Research article

The identification of a common different gene expression signature in patients with colorectal cancer

Zhongwei Zhao[1#], Xiaoxi Fan[1#], Lili Yang[2#], Jingjing Song[1], Shiji Fang[1], Jianfei Tu[1], Minjiang Chen[1], Liyun Zheng[1], Fazong Wu[1], Dengke Zhang[1], Xihui Ying[1] and Jiansong Ji[1,*]

[1] Department of Interventional Radiology, Zhejiang University Lishui Hospital, The Fifth Affiliated Hospital of Wenzhou Medical University, Lishui Central Hospital, Lishui 323000, China

[2] Department of Anesthesiology, Zhejiang University Lishui Hospital, The Fifth Affiliated Hospital of Wenzhou Medical University, Lishui Central Hospital, Lishui 323000, China

[#]These authors contributed equally to this study.

*Correspondence: Email: jjstcty1@163.com; Tel: +86-13857088508.

Abstract: Colorectal cancer (CRC) is one of the most common malignancies, giving rise to serious financial burden globally. This study was designed to explore the potential mechanisms implicated with CRC and identify some key biomarkers. CRC-associated gene expression dataset (GSE32323) was downloaded from GEO database. The differentially expressed genes (DEGs) were selected out based on the GEO2R tool. Gene ontology (GO) and Kyoto Encyclopedia of Genes and Genomes (KEGG) analyses were employed to search the enriched pathways of these DEGs. Additionally, a protein-protein interaction (PPI) network was also constructed to visualize interactions between these DEGs. Quantitative Real-time PCR (qPCR) was further performed to valid the top5 up-regulated and top5 down-regulated genes in patients with CRC. Finally, the survival analysis of the top5 up-regulated and top5 down-regulated genes was conducted using GEPIA, aiming to clarify their potential effects on CRC. In this study, a total of 451 DEGs were captured (306 down-regulated genes and 145 up-regulated genes). Among these DEGs, the top5 up-regulated genes were DPEP1, KRT23, CLDN1, LGR5 and FOXQ1 while the top5 down-regulated genes were CLCA4, ZG16, SLC4A4, ADH1B and GCG. GO analysis revealed that these DEGs were mainly enriched in cell adhesion, cell proliferation, RNA polymerase II promoter and chemokine activity. KEGG analysis disclosed that the enriched pathway included mineral absorption, chemokine signaling pathway, transcriptional misregulation in cancer, pathways in cancer and PPAR signaling pathway. Survival analysis showed that the expression level of ZG16 may correlate with the prognosis of CRC patients.

Furthermore, according to the connectivity degree of these DEGs, we selected out the top15 hub genes, namely MYC, CXCR1, TOP2A, CXCL12, SST, TIMP1, SPP1, PPBP, CDK1, THBS1, CXCL1, PYY, LPAR1, BMP2 and MMP3, which were expected to be promising therapeutic target in CRC. Collectively, our analysis unveiled potential biomarkers and candidate targets in CRC, which could be helpful to the diagnosis and treatment of CRC.

1. Introduction

Colorectal cancer (CRC) is one of the leading causes of death associated with cancer worldwide, which has been estimated to cause 50,000 deaths per year in America [1,2]. The pathogenesis of CRC is a complicated process involved in mutations of various oncogenes and tumor suppressor genes that pinpoint multiple cellular events, such as endoplasmic reticulum stress [3], oxidative stress [4], epithelial-mesenchymal transition (EMT) [5], abnormal cell proliferation and apoptosis [6].

Although surgery combined with adjuvant radiotherapy or chemotherapy has effectively improved the prognosis of CRC, metastasis or neoplasm recurrence still occur in more than 50% of the CRC patients who have received a definite diagnosis or underwent intended curative treatment, contributing to a relatively poor 5-year survival rate [7]. Additionally, it is difficult for doctors to achieve early diagnosis of CRC for the reason that specificity of auxiliary examination indices is lacking in clinic, so that many patients may eventually miss the optional chances for surgery, thereby increasing the risk of death [8]. Hence, the detection of early-stage biomarkers and identification of core therapeutic target appear significant to decrease CRC-related deaths.

Recent years, the high-throughput gene microarray has been widely used to screen the differentially expressed genes (DEGs) between normal samples and tumor samples in human beings and animal models, which makes it accessible for us to further explore the entire molecular alterations of tumors at multiple levels involving DNA, RNA, proteins, epigenetic alterations, and metabolism [9,10]. However, it still has a long way to go to put these microarrays in application in clinic because the DEGs identified by gene profiling were countless and the statistical analyses were also too complicated [11]. Therefore, it is urgent to verify a proper number of genes and develop a suitable approach which can be operated by routine assay in clinic.

In this study, we selected the GSE32323 from Gene Expression Omnibus (GEO) and applied bioinformatics analysis to screen the DEGs in CRC. Subsequently, we used STRING to construct the protein-protein interaction (PPI) network to identify the hub genes with top15 degree of connectivity in CRC. Moreover, we made the analysis of these DEGs, including biological process (BP), molecular function (MF), cellular component (CC) and KEGG pathways. We chose top5 up-regulated and top5 down-regulated DEGs to make the overall survival (OS) analysis, aiming to observe the associations between the DEGs and the prognosis of CRC. Finally, we confirmed the mRNA expression level of the top5 up-regulated and top5 down-regulated DEGs by Q-PCR in samples obtained from patients with CRC in our hospital. These genes will assist us to screen and identify significant biomarkers and therapeutic targets of CRC in the near future.

2.   Materials and method

2.1. Data source

The heading levels should not be more than 4 levels. The fond of heading and subheadings should be 12 point normal Times New Roman. The first letter of headings and subheadings should be capitalized. The gene expression profile of GSE32323 was downloaded from the GEO database, which was a free and publicly available database. 17 pairs of cancer and non-cancerous tissues from patients with CRC in this dataset were detected by Affymetrix HG-U133 Plus 2.0 arrays by Ahmed K [12]. We also downloaded the raw Series Matrix File of GSE32323 from the GEO database.

2.2. DEGs analysis

In our study, the online software GEO2R was employed to analyze the tissue samples in GSE32323 dataset. GEO2R is an online software by which users can divide the samples into two and more groups and select out the DEGs. We used the Benjamini and Hochberg methods by default to discover false rate and used the adjust P value to reduce the errors of false positive. The choice criterion contains the adjust P value < 0.05 and $|logFC| \geq 2$.

