

# Mining the key genes regulating colorectal cancer proliferation based on the GEO database

Kun Zhao<sup>1</sup>, Zhenhua Wu<sup>1</sup>, Ling Chen<sup>2</sup>, Hujun Shen<sup>2</sup>, and Mingsen Deng<sup>1</sup>

<sup>1</sup>Guizhou University of Finance and Economics

<sup>2</sup>Guizhou Education University

June 13, 2020

## Abstract

Colorectal cancer (CRC) is a common tumor in digestive system with high morbidity and mortality. In this work, three datasets (GSE68468, GSE21510, and GSE9348) were selected from the public GEO database, and the differentially expressed genes (DEGs) related to CRC were screened out by bioinformatics approach. The GO (gene ontology) analysis on DEGs was carried out and followed by the KEGG (kyoto encyclopedia of genes and genomes) analysis. Finally, protein-protein interaction (PPI) networks of DEGs were respectively constructed by using the STRING database. Our results showed that a total of 849 DEGs in common were identified from the three datasets, including 406 up-regulated genes and 443 down-regulated genes. The GO analysis demonstrated that the DEGs are mainly related to cell division, cell proliferation, cell signaling, and immune response. The KEGG analysis revealed that DEGs are mainly enriched in cell cycle and cell signaling pathway. Based on the PPI network analysis, we identified a total of 86 key genes: 72 up-regulated genes mainly consist of CD and KIF genes while 14 down-regulated genes are mainly composed of various members of CC genes. In particular, we found that the CRC proliferation should be enhanced by the interaction of CDK1 with CDKN3 and KIF15. On the one hand, CRC proliferation is enhanced by over-expressing the CD and KIF genes involving in cell division and cell proliferation. On the other hand, CRC proliferation is strengthened through silencing the expression of the CC genes associated with immune response.

## Keywords

Colorectal cancer proliferation, GO analysis, KEGG analysis, differentially expressed genes, PPI network.

## Introduction

Colorectal cancer (CRC) is a common malignant tumor of the digestive system. Its morbidity ranks third among all tumors, and the mortality rate ranks second and third among male and female patients respectively.<sup>1,2</sup> In the past 30 years, the morbidity and mortality rates of CRC in developed countries remain unchanged and have been stabilized. However, increasing morbidity and mortality of CRC poses a great threat on developing countries, such as China. Interestingly, the incidence and mortality of CRC have shown significantly geographical differences. For instance, Europe and North America are areas with higher incidence of CRC as compared with Central and South Asia or Africa. On the other side, compared with central Africa, Eastern Europe and Central Europe are regions with higher CRC mortality.<sup>3</sup>

At present, many diagnostic and therapeutic methods, including fiber colonoscopy, surgical resection, chemotherapy and radiation therapy, are widely used to treat colon cancer. Vining et al.<sup>4</sup> proposed the CT colonography as a new method for the examination of CRC for the first time in 1994. This technology is one of the effective methods for diagnosing CRC owing to high sensitivity and noninvasive advantages. Since then, great progresses have been made in the screening, diagnosis, and treatment of CRC. However, there are still great challenges in early diagnosis and treatment of tumor development and metastasis. Therefore,

new methods have emerged for the prevention and treatment of CRC through regulating the expression of key genes strongly influencing CRC development and proliferation.

Micro-array technology, which has been widely employed to study the expression of genes, can be used as an effective method not only for revealing tumor-related molecular mechanisms but also for identifying molecular therapeutic targets. Williams et al.<sup>5</sup> used cDNA chip analysis to determine key genes in CRC, and found that survivin genes can significantly inhibit tumor growth through RNA interference experiments. Tsavaris et al.<sup>6</sup> showed that the high expression of TOP2A is closely related to the occurrence and development of CRC. Zhou et al.<sup>7</sup> suggested that the TOP2A expression may be used to identify the progression and prognosis of pancreatic cancer by regulating the cell cycle and p53 signaling pathway. Gao et al.<sup>8</sup> demonstrated that KPNA2 can promote the proliferation of tumor cells by increasing the expression of CDK1. Han et al.<sup>9</sup> found that inhibiting the expression of transcription factor sp1 can promote the apoptosis of CRC cells and the arrest of cell cycle, thereby inhibiting the proliferation of CRC. Zheng et al.<sup>10</sup> found that the over-expression of parafibromin (the cell division cycle protein) can inhibit the progression of CRC division. Chen et al.<sup>11</sup> found that the over-expressed ribosomal protein S15A can induce malignant transformation of CRC cells through the p53 signaling pathway, consequently resulting in erroneous regulation of the p53 signaling pathway.

As a large amount of gene chip data are generated and stored in database, it is possible to obtain more understanding of the mechanism of CRC incidence and development through mining and analyzing the gene chip data stored in a public database. In this study, three datasets (GSE68468, GSE21510, and GSE9348) were selected from the GEO (Gene Expression Omnibus) database. From these datasets, 849 common differentially expressed genes (DEGs) of CRC were screened out using traditional bioinformatics methods, including 406 up-regulated genes and 443 down-regulated genes. Our results revealed that DEGs are mainly associated with cell division, proliferation, cell signaling, and immune response. Based on the constructed protein-protein interaction (PPI) network, 86 key genes were identified, including which 72 up-regulated genes and 14 down-regulated genes. Interestingly, the up-regulated genes are mainly related to cell division, proliferation, and metastasis, and the down-regulated genes are mainly associated with cell signaling and immune response. In addition, we discovered that the gene expression associated with the CD and KIF families were up-regulated while that associated with the CC family were down-regulated. In particular, we found that the interactions between CD genes with KIF genes should be strongly correlated with the CRC proliferation, which hasn't been reported previously. Furthermore, our work revealed that CRC proliferation should be influenced by the interactions between cell-cell signaling genes and immune response genes .

