


Research Article

Identification and verification of three key genes associated with survival and prognosis of COAD patients via integrated bioinformatics analysis

Yong Liu^{1,*}, Chao Li^{2,*}, Lijin Dong³, Xuewei Chen² and  Rong Fan^{2,4}

¹Department of General Surgery, Tianjin Xiqing Hospital, Tianjin 300380, P.R. China; ²Department of Operational Medicine, Tianjin Institute of Environmental and Operational Medicine, Tianjin 300050, P.R. China; ³Editorial Department of Education and Research Security Centre, Logistic University of Chinese People's Armed Police Force, Tianjin 300309, P.R. China; ⁴Central Laboratory, Tianjin Xiqing Hospital, Tianjin 300380, P.R. China

Correspondence: Xuewei Chen (chenxuewei11@sina.com) or Rong Fan (prosperity_39@sina.cn)



Background: Colorectal cancer (CRC) is the third most lethal malignancy in the world, wherein colon adenocarcinoma (COAD) is the most prevalent type of CRC. Exploring biomarkers is important for the diagnosis, treatment, and prevention of COAD. **Methods:** We used GEO2R and Venn online software for differential gene screening analysis. Hub genes were screened via Search Tool for the Retrieval of Interacting Genes/Proteins (STRING) and Cytoscape, following Gene Ontology (GO) and Kyoto Encyclopedia of Genes and Genomes (KEGG) enrichment analysis. Finally, survival analysis and RNA expression validation were performed via UALCAN online software and real-time PCR. Immunohistochemistry (IHC) was performed to verify the protein expression level of hub genes from tissues of COAD patients. **Results:** In the present study, we screened 323 common differentially expressed genes (DEGs) from four GSE datasets. Furthermore, four hub genes were selected for survival correlation analysis and expression level verification, three of which were shown to be statistically significant. **Conclusion:** Our study suggests that Serpin Family E Member 1 (SERPINE1), secreted phosphoprotein 1 (SPP1) and tissue inhibitor of metalloproteinase 1 (TIMP1) may be biomarkers closely related to the prognosis of CRC patients.

Introduction

Colorectal cancer (CRC) originates in the colon or rectum [1], and typical symptoms include alternating diarrhea, constipation, mucus, bloody stools, as well as weight loss [2]. According to annual reports on global cancer statistics, nearly 1.4 million newly diagnosed cases and 700000 deaths occur worldwide due to CRC [3]. The 5-year survival rate of patients with CRC at its early stage is 90%, while those at later stages show a rate of no more than 12% [4,5]. Colon adenocarcinoma (COAD) is one of the most prevalent types of CRC, the incidence and mortality of which was 10.2 and 9.2%, respectively, in 2018 [6,7]. Therefore, it is very important to select and identify the specific biomarkers of COAD for its early diagnosis, development of an efficient treatment strategy, and assessment of patient prognosis.

Gene microarray technology has been extensively applied in several fields of medicine and biology. This high-throughput technology can effortlessly acquire gene expression profile information, which is very useful for exploring various molecular mechanisms [8]. Recently, integrated bioinformatics analysis has also been widely applied; it includes mathematics and biology, which makes it possible to analyze large-scale microarray data. In addition, it is advantageous as an efficient method for the systematic screening of tumor-associated genes [9,10].

*These authors contributed equally to this work.

Received: 23 January 2020

Revised: 21 August 2020

Accepted: 10 September 2020

Accepted Manuscript online:

15 September 2020

Version of Record published:

28 September 2020

In the present study, we screened the common differentially expressed genes (DEGs) in four GSE datasets from the Gene Expression Omnibus (GEO) repository and constructed a protein–protein interaction (PPI) network. Via integrated bioinformatics analysis, we identified the four hub genes associated with survival of COAD patients and verified the mRNA expression levels in CRC cells. Finally, our results showed that secreted phosphoprotein 1 (SPP1), Serpin Family E Member 1 (SERPINE1) and tissue inhibitor of metalloproteinase 1 (TIMP1) could be potential biomarkers for prognostics of COAD in patients.

Materials and methods

Download of microarray data

The high-throughput gene expression profiles of normal colorectal tissues and COAD tissues were extracted from the GEO database. The independent datasets of GSE37364, GSE41328, GSE81558, and GSE110225 were selected which included 14 COAD tissues and 37 normal tissues, 10 COAD tissues and 10 normal tissues, 23 COAD tissues and 9 normal tissues, and 17 COAD tissues and 17 normal tissues, respectively.

Identification of DEGs

The DEGs between normal colorectal tissues and COAD tissues were screened via the GEO2R tool. We limited the cut-off criterion of $|\log_{2}FC| \geq 1$ to be statistically significant. Then, the DEGs in the four datasets were screened out using online Venn software. The DEGs with $\log_{2}FC \geq 1$ were considered as up-regulated genes, while DEGs with $\log_{2}FC \leq -1$ were considered as down-regulated genes.

Enrichment analysis via Gene Ontology and Kyoto Encyclopedia of Genes and Genomes pathways

In order to characterize the functional roles of the DEGs, we used Database for Annotation, Visualization and Integrated Discovery (DAVID; version 6.8 [11]) database for Gene Ontology (GO) and signal pathway enrichment analysis, which comprised biological process (BP), molecular function (MF) and cellular component (CC), and Kyoto Encyclopedia of Genes and Genomes (KEGG) pathways analysis.

Construction of PPI network and analysis of module

The PPI network was built via the Search Tool for the Retrieval of Interacting Genes/Proteins (STRING) [12] database to uncover the relationships of DEGs, based on a minimum required interaction score of 0.4. We used Cytoscape (version 3.7.1) [13] software to complete the analysis and visualization of the PPI network. In addition, the remarkable modules of PPI network were selected via the Molecular Complex Detection (MCODE) app (node score cutoff = 0.2, degree cutoff = 2, max. depth = 100, and k-score = 2).

Survival and RNA expression analysis of hub genes

UALCAN [14] was used for analyzing the relationship between hub genes expression and survival of patients with COAD, as it is a good tool for analyzing transcriptome data of cancers from The Cancer Genome Atlas (TCGA). Simultaneously, UALCAN was used to analyze RNA expression data based on TCGA. *P*-value < 0.05 was considered statistically significant.

Cell lines and culture

The normal colon epithelial cell line, NCM460, and COAD cell lines, HCT116, HCA-7, HT-29, and SW480 were maintained by our laboratory. And all cells were cultured in a 5% CO₂ humidified incubator in Dulbecco's modified Eagle's medium (HyClone) containing 10% fetal bovine serum (FBS, Gibco) at 37°C.

Quantitative real-time polymerase chain reaction (PCR)

Total RNA from cell cultures was collected by using Total RNA Extraction Kit (Solarbio Life Sciences, China). The specific primers were designed by using Primer Express 3.0 software. Each sample was used in triplicate for the qPCR using PowerUp™ SYBR™ Green Master Mix (Thermo Fisher Scientific, U.S.A.) on Applied Biosystems™ StepOnePlus instruments, and analyzed with the StepOne software, using GAPDH as a control. The primers used in the present study are designed as follows: TIMP1: Forward: 5'-GGGACACCAGAAGTCAACCAG-3'; Reverse: 5'-CAATGAGAACTCCTCGCTGC-3'; C-X-C motif chemokine ligand 2 (CXCL2): Forward: 5'-CCAAACCGAAGTCATAGCCAC-3'; Reverse: 5'-TGCCATTTTTCAGCATCTTTTC-3'; SPP1: Forward:

Table 1 A total of 323 commonly DEGs were screened from four profile datasets, including 111 up-regulated and 212 down-regulated genes in COAD tissues compared with normal colorectal tissues

