying Cause of Death in Patients with Inflammatory Bowel Disease in England: A Record Linkage Study. J Crohns Colitis 2017, 11(5):578-585.
3. Gong W, Lv N, Wang B, Chen Y, Huang Y, Pan W, Jiang B: Risk of ulcerative colitis-associated colorectal cancer in China: a multi-center retrospective study. Dig Dis Sci 2012, 57(2):503-507.
4. Eaden JA, Abrams KR, Mayberry JF: The risk of colorectal cancer in ulcerative colitis: a meta-analysis. Gut 2001, 48(4):526-535.
5. Dobbins WO, 3rd: Dysplasia and malignancy in inflammatory bowel disease. Annu Rev Med 1984, 35:33-48.
6. Fan L, Hui X, Mao Y, Zhou J: Identification of Acute Pancreatitis-Related Genes and Pathways by Integrated Bioinformatics Analysis. Dig Dis Sci 2020, 65(6):1720-1732.
7. Shi W, Zou R, Yang M, Mai L, Ren J, Wen J, Liu Z, Lai R: Analysis of Genes Involved in Ulcerative Colitis Activity and Tumorigenesis Through Systematic Mining of Gene Co-expression Networks. Front Physiol 2019, 10:662.
8. Zhou J, Xie Z, Cui P, Su Q, Zhang Y, Luo L, Li Z, Ye L, Liang H, Huang J: SLC1A1, SLC16A9, and CNTN3 Are Potential Biomarkers for the Occurrence of Colorectal Cancer. Biomed Res Int 2020, 2020:1204605.
9. Shawki S, Ashburn J, Signs SA, Huang E: Colon Cancer: Inflammation-Associated Cancer. Surg Oncol Clin N Am 2018, 27(2):269-287.
10. Colliver DW, Crawford NP, Eichenberger MR, Zacharius W, Petras RE, Stromberg AJ, Galandiuk S: Molecular profiling of ulcerative colitis-associated neoplastic progression. Exp Mol Pathol 2006, 80(1):1-10.
11. Ritchie ME, Phipson B, Wu D, Hu Y, Law CW, Shi W, Smyth GK: limma powers differential expression analyses for RNA-sequencing and microarray studies. Nucleic Acids Res 2015, 43(7):e47.

12. Szklarczyk D, Franceschini A, Kuhn M, Simonovic M, Roth A, Minguez P, Doerks T, Stark M, Muller J, Bork P et al: The STRING database in 2011: functional interaction networks of proteins, globally integrated and scored. Nucleic Acids Res 2011, 39(Database issue):D561-568.
13. Huang da W, Sherman BT, Lempicki RA: Systematic and integrative analysis of large gene lists using DAVID bioinformatics resources. Nat Protoc 2009, 4(1):44-57.
14. Huang da W, Sherman BT, Lempicki RA: Bioinformatics enrichment tools: paths toward the comprehensive functional analysis of large gene lists. Nucleic Acids Res 2009, 37(1):1-13.
15. Zhou G, Soufan O, Ewald J, Hancock REW, Basu N, Xia J: NetworkAnalyst 3.0: a visual analytics platform for comprehensive gene expression profiling and meta-analysis. Nucleic Acids Res 2019, 47(W1):W234-W241.
16. Tang Z, Li C, Kang B, Gao G, Li C, Zhang Z: GEPIA: a web server for cancer and normal gene expression profiling and interactive analyses. Nucleic Acids Res 2017, 45(W1):W98-W102.
17. Chandrashekar DS, Bashel B, Balasubramanya SAH, Creighton CJ, Ponce-Rodriguez I, Chakravarthi B, Varambally S: UALCAN: A Portal for Facilitating Tumor Subgroup Gene Expression and Survival Analyses. Neoplasia 2017, 19(8):649-658.
18. Gao Y, Li X, Yang M, Zhao Q, Liu X, Wang G, Lu X, Wu Q, Wu J, Yang Y et al: Colitis-accelerated colorectal cancer and metabolic dysregulation in a mouse model. Carcinogenesis 2013, 34(8):1861-1869.
19. Lu Y, Wang J, Ji Y, Chen K: Metabonomic Variation of Exopolysaccharide from Rhizopus nigricans on AOM/DSS-Induced Colorectal Cancer in Mice. Onco Targets Ther 2019, 12:10023-10033.
20. Pekow J, Hutchison AL, Meckel K, Harrington K, Deng Z, Talasila N, Rubin DT, Hanauer SB, Hurst R, Umanskiy K et al: miR-4728-3p Functions as a Tumor Suppressor in Ulcerative Colitis-associated Colorectal Neoplasia Through Regulation of Focal Adhesion Signaling. Inflamm Bowel Dis 2017, 23(8):1328-1337.
21. Li J, Lu Y, Wang D, Quan F, Chen X, Sun R, Zhao S, Yang Z, Tao W, Ding D et al: Schisandrin B prevents ulcerative colitis and colitis-associated-cancer by activating focal adhesion kinase and influence on gut microbiota in an in vivo and in vitro model. Eur J Pharmacol 2019, 854:9-21.
22. Sharif T, Martell E, Dai C, Kennedy BE, Murphy P, Clements DR, Kim Y, Lee PW, Gujar SA: Autophagic homeostasis is required for the pluripotency of cancer stem cells. Autophagy 2017, 13(2):264-284.
23. Davidson LA, Callaway ES, Kim E, Weeks BR, Fan YY, Allred CD, Chapkin RS: Targeted Deletion of p53 in Lgr5-Expressing Intestinal Stem Cells Promotes Colon Tumorigenesis in a Preclinical Model of Colitis-Associated Cancer. Cancer Res 2015, 75(24):5392-5397.
24. Josse C, Bouznad N, Geurts P, Irrthum A, Huynh-Thu VA, Servais L, Hego A, Delvenne P, Bours V, Oury C: Identification of a microRNA landscape targeting the PI3K/Akt signaling pathway in inflammation-induced colorectal carcinogenesis. Am J Physiol Gastrointest Liver Physiol 2014, 306(3):G229-243.
25. Khan MW, Keshavarzian A, Gounaris E, Melson JE, Cheon EC, Blatner NR, Chen ZE, Tsai FN, Lee G, Ryu H et al: PI3K/AKT signaling is essential for communication between tissue-infiltrating mast cells, macrophages, and epithelial cells in colitis-induced cancer. Clin Cancer Res 2013, 19(9):2342-2354.

