

# **Identification and verification of 3 key genes associated with survival and prognosis of patients with colon adenocarcinoma via integrated bioinformatics analysis**

Yong Liu<sup>1\*</sup>, Chao Li<sup>3\*</sup>, Lijin Dong<sup>4</sup>, Ping Li<sup>5</sup>, Xuewei Chen<sup>3#</sup>, Rong Fan<sup>2,3#</sup>

1. Department of General Surgery, Tianjin XiQing hospital, Tianjin 300380, PR China

2. Central laboratory, Tianjin XiQing hospital, Tianjin 300380, PR China

3. Department of Operational Medicine, Tianjin Institute of Environmental and Operational Medicine, Tianjin, 300050, PR China

4. Editorial Department of Education and Research Security Centre, Logistic University of Chinese People's Armed Police Force, Tianjin 300309, PR China

5. Southwest Medical University, Luzhou 646000, Sichuan, PR China

\*These authors contributed equally to this work.

#Correspondence:

Xuewei Chen

Department of Operational Medicine, Tianjin Institute of Environmental and Operational Medicine, Tianjin, 300050, PR China

Tel: +86-22-84655332

Fax: +86-22-84655332

E-mail address: chenxuewei11@sina.com

Rong Fan

Central Lab, Tianjin XiQing hospital, Tianjin 300380, PR China

Tel: +86-22-27391697;

Fax: +86-22-27391697;

E-mail address: [prosperity\\_39@sina.cn](mailto:prosperity_39@sina.cn)

## **Abstract**

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 STRING and Cytoscape, following Gene Ontology and KEGG enrichment analysis. Finally, survival analysis and expression validation were performed via UALCAN online software, real-time PCR and immunohistochemistry. Results: In this study, we screened 323 common differentially expressed genes 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 SERPINE1, SPP1 and TIMP1 may be biomarkers closely related to the prognosis of CRC patients.

Key words: colorectal cancer; bioinformatics analysis; biomarkers; colon adenocarcinoma

## **1. 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 700,000 deaths occur worldwide due to CRC [3]. The five year survival rate of patients with CRC at its early stage is 90 percent, while those at later stages show a rate of no more than 12 percent [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 systematical screening of tumor-associated genes [9, 10].

In this study, we screened the common differentially expressed genes (DEGs) in 4 GSE datasets from the Gene Expression Omnibus (GEO) repository and constructed a protein-protein interaction (PPI) network. Via integrated bioinformatics analysis, we identified the 4 hub genes associated with survival of COAD patients and verified the mRNA expression levels in CRC cells. Finally, our results showed that SPP1, SERPINE1 and TIMP1 could be potential biomarkers for prognostics of COAD in patients.

## **2. Materials and Methods**

### **2.1 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.

## 2.2 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  $|\log FC| \geq 1$  to be statistically significant. Then, the DEGs in the four datasets were screened out using online Venn software. The DEGs with  $\log FC \geq 1$  were considered as upregulated genes, while DEGs with  $\log FC \leq -1$  were considered as downregulated genes.

## 2.3 Enrichment analysis via GO and KEGG pathway

In order to characterize the functional roles of the DEGs, we used DAVID (version 6.8 [11]) database for gene ontology and signal pathway enrichment analysis, which comprised biological process (BP), molecular function (MF) and cellular component (CC), and KEGG pathways analysis.

## 2.4 Construction of PPI network and analysis of module

The PPI network was built via the 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 MCODE app (node score cutoff=0.2, degree cutoff=2, max. Depth=100, and k-score=2).

## 2.5 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 the TCGA. *P*-value <0.05 was considered statistically significant.

## 2.6 Cell lines and culture

The normal colon epithelial cell line, NCM460, and colon adenocarcinoma cell

lines, HCT116, HCA-7, HT-29, and SW480 were maintained by our laboratory. And all cells were cultured in a 5% CO<sub>2</sub> humidified incubator in Dulbecco's Modified Eagle's medium (Hyclone) containing 10% fetal bovine serum (FBS, Gibco) at 37°C.

## 2.7 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, USA) on Applied Biosystems™ StepOnePlus instruments, and analyzed with the StepOne software, using GAPDH as a control. The primers used in this study are designed as follows:

TIMP1: Forward: 5'- GGGACACCAGAAGTCAACCAG-3'; Reverse: 5'-

CAATGAGAACTCCTCGCTGC-3'; CXCL2: Forward: 5'-

CCAAACCGAAGTCATAGCCAC-3'; Reverse: 5'-

TGCCATTTTTCAGCATCTTTTC-3'; SPP1: Forward: 5'-

TGAGCATTCGATGTGATTGAT-3'; Reverse: 5'-

TTATCTTCTTCCTTACTTTTGGGGT-3'; SERPINE1: Forward: 5'-

CTGGGTGAAGACACACAAAAG-3'; Reverse: 5'-

CACAGAGACAGTGCTGCCGT-3'; GAPDH: Forward: 5'-

TCAAGGGCATCCTGGGCTACAC-3'; Reverse: 5'-

TCAAAGGTGGAGGAGTGGGTGTC-3'.

## 2.8 Tissue samples and immunohistochemistry (IHC)

Human colon adenocarcinoma tissue microarray sections (HColA160CS01) were obtained from Shanghai Outdo Biotech Co. Ltd. (Shanghai, China). The tissue samples were procured from 80 patients with colorectal cancer. Every specimen was provided with cancer tissue and adjacent-carcinoma tissue. 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 study was approved by the Ethics Committee of Shanghai Outdo

Biotech Company (Shanghai, China). The slides were analyzed using the Image-Pro PLUS software program (Media Cybernetics, Inc. USA).