## METHODS

### GEO database

As the largest public database for gene chips, the GEO (Gene Expression Omnibus, <https://www.ncbi.nlm.nih.gov/geo>) database of the National Center for Biotechnology Information (NCBI) contain high through-out data for gene expression, comparative genomics, single nucleotide polymorphism, genomic methylation, proteomics data, non-coding RNA, and so on. The GEO database provides a flexible and open data platform, allowing users or researchers to easily submit, save and retrieve many different types of data. The three datasets (GSE68468, GSE21510, and GSE9348 datasets) to be studied in this work were downloaded from the GEO database. Among them, GSE68468 contains 186 CRC tissues and 55 normal tissues, GSE21510 contains 123 CRC tissues and 25 normal tissues, and GSE9348 contains 70 CRC tissues and 12 normal tissues.

The Affy package in R was used to standardize the original data from the three datasets (see Figure S1 of Supporting Information). Then, the probe names were converted into gene names accordingly based on the annotation information of the GPL570 platform. As for the situation that multiple probes correspond to one gene, an average value of these probes was considered as the expression level of the gene.

### 2.2 Differentially expressed genes (DEGs)

The analysis of differentially expressed genes (DEGs) was performed by determining the genes whose expression levels have been changed (up-regulated or down-regulated) between the normal and CRC tissues. Bioconductor (<http://www.bioconductor.org/>), an open source platform for genomic information analysis based on the R language, integrates a large number of various software packages. In this work, the limma package of bioconductor was used for determining and analyzing DEGs, and the DEG factor was set to 2 ( $|\log_{2}FC| > 1$ ). P value, an important indicator for the selection of DEGs, was set to 0.05. Fun Rich (<http://www.funrich.org/>) software was used to analyze the common DEGs in the three datasets, and the MeV software (<http://mev.tm4.org>) was used for the hierarchical clustering analysis on the common differentially expressed genes (see Figure S2 of Supporting Information).

### Analysis of gene ontology (GO) enrichment

The gene ontology (GO) database can provide the description and definition of genes and protein functions based on molecular functions, biological processes, and cellular components. Among them, the molecular function describes protein functions, such as RNA binding, ATP binding, oxidoreductase activity, etc. The biological process is a combination of molecular functions, such as cell cycle process, spindle construction, etc. The cellular component presents the composition of proteins in cells, including sub-cellular structure, macro-molecular complex, the late cell division promotion complex, nuclear ubiquitin ligase complex, etc. Here, we can search the corresponding ID number in the GO entry through a gene symbol, and then the information for functional category or cell location can be found. In this study, the DAVID database (<https://david.ncifcrf.gov/>) was used for the GO enrichment analysis of DEGs and adjusted  $P < 0.05$  was used for the inclusion criterion.

**Analysis of Genome Encyclopedia (KEGG) signal pathway** The Kyoto Encyclopedia of Genes and Genomes database, known as the KEGG database (<http://www.kegg.jp/>), was jointly constructed by the University of Tokyo and Kyoto University in Japan. The KEGG pathway enrichment analysis is mainly to determine the hyper-geometric distribution relationship between DEGs and the KEGG pathway by estimating the P value of each DEG and the existing pathway. The smaller the P value, the more significant the pathway enrichment. Based on the results from the analysis of the KEGG pathway enrichment, the pathway entries of enriched DEGs can be found, such that the functional roles of DEGs can be determined from a biological perspective.

**Protein-protein interaction (PPI) network** STRING (Search Tool for the Retrieval of Interacting Genes/Proteins, <http://www.string-db.org>) is a database providing detailed information on known proteins, and it has been widely used to construct the protein-protein interaction (PPI) network, including direct or indirect interactions among proteins. The PPI network can be derived based on high-throughput sequencing data, genomic environment, conservative co-expression relationships, and published research results. A specific scoring mechanism in STRING was employed to assign weights to the PPI network and then comprehensive scores were calculated. Here, we used the Cytoscape software for the visualization of the constructed PPI networks, from which the MCODE analysis was performed to extract the sub-networks. During the process, the “Degree” value of each protein node was estimated in order to identify key genes.

## RESULTS

**3.1 Identification of differentially expressed genes (DEGs) in CRC** The Affy package in R was used to standardize raw data from three different datasets (see Figure S1 of Supporting Information). Then, the Limma package of R was used to analyze the DEGs between CRC and normal tissues. From the three datasets, 2963 DEGs from GSE68468, 3327 DEGs from GSE21510, and 3021 DEGs from GSE9348 were determined respectively. Volcano plots were used to display the DEGs in three datasets (shown in Figure 1). In each figure, the up-regulated genes are represented by red dots and the down-regulated genes by green dots. Figure 2 shows that there are 849 common DEGs in total. Finally, a total of 406 up-regulated DEGs and 443 down-regulated DEGs were identified.

**3.2 Analysis of GO function and KEGG signaling pathway** In order to explore the roles of DEGs in CRC proliferation, the DAVID database was used for the GO enrichment analysis. It was found that

the up-regulated genes are mainly associated with cell division, proliferation, and metastasis processes, see details in Table 1. The down-regulated genes are mainly related to cell signaling and immune response, see details in Table 2. Analysis of the KEGG signaling pathway revealed that up-regulated genes mainly contribute to the formation of the cell cycle, ECM-receptor interaction, and the transmission of p53 signaling pathway (see details in Table 3). Down-regulated genes were mainly involved in the absorption of minerals in cells, the metabolism of nitrogen in cells, the re-absorption of bicarbonate by proximal tubules, and the signaling pathways of retinol metabolism and bile secretion (see details in Table 4).