26. Youssif C, Cubillos-Rojas M, Comalada M, Llonch E, Perna C, Djouder N, Nebreda AR: Myeloid p38alpha signaling promotes intestinal IGF-1 production and inflammation-associated tumorigenesis. EMBO Mol Med 2018, 10(7).
27. Wang SQ, Yang XY, Cui SX, Gao ZH, Qu XJ: Heterozygous knockout insulin-like growth factor-1 receptor (IGF-1R) regulates mitochondrial functions and prevents colitis and colorectal cancer. Free Radic Biol Med 2019, 134:87-98.
28. Chen G, Han Y, Feng Y, Wang A, Li X, Deng S, Zhang L, Xiao J, Li Y, Li N: Extract of Ilex rotunda Thunb alleviates experimental colitis-associated cancer via suppressing inflammation-induced miR-31-5p/YAP overexpression. Phytomedicine 2019, 62:152941.
29. Kim HB, Kim M, Park YS, Park I, Kim T, Yang SY, Cho CJ, Hwang D, Jung JH, Markowitz SD et al: Prostaglandin E2 Activates YAP and a Positive-Signaling Loop to Promote Colon Regeneration After Colitis but Also Carcinogenesis in Mice. Gastroenterology 2017, 152(3):616-630.
30. Chou YC, Suh JH, Wang Y, Pahwa M, Badmaev V, Ho CT, Pan MH: Boswellia serrata resin extract alleviates azoxymethane (AOM)/dextran sodium sulfate (DSS)-induced colon tumorigenesis. Mol Nutr Food Res 2017, 61(9).
31. Mattaveewong T, Wongkrasant P, Chanchai S, Pichyangkura R, Chatsudthipong V, Muanprasat C: Chitosan oligosaccharide suppresses tumor progression in a mouse model of colitis-associated colorectal cancer through AMPK activation and suppression of NF-kappaB and mTOR signaling. Carbohydr Polym 2016, 145:30-36.
32. Hardwick JC, Kodach LL, Offerhaus GJ, van den Brink GR: Bone morphogenetic protein signalling in colorectal cancer. Nat Rev Cancer 2008, 8(10):806-812.
33. Karagiannis GS, Afaloniati H, Karamanavi E, Poutahidis T, Angelopoulou K: BMP pathway suppression is an early event in inflammation-driven colon neoplasmatogenesis of uPA-deficient mice. Tumour Biol 2016, 37(2):2243-2255.
34. Meng S, Li Y, Zang X, Jiang Z, Ning H, Li J: Effect of TLR2 on the proliferation of inflammation-related colorectal cancer and sporadic colorectal cancer. Cancer Cell Int 2020, 20:95.
35. Bahri R, Pateras IS, D'Orlando O, Goyeneche-Patino DA, Campbell M, Polansky JK, Sandig H, Papaioannou M, Evangelou K, Foukas PG et al: IL-15 suppresses colitis-associated colon carcinogenesis by inducing antitumor immunity. Oncoimmunology 2015, 4(9):e1002721.
36. Yang VW, Liu Y, Kim J, Shroyer KR, Bialkowska AB: Increased Genetic Instability and Accelerated Progression of Colitis-Associated Colorectal Cancer through Intestinal Epithelium-specific Deletion of Klf4. Mol Cancer Res 2019, 17(1):165-176.
37. Yan P, Wang Y, Meng X, Yang H, Liu Z, Qian J, Zhou W, Li J: Whole Exome Sequencing of Ulcerative Colitis-associated Colorectal Cancer Based on Novel Somatic Mutations Identified in Chinese Patients. Inflamm Bowel Dis 2019, 25(8):1293-1301.
38. Subramaniam V, Vincent IR, Gardner H, Chan E, Dhamko H, Jothy S: CD44 regulates cell migration in human colon cancer cells via Lyn kinase and AKT phosphorylation. Exp Mol Pathol 2007, 83(2):207-215.