## **2.9 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 a Student's t-test, wherein \*, \*\*, and \*\*\* represented  $P < 0.05$ ,  $P < 0.01$  and  $P < 0.001$ , respectively.

## **3. Results**

### **3.1 Identification of DEGs in COAD**

In this study, we selected and download 4 GEO datasets and extracted DEGs on the basis of the cut-off criteria. This 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 upregulated 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 & Figure 1).

### **3.2 GO and signaling pathway enrichment analysis**

We analyzed the DEGs via the DAVID software and identified 125 significant enrichments 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 2 A-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, PI3K-Akt signaling pathway, PPAR signaling pathway, etc. (Figure 2D).

### 3.3 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 downregulated and 171 upregulated genes. Then we conducted further analyses by applying Cytoscape MCODE plus, and found 32 central nodes that included 20 upregulated genes and 12 downregulated genes (Figure 3, 4 & Table 2).

### 3.4 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 4 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 4 hub genes in COAD and normal tissues via UALCAN. Results showed that expression levels of the 4 genes were significantly increased in the COAD samples as compared to 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 to 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 to 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 colorectal cancer tissues than in the adjacent tissues, while the expression of CXCL2 was not statistically different between cancer tissues and adjacent tissues (Figure 8).

#### 4. Discussion

CRC is the third main cause of death 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].

In this 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 biological processes, 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, 4 hub genes, *CXCL2*, *SERPINE1*, *SPP1*, and *TIMP1*, were selected. Furthermore, we tested the expression of these 4 genes at the mRNA as



well as protein levels and found that SPP1, TIMP1, and SERPINE1 had higher expression levels in colon adenocarcinoma cell lines than the colon epithelial cell line as well as in colorectal cancer tissue compared to 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 to 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 MMP9 [26]. As one of the transcription factors of EMT, snail can inhibit the expression of E-cadherin, thereby promotes the activation of EMT in tumor cells [27]. Matrix metalloproteinase 9 (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 p38MAPK 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 [40], Zeng [41],

and Zuo [42] et al. 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 (VSMC) compared to 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 SEPINE1 was remarkably upregulated 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 to 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 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 this study, we screened 323 DEGs from four COAD GEO datasets and identified *SERPINE1*, *SPPI*, and *TIMPI* 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.

### **Author's contributions**

Yong Liu conceived the idea behind this 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.

### **Acknowledgements**

This research was funded by the National Natural Science Foundation of China, No. 81201757 and 81973237; as well as the Basic research projects of Logistic University of Chinese People's Armed Police Force, No.WHJ201702; and 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>).

### **Conflict of interests**

The authors declare that they have no conflict of interests.

### **Ethics approval and consent to participate**

Not applicable.

### **Abbreviation**

CRC: colorectal cancer; GEO: Gene Expression Omnibus; DEGs: differentially

expressed genes; DAVID: Database for Annotation, Visualization and Integrated Discovery; GO: Gene ontology; KEGG: Kyoto Encyclopedia of Gene and Genome; BP: biological process; CC: cellular component; MF: molecular function; STRING: Search Tool for the Retrieval of Interacting Genes/Proteins; MCODE: Molecular Complex Detection; TCGA: The Cancer Genome Atlas; CXCL2: C-X-C Motif Chemokine Ligand 2; SPP1: Secreted Phosphoprotein 1; SERPINE1: Serpin Family E Member 1; TIMP1: Tissue Inhibitor Of Metalloproteinases 1 ; MAPK: Mitogen-Activated Protein Kinase; PI3K: Phosphoinositide-3-Kinase; HUVEC: human umbilical vein endothelial cells; EMT: epithelial-mesenchymal transition; MMPs: matrix metalloproteinases; OS: over survival.

## Reference

- 1.Weitz J, Koch M, Debus J, Hohler T, Galle PR, Buchler MW. Colorectal cancer. *Lancet*. 2005;365(9454):153-65.
- 2.Walker MS, Pharm EY, Kerr J, Yim YM, Stepanski EJ, Schwartzberg LS. Symptom burden & quality of life among patients receiving second-line treatment of metastatic colorectal cancer. *BMC research notes*. 2012;5:314.
- 3.Torre LA, Bray F, Siegel RL, Ferlay J, Lortet-Tieulent J, Jemal A. Global cancer statistics, 2012. *CA: a cancer journal for clinicians*. 2015;65(2):87-108.
- 4.Brenner H, Kloor M, Pox CP. Colorectal cancer. *Lancet*. 2014;383(9927):1490-502.
- 5.Cao H, Xu E, Liu H, Wan L, Lai M. Epithelial-mesenchymal transition in colorectal cancer metastasis: A system review. *Pathology, research and practice*. 2015;211(8):557-69.
- 6.Di Como JA, Mahendraraj K, Lau CSM, Chamberlain RS. 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). *Journal of the American College of Surgeons*. 2015;221(4):e56.
- 7.Bray F, Ferlay J, Soerjomataram I, Siegel RL, Torre LA, Jemal A. Global cancer statistics 2018: GLOBOCAN estimates of incidence and mortality worldwide for 36 cancers in 185 countries. *CA: a cancer journal for clinicians*. 2018;68(6):394-424.
- 8.Shih W, Chetty R, Tsao MS. Expression profiling by microarrays in colorectal cancer (Review). *Oncology reports*. 2005;13(3):517-24.
- 9.Rifai N, Gillette MA, Carr SA. Protein biomarker discovery and validation: the long and uncertain path to clinical utility. *Nature biotechnology*. 2006;24(8):971-83.
- 10.Segata N, Izard J, Waldron L, Gevers D, Miropolsky L, Garrett WS, et al. Metagenomic biomarker discovery and explanation. *Genome biology*. 2011;12(6):R60.
- 11.Dennis G, Jr., Sherman BT, Hosack DA, Yang J, Gao W, Lane HC, et al. DAVID: Database for Annotation, Visualization, and Integrated Discovery. *Genome biology*. 2003;4(5):P3.
- 12.Szklarczyk D, Morris JH, Cook H, Kuhn M, Wyder S, Simonovic M, et al. The STRING database in