DEGs	Gene symbol
Up-regulated	<i>CD44, ADAM12, CXCL3, PLAU, GDF15, MMP7, TRIB3, SERPINB5, FOXQ1, MET, SFTA2, RFC3, RGS16, CDC25B, COL1A1, ANLN, DUSP4, SLC6A6, TRIM29, CLDN2, VSNL1, PHLDA1, CHI3L1, NEBL, CXCL11, CXCL8, MSX2, PPM1H, SOX9, STC1, GRHL1, CADPS, THY1, OSBPL3, HS6ST2, RAD54B, SPP1, CA9, BMP7, OLR1, EPHX4, NOX4, SFRP4, SLC7A11, LRRC8E, TIMP1, BGN, TNFRSF12A, IBSF, MMP1, S100A11, CBX2, COL8A1, CRNDE, TEAD4, CDH3, SLC04A1, MMP3, UBE2S, KRT80, FAP, INHBA, SLC7A5, CXCL2, GALNT6, UBE2C, MYEOV, FXYD5, SRPX2, HOMER1, PSAT1, TMPPRSS3, LRP8, NFE2L3, CEMIP, AZGP1, TACSTD2, TNS4, CELSR1, PMAIP1, S100A2, CXCL1, SORD, DPEP1, FABP6, PPAT, SERPINE1, MSLN, KLK10, PLEKHS1, EVA1A, SIM2, AJUBA, HILPDA, SHISA2, CLDN1, ENC1, STC2, COL10A1, KRT6B, SP5, PMEPA1, TNFSF9, COL11A1, CTHRC1, TESC, TGFB1, COL1A2, ZAK, C2CD4A, KRT23</i>
Down-regulated	<i>PLCE1, XDH, HSD17B2, NAP1L2, EPB41L4B, ZG16, TMCC3, DHRS11, HSD11B2, TRPM6, IGH, CES2, PYY, VWA5A, MAOA, BCHE, THRB, TSPAN7, HHLA2, AHCYL2, IL6R, PKIB, CDHR5, PBLD, FHL1, PTPRH, SORBS2, KLF4, TCF21, NEDD4L, SLC51B, AQP8, LYPD8, TNFRSF17, CD177, C1orf115, SLC25A34, HPGDS, ADH1C, SGK2, TMEM72, SLC17A4, NPY1R, UGT2A3, GCNT2, LRRC19, SST, SCNN1B, DENND2A, CLIC5, PRKACB, TRHDE, EPB41L3, LAMA1, CWH43, NRXN1, MEP1A, PDK4, GUCA2A, ANGPTL1, SLC22A5, GREM2, LGALS2, NR3C2, CHGB, OGN, SLC22A23, CA4, NKX2-3, APPL2, PRIMA1, NAAA, RGS13, SEMA6D, LINC01082, FUCA1, LRMP, SLC26A2, ADTRP, VSTM2A, MT1F, PLPP1, SOSTDC1, ACACB, MS4A12, CDH19, VSIG2, ABCC13, CEACAM1, C2orf88, PLP1, BEST4, STMN2, PPP2R3A, TMEM56, TEX11, SLC4A4, C2orf40, MEP1B, SI, TMEM37, EYA2, PDE6A, CHRDL1, LDHD, PHLPP2, AKR1B10, TUBAL3, ARL14, RMDN2, PTGDR, SRI, POU2AF1, TFCP2L1, DNASE1L3, MIER3, SCIN, CKB, SLC26A3, ACOX1, PIGR, MYOT, RCAN2, CHGA, ITGA8, MT1M, TP53INP2, CNR1, LRRC66, PDE9A, ABCA8, SLC13A2, SLC16A9, CLDN23, EDN3, ANK2, HIGD1A, SCGN, PCK1, CPM, SMPDL3A, ABI3BP, WDR78, C1orf210, RUNDC3B, USP2, HPGD, FKBP1B, C7, MOGAT2, EFHC2, ECI2, EPHX2, SEMA6A, TMEM100, DHRS9, CA7, ENTPD5, SPINK2, CPNE8, CNTN3, ABCG2, MMP28, GUCA2B, CHP2, SCARA5, CAPN13, CLCA4, ATP1A2, MYO1A, NXPE4, ETFDH, MB, LIFR, MX11, GDDP3, CXCL12, NR5A2, C16orf89, FMO5, PAQR5, CCL28, CA2, BCAS1, SMIM14, ANPEP, KCNIP4, CEACAM7, PADI2, METTL7A, ITM2A, ARHGAP44, NR1H4, TMEM220, BEST2, ADH1B, ZNF575, GCG, GBA3, GPM6B, ANK3, TCEA3, NXPE1, LPAR1, MAMDC2, CA1, SULT1B1, DPP10, SFRP1, SELENBP1, GPAT3, HAPLN1</i>

5'-TGAGCATTCGGATGTGATTGAT-3'; Reverse: 5'-TTATCTTCTTCCTTACTTTTTGGGGT-3'; SERPINE1: Forward: 5'-CTGGGTGAAGACACACAAAAAG-3'; Reverse: 5'-CACAGAGACAGTGCTGCCGT-3'; GAPDH: Forward: 5'-TCAAGGGCATCCTGGGCTACAC-3'; Reverse: 5'-TCAAAGGTGGAGGAGTGGGTGTC-3'.

Tissue samples and immunohistochemistry

Sixty paired primary cancer tissues and adjacent tissues from colon cancer patients who underwent surgery between September 2011 and April 2020, were collected at the Tianjin Xiqing Hospital. No chemotherapy and radiotherapy was given to patients before surgery. All the pathology materials were reviewed independently by two pathologists. Informed consent was obtained from each patient, and the collection of tissue specimens was approved by the Institutional Review and Ethics Boards of the Tianjin Xiqing Hospital. The two-step EnVision method was used to perform immune histochemical experiments, along with different primary antibodies against CXCL2 (1:50), SERPINE1 (1:50), TIMP1 (1:50) and SPP1 (1:50). The slides were analyzed using the Image-Pro Plus software program (Media Cybernetics, Inc. U.S.A.).

Statistical analysis

Statistical analysis was performed using SPSS 22.0 and GraphPad Prism 8.0 software. Results were presented as the mean \pm standard deviation. The statistically significant changes were indicated with asterisks and the *P*-values were calculated via Student's *t* test, wherein *, **, and *** represented $P < 0.05$, $P < 0.01$ and $P < 0.001$, respectively.

Results

Identification of DEGs in COAD

In the present study, we selected and downloaded four GEO datasets and extracted DEGs on the basis of the cut-off criteria. The present study covered 64 COAD and 73 normal colorectal tissues in total. The results showed that 2729, 957, 1864, and 617 genes were down-regulated while 1917, 770, 1282, and 469 genes were up-regulated in GSE37364, 41328, 81558, and 110250, respectively. By using the Venn diagram software, we detected a total of 323 common DEGs in the COAD samples, composed of 111 up-regulated genes and 212 down-regulated genes (Table 1 and Figure 1).

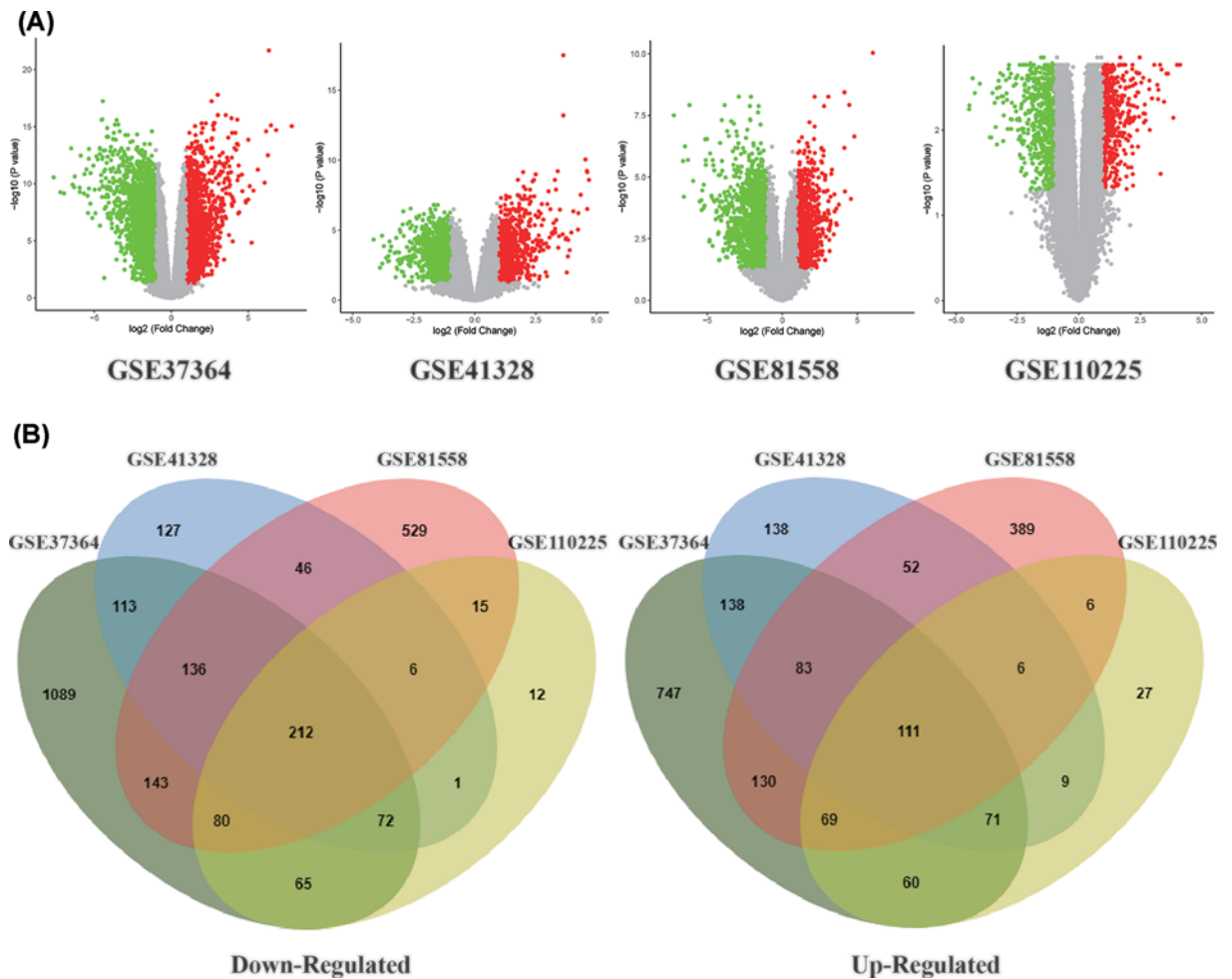


Figure 1. Selection of 323 common DEGs from the four datasets (GSE37364, GSE41328, GSE81558, and GSE110225) (A) Volcano plot of DEGs from the four datasets. (B) On the left, 136 DEGs are down-regulated ($\log_{2}FC \leq 1$). On the right, 198 DEGs are up-regulated ($\log_{2}FC \geq 1$).