39. Mikami T, Mitomi H, Hara A, Yanagisawa N, Yoshida T, Tsuruta O, Okayasu I: Decreased expression of CD44, alpha-catenin, and deleted colon carcinoma and altered expression of beta-catenin in ulcerative colitis-associated dysplasia and carcinoma, as compared with sporadic colon neoplasms. Cancer 2000, 89(4):733-740.
40. Zhang HX, Xu ZS, Lin H, Li M, Xia T, Cui K, Wang SY, Li Y, Shu HB, Wang YY: TRIM27 mediates STAT3 activation at retromer-positive structures to promote colitis and colitis-associated carcinogenesis. Nat Commun 2018, 9(1):3441.
41. Callejas BE, Mendoza-Rodriguez MG, Villamar-Cruz O, Reyes-Martinez S, Sanchez-Barrera CA, Rodriguez-Sosa M, Delgado-Buenrostro NL, Martinez-Saucedo D, Chirino YI, Leon-Cabrera SA et al: Helminth-derived molecules inhibit colitis-associated colon cancer development through NF-kappaB and STAT3 regulation. Int J Cancer 2019, 145(11):3126-3139.
42. Liu ZY, Wu B, Guo YS, Zhou YH, Fu ZG, Xu BQ, Li JH, Jing L, Jiang JL, Tang J et al: Necrostatin-1 reduces intestinal inflammation and colitis-associated tumorigenesis in mice. Am J Cancer Res 2015, 5(10):3174-3185.
43. Kang DW, Choi CY, Cho YH, Tian H, Di Paolo G, Choi KY, Min do S: Targeting phospholipase D1 attenuates intestinal tumorigenesis by controlling beta-catenin signaling in cancer-initiating cells. J Exp Med 2015, 212(8):1219-1237.
44. Zhong L, Huot J, Simard MJ: p38 activation induces production of miR-146a and miR-31 to repress E-selectin expression and inhibit transendothelial migration of colon cancer cells. Sci Rep 2018, 8(1):2334.
45. Laissue P: The forkhead-box family of transcription factors: key molecular players in colorectal cancer pathogenesis. Mol Cancer 2019, 18(1):5.
46. Han B, Bhowmick N, Qu Y, Chung S, Giuliano AE, Cui X: FOXC1: an emerging marker and therapeutic target for cancer. Oncogene 2017, 36(28):3957-3963.