2017: quality-controlled protein-protein association networks, made broadly accessible. *Nucleic acids research*. 2017;45(D1):D362-D8.

13.Kohl M, Wiese S, Warscheid B. Cytoscape: software for visualization and analysis of biological networks. *Methods in molecular biology*. 2011;696:291-303.

14.Chandrashekar DS, Bashel B, Balasubramanya SAH, Creighton CJ, Ponce-Rodriguez I, Chakravarthi B, et al. UALCAN: A Portal for Facilitating Tumor Subgroup Gene Expression and Survival Analyses. *Neoplasia* (New York, NY). 2017;19(8):649-58.

15.McGuire S. World Cancer Report 2014. Geneva, Switzerland: World Health Organization, International Agency for Research on Cancer, WHO Press, 2015. *Advances in nutrition*. 2016;7(2):418-9.

16.Gu MJ, Huang QC, Bao CZ, Li YJ, Li XQ, Ye D, et al. Attributable causes of colorectal cancer in China. *BMC cancer*. 2018;18(1):38.

17.Bogaert J, Prenen H. Molecular genetics of colorectal cancer. *Annals of gastroenterology*. 2014;27(1):9-14.

18.Koveitypour Z, Panahi F, Vakilian M, Peymani M, Seyed Forootan F, Nasr Esfahani MH, et al. Signaling pathways involved in colorectal cancer progression. *Cell & bioscience*. 2019;9:97.

19.Frattini M, Saletti P, Molinari F, De Dosso S. EGFR signaling in colorectal cancer: a clinical perspective. *Gastrointestinal Cancer: Targets and Therapy*. 2015:21.

20.Hutchinson RA, Adams RA, McArt DG, Salto-Tellez M, Jasani B, Hamilton PW. Epidermal growth factor receptor immunohistochemistry: new opportunities in metastatic colorectal cancer. *Journal of translational medicine*. 2015;13:217.

21.Rajendran DT, Subramaniyan B, Ganeshan M. Role of Notch Signaling in Colorectal Cancer. 2017:307-14.

22.Jiang Z, Cao Q, Dai G, Wang J, Liu C, Lv L, et al. Celestrol inhibits colorectal cancer through TGF-beta1/Smad signaling. *OncoTargets and therapy*. 2019;12:509-18.

23.Xu Y, Pasche B. TGF-beta signaling alterations and susceptibility to colorectal cancer. *Human molecular genetics*. 2007;16 Spec No 1:R14-20.

24.Kaya Temiz T, Altun A, Turgut NH, Balci E. Investigation of the effects of drugs effective on PI3K-AKT signaling pathway in colorectal cancer alone and in combination. *Cumhuriyet Medical Journal*. 2014;36(2):167-77.

25.Cheng Y, Wen G, Sun Y, Shen Y, Zeng Y, Du M, et al. Osteopontin Promotes Colorectal Cancer Cell Invasion and the Stem Cell-Like Properties through the PI3K-AKT-GSK/3beta-beta/Catenin Pathway. *Medical science monitor : international medical journal of experimental and clinical research*. 2019;25:3014-25.

26.Wei R, Wong JPC, Lyu P, Xi X, Tong O, Zhang SD, et al. In vitro and clinical data analysis of Osteopontin as a prognostic indicator in colorectal cancer. *Journal of cellular and molecular medicine*. 2018;22(9):4097-105.

27.Zhang Y, Zou X, Qian W, Weng X, Zhang L, Zhang L, et al. Enhanced PAPSS2/VCAN sulfation axis is essential for Snail-mediated breast cancer cell migration and metastasis. *Cell death and differentiation*. 2019;26(3):565-79.

28.Hidalgo M, Eckhardt SG. Development of matrix metalloproteinase inhibitors in cancer therapy. *Journal of the National Cancer Institute*. 2001;93(3):178-93.

29.Huang RH, Quan YJ, Chen JH, Wang TF, Xu M, Ye M, et al. Osteopontin Promotes Cell Migration and Invasion, and Inhibits Apoptosis and Autophagy in Colorectal Cancer by activating the p38 MAPK

Signaling Pathway. Cellular physiology and biochemistry : international journal of experimental cellular physiology, biochemistry, and pharmacology. 2017;41(5):1851-64.

30.Choe EK, Yi JW, Chai YJ, Park KJ. Upregulation of the adipokine genes ADIPOR1 and SPP1 is related to poor survival outcomes in colorectal cancer. Journal of surgical oncology. 2018;117(8):1833-40.

31.Gardner J, Ghorpade A. Tissue inhibitor of metalloproteinase (TIMP)-1: the TIMPed balance of matrix metalloproteinases in the central nervous system. Journal of neuroscience research. 2003;74(6):801-6.