**3.3 Analysis of protein-protein interaction (PPI) network** In order to investigate the mechanism of DEGs in CRC proliferation, the protein-protein interaction (PPI) networks of DEGs were constructed through the STRING database. The PPI networks of up-regulated genes and down-regulated are presented respectively in Figure 3. In this figure, the nodes represent DEGs enriched in the STRING database and the edges reflect the interactions among DEGs. In addition, the PPI networks were further analyzed by using the MCODE method of the Cytoscape software. The results showed that the core sub-networks contains 72 key up-regulated genes (see Table S1 of Supporting Information) and that 14 key down-regulated genes (see Table S2 of Supporting Information).

## DISCUSSION

**4.1 The up-regulated genes are mainly associated with cell division and proliferation.** From Table S1 of Supporting Information, it is seen that the up-regulated genes are mainly associated with cell division and proliferation. Based on the PPI network of up-regulated genes, we found that FBXO5 is the most closely correlated with CDK1, CCNB1, and CCNA2, which participate in the process of cell cycle and division. Moreover, we discovered that FBXO5 is also closely connected with TRIP13 and TTK. TRIP13 is a member of ATPase related to various cell active protein families and is a thyroid hormone receptor. Some studies have revealed that TRIP13 is essential for the inactivation of spindle assembly checkpoints and is related to the progression of certain cancers.<sup>12,13</sup> Human protein kinase unipolar spindle 1 (hMPS1), also known as TTK, is a protein kinase. It has been reported that TTK participates in mitotic checkpoints. Silencing TTK leads to increased H2AX phosphorylation and caspase 3/7 activation, and the depletion of TTK in TNBC cells induces apoptosis<sup>14-16</sup>. Basically, the cell cycle is regulated by members of CD gene family, and here we can find some key CD genes highly expressed in CRC, such as CDC7, CDC25A, CDC45, CDCA5, CDCA7, CDKN3, and CDK1. Melling et al.<sup>17</sup> have shown that CDC7 is highly expressed in CRC tissues and can be used as an independent prognostic biomarker. CDC25A is an important member of the CDC25 family, participating in the transition of the cell cycle from G1 phase to S phase and promoting cell proliferation. Although the role of CDC25A remain unclear, a variety of studies have shown that silencing CDC25A or treating with CDC25A inhibitors can reverse the chemoresistance of B7-H3-induced cancer cell.<sup>18</sup> CDC45 forms eukaryotic replication CMG helicase with MCM and GINS, which plays an important role in DNA replication. Hu et al.<sup>19</sup> also found that CDC45 had abnormally high expression level in CRC tissues compared with normal tissues. Dai et al.<sup>20</sup> showed that CDKN3 promotes cell proliferation and invasion by activating AKT signaling pathway in cancer cells. However, there is still no relevant reports about the role of CDKN3 in CRC proliferation. Figure 4 shows the PPI sub-network diagram of highly expressed CDKN3 and its interactions with other genes, including CDK1. Interestingly, from Table S2 of Supporting Information, it is seen that both CDKN3 and CDK1 involve in G1/S transition of mitotic cell cycle, suggesting that the G1/S transition of mitotic cell cycle should be a check point to control the CRC proliferation. In addition, we found that some members of kinesin family (KIF genes) were also highly expressed in CRC tissues, such as KIF4A, KIF11, KIF14, KIF15, KIF18A, KIF20A, and KIF23. Matsumot et al.<sup>21</sup> reported that the high expression level of KIF4A would promote the development of CRC cells and the knockdown of KIF4A in CRC cells would inhibit CRC cell proliferation. By examining the expression of KIF11 in 100 CRC patients, Imai et al.<sup>22</sup> found that 62 cases had high expression level of KIF11 and inhibiting the expression of KIF11 could reduce the CRC proliferation. Wang et al.<sup>23</sup> observed that KIF14 enhances the proliferation of CRC cells and facilitate the cell cycle by activating protein kinase B. Xiong et al.<sup>24</sup> found that KIF20A regulates a series of malignant features in CRC by activating JAK/STAT3 pathway. It has been found that KIF15, participating in the transport of macromolecules in basic cells, was highly expressed in liver cancer tissues

<sup>25</sup>. However, the association of KIF15 with CRC has not been reported so far. Figure 5 shows the PPI sub-network diagram of highly expressed KIF15 and its interactions with other key genes. From Table S2 of Supporting Information, one can see that both KIF15 and CDK1 involve in mitotic nuclear division, and this might suggest that silencing mitotic nuclear division could inhibit CRC proliferation.

**4.2 Down-regulated genes are mainly related to cell signaling and immune response** Based on the PPI network of down-regulated genes (in Figure 3B), 14 key genes are screened out (see Table S2 of Supporting Information), and the down-regulated genes are mainly related to cell signaling and immune response. Among them, GPER1, ADRA2A, and GPR18 are correlated with G protein-coupled receptors (GPCRs). GPCRs constitute the largest family of cell surface molecules involving in signal transmission. According to the PPI network analysis of down-regulated genes (shown in Figure 6), PYY has the highest correlation with SST, GPR18 and NPY1R, which are important genes for cell-cell signaling (see Table S2 of Supporting Information). PYY is a naturally occurring intestinal hormone that has a major inhibitory effect on multiple tissue targets. It is known that PYY coordinates with SST and GPR18 to participate in signal transmission process involving GPCRs. Various studies have shown that PYY plays an important role in regulating cell growth<sup>26</sup> and the reduced expression of PYY may be related to the occurrence and development of colon adenocarcinoma<sup>27-29</sup>. Among the down-regulated genes, one can find the chemokine gene family (CC genes), including CCL13, CCL19, and CCL28. Ahluwalia et al.<sup>30</sup> investigated the prognostic value of pro-inflammatory M1 gene expression in CRC patient through Kaplan-Meier curve analysis, finding that CCL13 can determine the overall survival rate of CRC patients. Xu et al.<sup>31</sup> reported that CCL19 is substantially under-expressed in CRC tissues and is positively correlated with high tumor microvessel density, suggesting that CCL19 can prevent the angiogenesis in CRC by regulating the Met/ERK/Elk-1/HIF-1 $\alpha$ /VEGF-A pathway. Based on the PPI network in Figure 6, we can see the interactions among CCL19, CCL13, GPER1, and INSL5, which play a critical role in positive regulation of ERK1 and ERK2 cascade as well as inflammatory response (see Table S2 of Supporting Information). This suggests that preventing the CRC proliferation can be achieved by regulating the correlation between cell-cell signaling genes and immune response genes.