2.3. Gene ontology and KEGG pathway analysis of DEGs

Gene ontology analysis (GO), serving as a useful approach to annotate genes and gene products, could also identify characteristic biological attributing to high-throughput genome or transcriptome data. Kyoto Encyclopedia of Genes and Genomes (KEGG) is a collection of databases, aiming to handle biological pathways, genomes, diseases, chemical substances and drugs. We divided these DEGs into up-regulated group and down-regulated group. The Database for Annotation, Visualization and Integrated Discovery (DAVID, https://david.ncifcrf.gov/), an online web based on the bioinformatics, is routinely applied for annotating genes and protein function [13]. We input these DEGs separately into DAVID to acquire the GO and KEGG analysis, $P < 0.05$ as choice criterion.   Finally,   we   used   the   ImageGP (http://www.ehbio.com/ImageGP/index.php/Home/Index/index.html)   to   visualize   the   major biological processes (BP), molecular function (MF) and cell composition (CC) of these DEGs.

2.4. Comparison of the top5 upregulated and top5 downregulated DEGs

GEPIA (http://gepia.cancer-pku.cn/index.html), designed by Chenwei Li, Zefang Tang, and Boxi Kang of Zhang Lab, Peking University, is a newly developed interactive web server aiming at analyzing the RNA sequencing expression data of 9736 tumors and 8587 normal samples from the GTEx and TCGA projects in a standard processing pipeline [14]. In this study, we employed the boxplot to visualize the mRNA expression of top5 up-regulated and top5 down-regulated DEGs in CRC tissues and and normal colorectum tissues. Transcripts per million (TPM) is the number of transcripts from this particular gene normalized first by gene length, and then by sequencing depth (in millions) in the sample. GEPIA also provides the transcripts per million (TPM) of these DEGs to display their relative expression levels.

2.5. Comparison of the top5 upregulated and top5 downregulated DEGs The overall survival (OS) of the top5 upregulated and top5 downregulated DEGs

Similarly, we used the GEPIA database to get the overall survival information of these DEGs. The logrank P value and hazard ratio (HR) with 95% confidence intervals were showed on the plot. $P < 0.05$ was statistically significant.

2.6. Comparison of the top5 upregulated and top5 downregulated DEGs Sample Collection and Quantitative Real-time PCR (Q-PCR)

We enrolled 15 male patients who were diagnosed with CRC by pathology reports in our hospital. The samples from patients with CRC for the validation of these DEGs were supported by the ethics committee in our hospital. The paired carcinoma and adjacent tissues were firstly isolated. Then total RNA was isolated using the TRIzol (Invitrogen, Carlsbad, CA, USA) kit, the concentrations and purities of which were quantified by an ultraviolet spectrophotometer. The RNA was then reversely transcribed according to the previous literature [15]. The expression levels of top5 up-regulated genes and top5 down-regulated genes were normalized to GAPDH. Relative mRNA expression levels were analyzed by 2-ΔΔcycle threshold (CT) method.

2.7. Comparison of the top5 upregulated and top5 downregulated DEGs Establishing the PPI network

Search Tool for the Retrieval of Interacting Genes (STRING) is an online app for evaluating PPI network [16]. For finding the potential correlation of DEGs, we used online app STRING to map the DEGs, subsequently the Cytoscape software was utilized to construct a PPI network. The confidence score ≥ 0.4 and maximum number of interactors $= 0$ were set as the criterion.

3. Results

3.1. Data source

The overall flow diagram of our study is presented in Figure 1. In our study, a total of 17 paired carcinoma and adjacent tissue from patients with CRC were analyzed. We applied the GEO2R online analysis tool with default parameters to screen the DEGs, using adjusted P value < 0.05 and logFC ≤ -2 or logFC ≥ 2 as the cut-off criteria. We captured 451 DEGs including 145 up-regulated DEGs and 306 downregulated DEGs. Whereafter, the DEGs were presented in the form of a heatmap and a volcano plot (Figure 2A–B). In the heatmap, the top25 up-regulated genes and top25 down-regulated genes between carcinoma and adjacent tissue were presented. Among the 451 DEGs, the top5 up-regulated genes involved DPEP1, KRT23, CLDN1, LGR5 and FOXQ1, while the top5 down-regulated genes were CLCA4, ZG16, SLC4A4, ADH1B and GCG. The gene tiles and biological functions of top5 upregulated and top5 down regulated genes were displayed in Table 1.

3.2. Re-identification of DEGs by Q-PCR

To ensure the credibility and reliability of the microarray of GSE32323 and get further credible

analysis, we re-identified the top5 up-regulated genes and top5 down-regulated genes via Q-PCR. Q-PCR (Figure 5 A-B) showed that the mRNA expression levels of DPEP1, KRT23, CLDN1, LGR5 and FOXQ1 were significantly higher in carcinoma group compared with adjacent tissue group ($P<$ 0.05). Meantime, the mRNA expression levels of CLCA4, ZG16, SLC4A4, ADH1B and GCG were obviously down-regulated in carcinoma tissues from patients with CRC ($P<0.05$).
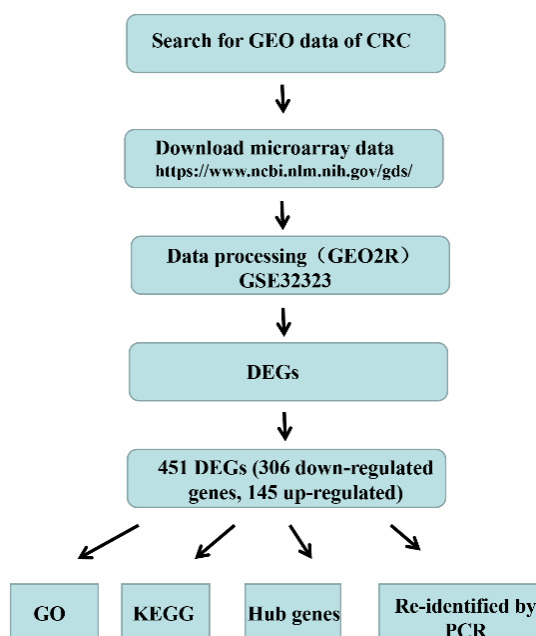

Figure 1. Flow diagram of the analysis procedure: data collection, preprocessing, analysis and validation.