32.Grunnet M, Mau-Sorensen M, Brunner N. Tissue inhibitor of metalloproteinase 1 (TIMP-1) as a biomarker in gastric cancer: a review. Scandinavian journal of gastroenterology. 2013;48(8):899-905.

33.Roy R, Zurakowski D, Wischhusen J, Frauenhoffer C, Hooshmand S, Kulke M, et al. Urinary TIMP-1 and MMP-2 levels detect the presence of pancreatic malignancies. British journal of cancer. 2014;111(9):1772-9.

34.Wang YY, Li L, Zhao ZS, Wang HJ. Clinical utility of measuring expression levels of KAP1, TIMP1 and STC2 in peripheral blood of patients with gastric cancer. World journal of surgical oncology. 2013;11:81.

35.Bjerre C, Vinther L, Belling KC, Wurtz SO, Yadav R, Lademann U, et al. TIMP1 overexpression mediates resistance of MCF-7 human breast cancer cells to fulvestrant and down-regulates progesterone receptor expression. Tumour biology : the journal of the International Society for Oncodevelopmental Biology and Medicine. 2013;34(6):3839-51.

36.Peng L, Yanjiao M, Ai-guo W, Pengtao G, Jianhua L, Ju Y, et al. A fine balance between CCNL1 and TIMP1 contributes to the development of breast cancer cells. Biochemical and biophysical research communications. 2011;409(2):344-9.

37.Seubert B, Grunwald B, Kobuch J, Cui H, Schelter F, Schaten S, et al. Tissue inhibitor of metalloproteinases (TIMP)-1 creates a premetastatic niche in the liver through SDF-1/CXCR4-dependent neutrophil recruitment in mice. Hepatology (Baltimore, Md). 2015;61(1):238-48.

38.Mahner S, Woelber L, Eulenburg C, Schwarz J, Carney W, Jaenicke F, et al. TIMP-1 and VEGF-165 serum concentration during first-line therapy of ovarian cancer patients. BMC cancer. 2010;10:139.

39.Yang L, Jiang Q, Li DZ, Zhou X, Yu DS, Zhong J. TIMP1 mRNA in tumor-educated platelets is diagnostic biomarker for colorectal cancer. Aging. 2019;11(20):8998-9012.

40.Huang R, Wang K, Gao L, Gao W. TIMP1 Is A Potential Key Gene Associated With The Pathogenesis And Prognosis Of Ulcerative Colitis-Associated Colorectal Cancer. OncoTargets and therapy. 2019;12:8895-904.

41.Zeng C, Chen Y. HTR1D, TIMP1, SERPINE1, MMP3 and CNR2 affect the survival of patients with colon adenocarcinoma. Oncology letters. 2019;18(3):2448-54.

42.Zuo S, Dai G, Ren X. Identification of a 6-gene signature predicting prognosis for colorectal cancer. Cancer cell international. 2019;19:6.

43.Song G, Xu S, Zhang H, Wang Y, Xiao C, Jiang T, et al. TIMP1 is a prognostic marker for the progression and metastasis of colon cancer through FAK-PI3K/AKT and MAPK pathway. Journal of experimental & clinical cancer research : CR. 2016;35(1):148.

44.Giaginis C, Nikiteas N, Margeli A, Tzanakis N, Rallis G, Kouraklis G, et al. 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. The International journal of biological markers. 2009;24(4):245-52.

45.Troy AM, Sheahan K, Mulcahy HE, Duffy MJ, Hyland JM, O'Donoghue DP. Expression of Cathepsin B

and L antigen and activity is associated with early colorectal cancer progression. *European journal of cancer*. 2004;40(10):1610-6.

46.Mazzocchi G, Paziienza V, Panza A, Valvano MR, Benegiamo G, Vinciguerra M, et al. ARNTL2 and SERPINE1: potential biomarkers for tumor aggressiveness in colorectal cancer. *Journal of cancer research and clinical oncology*. 2012;138(3):501-11.

47.Chen Y, Kelm RJ, Jr., Budd RC, Sobel BE, Schneider DJ. Inhibition of apoptosis and caspase-3 in vascular smooth muscle cells by plasminogen activator inhibitor type-1. *Journal of cellular biochemistry*. 2004;92(1):178-88.

48.Kang K, Song DG, Lee EH, Lee KM, Park YG, Jung SH, et al. Secretome profiling reveals the signaling molecules of apoptotic HCT116 cells induced by the dietary polyacetylene gymnasterkoreayne B. *Journal of agricultural and food chemistry*. 2014;62(11):2353-63.

49.Chida S, Okayama H, Noda M, Saito K, Nakajima T, Aoto K, et al. Stromal VCAN expression as a potential prognostic biomarker for disease recurrence in stage II-III colon cancer. *Carcinogenesis*. 2016;37(9):878-87.

50.Wang T, Xu H, Liu X, Chen S, Zhou Y, Zhang X. Identification of Key Genes in Colorectal Cancer Regulated by miR-34a. *Medical science monitor : international medical journal of experimental and clinical research*. 2017;23:5735-43.

51.Liang Y, Zhang C, Ma MH, Dai DQ. Identification and prediction of novel non-coding and coding RNA-associated competing endogenous RNA networks in colorectal cancer. *World journal of gastroenterology*. 2018;24(46):5259-70.

52.Verbeke H, Struyf S, Laureys G, Van Damme J. The expression and role of CXC chemokines in colorectal cancer. *Cytokine & growth factor reviews*. 2011;22(5-6):345-58.