## Conclusion

In order to determine the key genes that affect the CRC proliferation, we selected three datasets (GSE68468, GSE21510, and GSE9348) from the GEO database and performed bioinformatics analysis on them. First, based on the R platform, we screened out differentially expressed genes (DEGs) from these three datasets. Second, the GO enrichment analysis of DEGs was carried out to obtain information about the protein function description about genes. Subsequently, the KEGG pathway enrichment analysis was performed to determine the biological pathways involving in these genes. Finally, the STRING database was used to establish protein-protein interaction (PPI) networks, including the sub-networks composed of hub genes. Our results show that 72 up-regulated genes (in particular genes from CD and KIF families) are mainly involved in the process of cell division and proliferation. Based on the PPI network of up-regulated genes, we discovered that the interactions between up-regulated CD and KIF genes should enhance the CRC proliferation. Furthermore, we screened out 14 key genes based on the PPI network of down-regulated genes. It is shown that the down-regulated genes are mainly associated with cell signaling and immune response. Our work suggests that suppressing CRC proliferation can be achieved by activating the interactions between cell-cell signaling genes (PYY etc.) and immune response genes (CC genes).

## REFERENCES

1. Donovan MG, Selmin OI, Doetschman TC, et al. Mediterranean Diet: Prevention of Colorectal Cancer. *Frontiers in Nutrition* 2017;p. 59.
2. Torre LA, Bray F, Siegel RL, et al. Global cancer statistics, 2012. *CA: a cancer journal for clinicians* 2015;p. 87-108.
3. Jemal A, Bray F, Center M M, et al. Global cancer statistics. *CA: a cancer journal for clinicians* 2011; p. 69-90.
4. Vining D, Galfand D, Bechtold R. Technical feasibility of colon imaging with helical CT and virtual

- reality. *AJR Am J Roentgenol* 1994;p.104.
5. Williams NS, Gaynor RB, Scoggin S, et al. Identification and validation of genes involved in the pathogenesis of colorectal cancer using cDNA microarrays and RNA interference. *Clinical Cancer Research* 2003;p.931-946.
  6. Tsavaris N, Lazaris A, Kosmas C, et al. Topoisomerase I and IIalpha protein expression in primary colorectal cancer and recurrences following 5-fluorouracil-based adjuvant chemotherapy. *Cancer Chemother Pharmacol* 2009;p.391-398.
  7. Zhou Z, Liu S, Zhang M, et al. Overexpression of Topoisomerase 2-Alpha Confers a Poor Prognosis in Pancreatic Adenocarcinoma Identified by Co-Expression Analysis. *Dig Dis Sci* 2017;p.2790-2800.
  8. Gao CL, Wang GW, Yang GQ, et al. Karyopherin subunit-alpha 2 expression accelerates cell cycle progression by upregulating CCNB2 and CDK1 in hepatocellular carcinoma[J]. *Oncol Lett* 2018;p.2815-2820.
  9. Han D, Cho JH, Lee RH, et al. Antitumorigenic effect of atmospheric-pressure dielectric barrier discharge on human colorectal cancer cells via regulation of Sp1 transcription factor[J]. *Sci Rep* 2017;p.43081.
  10. Zheng HC, Liu JJ, Li J, et al. The in vitro and vivo effects of nuclear and cytosolic parafibromin expression on the aggressive phenotypes of colorectal cancer cells: a search of potential gene therapy target. *Oncotarget* 2017;p.23603-23612.
  11. Chen J, Wei Y, Feng Q, et al. Ribosomal protein S15A promotes malignant transformation and predicts poor outcome in colorectal cancer through misregulation of p53 signaling pathway. *Int J Oncol* 2016;p.1628-1638.
  12. Kurita K, Maeda M, Mansour M A, et al. TRIP13 is expressed in colorectal cancer and promotes cancer cell invasion[J]. *Oncology letters*, 2016, 12(6): 5240-5246.
  13. Banerjee R, Russo N, Liu M, et al. TRIP13 promotes error-prone nonhomologous end joining and induces chemoresistance in head and neck cancer[J]. *Nature communications*, 2014, 5(1): 1-18.
  14. Maia A R R, de Man J, Boon U, et al. Inhibition of the spindle assembly checkpoint kinase TTK enhances the efficacy of docetaxel in a triple-negative breast cancer model[J]. *Annals of Oncology*, 2015, 26(10): 2180-2192.
  15. Maire V, Baldeyron C, Richardson M, et al. TTK/hMPS1 is an attractive therapeutic target for triple-negative breast cancer[J]. *PloS one*, 2013, 8(5).
  16. Olesen S H, Thykjaer T, Ørntoft T F. Mitotic checkpoint genes hBUB1, hBUB1B, hBUB3 and TTK in human bladder cancer, screening for mutations and loss of heterozygosity[J]. *Carcinogenesis*, 2001, 22(5): 813-815.
  17. Melling N, Muth J, Simon R, et al. Cdc7 overexpression is an independent prognostic marker and a potential therapeutic target in colorectal cancer[J]. *Diagnostic pathology*, 2015, 10(1): 125.
  18. Ma Y, Wang R, Lu H, et al. B7-H3 promotes the cell cycle-mediated chemoresistance of colorectal cancer cells by regulating CDC25A[J]. *Journal of Cancer*, 2020, 11(8): 2158.
  19. Hu Y, Wang L, Li Z, et al. Potential Prognostic and Diagnostic Values of CDC6, CDC45, ORC6 and SNHG7 in Colorectal Cancer[J]. *OncoTargets and therapy*, 2019, 12: 11609.
  20. Dai W, Miao H, Fang S, et al. CDKN3 expression is negatively associated with pathological tumor stage and CDKN3 inhibition promotes cell survival in hepatocellular carcinoma[J]. *Molecular medicine reports*, 2016, 14(2): 1509-1514.
  21. Matsumoto, Y;Saito, M;Saito, K;Kanke, Y;Watanabe, Y;Onozawa, H;Hayase, S;Sakamoto, W;Ishigame, T;Momma, T. Enhanced expression of KIF4A in colorectal cancer is associated with lymph node metastasis *ONCOLOGY LETTERS* 2018,15,2188-2194.
  22. Imai T, Oue N, Sentani K, et al. KIF11 is required for spheroid formation by oesophageal and colorectal cancer cells[J]. *Anticancer research*, 2017, 37(1): 47-55.
  23. Wang Z Z, Yang J, Jiang B H, et al. KIF14 promotes cell proliferation via activation of Akt and is directly targeted by miR-200c in colorectal cancer[J]. *International journal of oncology*, 2018, 53(5): 1939-1952.
  24. Xiong M, Zhuang K, Luo Y, et al. KIF20A promotes cellular malignant behavior and enhances resistan-