Table 1. The top 5 up-regulated and down-regulated differentially expressed genes in patients with colorectal cancer.

| DEGs	Gene title	Gene symbol	LogFC	Biological function
Down-regulated	Chloride channel accessory 4	CLCA4	4.91	Calcium sensitive chloride conductance protein
	Zymogen granule protein 16	ZG16	4.45	Promoting cell growth and stemness.
	Solute carrier family 4 member 4	SLC4A4	4.43	Sodium bicarbonate cotransporter
	Alcohol dehydrogenase 1B (class I), beta polypeptide	ADH1B	4.29	Metabolizing ethanol, retinol, and aliphatic alcohol
	Glucagon	GCG	4.15	Counteracting the action of insulin in the bloodstream.
Up-regulated	Dipeptidase 1	DPEP1	−4.14	Kidney membrane enzyme
	Keratin 23	KRT23	−4.28	Intermediate filament proteins
	Claudin 1	CLDN1	−4.46	Tight junctions
	Leucine rich repeat containing G protein-coupled receptor 5	LGR5	−4.51	Leucine-rich repeat-containing receptor
	Forkhead box Q1	FOXQ1	−5.55	Embryonic development and cell cycle regulation
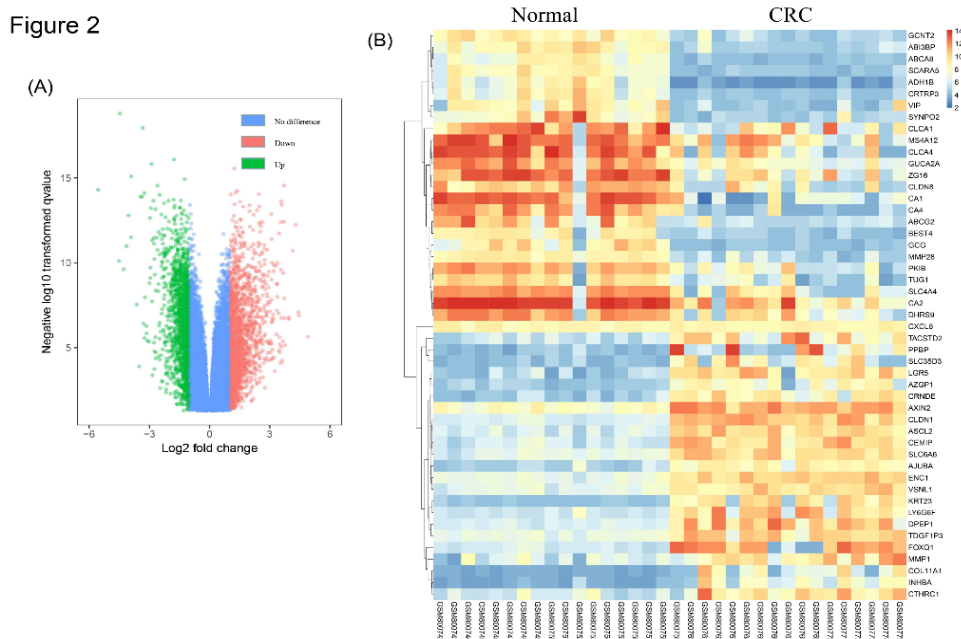
Figure 2. Volcano plot and heatmap of the differentially expressed genes (DEGs) between adjacent mucosa and carcinoma tissues from patients with colorectal cancer (CRC). (A) Volcano plot of genes detected in CRC. green means up-regulated DEGs; Red means down-regulated DEGs; blue means no difference. (B) Heatmap of top 25 up-regulated DEGs and top25 down-regulated DEGs between normal and CRC tissues.
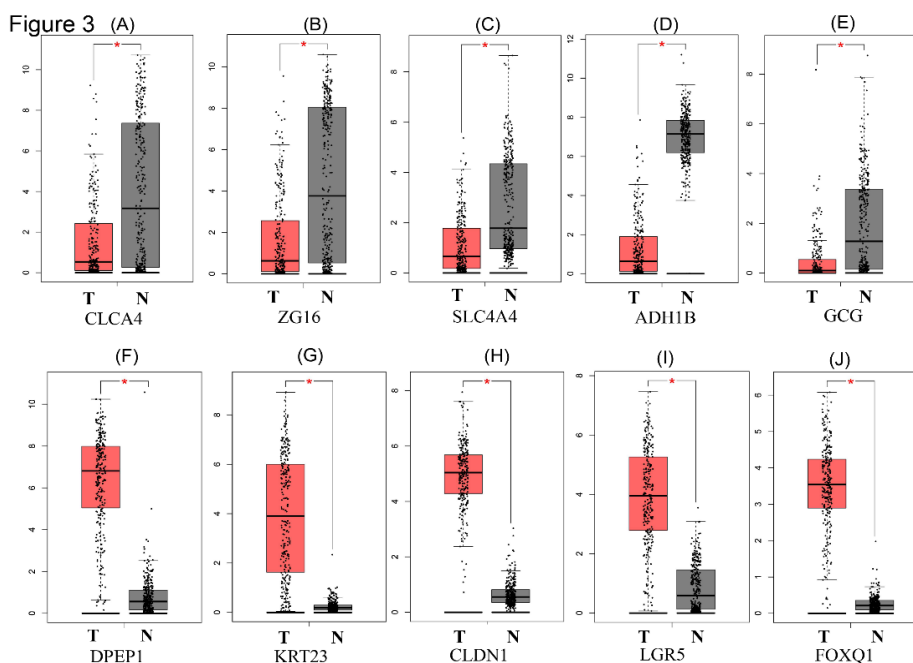

Figure 3. The mRNA expression of top5 up-regulated and top5 down-regulated genes based on TCGA database. (A–J) represents CLCA4, ZG16, SLC4A4, ADH1B, GCG, DPEP1, KRT23, CLDN1, LGR5, and FOXQ1. T: tumor；N: Normal. * $P < 0.05$ versus normal group. $P < 0.05$ was regarded statistically different.

3.3. Validation of DEGs by TCGA

To ensure the credibility of the microarray of GSE32323 and proceed further credible analysis, we validated the top5 up-regulated genes and top5 down-regulated genes based on TCGA database via GEPIA. The results showed that the mRNA expression levels of CLCA4, ZG16, SLC4A4, ADH1B and GCG were significantly lower in carcinoma group compared to adjacent tissue group while the mRNA expression level of DPEP1, KRT23, CLDN1, LGR5 and FOXQ1 in carcinoma group were statistically higher than the adjacent tissue group ($P<0.05$) (Figure 3A–J). Meanwhile, the TPM analysis demonstrated that the relative expressions of these DEGs were consist with those based on GSE32323 ($P<0.05$) (Figure 4A–J).
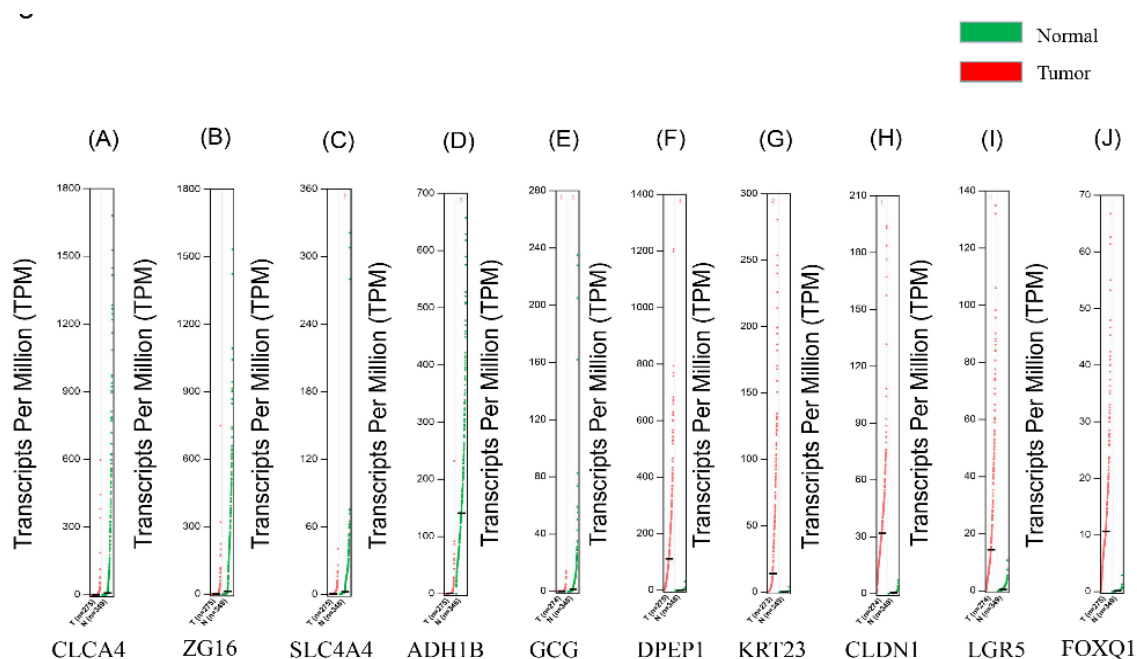