53.Doll D, Keller L, Maak M, Boulesteix AL, Siewert JR, Holzmann B, et al. Differential expression of the chemokines GRO-2, GRO-3, and interleukin-8 in colon cancer and their impact on metastatic disease and survival. *International journal of colorectal disease*. 2010;25(5):573-81.

54.Chen MC, Baskaran R, Lee NH, Hsu HH, Ho TJ, Tu CC, et al. CXCL2/CXCR2 axis induces cancer stem cell characteristics in CPT-11-resistant LoVo colon cancer cells via Galphai-2 and Galphaq/11. *Journal of cellular physiology*. 2019;234(7):11822-34.

55.Zhang Z, Tan X, Luo J, Cui B, Lei S, Si Z, et al. GNA13 promotes tumor growth and angiogenesis by upregulating CXC chemokines via the NF-kappaB signaling pathway in colorectal cancer cells. *Cancer medicine*. 2018;7(11):5611-20.

56.Ma JC, Sun XW, Su H, Chen Q, Guo TK, Li Y, et al. Fibroblast-derived CXCL12/SDF-1alpha promotes CXCL6 secretion and co-operatively enhances metastatic potential through the PI3K/Akt/mTOR pathway in colon cancer. *World journal of gastroenterology*. 2017;23(28):5167-78.