GO and signaling pathway enrichment analysis

We analyzed the DEGs via the DAVID software and identified 125 significant enrichment terms, including BP (71), MF (24), and CC (26). With regard to BP, DEGs were seen to be significantly enriched during cell adhesion, negative regulation of cell proliferation, angiogenesis, cytoskeleton organization, steroid hormone-mediated signaling pathway, inflammatory response, negative regulation of protein kinase activity, and so on. Whereas, for MF, DEGs were significantly enriched during chemokine activity, hormone activity, CXCR chemokine receptor binding, receptor binding, calcium ion binding, structural constituent of cytoskeleton, protein kinase inhibitor activity, oxidoreductase activity etc. With regard to CC, DEGs were significantly enriched in extracellular space, apical plasma membrane, extracellular exosome, extracellular region, cell surface, plasma membrane, basal plasma membrane, and so forth. The top 20 GO terms are depicted in Figure 2A–C. Furthermore, KEGG pathway analysis indicated that the DEGs were enriched in 11 pathways including nitrogen metabolism, bile secretion, proximal tubule bicarbonate reclamation, cytokine–cytokine receptor interaction, aldosterone-regulated sodium reabsorption, mineral absorption, phosphoinositide-3-kinase (PI3K)-Akt signaling pathway, PPAR signaling pathway etc. (Figure 2D).

Construction of PPI network and module analysis

Via the STRING and Cytoscape analysis, we drew the PPI network complex constructed with 268 nodes and 709 edges, consisting of 97 down-regulated and 171 up-regulated genes. Then we conducted further analyses by applying Cytoscape MCODE plus, and found 32 central nodes that included 20 up-regulated genes and 12 down-regulated genes (Figures 3 and 4 and Table 2).

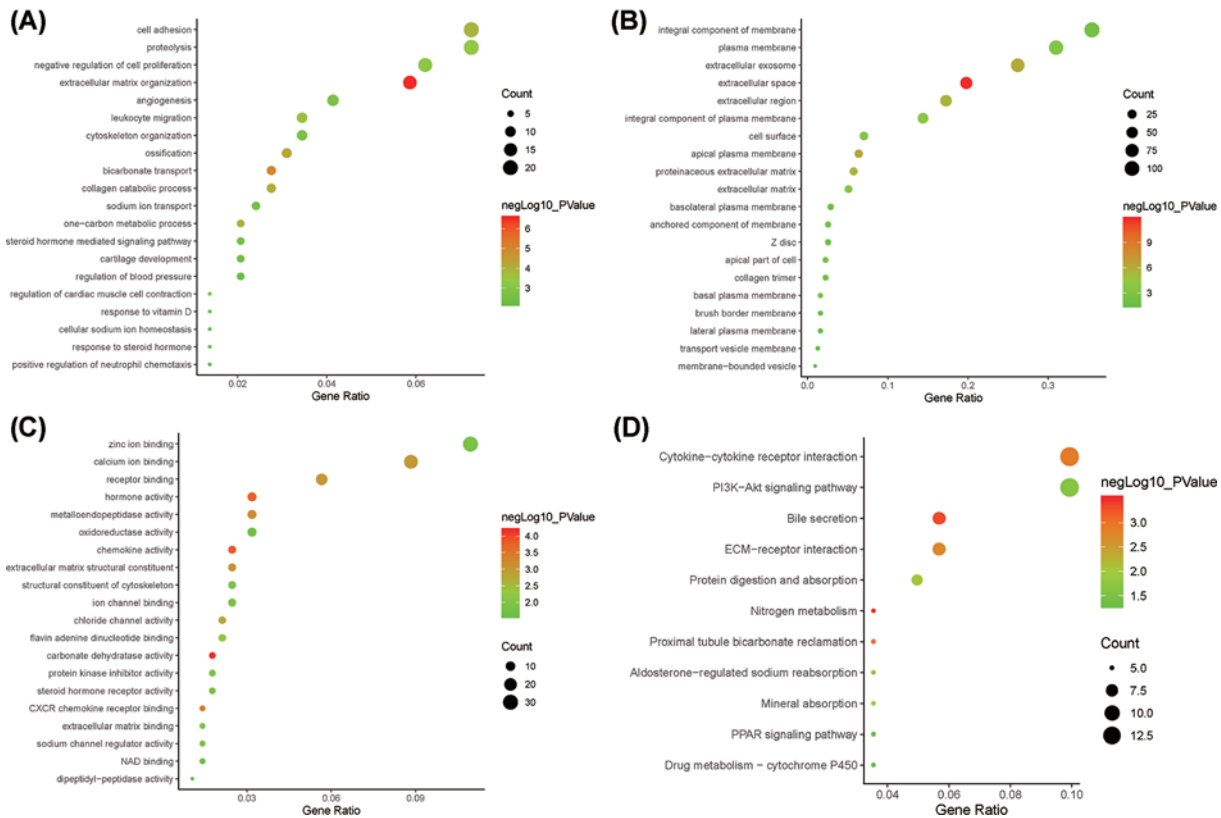


Figure 2. The top 20 GO and significantly enriched KEGG pathways (A) BP; (B) CC; (C) MF; (D) KEGG pathways.

Table 2 A total of 32 central genes were selected by STRING and Cytoscape software from PPI network including 20 up-regulated and 12 down-regulated genes

DEGs	Gene symbol
Up-regulated	<i>BGN, BMP7, COL1A1, COL1A2, CXCL1, CXCL11, CXCL2, CXCL3, CXCL8, EVA1A, IBSP, MMP1, MMP3, MMP7, MSLN, SERPINE1, SPP1, STC2, THY1, TIMP1</i>
Down-regulated	<i>CCL28, CD44, CHGB, CHRDL1, CNR1, CXCL12, LPAR1, NPY1R, PLAU, PYY, SOX9, SST</i>

Selection of hub genes and validation of the expression levels

We used UALCAN to identify the survival data of 32 central genes and found that the expression levels of four hub genes, namely *CXCL2*, *SERPINE1*, *SPP1*, and *TIMP1*, were remarkably related to the survival of COAD patients (Figure 5). Furthermore, we validated the expression level of these four hub genes in COAD and normal tissues via UALCAN. Results showed that expression levels of the four genes were significantly increased in the COAD samples as compared with the normal samples (Figure 6). Furthermore, we verified the mRNA levels of the four hub genes in NCM460, HCT116, HCA-7, HT-29, and SW480 cells. The real-time PCR results showed that the expression levels of *SPP1*, *TIMP1*, and *SERPINE1* mRNAs were significantly increased in the HCT116, HCA-7, HT-29, and SW480 cell lines as compared with the NCM460 cell line (Figure 7). Moreover, *CXCL2* expression was higher in the HCA-7, HT-29, and SW480 cells, but not in the HCT116 cells when compared with NCM460. In addition, we also analyzed the protein expressions of the four hub genes by using tissue microarray. The results indicated that *SPP1*, *TIMP1*, and *SERPINE1* expression appeared to be remarkably higher in CRC tissues than in the adjacent tissues, while the expression of *CXCL2* was not statistically different between cancer tissues and adjacent tissues (Figure 8).