Figures

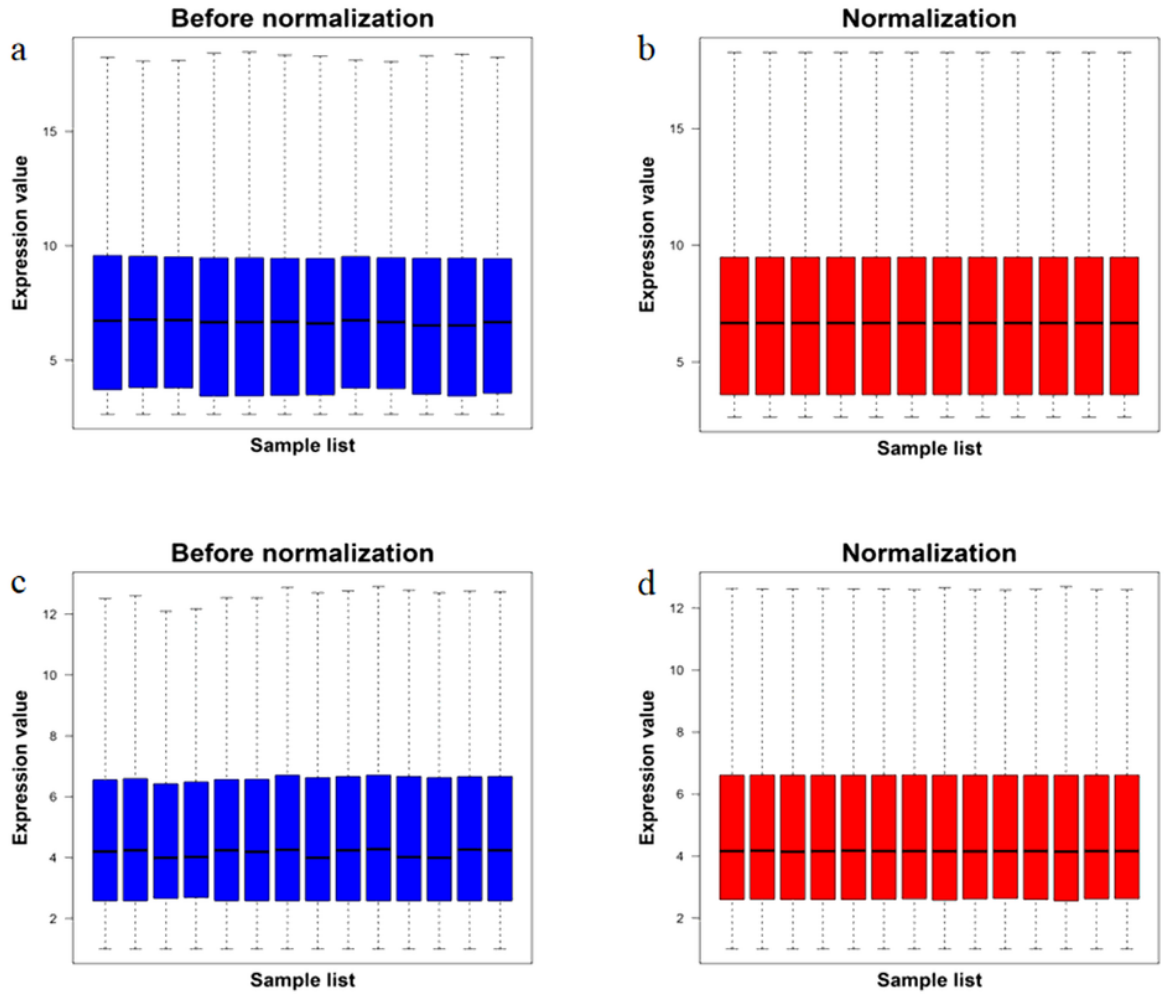

Figure 1

Standardized gene expression was showed in fig.1. The standardization of GSE44904 data set was showed in fig1.a and b. The normalization of GSE43338 dataset was showed in fig 1 c and d.