## Figure legends

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 downregulated ( $\log_{2}FC \leq -1$ ); On the right, 198 DEGs are upregulated ( $\log_{2}FC \geq 1$ ).



Figure 2. The top 20 gene ontology (GO) and significantly enriched KEGG pathways. A. Biological process (BP); B. Cellular components (CC); C. Molecular functions (MF); D. KEGG pathways.

Figure 3. PPI network constructed via STRING and Cytoscape. Each node indicates a protein module; the edges represent proteins interaction; blue circles denote upregulated DEGs while red circles represent downregulated DEGs.

Figure 4. Module analysis by MCODE in the Cytoscape software. A. module1; B. module2. (Blue circles signify upregulated DEGs and red circles denoted downregulated DEGs)

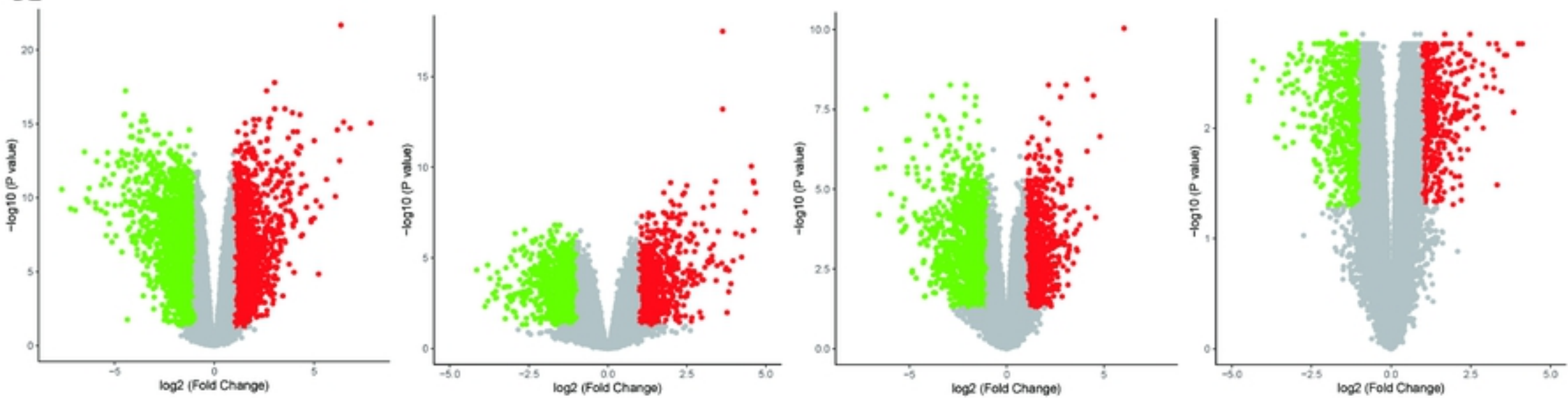
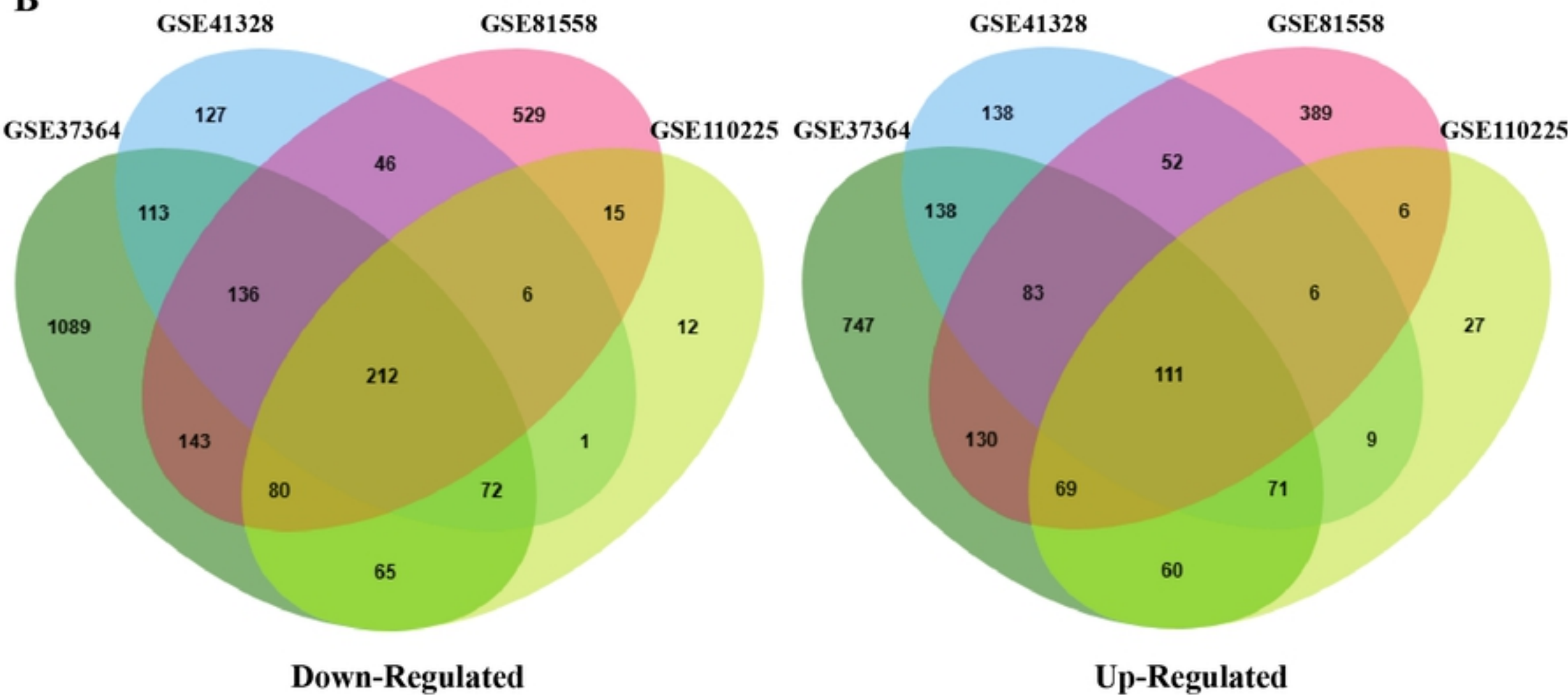
Figure 5. Correlation between the 4 hub genes expression and the overall survival of COAD patients, via UALCAN analysis.