Figure 4. The transcripts per million (TPM) of top5 up-regulated and top5 down-regulated genes based on TCGA database. (A‑J) represents CLCA4, ZG16, SLC4A4, ADH1B, GCG, DPEP1, KRT23, CLDN1, LGR5, and FOXQ1. T: tumor；N: Normal. * $P < 0.05$ versus normal group. $P < 0 05$ was regarded statistically different.

3.4. Overall survival (OS)

Furthermore, we analyzed the potential association between the expression levels of top5 upregulated genes as well as top5 downregulated genes and the OS of patients with CRC. The Kaplan-Meier showed that only 1(ZG16) of the 10 genes displayed significantly correlation with the OS of patients with CRC (Figure 6A–J). The high level of ZG16 may contribute to a poorer prognosis of CRC (Logrank $p = 0.044$, HR = 0.61).
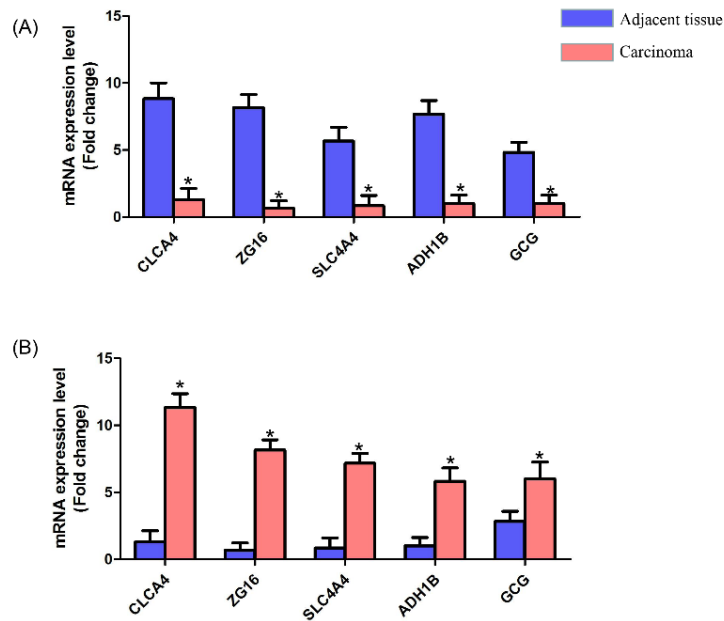
Figure 5. Validation of top5 up-regulated and top5 down-regulated DEGs in the samples from patients with CRC. (A) CLCA4, ZG16, SLC4A4, ADH1B and GCG were significantly down-regulated in carcinoma group. (B) DPEP1, KRT23, CLDN1, LGR5, and FOXQ1 were significantly up-regulated in carcinoma group. * $P < 0.05$ versus adjacent tissue group. $P < 0.05$ was regarded statistically different.
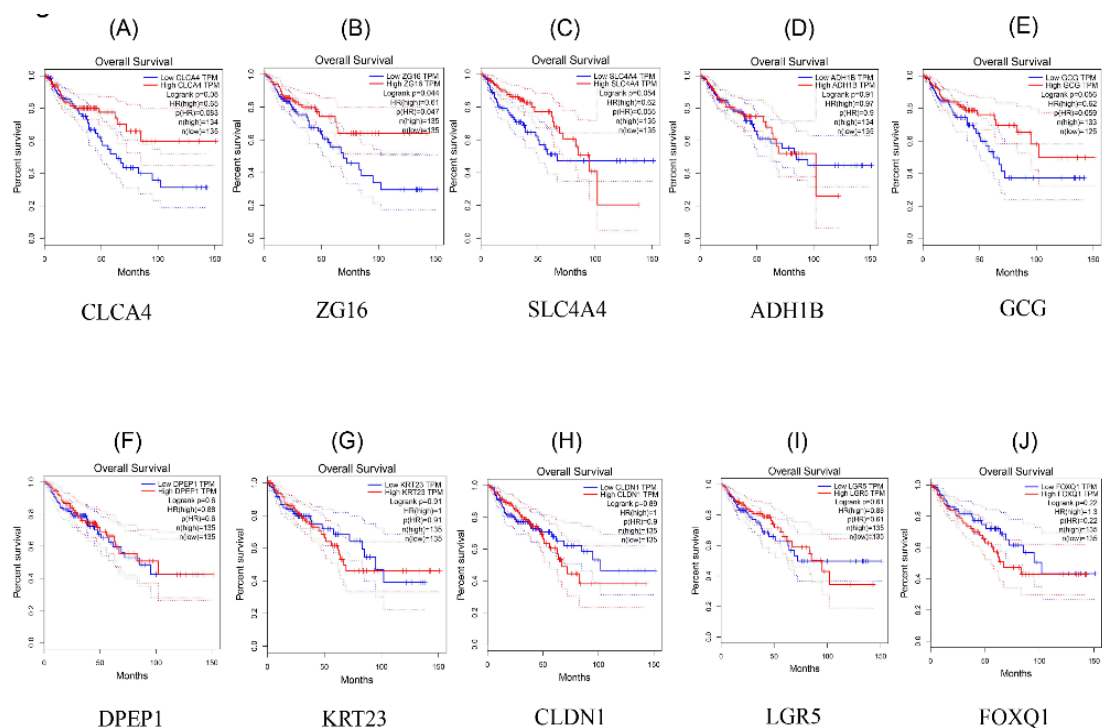

Figure 6. Prognostic value of top5 up-regulated and top5 down-regulated genes. (A–J) represents CLCA4, ZG16, SLC4A4, ADH1B, GCG, DPEP1, KRT23, CLDN1, LGR5, and FOXQ1. $P < 0.05$ was regarded statistically different.