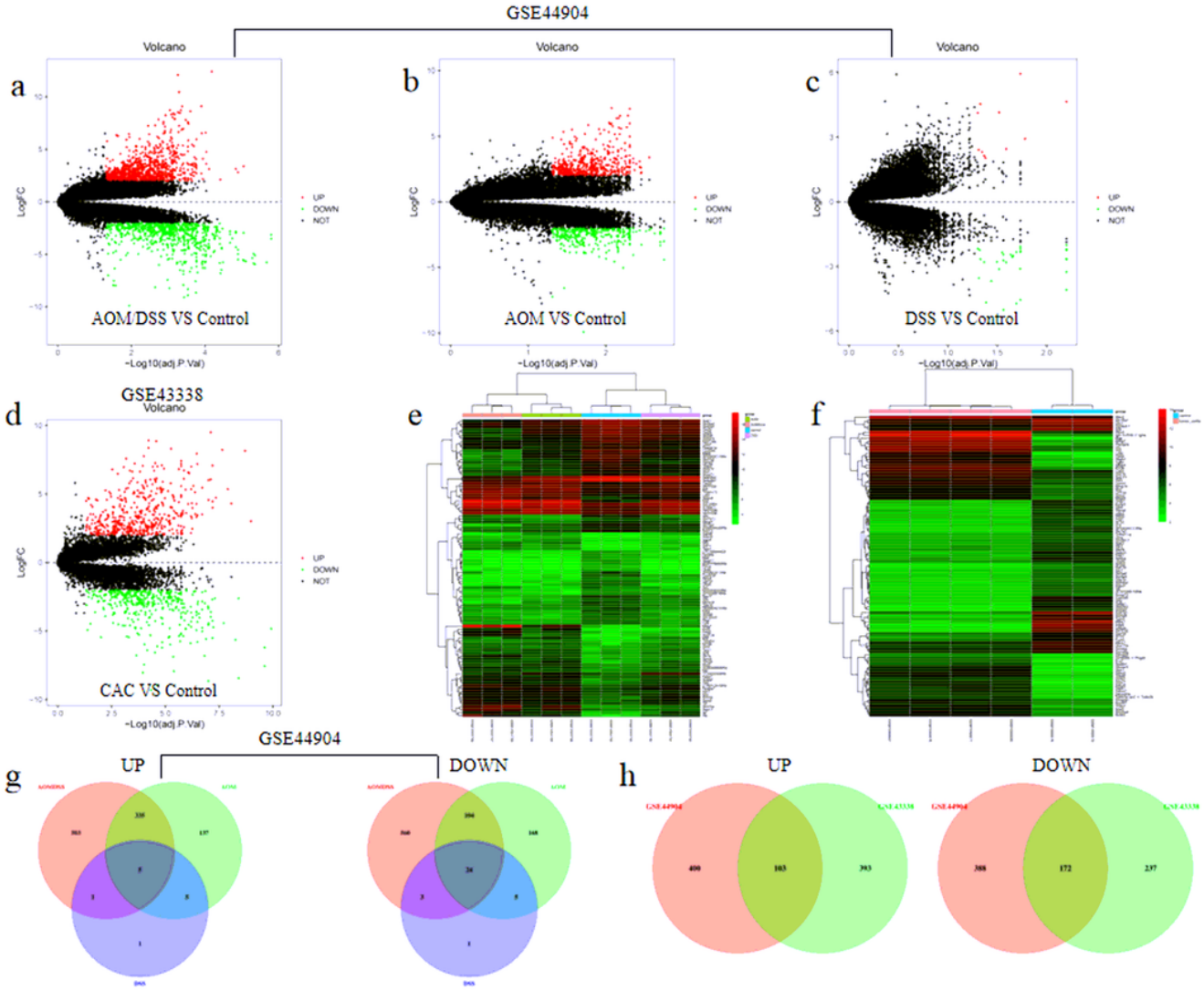

Figure 2

Identification of differentially expressed genes from two dataset chips was showed in figure 2. Fig a, fig b , fig c and fig d showed different groups in GSE44904 dataset, AOM/DSS VS Control group, AOM VS Control group, DSS VS Control group, CAC VS Control group respectively. Adjusted P <0.05 and | log a fold change | & gt;2. Differential genes were screened. Red dots represented up-regulated genes, green dots represented down-regulated genes, and black dots represented genes with no significant difference. (e) Heat maps of the top 100 differentially expressed genes in GSE44904 and (f) GSE43338 datasets. The red represents a relatively up-regulated gene, the green represents a relatively down-regulated gene, the black represents no significant change in the gene, and the depth of the color indicates the degree of gene expression. It was showed that in the Venn diagrams between different groups in dataset GSE44904, 503 up-regulated genes and 560 down-regulated genes were found in the AOM/DSS group alone in fig 2g. Meanwhile it was showed that through the Venn diagrams of differentially expressed genes in datasets GSE44904 and GSE43338, 103 up-regulated genes and 172 down-regulated genes were found in the two datasets.