ce to chemotherapy in colorectal cancer through regulation of the JAK/STAT3 signaling pathway[J]. *Aging* (Albany NY), 2019, 11(24): 11905.

25. Kitagawa A, Masuda T, Takahashi J, et al. KIF15 Expression in Tumor-associated Monocytes Is a Prognostic Biomarker in Hepatocellular Carcinoma[J]. *Cancer Genomics-Proteomics*, 2020, 17(2): 141-149.
26. Tseng W W, Liu C D. Peptide YY and cancer: current findings and potential clinical applications[J]. *Peptides*, 2002, 23(2): 389-395.
27. Fiocca R, Rindi G, Capella C, et al. Glucagon, glicentin, proglucagon, PYY, PP and proPP-icosapeptide immunoreactivities of rectal carcinoid tumors and related non-tumor cells[J]. *Regulatory peptides*, 1987, 17(1): 9-29.
28. Iwafuchi M, Watanabe H, Ishihara N, et al. Peptide YY immunoreactive cells in gastrointestinal carcinoids: immunohistochemical and ultrastructural studies of 60 tumors[J]. *Human pathology*, 1986, 17(3): 291-296.
29. Wilander E, El-Salhy M, Lundqvist M, et al. Polypeptide YY (PYY) and pancreatic polypeptide (PP) in rectal carcinoids[J]. *Virchows Archiv A*, 1983, 401(1): 67-72.
30. Ahluwalia P.;Bloomer C.;Pundkar C.;Jones K.;Mondal A.;Gahlay G.;Kolhe R. Investigating prognostic value of proinflammatory M1 gene expression in colorectal cancer patients. *Cancer Research*, 2019, 79.
31. Zhuoqing Xu; Congcong Zhu; Chun Chen; Yaping Zong; Hao Feng; Di Liu; Wenqing Feng; Jingkun Zhao; Aiguo Lu. CCL19 suppresses angiogenesis through promoting miR-206 and inhibiting Met/ERK/Elk-1/HIF-1 $\alpha$ /VEGF-A pathway in colorectal cancer. *Cell Death Dis*, 2018, 9(10): 974.

### Acknowledgments

This work is supported by the National Natural Science Foundation of China (No. 21863002), the Science and Technology Foundation of Guizhou Province (No. QKHJC-[2020]1Y040), the start-up fund from the Guizhou Education University, the construction project for Guizhou Provincial Key Disciplines (No. ZDXK[2015]10).

### TABLES

**Table 1.** GO enrichment analysis of up-regulated genes.

GO ID	GO functions	Number of DEGs	P value	Key Genes
GO:0051301	Cell division	41	1.33E-18	ITGB3BP, ASUN, NEK2
GO:0007067	Nuclear mitotic protein	34	1.64E-17	ITGB3BP, ASUN, NEK2
GO:0008283	Cell Proliferation	23	9.69E-06	CXCL1, STIL, CDK1
GO:0000086	Mitotic cell cycle, G2/M transition	17	2.82E-08	CEP131, CDK1, CEP72
GO:0000082	Mitotic cell cycle, G1/S transition	15	2.65E-08	CDC7, CDK1, CDKN3
GO:0006260	DNA replication	15	4.75E-06	CDC7, CDK1, DTL, DBF4
GO:0007062	Sister chromatid adhesion	15	3.01E-08	ITGB3BP, CENPQ

**Table 2.** GO enrichment analysis of down-regulated genes

GO ID	GO functions	Number of DEGs	P value	Key Genes
GO:0015701	Bicarbonate transportation	9	4.94E-06	SLC26A3, CA12, CA7
GO:0006805	Heterogeneous metabolic process	11	8.72E-06	CES2, SULT1B1, CYP2C18,
GO:1902476	Chloride ion transport across membranes	8	0.004725	BEST2, SLC26A3, CLCA1
GO:0007166	Cell surface receptor signaling pathway	14	0.008677	EDN3, BSG, TSPAN1
GO:0006814	Sodium ion transport	8	0.002174	SGK1, SCN3A, SLC17A4
GO:0090244	Wnt signaling pathway participant synthesis	3	0.016059	PPP2R3A, SFRP1, SFRP2
GO:0006810	Cell Signaling	15	0.024666	CLCA1, CLCA4, AQP8