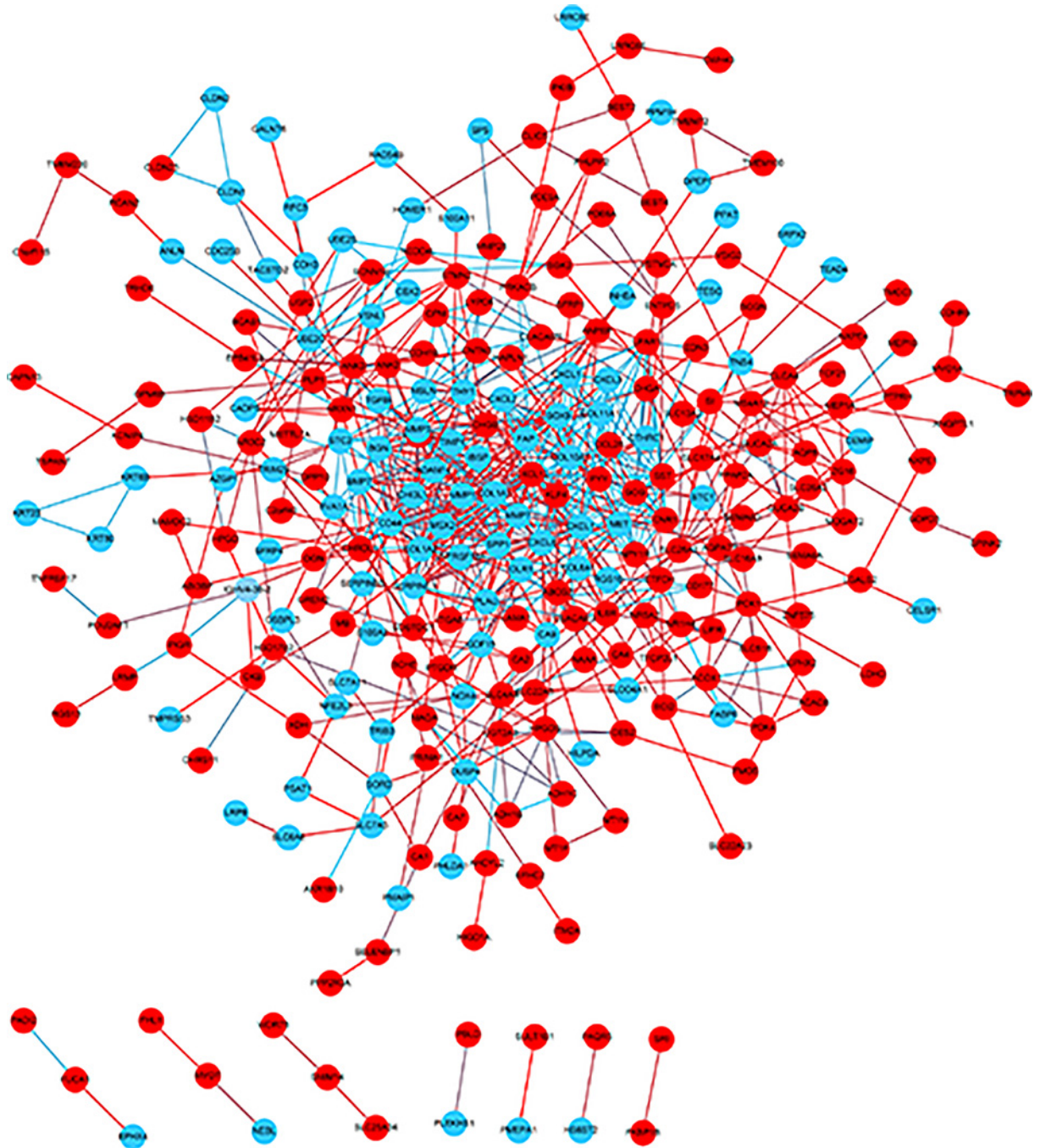


Figure 3. PPI network constructed via STRING and Cytoscape

Each node indicates a protein module; the edges represent proteins interaction; blue circles denote up-regulated DEGs while red circles represent down-regulated DEGs.

Discussion

CRC is the third main cause of deaths from malignant cancers worldwide, and the fourth most common cancer in China [15,16]. One of the most common types of CRC is COAD, which originates in the superficial glandular epithelial cells [6]. Previous studies have indicated that the mechanism behind the tumorigenesis and development of COAD is a complicated process, including chromosomal instability, dysregulation of oncogenes and tumor suppressor, epigenetic alterations, metabolic alterations, abnormal immune response etc. [17].

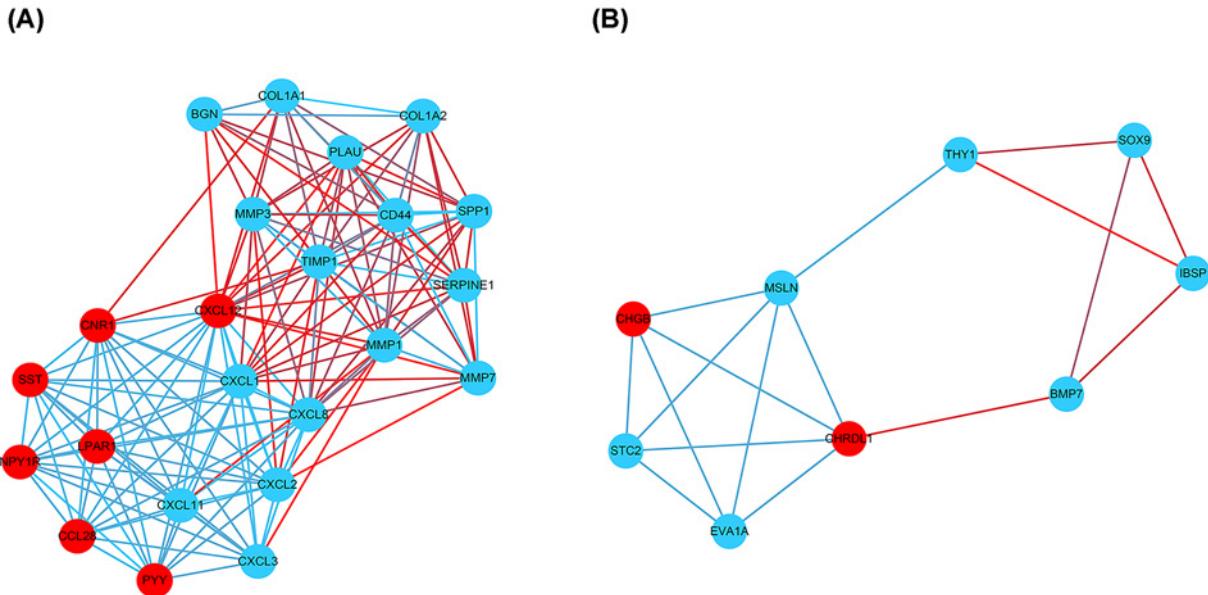


Figure 4. Module analysis by MCODE in the Cytoscape software
(A) Module1; (B) module2 (blue circles signify up-regulated DEGs and red circles denoted down-regulated DEGs).

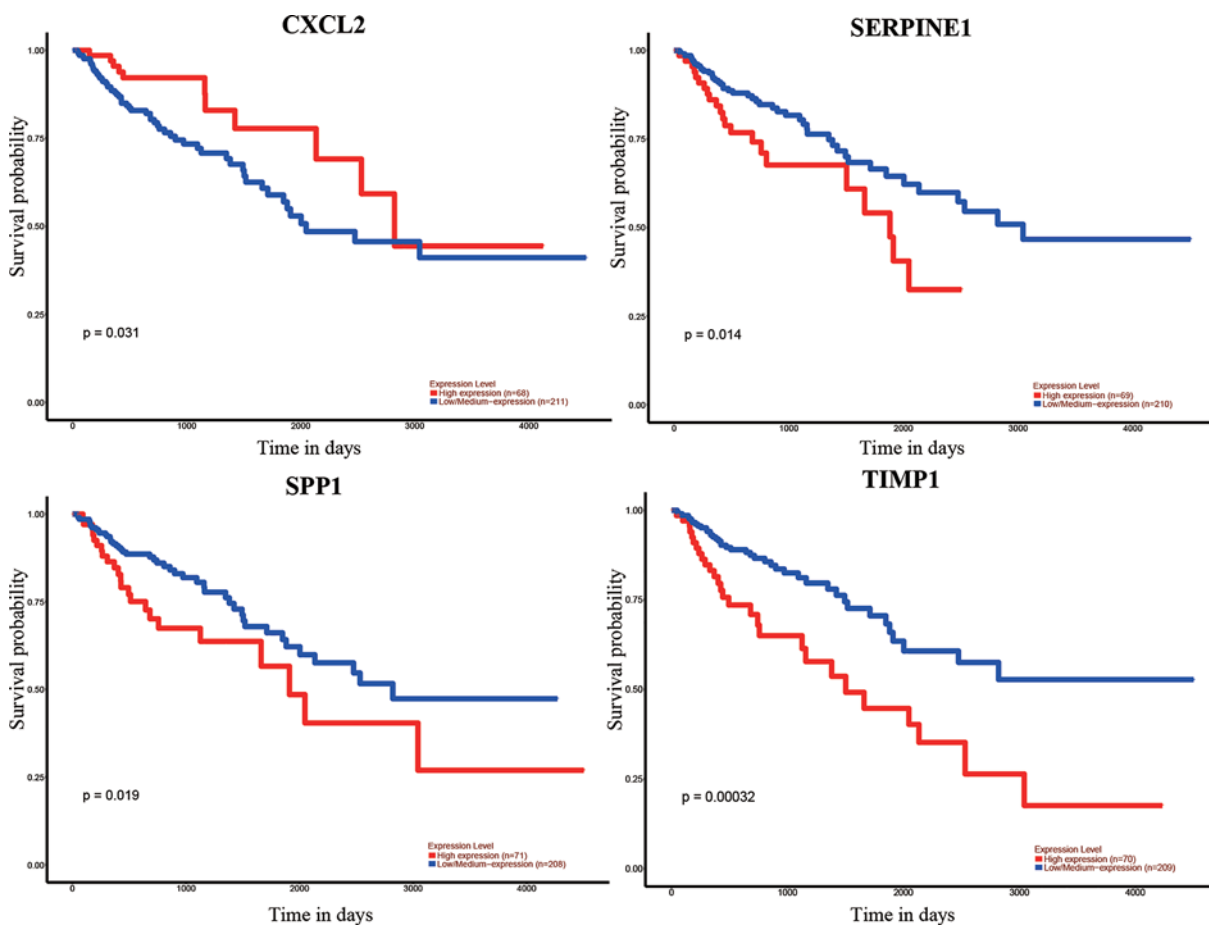


Figure 5. Correlation between the four hub genes expression and the overall survival of COAD patients, via UALCAN analysis