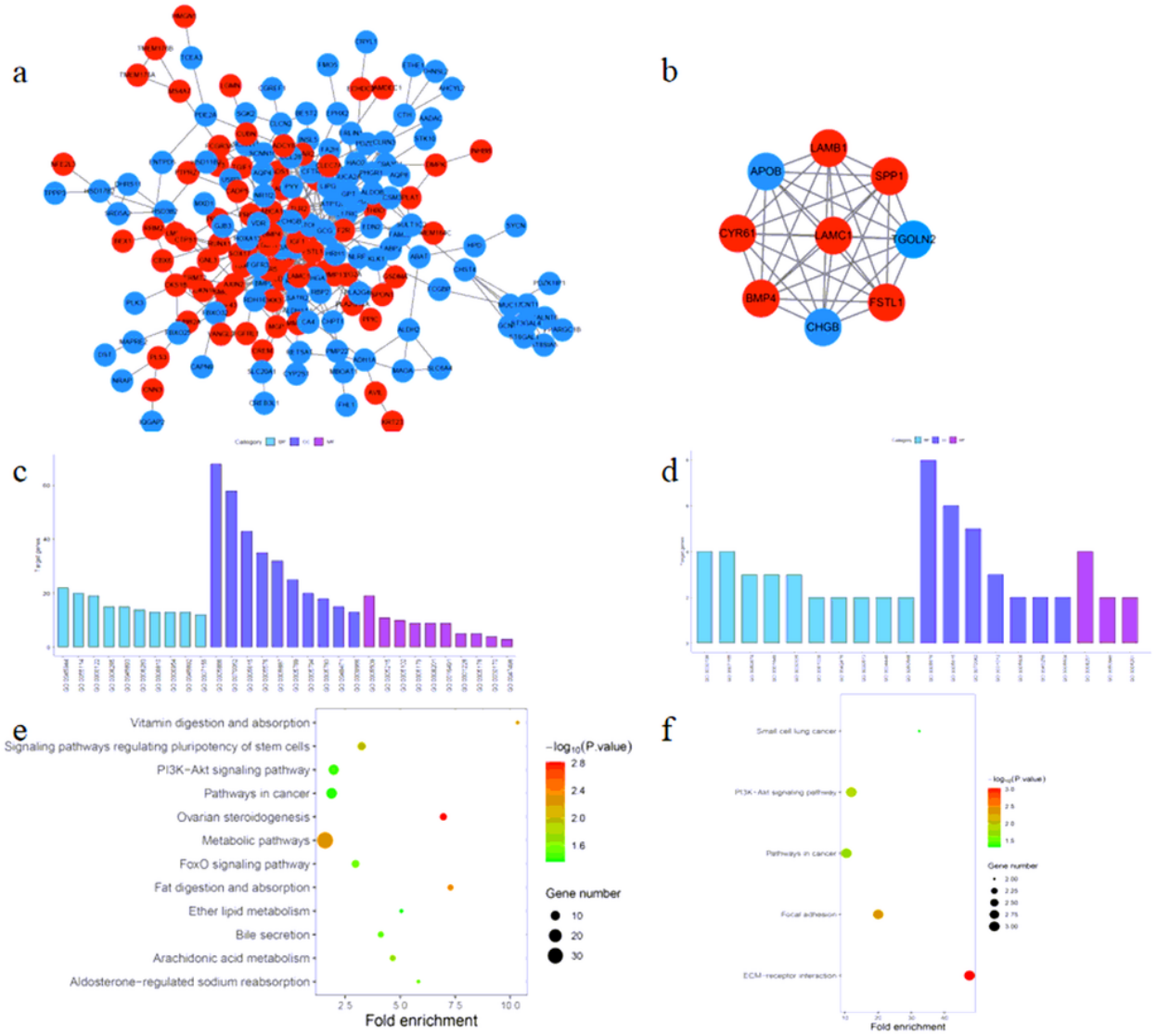

Figure 3

Protein-protein network and PPI module analysis of differential genes. the network map of differentially expressed genes was constructed by the STRING in fig 3a. Meanwhile the modular analysis was carried out on the network map to screen out the module B (fig.3b) with the highest score (MCODE score = 9.0). The red represents up-regulated genes and the blue represents down-regulated genes. Gene function and pathway enrichment analysis in DEGS and modules genes are performed respectively. Gene ontology (GO) enrichment analysis is performed using the Database for Annotation, Visualization, and Integration Discovery (DAVID) Database(fig.3d and 3e), c: DEGS, D: Module genes; Classification: Biological Process (BP), B: Cellular Component (CC), C: Molecular Function (MF); Kyoto Encyclopedia of Genes and Genomes (KEGG) Pathway Enrichment was Coarded by Metascape Database(fig.3f).The size of the dot represents the amount of gene enrichment, and the color of the dot represents P.Value.