**Table 3.** Analysis of the KEGG pathway of up-regulated genes.

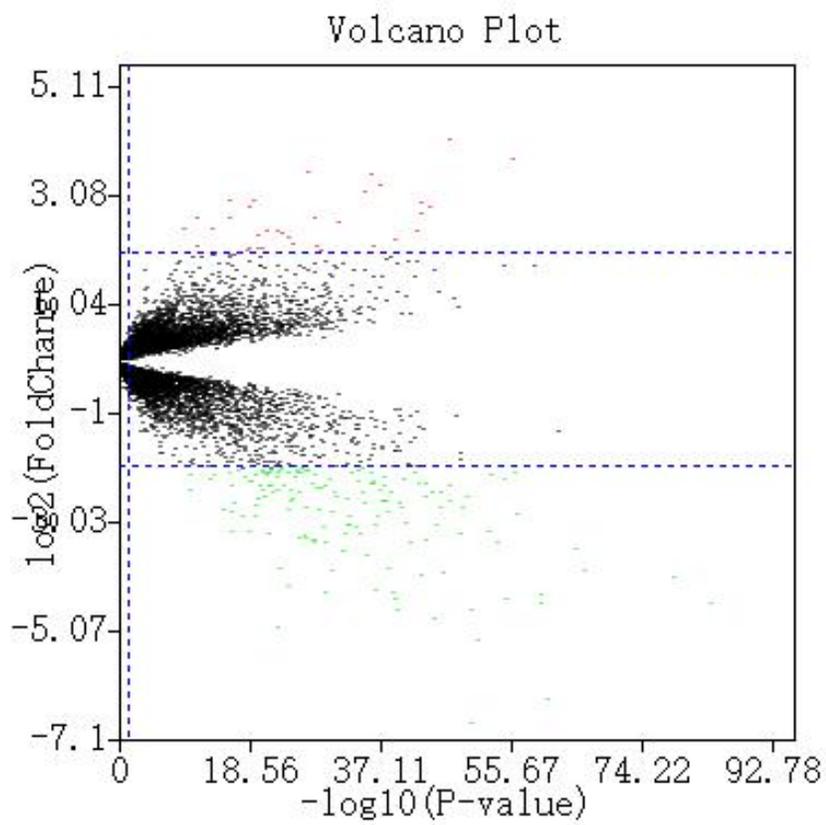
KEGG ID	KEGG Pathway	Number of DEGs	P-value	Key Genes
hsa04110	cell cycle	16	8.67E-09	CDC7, CDK1, DBF4, TTF
hsa04512	ECM-receptor interaction	9	2.19E-04	COL4A1, CD44, COMP,
hsa04115	p53 signaling pathway	8	2.48E-04	CCNB1, CDK1, CCND1
hsa04914	Progesterone-mediated maturation of oocytes	7	0.005872	CCNB1, CDK1, MAD2L1
hsa04114	Oocyte meiosis	7	0.018277	CCNB1, CDK1, MAD2L1
hsa03430	Mismatch repair	3	0.069263	RFC3, RFC4, MSH2

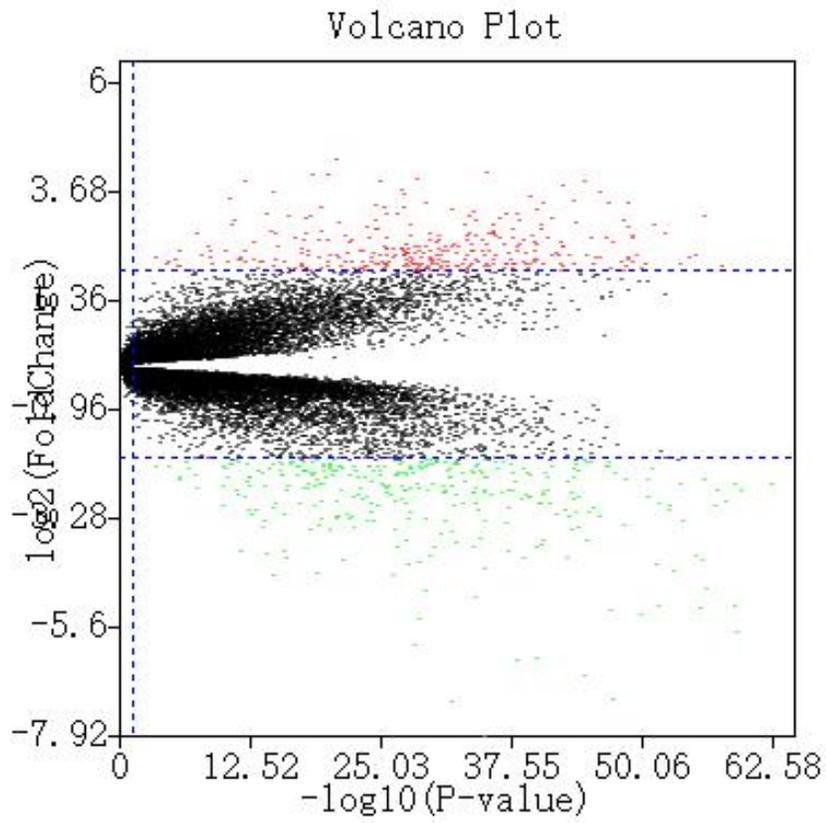
**Table 4.** Analysis of the KEGG pathway of down-regulated genes.