Figure 6. Analysis of the 4 hub genes expression levels via UALCAN analysis.

Figure 7. Relative mRNA levels of the 4 hub genes in HCT116, HCA-7, HT-29, and SW480 cells compared to NCM460 cells, via real-time PCR. (\*\* represents  $P < 0.01$ , \*\*\* represents  $P < 0.001$ , ns represents no significance)

Figure 8. Expression of CXCL2 (A), SERPINE1 (B), TIMP1 (C), and SPP1 (D) in colorectal adenocarcinoma tissue and adjacent tissue in tissue microarrays. (Scale bar: 100 $\mu$ m)



**A****GSE37364****GSE41328****GSE81558****GSE110225****B****Figure1**

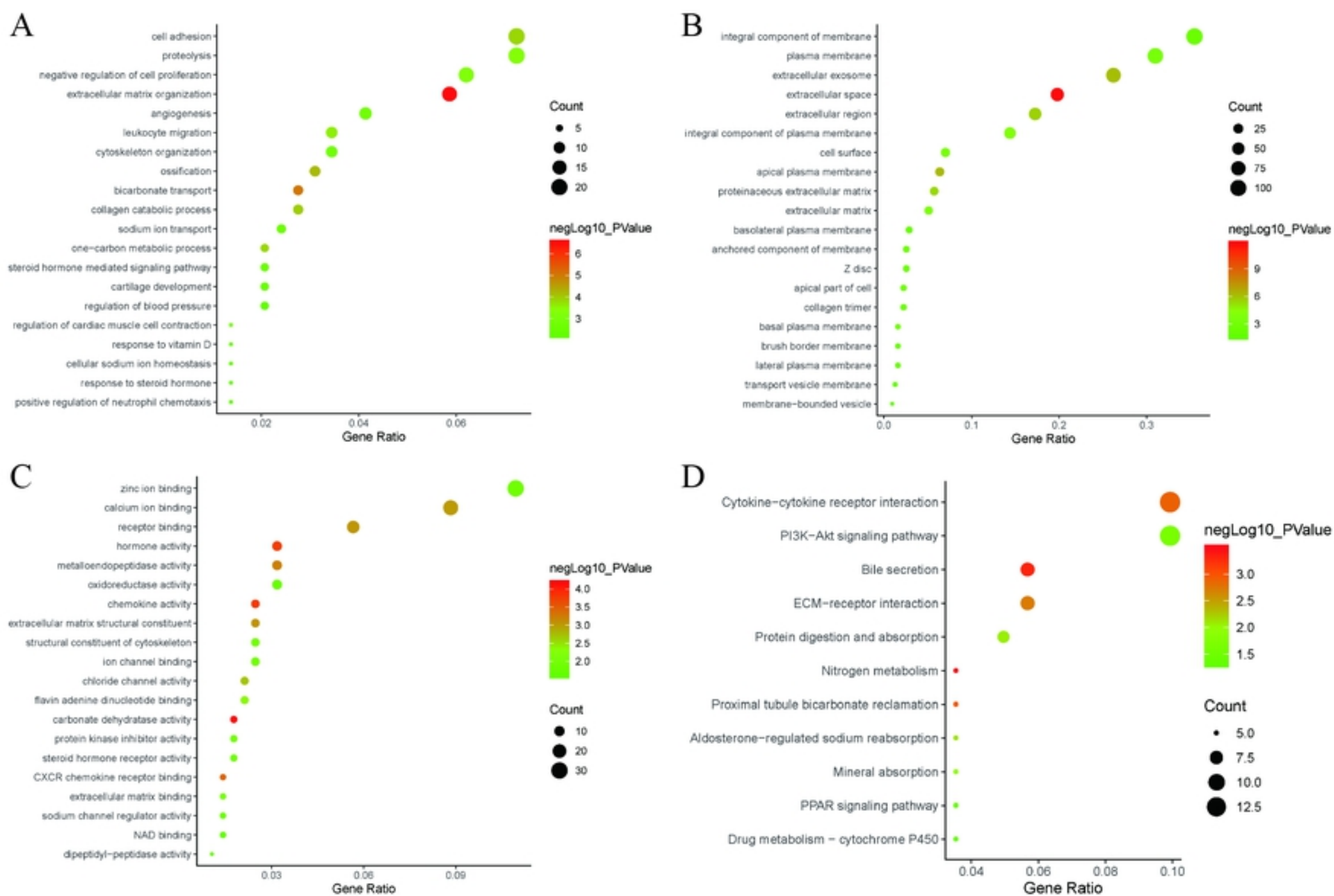
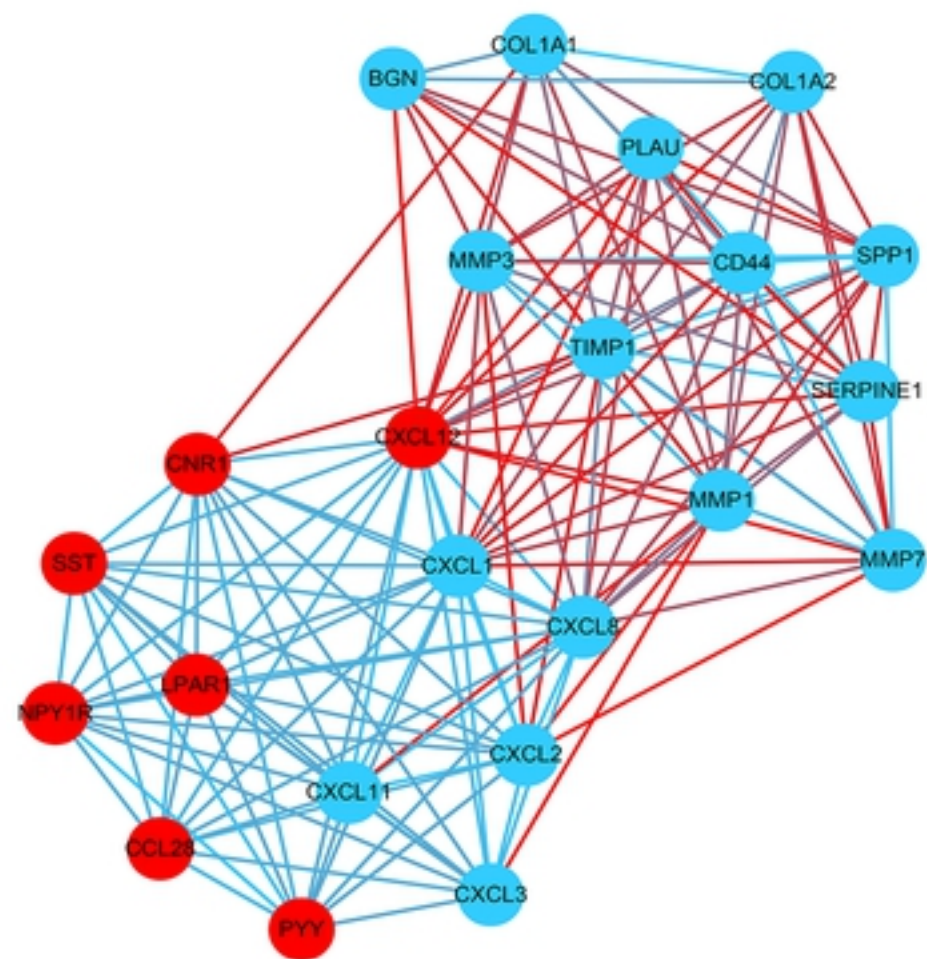