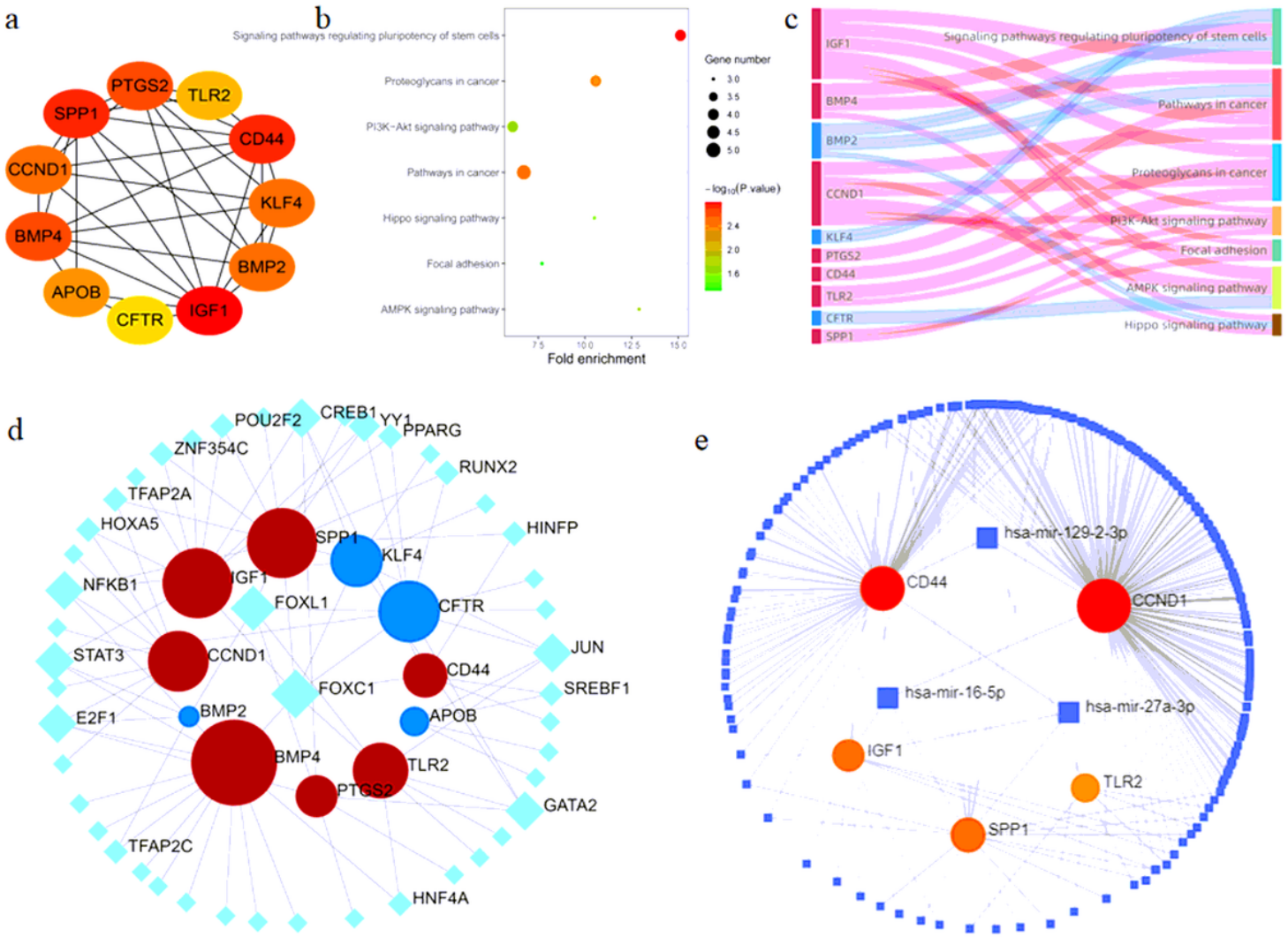

Figure 4

The core genes were screened and analyzed by KEGG and methylation. It was described that the degree score was calculated by the Cytohubba plug-in on the Cytoscape software, and the top 11 genes with the most significance were selected as hub genes according to the score in. KEGG pathway analysis of HUB gene was analysised by KEGG pathway in f fig 4a ig 4b. It was showed that the relationship between core genes and pathways in fig 4c. Red represents up-regulated genes and blue represents down-regulated genes. In the gene-related miRNA network of partial pathway (fig.4d), the circle represents the gene, the size represents the degree, and the square represents the miRNA(fig.4e).