3.5. GO enrichment analysis

The results (Table2 & Figure 7A–C) from GO term enrichment analysis varied from expression levels and GO classification of the DEGs. By analyzing GO enrichment of these up-regulated and down-regulated DEGs via DAVID, we found that the up-regulated DEGs in BP were mainly enriched in negative regulation of cell proliferation, positive regulation of transcription from RNA polymerase II promoter, proteolysis, positive regulation of transcription, DNA-templated and negative regulation of transcription from RNA polymerase II promoter while the down-regulated DEGs in BP were mainly concentrated on cell adhesion, negative regulation of cell proliferation, positive regulation of cell proliferation, response to drug and nervous system development. As for CC, the up-regulated DEGs were principally enriched in extracellular region, extracellular space, integral component of plasma membrane, proteinaceous extracellular matrix and cell surface while the down-regulated DEGs were enriched in integral component of membrane, plasma membrane, extracellular exosome, extracellular space and extracellular region. Additionally, MF analysis uncovered that the up-regulated DEGs were mainly enriched in chemokine activity, CXCR chemokine receptor binding, serine-type endopeptidase activity, extracellular matrix binding and Wnt-protein binding while the down-regulated DEGs are responsible for carbonate dehydratase activity, heparin binding, structural constituent of cytoskeleton, hormone activity and zinc ion binding.
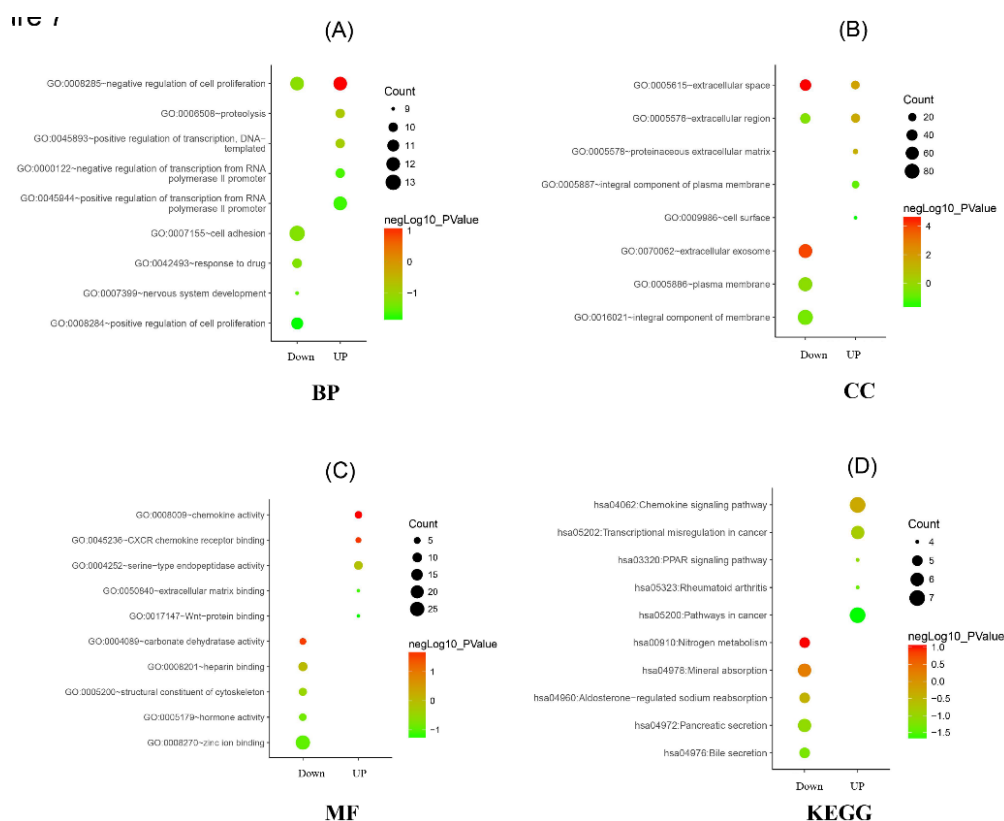

Figure 7. Gene ontology (GO) and Kyoto Encyclopedia of Genes and Genomes (KEGG) pathway analysis of CRC. (A) The enriched GO terms in the biological process (BP); (B) The enriched GO terms in the cellular component (CC); (C) The enriched GO terms in the molecular function (MF); (D) The enriched KEGG pathway in CRC.

Table 2. Gene ontology analysis of differentially expressed genes associated with colorectal cancer.

Expression	Category	Term	Count	%	*P*-Value	FDR
	GOTERM_BP_DIRECT	GO:0007155~cell adhesion	13	0.038728513	0.013503078	19.70156803
	GOTERM_BP_DIRECT	GO:0008285~negative regulation of cell proliferation	12	0.035749397	0.011658571	17.24378143
	GOTERM_BP_DIRECT	GO:0008284~positive regulation of cell proliferation	11	0.03277028	0.07067658	69.36458376
	GOTERM_BP_DIRECT	GO:0042493~response to drug	10	0.029791164	0.014890967	21.50558949
	GOTERM_BP_DIRECT	GO:0007399~nervous system development	9	0.026812048	0.028891454	37.69742039
	GOTERM_CC_DIRECT	GO:0016021~integral component of membrane	83	0.247266661	0.004198766	5.157367461
	GOTERM_CC_DIRECT	GO:0005886~plasma membrane	71	0.211517264	0.001689338	2.105290204
	GOTERM_CC_DIRECT	GO:0070062~extracellular exosome	67	0.199600798	6.59E-08	8.30E-05
	GOTERM_CC_DIRECT	GO:0005615~extracellular space	43	0.128102005	1.82E-08	2.29E-05
	GOTERM_CC_DIRECT	GO:0005576~extracellular region	34	0.101289957	0.002466544	3.060090837
Down-regulated	GOTERM_MF_DIRECT	GO:0004089~carbonate dehydratase activity	5	0.014895582	1.89E-05	0.0259938
	GOTERM_MF_DIRECT	GO:0008201~heparin binding	9	0.026812048	7.38E-04	1.012297275
	GOTERM_MF_DIRECT	GO:0005200~structural constituent of cytoskeleton	7	0.020853815	0.002187539	2.974720065
	GOTERM_MF_DIRECT	GO:0005179~hormone activity	6	0.017874698	0.005649705	7.515474579
	GOTERM_MF_DIRECT	GO:0008270~zinc ion binding	25	0.07447791	0.007595654	9.980321245
	GOTERM_BP_DIRECT	GO:0008285~negative regulation of cell proliferation	12	0.072115385	4.45E-05	0.070505424
	GOTERM_BP_DIRECT	GO:0045944~positive regulation of transcription from RNA polymerase II promoter	12	0.072115385	0.047923342	54.04862787
	GOTERM_BP_DIRECT	GO:0006508~proteolysis	10	0.060096154	0.004653246	7.118885476
	GOTERM_BP_DIRECT	GO:0045893~positive regulation of transcription, DNA-templated	10	0.060096154	0.005625306	8.544770506
	GOTERM_BP_DIRECT	GO:0000122~negative regulation of transcription from RNA polymerase II promoter	10	0.060096154	0.040288919	47.85424336
	GOTERM_CC_DIRECT	GO:0005576~extracellular region	25	0.150240385	2.24E-05	0.026796109