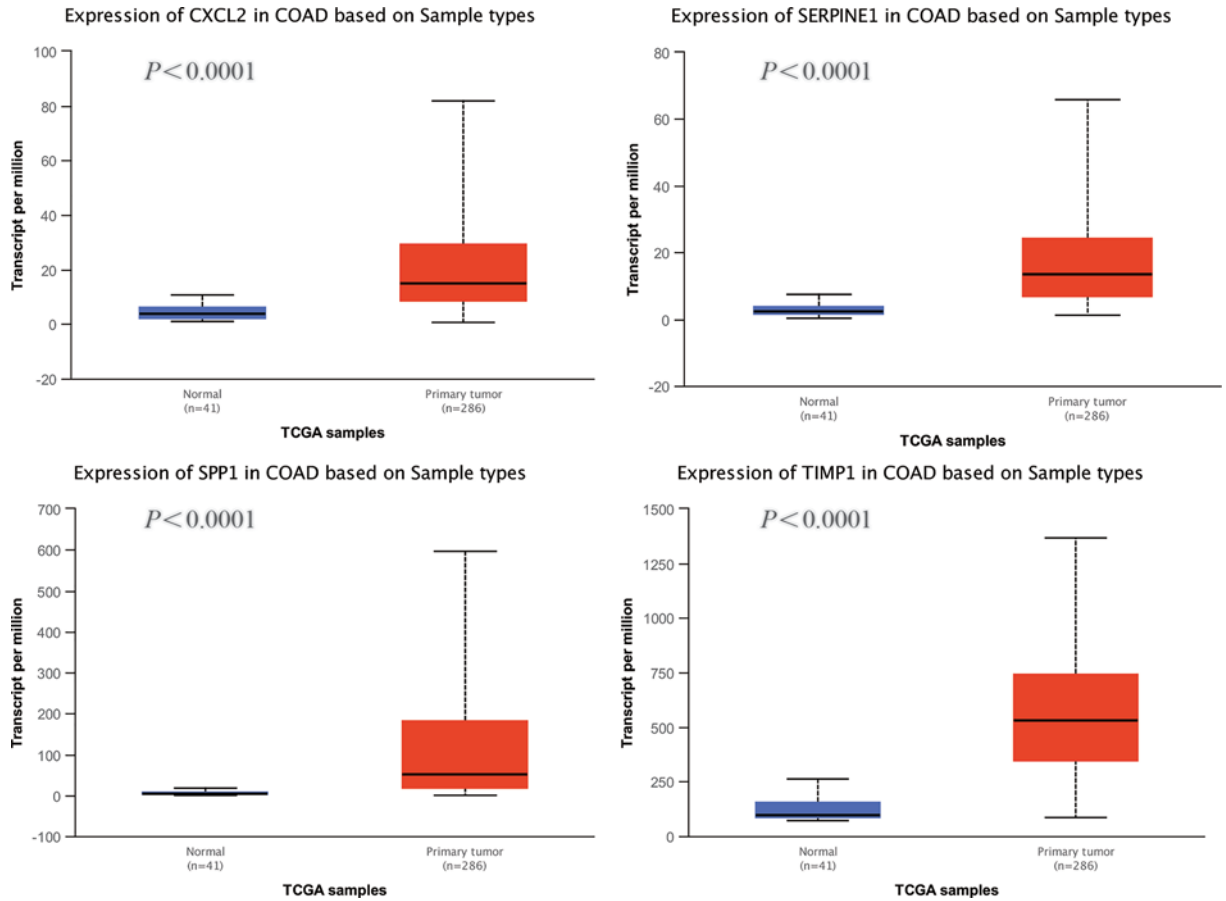


Figure 6. Analysis of the four hub genes expression levels via UALCAN analysis

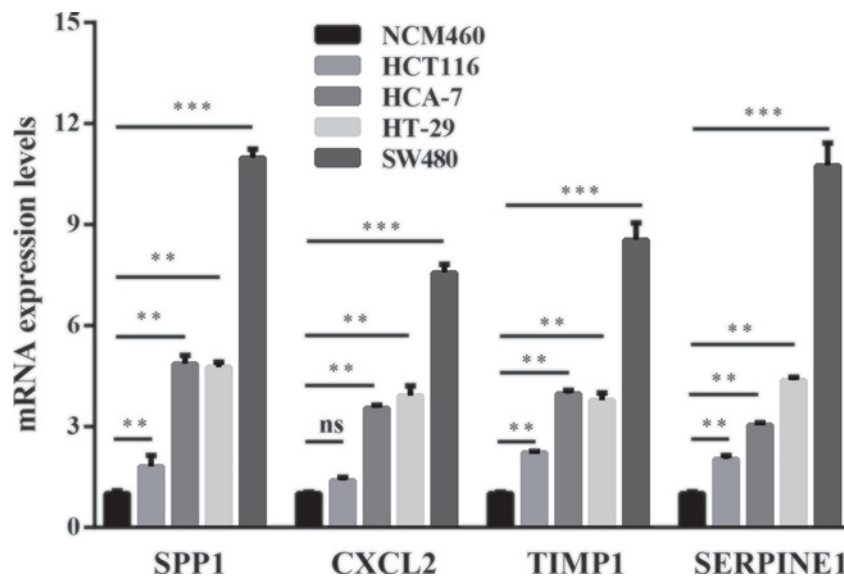


Figure 7. Relative mRNA levels of the four hub genes in HCT116, HCA-7, HT-29, and SW480 cells compared with NCM460 cells, via real-time PCR

** Represents $P < 0.01$, *** represents $P < 0.001$, ns represents no significance.

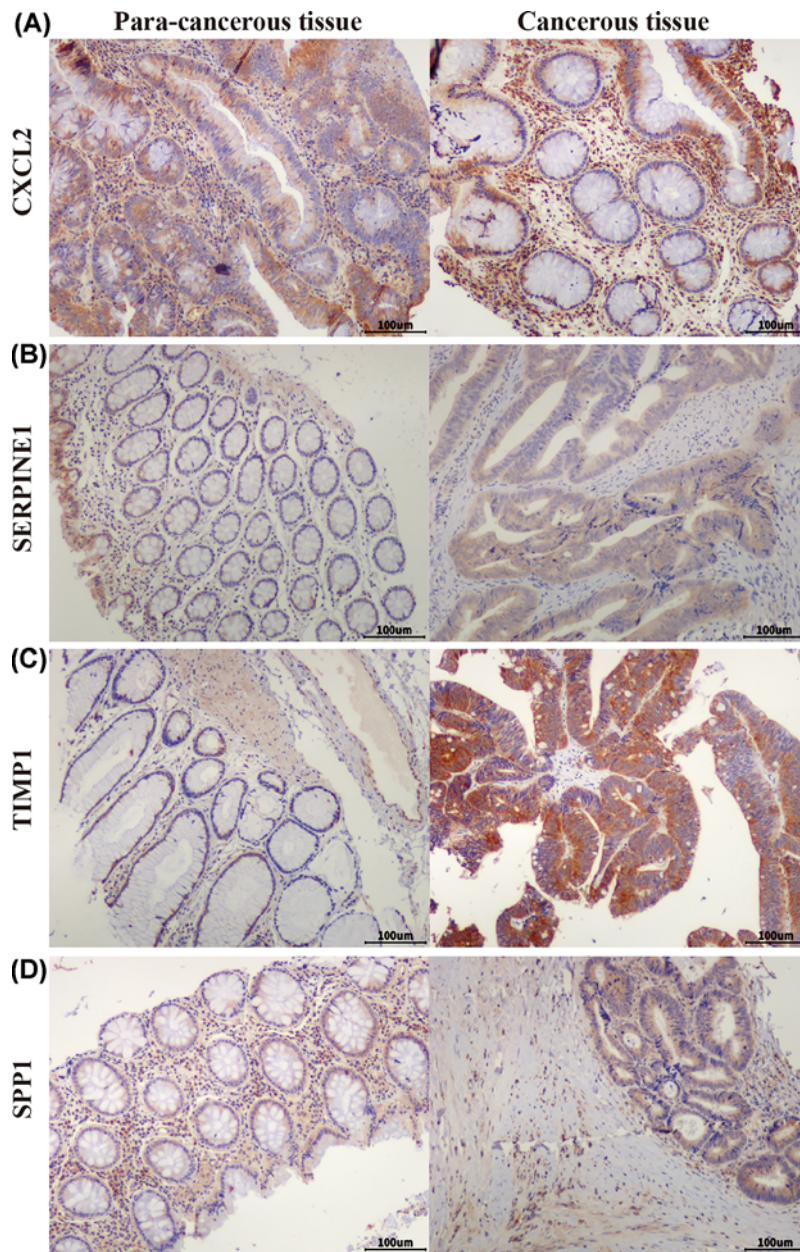


Figure 8. Expression of four hub genes in colorectal tissues

Expression of CXCL2 (A), SERPINE1 (B), TIMP1 (C), and SPP1 (D) in colorectal adenocarcinoma tissue and adjacent tissue in tissue microarrays (scale bar: 100 μm).