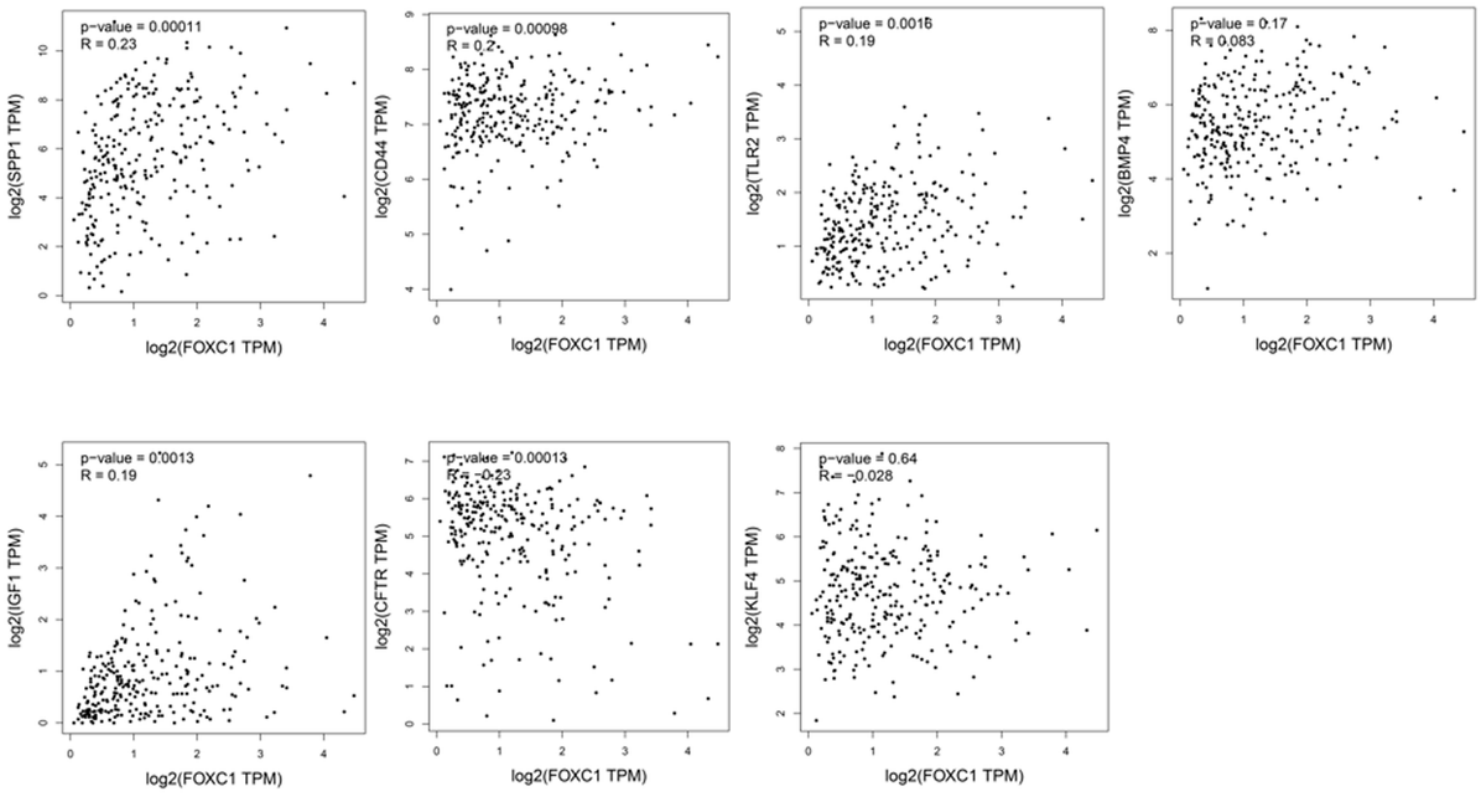

Figure 5

Correlation analysis between FOXC1 and genes,SPP1, IGF1, BMP4, TLR2, CD44, KLF4, CFTR,which were regulated by FOXC1.

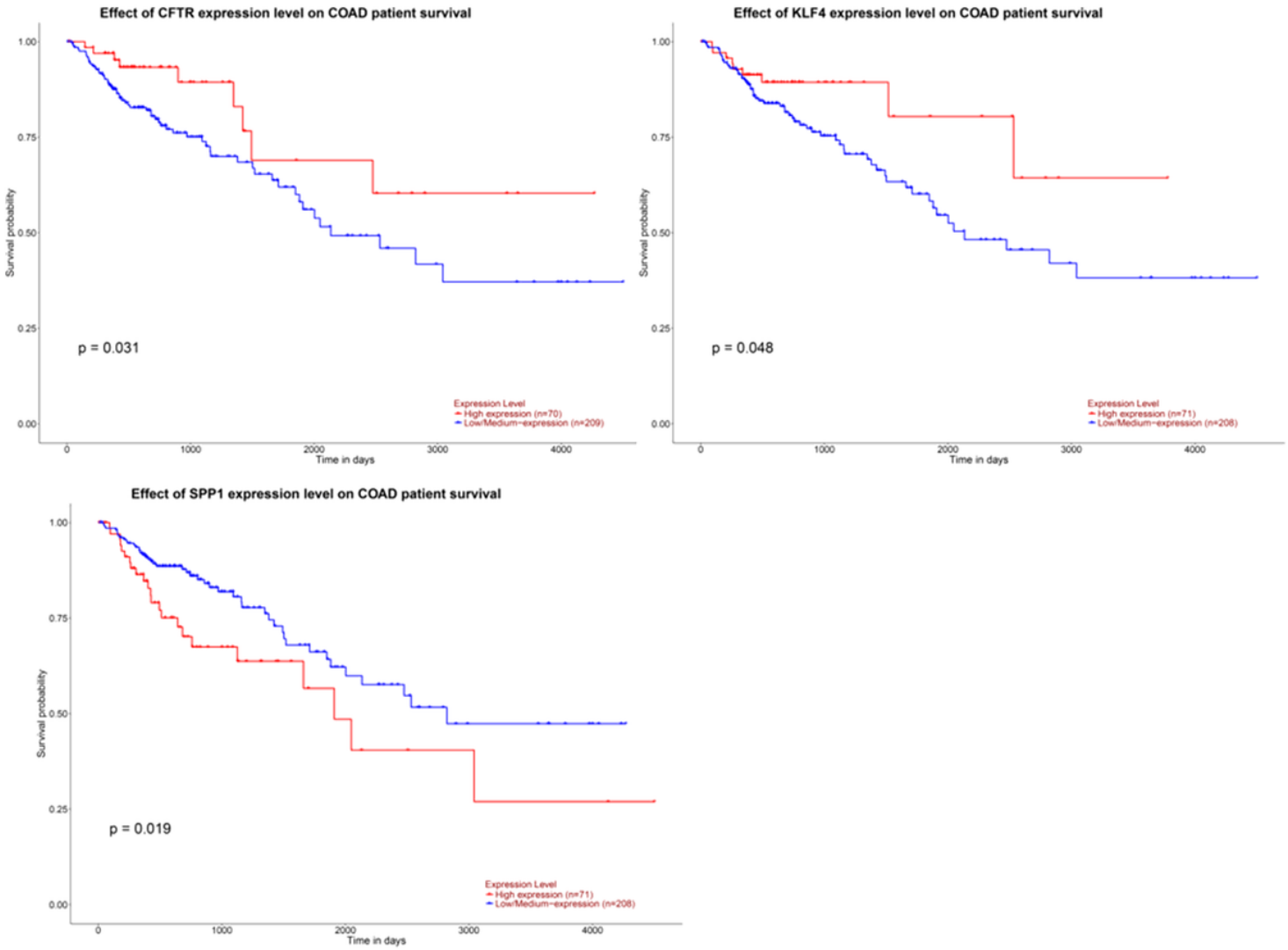

Figure 6

Survival analysis of core genes in colon cancer.

Supplementary Files

This is a list of supplementary files associated with this preprint. Click to download.

- DOC.doc