Category		Term	Count	%	P-Value	FDR
Up-regulated	GOTERM_CC_DIRECT	GO:0005615~extracellular space	23	0.138221154	1.30E-05	0.015593466
	GOTERM_CC_DIRECT	GO:0005887~integral component of plasma membrane	17	0.102163462	0.0098955	11.22182273
	GOTERM_CC_DIRECT	GO:0005578~proteinaceous extracellular matrix	10	0.060096154	3.33E-05	0.039831238
	GOTERM_CC_DIRECT	GO:0009986~cell surface	8	0.048076923	0.044084961	41.70407607
	GOTERM_MF_DIRECT	GO:0008009~chemokine activity	6	0.036057692	1.24E-05	0.016078978
	GOTERM_MF_DIRECT	GO:0045236~CXCR chemokine receptor binding	4	0.024038462	1.86E-05	0.024112254
	GOTERM_MF_DIRECT	GO:0004252~serine-type endopeptidase activity	8	0.048076923	0.00101144	1.305200327
	GOTERM_MF_DIRECT	GO:0050840~extracellular matrix binding	3	0.018028846	0.01109514	13.48485303
	GOTERM_MF_DIRECT	GO:0017147~Wnt-protein binding	3	0.018028846	0.01556121	18.42243874

GO: Gene Ontology; FDR: False Discovery Rate.

Table 3. KEGG pathway analysis of differentially expressed genes associated with colorectal cancer

Category	Term	Count	%	P-Value	Genes	FDR
	hsa04978:Mineral absorption	6	0.017874698	4.69E-04	TRPM6, MT1M, MT1E, MT1H, MT1X, MT1F	0.546769342
Down-regulated DEGs	hsa04972:Pancreatic secretion	6	0.017874698	0.010495925	KCNMA1, CLCA1, CLCA4, ATP2A3, CA2, SLC4A4	11.59221263
	hsa00910:Nitrogen metabolism	5	0.014895582	8.18E-05	CA12, CA7, CA4, CA2, CA1	0.095443878
	hsa04960:Aldosterone-regulated sodium reabsorption	5	0.014895582	0.002219912	SGK1, NR3C2, HSD11B2, SCNN1G, SCNN1B	2.561720438
	hsa04976:Bile secretion	5	0.014895582	0.016846119	AQP8, SLC51B, CA2, SLC4A4, ABCG2	17.99502607
Up-regulated DEGs	hsa04062:Chemokine signaling pathway	7	0.042067308	0.001681492	CXCL1, PLCB4, PPBP, CXCL5, CXCL3, CXCL8, CXCL11	1.951113014
	hsa05200:Pathways in cancer	7	0.042067308	0.053070195	BMP4, PLCB4, MET, CXCL8, AXIN2, MYC, MMP1	47.1893587
	hsa05202:Transcriptional misregulation in cancer	6	0.036057692	0.005889645	MET, CXCL8, MMP3, MYC, WT1, PLAU	6.682415976
	hsa03320:PPAR signaling pathway	4	0.024038462	0.011041824	SCD, ACSL6, MMP1, FABP6	12.19046418
	hsa05323:Rheumatoid arthritis	4	0.024038462	0.022861328	CXCL5, CXCL8, MMP3, MMP1	23.72114711

KEGG: Kyoto Encyclopedia of Genes and Genomes; FDR: False Discovery Rat

3.6. KEGG pathway analysis

To obtain a more comprehensive information regarding to the critical pathways of those selected DEGs, KEGG pathways analysis were also carried out via DAVID. The results in Table3 and Figure 7D disclosed the most vital KEGG pathways of the down-regulated and up-regulated DEGs. The down-regulated DEGs were mainly enriched in mineral absorption, pancreatic secretion, nitrogen metabolism, aldosterone-regulated sodium reabsorption and bile secretion. By contrast, the up-regulated DEGs were mainly responsible for chemokine signaling pathway, pathways in cancer, transcriptional misregulation in cancer, PPAR signaling pathway and rheumatoid arthritis.

3.7. Identification of hub genes

Applying the STRING online tool, 213 nodes with 264 PPI relationships were found, accounting for about 77.8% of these selected DEGs. According to the degree of connectivity of these DEGs, we constructed the PPI network and selected the top 15 hub genes (Table 4). The top 15 hub genes, possessing high degree of connectivity in CRC are as follows, MYC, CXCR1, TOP2A, CXCL12, SST, TIMP1, SPP1, PPBP, CDK1, THBS1, CXCL1, PYY, LPAR1, BMP2 and MMP3. Among these 15 hub genes: MYC, CXCR1, TOP2A, SPP1, PPBP, CDK1,CXCL1 and MMP3 were significantly up-regulated while CXCL12, SST, TIMP1,THBS1, PYY, LPAR1 and BMP2 significantly down-regulated ($P<0.05$). The 15 hub genes could interact with 243 genes directly, and MYC acted as the most intensive gene which could interact with 46 up-regulated genes and 31 down-regulated genes. Intriguingly, among these hub genes, there also displayed very strong interactions (Figure 8).

Table 4. Top 15 hub genes with higher degree of connectivity.

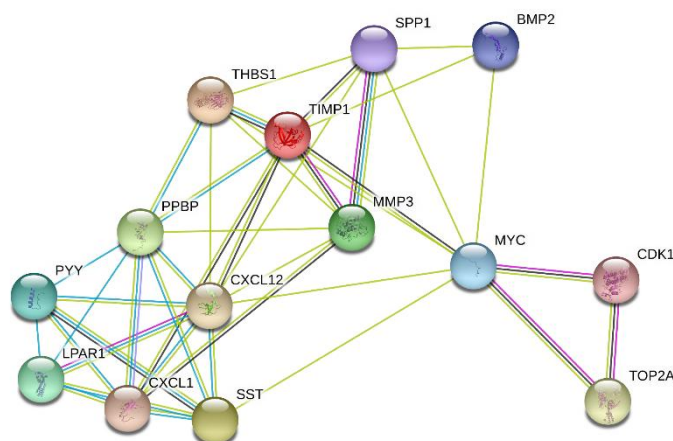| Gene	Degree of connectivity	Adjusted P value
MYC	34	9.25E-11
CXCR1	28	1.87E-03
TOP2A	23	2.90E-02
CXCL12	20	1.95E-08
SST	18	7.36E-09
TIMP1	17	1.05E-02
SPP1	17	3.95E-02
PPBP	17	9.73E-04
CDK1	16	1.13E-02
THBS1	16	2.99E-08
CXCL1	15	1.91E-03
PYY	15	1.77E-08
LPAR1	15	8.36E-11
BMP2	15	1.06E-05
MMP3	14	7.22E-05

Figure 8. Protein-protein interaction (PPI) network. The PPI network of top15 hub genes with high connectivity degree.