In the present study, we screened DEGs via GEO2R, and selected 111 up-regulated genes and 212 down-regulated genes found in COAD tissues. Then we analyzed their enrichment in GO and signaling pathways and found that the DEGs were enriched in 11 pathways, including the PI3K-Akt signaling pathway. Previous studies showed that several signal pathways could be involved in the progress and development of CRC [18–23]. PI3K/AKT is one of the classic pathways responsible for multiple BPs, such as cell proliferation, differentiation, and migration. As a key effector of PI3K, Akt phosphorylation is closely related to CRC cell proliferation and apoptosis inhibition [18]. The specific inhibition of the PI3K/AKT signaling pathway could result in a decrease in proliferation and increase in apoptosis of the cells [24].

Furthermore, we constructed a PPI network and performed survival analysis along with verification of the RNA expression levels via the GEO database, DAVID and STRING online tools, Cytoscape software, and UALCAN online

analysis. After the above series of operations, four hub genes, *CXCL2*, *SERPINE1*, *SPP1*, and *TIMP1*, were selected. Furthermore, we tested the expression of these four genes at the mRNA as well as protein levels and found that *SPP1*, *TIMP1*, and *SERPINE1* had higher expression levels in COAD cell lines than the colon epithelial cell line as well as in CRC tissue compared with para-cancerous tissue.

SPP1, named as osteopontin, is a phosphoprotein secreted by malignant cells. Cheng et al. [25] found that *SPP1* has a higher expression level in CRC tissues as compared with normal tissue, wherein its expression is closely associated with lymph node metastasis and the TNM stage. Inhibition of *SPP1* in the HCT116 cells consequently resulted in the suppression of cellular growth, migration, and invasion. *SPP1* overexpression could thus enhance the proliferation, migration, and invasion of CRC cells as well as the activation of the PI3K pathway, accompanied by the increased expression of Snail and matrix metalloproteinase 9 (MMP9) [26]. As one of the transcription factors of epithelial–mesenchymal transition (EMT), snail can inhibit the expression of E-cadherin, thereby promotes the activation of EMT in tumor cells [27]. MMP9 is one of the members in the MMP family which is a class of zinc-dependent proteases with the main function of participating in the degradation of extracellular matrix. MMP9 is often overexpressed in malignant tumors and is associated with aggressive malignant phenotypes and poor prognosis in cancer patients [28]. Similarly, *SPP1* could not only promote CRC cell growth and migration, but it also inhibited autophagy, which may be thus achieved by activating the p38 mitogen-activated protein kinase (MAPK) signaling pathway [29]. In addition, a study by Choe et al. [30] also confirmed higher *SPP1* expression along with shorter survival times in CRC patients.

TIMP1 is one of the members of the TIMP family which can bind to MMP substrates and mediate MMP degradation that is necessary for matrix renewal [31]. *TIMP1* has been shown to be highly expressed in many types of cancer, consequently correlated to their poor prognosis [32–36]. *TIMP1* may also promote liver metastasis in CRC [37]. After chemotherapy, high expression levels of *TIMP1* in the serum have been closely related to the decline over survival (OS) [38]. Moreover, knockdown of *TIMP1* leads to decreased proliferation of HT29 cells [39]. Huang et al. [40], Zeng et al. [41], and Zuo et al. [42] demonstrated that the expression levels of *TIMP1* may be negatively correlated with the prognosis of COAD patients. Song et al. [43] found that *TIMP1* knockdown could result in decreased proliferation, migration, and invasion as well as increased apoptosis *in vitro*, whereas there is weakened tumorigenesis and metastasis *in vivo* via the FAK-PI3K/Akt and MAPK pathways. The elevated *TIMP1* levels in serum were related to lower tumor grade and shorter overall survival times of CRC patients [44].

SERPINE1, also called PAI-1, can regulate plasmin formation from plasminogen [45]. The expression level of *SERPINE1* was seen to increase in CRC tissue and colon cancer cell lines showing active proliferation [46]. Moreover, overexpression of *SERPINE1* led to decreased apoptosis and caspase-3 amidolytic activity in vascular smooth muscle cells (VSMCs) compared with control [47]. LC-MS/MS analysis has shown that *SERPINE1* may be a peripheral participant in the apoptosis of CRC cells [48]. Shun et al. found that the expression level of *SERPINE1* was remarkably up-regulated in tumor stroma and associated with the relapse of CRC stage II–III patients [49]. Several recent studies have suggested that *SERPINE1* may be associated with prognosis of CRC patients [41,50,51].

CXCL2 belongs to the chemokine superfamily, and is responsible for the attraction of leukocytes to inflammatory sites apart from the regulation of angiogenesis, tumorigenesis [52]. Dietrich et al. found that there was significantly increased expression of *CXCL2* in CRC, in contrast with normal tissue [53]. Further, there was increased expression of mRNA and protein levels of *CXCL2* in colon cancer stem cells, whereas *CXCL2* knockdown led to the decreased expression of EMT markers, MMPs, and Gai-2, which consequently promoted tumor initiation and development [54]. Down-regulation of *CXCL2* destroyed the proliferation and colony formation in CRC cells [55]. Moreover, *CXCL2* remarkably promoted the proliferation and migration of human umbilical vein endothelial cells (HUVECs) [56]. But the qPCR results showed no significant difference in *CXCL2* expression between HCT116 and NCM460 cell lines. This may be due to the source of the cell lines; more cell lines should thus be selected for verification in future studies.

In summary, in the present study, we screened 323 DEGs from four COAD GEO datasets and identified *SERPINE1*, *SPP1*, and *TIMP1* as key genes which may be closely related to the prognosis of COAD in patients. Further investigations need to be performed to explore the mechanisms of these genes. These results would consequently provide candidate targets for the diagnosis and prognosis of CRC in patients, as well as new ideas for treatment.

Competing Interests

The authors declare that there are no competing interests associated with the manuscript.

Funding

This work was supported by the National Natural Science Foundation of China [grant number 81201757]; and the Basic Research Projects of Logistic University of Chinese People's Armed Police Force [grant numbers WHJ201702, BWS17J025, AWS16J004].

The visualization of GO analysis, including BP, MF and CC of DEGs, was performed via the online software ImageGP (<http://www.ehbio.com/ImageGP/index.php/Home/Index/index>).

Author Contribution

Yong Liu conceived the idea behind the present paper. Chao Li and Xuewei Chen collected and analyzed the data. Lijin Dong and Ping Li drafted the paper. Rong Fan revised the final paper.

Ethics Approval

Clinical colon cancer samples were obtained from the Tianjin Xiqing Hospital, with informed written consents from the patients or their guardians and the study protocol was approved by the Institutional Review Board of Tianjin Xiqing Hospital.

Abbreviations

BP, biological process; CC, cellular component; COAD, colon adenocarcinoma; CRC, colorectal cancer; CXCL2, C-X-C motif chemokine ligand 2; DAVID, Database for Annotation, Visualization and Integrated Discovery; DEG, differentially expressed gene; EMT, epithelial-mesenchymal transition; FAK, focal adhesion kinase-related nonkinase; Gai-2, $G\alpha i-2$; GAPDH, glyceraldehyde-3-phosphate dehydrogenase; GEO, Gene Expression Omnibus; GO, Gene Ontology; GSE, Gene Expression Omnibus Series; HUVEC, human umbilical vein endothelial cell; KEGG, Kyoto Encyclopedia of Genes and Genomes; MAPK, mitogen-activated protein kinase; MCODE, Molecular Complex Detection; MF, molecular function; MMP, matrix metalloproteinase; PAI-1, plasminogen activator inhibitor 1; PI3K, phosphoinositide-3-kinase; PPAR, peroxisome proliferator activated receptor; SERPINE1, Serpin Family E Member 1; SPP1, secreted phosphoprotein 1; STRING, Search Tool for the Retrieval of Interacting Genes/Proteins; TCGA, The Cancer Genome Atlas; TIMP1, tissue inhibitor of metalloproteinase 1; TNM, tumor node metastasis.