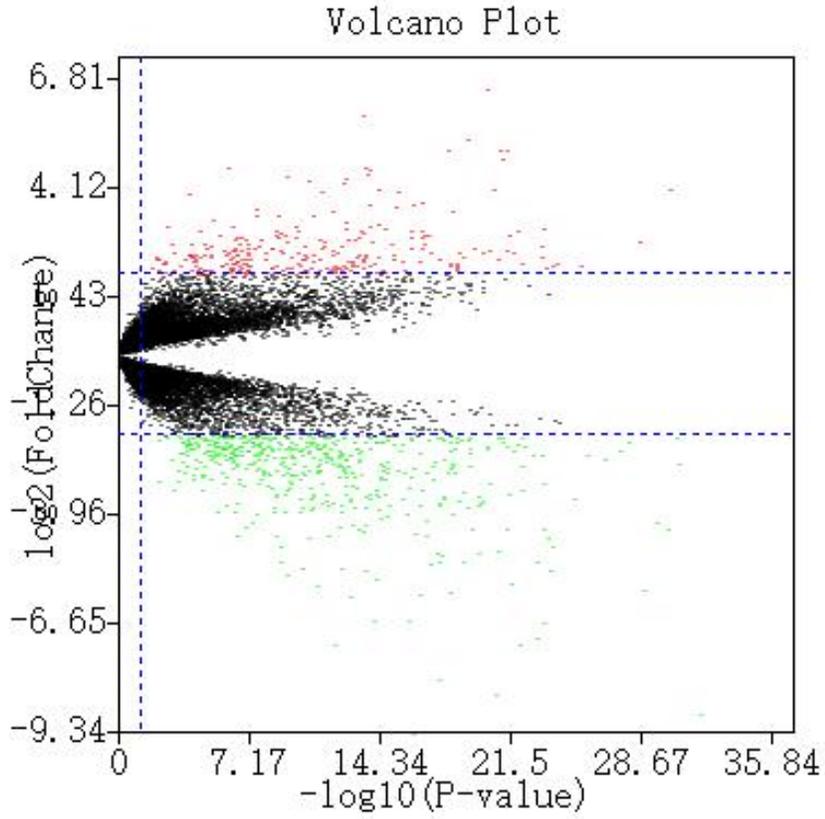
KEGG ID	KEGG Pathway	Number of DEGs	P-value	Key Genes
hsa04978	Mineral absorption	11	2.24E-07	SLC26A3, TRPM6, MT1M
hsa04976	Bile secretion	11	1.68E-05	ABCB11, AQP8, EPHX1
hsa04976	Retinol metabolism	11	8.45E-06	ALDH1A1,UGT2B17,CYP2C18
hsa00982	Drug Metabolism-Cytochrome P450	10	9.33E-05	GSTA1, UGT2B17, MAOA
hsa00983	Drug metabolism-other enzymes	9	3.10E-05	XDH, CES2, UGT2B17
hsa04964	Proximal tubule bicarbonate recovery	5	0.0033074	CA4, ATP1A2, CA2

## FIGURES

**Figure 1. Volcano plots of DEGs in three datasets.** GSE68468 on the left panel, GSE21510 in the middle, GSE9348 on the right. The upregulated genes are represented by red dots and the down-regulated genes by green dots.  $|\logFC| > 1$  and  $\text{adj.}P < 0.05$  were considered to be the cutoff values