4. Discussion

CRC is a one of the serious public health problems, which is also being the third most commonly diagnosed cancer and the major factor resulting in cancer-related deaths around the world [17,18]. CRC screening has been demonstrated to greatly decrease the morbidity and the mortality in a great many longstanding or newly economically developed countries [19]. However, at present, there is no an efficient and specific diagnostic methodology and treatment strategy for CRC, which is mainly attributed to the intricate pathogenesis, and its symptoms that are difficult to diagnose in the first several years [20]. Our study mainly focused on the expression profiling selected from microarray studies of CRC. In this study, a total of 17 pairs of cancer and non-cancerous tissues isolated from patients with CRC were analyzed. 451 DEGs including 145 up-regulated DEGs and 306 downregulated DEGs were screened. To obtain a comprehensive understanding of these DEGs, we performed GO function and KEGG pathway analysis.

4.1. DEGs may serve as potential biomarkers for the diagnosis of CRC

Our analysis screened 451 DEGs with a quadruple fold change between carcinoma and adjacent tissue from patients with CRC. In our heat map, a total of 50 DEGs including top25 up-regulated genes and top25 down-regulated genes were displayed. From our perspective, these DEGs would be promised to be possible candidates for the diagnosis of CRC in future. Currently, some of these DEGs, in fact, have been already disclosed to be novel indicators of CRC. H Meng et al. [21] found that ZG16 exhibited an extremely high expression in normal epithelial cells from rectum, colon and small intestine in an organ-specific fashion while its protein expression was completely lost in CRC tissues. Furthermore, another study proved that the gene expression and copy number alterations of ZG16 significantly correlated with MLH1 silencing, microsatellite instability (MSI), hyper-mutation status, presence of synchronous adenomas, CpG island methylator phenotype, histological type and gender. Survival analysis showed that patients with higher ZG16 gene expression level displayed longer overall survival and progression-free survival compared with those having relatively lower

level. Another study showed that the expression of DPEP1 was strikingly up-regulated in human CRC tissue than that in normal mucosa. Meanwhile, high level of DPEP1 was found to be significantly associated with localisation and histological stage of CRC [22]. Our study screened the DEGs of CRC from the angle of bioinformatics for the first time, however, the clinical practice of these DEGs needs to be further investigated.

4.2. Hub genes are possible therapeutic targets in CRC

In our study, we picked out 15 hub genes of CRC, all of which were located in the core nodes in PPI network, meaning that these genes could be critical therapeutic targets to protect against CRC. For instance, somatostatin (SST), encoding a well-characterized growth-regulatoryand gastrointestinal neuroendocrine peptide, has been proved to be a tumor suppressor gene and own potent anti-tumor effects. SST could suppress tumors through direct and indirect mechanisms. Direct anti-tumor effects involve inhibiting cell proliferation; inducing cell cycle arrest; promoting apoptosis; and suppressing cell invasion while indirect anti-tumor effects are blocking the synthesis and secretion of growth factors as well as growth-promoting hormones, meanwhile stopping angiogenesis which is vital in neoplasm growth and spread [23]. Corroborative evidence has unveiled the decreased protein and mRNA expression of SST in CRC tissues [24]. One of the possible reason contributing to the down-regulation of SST was attributed to the higher SST methylation levels in tumor tissues. As expected, another study showed that pretreatment of demethylation agents into SST methylated CRC cells induced the up-regulation of SST expression in vitro [25]. Clinical research demonstrated that serum methylation level of SST in patients with CRC were obviously predictive of tumor recurrence and cancer-related death [26]. Collectively, our study proved that up-regulating SST in CRC cells may exert potent anti-tumor activity.

4.3. Chemokines may play essential roles in the development of CRC

An interesting finding of our study was that many DEGs in CRC were related with chemokines and their receptors. To be more specific, GO analysis showed that up-regulated DEGs enriched in MF were mainly responsible for chemokine activity and CXCR chemokine receptor binding. Also, KEGG pathways analysis also disclosed that the up-regulated genes including CXCL1, PLCB4, PPBP, CXCL5, CXCL3, CXCL8, CXCL11 are implicated with chemokine signaling pathway. Additionally, 3 of the top15 hub genes based PPI network were also chemokines and their receptors, namely CXCR1, CXCL12 and CXCL1. Mountainous evidence proved that the interactions between chemokines as well as chemokine receptors could mediate the recruitment of different immune cell subsets into tumour microenvironment where the host immune system and tumour cells could interact [27]. On the one hand, chemokines regulate the molecules affecting tumour immunity as well as therapeutic outcomes in patients [28,29]. On the other one hand, chemokines could also target stromal cells and tumor cells, thereby regulating tumour stemness, angiogenesis [30], proliferation [31], apoptosis [32], and survival [33]. These findings suggested that therapies targeting chemokines in CRC may synergize with current cancer therapies, especially immunotherapies by affecting tumour immunological phenotypes, and improving antitumour immune responses.

5. Conclusion

In conclusion, we provided a comprehensive and novel analysis of gene expression profiles patients with CRC. Particularly, the top5 down-regulated genes including CLCA4, ZG16, SLC4A4, ADH1B and GCG and the top5 up-regulated genes including DPEP1, KRT23, CLDN1, LGR5 and FOXQ1, which are expected to sensitive biomarkers in diagnosis of CRC. Meanwhile, we also screened the top 15 hub genes involving MYC, CXCR1, TOP2A, CXCL12, SST, TIMP1, SPP1, PPBP, CDK1, THBS1, CXCL1, PYY, LPAR1, BMP2 and MMP3, which could be promising therapeutic targets of CRC. Additionally, genes involved in chemokines pathways were also significantly altered in patients with CRC. Anyway, this analysis may offer the powerful evidence and clues for the future genomic individualized treatment of CRC.

Acknowledgments

Conflict of interest

All authors declare no conflicts of interest in this paper.