References

- Weitz, J., Koch, M., Debus, J., Hohler, T., Galle, P.R. and Buchler, M.W. (2005) Colorectal cancer. *Lancet* **365**, 153–165, [https://doi.org/10.1016/S0140-6736\(05\)17706-X](https://doi.org/10.1016/S0140-6736(05)17706-X)
- Walker, M.S., Pharm, E.Y., Kerr, J., Yim, Y.M., Stepanski, E.J. and Schwartzberg, L.S. (2012) Symptom burden & quality of life among patients receiving second-line treatment of metastatic colorectal cancer. *BMC Res. Notes* **5**, 314
- Torre, L.A., Bray, F., Siegel, R.L., Ferlay, J., Lortet-Tieulent, J. and Jemal, A. (2015) Global cancer statistics, 2012. *CA Cancer J. Clin.* **65**, 87–108, <https://doi.org/10.3322/caac.21262>
- Brenner, H., Kloor, M. and Pox, C.P. (2014) Colorectal cancer. *Lancet* **383**, 1490–1502, [https://doi.org/10.1016/S0140-6736\(13\)61649-9](https://doi.org/10.1016/S0140-6736(13)61649-9)
- Cao, H., Xu, E., Liu, H., Wan, L. and Lai, M. (2015) Epithelial-mesenchymal transition in colorectal cancer metastasis: a system review. *Pathol. Res. Pract.* **211**, 557–569, <https://doi.org/10.1016/j.prp.2015.05.010>
- Di Como, J.A., Mahendraraj, K., Lau, C.S.M. and Chamberlain, R.S. (2015) Adenosquamous carcinoma of the colon and rectum: a population based clinical outcomes study involving 578 patients from the Surveillance Epidemiology and End Result (SEER) database (1973–2010). *J. Am. Coll. Surg.* **221**, e56, <https://doi.org/10.1016/j.jamcollsurg.2015.08.044>
- Bray, F., Ferlay, J., Soerjomataram, I., Siegel, R.L., Torre, L.A. and Jemal, A. (2018) Global cancer statistics 2018: GLOBOCAN estimates of incidence and mortality worldwide for 36 cancers in 185 countries. *CA Cancer J. Clin.* **68**, 394–424, <https://doi.org/10.3322/caac.21492>
- Shih, W., Chetty, R. and Tsao, M.S. (2005) Expression profiling by microarrays in colorectal cancer (review). *Oncol. Rep.* **13**, 517–524
- Rifai, N., Gillette, M.A. and Carr, S.A. (2006) Protein biomarker discovery and validation: the long and uncertain path to clinical utility. *Nat. Biotechnol.* **24**, 971–983, <https://doi.org/10.1038/nbt1235>
- Segata, N., Izard, J., Waldron, L., Gevers, D., Miropolsky, L., Garrett, W.S. et al. (2011) Metagenomic biomarker discovery and explanation. *Genome Biol.* **12**, R60, <https://doi.org/10.1186/gb-2011-12-6-r60>
- Dennis, Jr, G., Sherman, B.T., Hosack, D.A., Yang, J., Gao, W., Lane, H.C. et al. (2003) DAVID: Database for Annotation, Visualization, and Integrated Discovery. *Genome Biol.* **4**, P3, <https://doi.org/10.1186/gb-2003-4-5-p3>
- Szklarczyk, D., Morris, J.H., Cook, H., Kuhn, M., Wyder, S., Simonovic, M. et al. (2017) The STRING database in 2017: quality-controlled protein-protein association networks, made broadly accessible. *Nucleic Acids Res.* **45**, D362–D368, <https://doi.org/10.1093/nar/gkw937>
- Kohl, M., Wiese, S. and Warscheid, B. (2011) Cytoscape: software for visualization and analysis of biological networks. *Methods Mol. Biol.* **696**, 291–303, https://doi.org/10.1007/978-1-60761-987-1_18
- Chandrashekar, D.S., Bashel, B., Balasubramanya, S.A.H., Creighton, C.J., Ponce-Rodriguez, I., Chakravarthi, B. et al. (2017) UALCAN: a portal for facilitating tumor subgroup gene expression and survival analyses. *Neoplasia* **19**, 649–658, <https://doi.org/10.1016/j.neo.2017.05.002>
- McGuire, S. (2016) World Cancer Report 2014. Geneva, Switzerland: World Health Organization, International Agency for Research on Cancer, WHO Press, 2015. *Adv. Nutr.* **7**, 418–419, <https://doi.org/10.3945/an.116.012211>
- Gu, M.J., Huang, Q.C., Bao, C.Z., Li, Y.J., Li, X.Q., Ye, D. et al. (2018) Attributable causes of colorectal cancer in China. *BMC Cancer* **18**, 38, <https://doi.org/10.1186/s12885-017-3968-z>
- Bogaert, J. and Prenen, H. (2014) Molecular genetics of colorectal cancer. *Ann. Gastroenterol.* **27**, 9–14
- Koveitpour, Z., Panahi, F., Vakilian, M., Peymani, M., Seyed Forootan, F., Nasr Esfahani, M.H. et al. (2019) Signaling pathways involved in colorectal cancer progression. *Cell Biosci.* **9**, 97