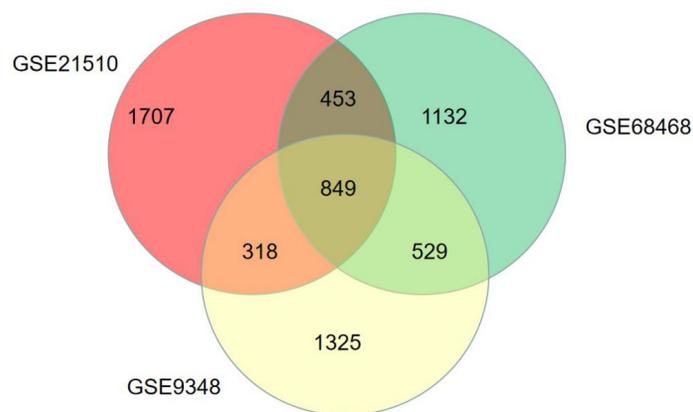




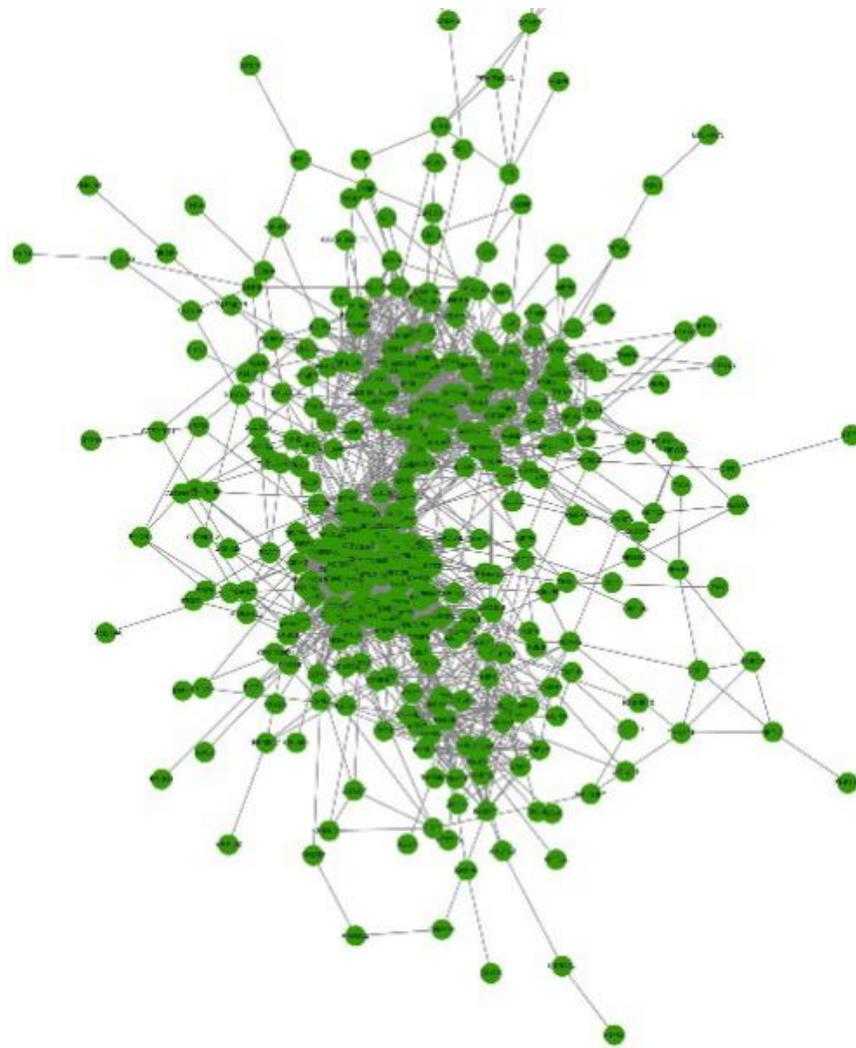


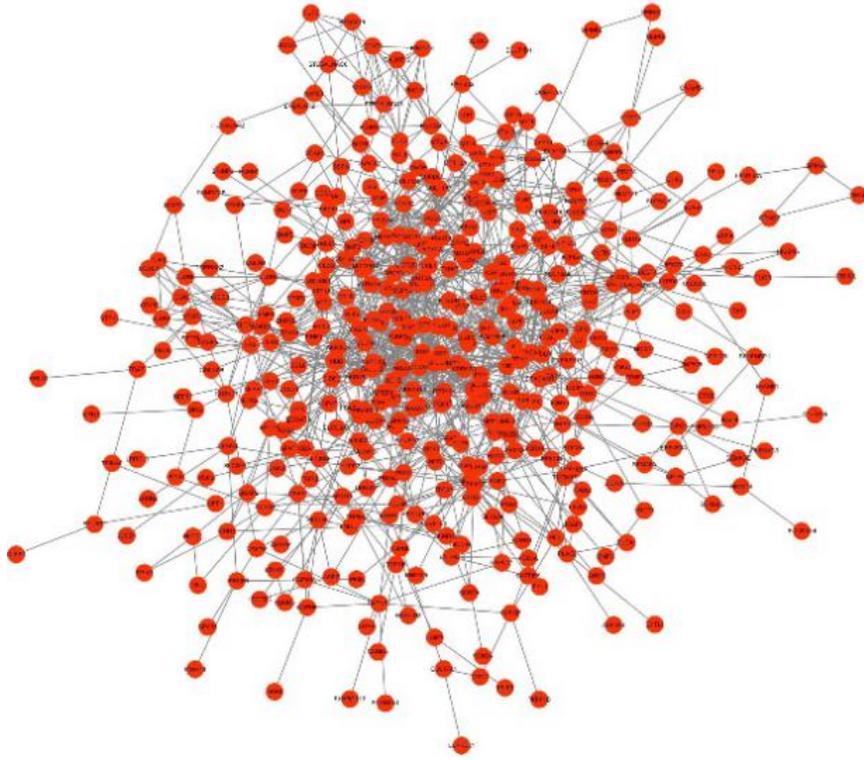
**GSE68468 GSE21510 GSE9348**

**Figure 2. Screening out DEGs from three GEO datasets: GSE68468, GSE21510, and GSE9348.**  $|\log_{2}FC| > 1$  and  $\text{adj. } P < 0.05$  were considered to be the cutoff values. There are 849 common DEGs in total.



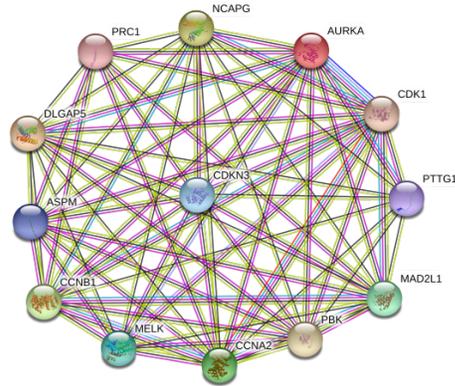
**Figure 3. Protein-protein interaction network (PPI) of DEGs .** Left panel represents the PPI network of up-regulated DEGs while right panel shows that of down-regulated DEGs.



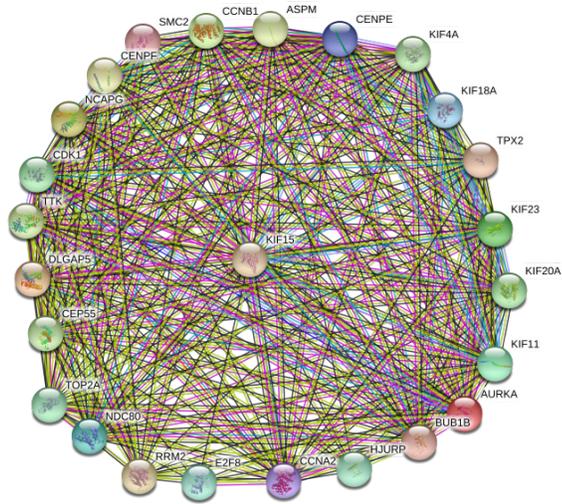


PPI network of up-regulated DEGs PPI network of down-regulated DEGs

**Figure 4.** PPI subnetwork diagram of up-regulated gene CDKN3 and the interactions with other highly expressed key genes



**Figure 5.** PPI subnetwork diagram of up-regulated gene KIF15 and the interactions with other highly expressed key genes



**Figure 6.** PPI sub-network diagram of down-regulated genes.