References

1. K. Yang, F. Zhang, P. Han, et al., Metabolomics approach for predicting response to neoadjuvant chemotherapy for colorectal cancer, *Metabolomics*, **9** (2018), 110–121.
2. R. L. Siegel, K. D. Miller, S. A. Fedewa, et al., Colorectal cancer statistics, 2017. *CA Cancer J. Clin.,* **3** (2017), 104–117.
3. P. M. Yang, Y. T. Li, C. T. Shun, et al., Zebularine inhibits tumorigenesis and stemness of colorectal cancer via p53-dependent endoplasmic reticulum stress, *Sci. Rep.,* **11** (2013), 3219.
4. S. A. Noonan, M. E. Morrissey, P. Martin, et al., Tumour vasculature immaturity, oxidative damage and systemic inflammation stratify survival of colorectal cancer patients on bevacizumab treatment, *Oncotarget,* **12** (2018), 10536–10548.
5. Y. Liu, G. Wang, Y. Yang, et al., Increased TEAD4 expression and nuclear localization in colorectal cancer promote epithelial-mesenchymal transition and metastasis in a YAP-independent manner, *Oncogene*, **21** (2016), 2789–2795.
6. A. G. Long, E. T. Lundsmith and K. E. Hamilton, Inflammation and colorectal cancer, *Curr. Colorectal Cancer Rep.,* **4** (2017), 341–351.
7. H. Zheng, L. Yu, S. Luo, et al., miR-296 inhibits the metastasis and epithelial-mesenchymal transition of colorectal cancer by targeting S100A4, *BMC Cancer,* **1** (2017), 140–147.
8. A. Mahasneh, F. Alshaheri and E. Jamal, Molecular biomarkers for an early diagnosis, effective treatment and prognosis of colorectal cancer: Current updates, *Exp. Mol. Pathol*., **3** (2017), 475–483.
9. N. Li, L. Li and Y. Chen, The identification of core gene expression signature in hepatocellular carcinoma, *Oxid. Med. Cell Longev*., **4** (2018), 1–15.
10. L. Ling, L. Ning, C. He, et al., Proteomic analysis of differentially expressed proteins in kidneys

of brain dead rabbits, *Mol. Med. Rep*., **1** (2017), 215–223.

11. T. Braunschweig, J. Y. Chung, S. M. Hewitt, Tissue microarrays: bridging the gap between research and the clinic, *Expert Rev. Proteom.,* **3** (2005), 325–336.

12. A. Khamas, T. Ishikawa, K. Shimokawa, et al., Screening for epigenetically masked genes in colorectal cancer Using 5-Aza-2'-deoxycytidine, microarray and gene expression profile. *Cancer Genom. Proteom.,* **2** (2012), 67–75.

13. G. Dennis, B. T. Sherman, D. A. Hosack, et al., DAVID: Database for Annotation, visualization, and Integrated Discovery, *Genome Biol.,* **5** (2003), 3.

14. Z. Tang, C. Li, B. Kang, et al., GEPIA: a web server for cancer and normal gene expression profiling and interactive analyses, *Nucleic Acids Res*., **1** (2017), 98–102.

15. L. P. Sun, Y. Guo, Y. X. Zhang, et al., Antioxidant and anti-tyrosinase activities of phenolic extracts from rape bee pollen and inhibitory melanogenesis by cAMP/MITF/TYR pathway in B16 mouse melanoma cells, *Front. Pharmaco*., **176** (2017), 46–51.

16. A. Ikin, C. Riveros, P. Moscato, et al., The Gene Interaction Miner: a new tool for data mining contextual information for protein–protein interaction analysis, *Bioinformatics*, **2** (2010), 283–284.

17. M. Arnold, M. S. Sierra, M. Laversanne, et al., Global patterns and trends in colorectal cancer incidence and mortality, *Gut*, **4** (2017), 683–691.

18. D. K. Rex, C. R. Boland, J. A. Dominitz, et al., Colorectal cancer screening: recommendations for physicians and patients from the U.S. multi-society task force on colorectal cancer, *Gastroenterology*, **7** (2017), 307–323.

19. E. V. Cutsem, A. Cervantes, R. Adam, et al., ESMO consensus guidelines for the management of patients with metastatic colorectal cancer, *Ann. Oncol*., **8** (2016), 1386–1422.

20. V. P. Deenadayalu and D. K. Rex, Colorectal cancer screening: a guide to the guidelines, *Rev. Gasteroenterol. Di.,* **7** (2016), 204.

21. M. Hui, W. Li, L. A. Boardman, et al., Loss of ZG16 is associated with molecular and clinicopathological phenotypes of colorectal cancer, *BMC Cancer*, **1** (2018), 433–441.

22. P. A. Eisenach, E. Soeth, C. Röder, et al., Dipeptidase 1 (DPEP1) is a marker for the transition from low-grade to high-grade intraepithelial neoplasia and an adverse prognostic factor in colorectal cancer, *Brit. J. Cancer*., **3** (2013), 694–703.

23. S. Pyronnet, C. Bousquet, S. Najib, et al., Antitumor effects of somatostatin, *Mol. Cell Endocrinol*., **1** (2008), 230–237.

24. K. Leiszter, F. Sipos, O. Galamb, et al., Promoter Hypermethylation-Related Reduced Somatostatin Production Promotes Uncontrolled Cell Proliferation in Colorectal Cancer, *Plos One*, **2** (2015), e0118332.

25. Y. Mori, K. Cai, Y. Cheng, et al., A genome-wide search identifies epigenetic silencing of somatostatin, tachykinin-1, and 5 other genes in colon cancer, *Gastroenterology*, **3** (2006), 797–808.

26. Y. Liu, H. C. Min, C. K. Tham, et al., Methylation of serum SST gene is an independent prognostic marker in colorectal cancer, *Am. J. Cancer Res*., **9** (2016), 2098–3008.

27. N. Nagarsheth, M. S. Wicha and W. Zou, Chemokines in the cancer microenvironment and their relevance in cancer immunotherapy, *Nat. Rev. Immunol*., **9** (2017), 559–572.

28. H. Verbeke, S. Struyf, G. Laureys, et al., The expression and role of CXC chemokines in colorectal cancer, *Cytokine Growth F. R.,* **5** (2011), 345–358.

29. O. Oladipo, S. Conlon, A. O'Grady, et al., The expression and prognostic impact of CXC-chemokines in stage II and III colorectal cancer epithelial and stromal tissue, *Br. J. Cancer*, **3** (2011),480–487.

30. K. Rupertus, J. Sinistra, C. Scheuer, et al., Interaction of the chemokines I-TAC (CXCL11) and SDF-1 (CXCL12) in the regulation of tumor angiogenesis of colorectal cancer, *Clin. Exp. Metastasis*, **4** (2014), 447–459.

31. X. Shen, A. Artinyan, D. Jackson, et al., Chemokine receptor CXCR4 enhances proliferation in pancreatic cancer cells through AKT and ERK dependent pathways, *Pancreas*, **1**(2010), 81–87.

32. Y. Fang, F. C. Henderson, Q. Yi, et al., Chemokine CXCL16 expression suppresses migration and invasiveness and induces apoptosis in breast cancer cells, *Mediat. Inflamm*., **8** (2014), 478641.

33. J. Ehling and F. Tacke, Role of chemokine pathways in hepatobiliary cancer, *Cancer Lett*., **2** (2016), 173–183.