- 19 Frattini, M., Saletti, P., Molinari, F. and De Dosso, S. (2015) EGFR signaling in colorectal cancer: a clinical perspective. *Gastrointest. Cancer Targets Ther.* **21**, 21–38, <https://doi.org/10.2147/GICTT.S49002>
- 20 Hutchinson, R.A., Adams, R.A., McArt, D.G., Salto-Tellez, M., Jasani, B. and Hamilton, P.W. (2015) Epidermal growth factor receptor immunohistochemistry: new opportunities in metastatic colorectal cancer. *J. Transl. Med.* **13**, 217, <https://doi.org/10.1186/s12967-015-0531-z>
- 21 Rajendran, D.T., Subramaniam, B. and Ganeshan, M. (2017) Role of Notch signaling in colorectal cancer. *Role of Transcription Factors in Gastrointestinal Malignancies* 307–314, https://doi.org/10.1007/978-981-10-6728-0_21
- 22 Jiang, Z., Cao, Q., Dai, G., Wang, J., Liu, C., Lv, L. et al. (2019) Celestrol inhibits colorectal cancer through TGF-beta1/Smad signaling. *Onco Targets Ther.* **12**, 509–518, <https://doi.org/10.2147/OTT.S187817>
- 23 Xu, Y. and Pasche, B. (2007) TGF-beta signaling alterations and susceptibility to colorectal cancer. *Hum. Mol. Genet.* **16**, R14–R20, <https://doi.org/10.1093/hmg/ddl486>
- 24 Kaya Temiz, T., Altun, A., Turgut, N.H. and Balç, E. (2014) Investigation of the effects of drugs effective on PI3K-AKT signaling pathway in colorectal cancer alone and in combination. *Cumhuriyet Med. J.* **36**, 167–177, <https://doi.org/10.7197/cmj.v36i2.5000033144>
- 25 Cheng, Y., Wen, G., Sun, Y., Shen, Y., Zeng, Y., Du, M. et al. (2019) Osteopontin promotes colorectal cancer cell invasion and the stem cell-like properties through the PI3K-AKT-GSK3beta-beta/catenin pathway. *Med. Sci. Monit.* **25**, 3014–3025, <https://doi.org/10.12659/MSM.913185>
- 26 Wei, R., Wong, J.P.C., Lyu, P., Xi, X., Tong, O., Zhang, S.D. et al. (2018) In vitro and clinical data analysis of osteopontin as a prognostic indicator in colorectal cancer. *J. Cell. Mol. Med.* **22**, 4097–4105, <https://doi.org/10.1111/jcmm.13686>
- 27 Zhang, Y., Zou, X., Qian, W., Weng, X., Zhang, L., Zhang, L. et al. (2019) Enhanced PAPSS2/VCAN sulfation axis is essential for Snail-mediated breast cancer cell migration and metastasis. *Cell Death Differ.* **26**, 565–579, <https://doi.org/10.1038/s41418-018-0147-y>
- 28 Hidalgo, M. and Eckhardt, S.G. (2001) Development of matrix metalloproteinase inhibitors in cancer therapy. *J. Natl. Cancer Inst.* **93**, 178–193, <https://doi.org/10.1093/jnci/93.3.178>
- 29 Huang, R.H., Quan, Y.J., Chen, J.H., Wang, T.F., Xu, M., Ye, M. et al. (2017) Osteopontin promotes cell migration and invasion, and inhibits apoptosis and autophagy in colorectal cancer by activating the p38 MAPK signaling pathway. *Cell. Physiol. Biochem.* **41**, 1851–1864, <https://doi.org/10.1159/000471933>
- 30 Choe, E.K., Yi, J.W., Chai, Y.J. and Park, K.J. (2018) Upregulation of the adipokine genes ADIPOR1 and SPP1 is related to poor survival outcomes in colorectal cancer. *J. Surg. Oncol.* **117**, 1833–1840, <https://doi.org/10.1002/jso.25078>
- 31 Gardner, J. and Ghorpade, A. (2003) Tissue inhibitor of metalloproteinase (TIMP)-1: the TIMPed balance of matrix metalloproteinases in the central nervous system. *J. Neurosci. Res.* **74**, 801–806, <https://doi.org/10.1002/jnr.10835>
- 32 Grunnet, M., Mau-Sorensen, M. and Brunner, N. (2013) Tissue inhibitor of metalloproteinase 1 (TIMP-1) as a biomarker in gastric cancer: a review. *Scand. J. Gastroenterol.* **48**, 899–905, <https://doi.org/10.3109/00365521.2013.812235>
- 33 Roy, R., Zurakowski, D., Wischhusen, J., Frauenhoffer, C., Hooshmand, S., Kulke, M. et al. (2014) Urinary TIMP-1 and MMP-2 levels detect the presence of pancreatic malignancies. *Br. J. Cancer* **111**, 1772–1779, <https://doi.org/10.1038/bjc.2014.462>
- 34 Wang, Y.Y., Li, L., Zhao, Z.S. and Wang, H.J. (2013) Clinical utility of measuring expression levels of KAP1, TIMP1 and STC2 in peripheral blood of patients with gastric cancer. *World J. Surg. Oncol.* **11**, 81, <https://doi.org/10.1186/1477-7819-11-81>
- 35 Bjerre, C., Vinther, L., Belling, K.C., Wurtz, S.O., Yadav, R., Lademann, U. et al. (2013) TIMP1 overexpression mediates resistance of MCF-7 human breast cancer cells to fulvestrant and down-regulates progesterone receptor expression. *Tumour Biol.* **34**, 3839–3851, <https://doi.org/10.1007/s13277-013-0969-7>
- 36 Peng, L., Yanjiao, M., Ai-guo, W., Pengtao, G., Jianhua, L., Ju, Y. et al. (2011) A fine balance between CCNL1 and TIMP1 contributes to the development of breast cancer cells. *Biochem. Biophys. Res. Commun.* **409**, 344–349, <https://doi.org/10.1016/j.bbrc.2011.05.021>
- 37 Seubert, B., Grunwald, B., Kobuch, J., Cui, H., Schelter, F., Schaten, S. et al. (2015) Tissue inhibitor of metalloproteinases (TIMP)-1 creates a premetastatic niche in the liver through SDF-1/CXCR4-dependent neutrophil recruitment in mice. *Hepatology* **61**, 238–248, <https://doi.org/10.1002/hep.27378>
- 38 Mahner, S., Woelber, L., Eulenburg, C., Schwarz, J., Carney, W., Jaenicke, F. et al. (2010) TIMP-1 and VEGF-165 serum concentration during first-line therapy of ovarian cancer patients. *BMC Cancer* **10**, 139, <https://doi.org/10.1186/1471-2407-10-139>
- 39 Yang, L., Jiang, Q., Li, D.Z., Zhou, X., Yu, D.S. and Zhong, J. (2019) TIMP1 mRNA in tumor-educated platelets is diagnostic biomarker for colorectal cancer. *Aging* **11**, 8998–9012, <https://doi.org/10.18632/aging.102366>
- 40 Huang, R., Wang, K., Gao, L. and Gao, W. (2019) TIMP1 is a potential key gene associated with the pathogenesis and prognosis of ulcerative colitis-associated colorectal cancer. *Onco Targets Ther.* **12**, 8895–8904, <https://doi.org/10.2147/OTT.S222608>
- 41 Zeng, C. and Chen, Y. (2019) HTR1D, TIMP1, SERPINE1, MMP3 and CNR2 affect the survival of patients with colon adenocarcinoma. *Oncol. Lett.* **18**, 2448–2454
- 42 Zuo, S., Dai, G. and Ren, X. (2019) Identification of a 6-gene signature predicting prognosis for colorectal cancer. *Cancer Cell Int.* **19**, 6, <https://doi.org/10.1186/s12935-018-0724-7>
- 43 Song, G., Xu, S., Zhang, H., Wang, Y., Xiao, C., Jiang, T. et al. (2016) TIMP1 is a prognostic marker for the progression and metastasis of colon cancer through FAK-PI3K/AKT and MAPK pathway. *J. Exp. Clin. Cancer Res.* **35**, 148, <https://doi.org/10.1186/s13046-016-0427-7>
- 44 Giaginis, C., Nikiteas, N., Margeli, A., Tzanakis, N., Rallis, G., Kouraklis, G. et al. (2009) Serum tissue inhibitor of metalloproteinase 1 and 2 (TIMP-1 and TIMP-2) levels in colorectal cancer patients: associations with clinicopathological variables and patient survival. *Int. J. Biol. Markers* **24**, 245–252, <https://doi.org/10.1177/172460080902400405>
- 45 Troy, A.M., Sheahan, K., Mulcahy, H.E., Duffy, M.J., Hyland, J.M. and O'Donoghue, D.P. (2004) Expression of Cathepsin B and L antigen and activity is associated with early colorectal cancer progression. *Eur. J. Cancer* **40**, 1610–1616, <https://doi.org/10.1016/j.ejca.2004.03.011>
- 46 Mazzoccoli, G., Paziienza, V., Panza, A., Valvano, M.R., Benegiamo, G., Vinciguerra, M. et al. (2012) ARNTL2 and SERPINE1: potential biomarkers for tumor aggressiveness in colorectal cancer. *J. Cancer Res. Clin. Oncol.* **138**, 501–511, <https://doi.org/10.1007/s00432-011-1126-6>

- 47 Chen, Y., Kelm, Jr, R.J., Budd, R.C., Sobel, B.E. and Schneider, D.J. (2004) Inhibition of apoptosis and caspase-3 in vascular smooth muscle cells by plasminogen activator inhibitor type-1. *J. Cell. Biochem.* **92**, 178–188, <https://doi.org/10.1002/jcb.20058>
- 48 Kang, K., Song, D.G., Lee, E.H., Lee, K.M., Park, Y.G., Jung, S.H. et al. (2014) Secretome profiling reveals the signaling molecules of apoptotic HCT116 cells induced by the dietary polyacetylene gymnasterkoreayne B. *J. Agric. Food Chem.* **62**, 2353–2363, <https://doi.org/10.1021/jf404047z>
- 49 Chida, S., Okayama, H., Noda, M., Saito, K., Nakajima, T., Aoto, K. et al. (2016) Stromal VCAN expression as a potential prognostic biomarker for disease recurrence in stage II-III colon cancer. *Carcinogenesis* **37**, 878–887, <https://doi.org/10.1093/carcin/bgw069>
- 50 Wang, T., Xu, H., Liu, X., Chen, S., Zhou, Y. and Zhang, X. (2017) Identification of key genes in colorectal cancer regulated by miR-34a. *Med. Sci. Monit.* **23**, 5735–5743, <https://doi.org/10.12659/MSM.904937>
- 51 Liang, Y., Zhang, C., Ma, M.H. and Dai, D.Q. (2018) Identification and prediction of novel non-coding and coding RNA-associated competing endogenous RNA networks in colorectal cancer. *World J. Gastroenterol.* **24**, 5259–5270, <https://doi.org/10.3748/wjg.v24.i46.5259>
- 52 Verbeke, H., Struyf, S., Laureys, G. and Van Damme, J. (2011) The expression and role of CXC chemokines in colorectal cancer. *Cytokine Growth Factor Rev.* **22**, 345–358, <https://doi.org/10.1016/j.cytogfr.2011.09.002>
- 53 Doll, D., Keller, L., Maak, M., Boulesteix, A.L., Siewert, J.R., Holzmann, B. et al. (2010) Differential expression of the chemokines GRO-2, GRO-3, and interleukin-8 in colon cancer and their impact on metastatic disease and survival. *Int. J. Colorectal Dis.* **25**, 573–581, <https://doi.org/10.1007/s00384-010-0901-1>
- 54 Chen, M.C., Baskaran, R., Lee, N.H., Hsu, H.H., Ho, T.J., Tu, C.C. et al. (2019) CXCL2/CXCR2 axis induces cancer stem cell characteristics in CPT-11-resistant LoVo colon cancer cells via Galphai-2 and Galphaq/11. *J. Cell. Physiol.* **234**, 11822–11834, <https://doi.org/10.1002/jcp.27891>
- 55 Zhang, Z., Tan, X., Luo, J., Cui, B., Lei, S., Si, Z. et al. (2018) GNA13 promotes tumor growth and angiogenesis by upregulating CXC chemokines via the NF-kappaB signaling pathway in colorectal cancer cells. *Cancer Med.* **7**, 5611–5620, <https://doi.org/10.1002/cam4.1783>
- 56 Ma, J.C., Sun, X.W., Su, H., Chen, Q., Guo, T.K., Li, Y. et al. (2017) Fibroblast-derived CXCL12/SDF-1alpha promotes CXCL6 secretion and co-operatively enhances metastatic potential through the PI3K/Akt/mTOR pathway in colon cancer. *World J. Gastroenterol.* **23**, 5167–5178, <https://doi.org/10.3748/wjg.v23.i28.5167>