Figure2





A



B

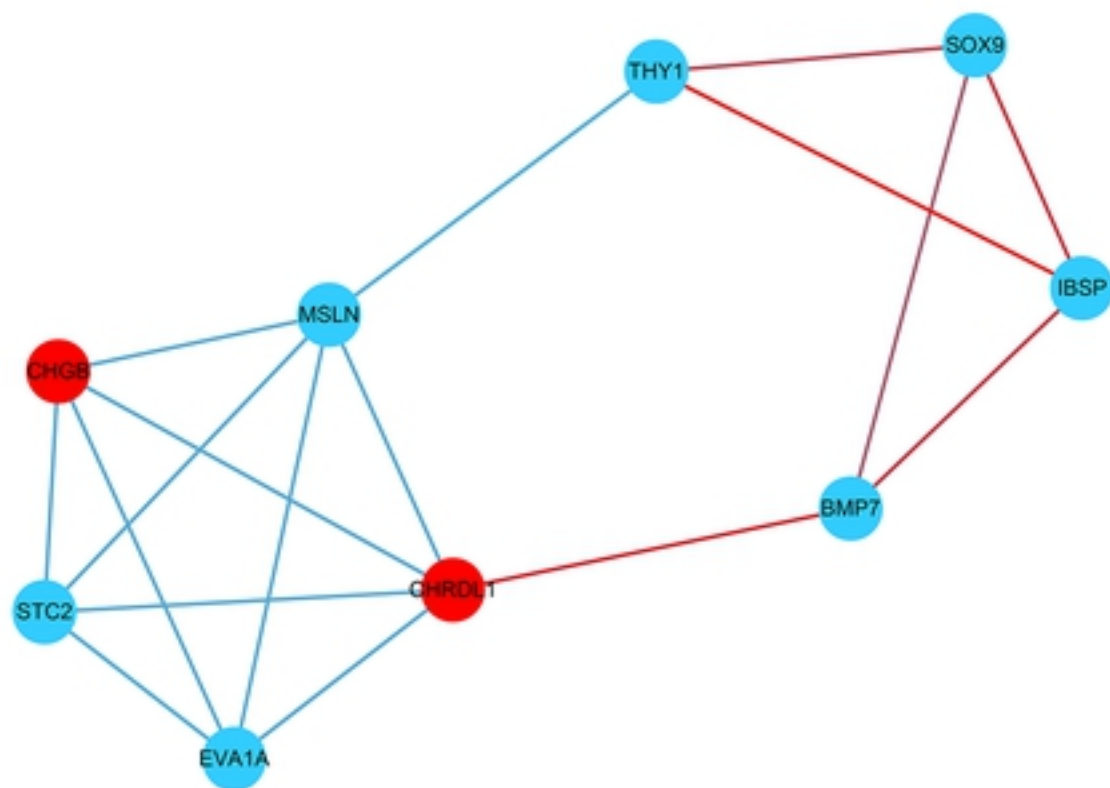


Figure4



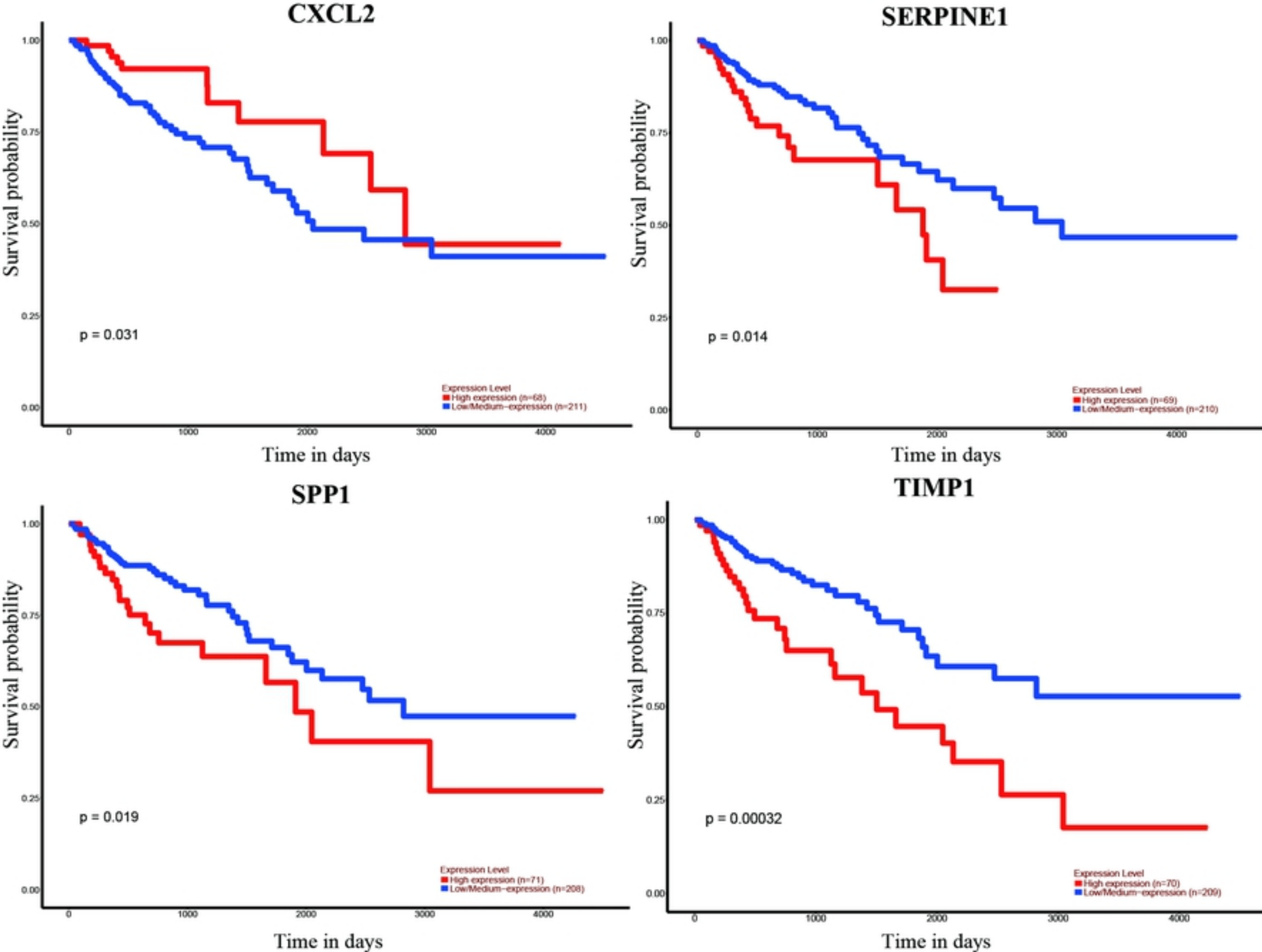
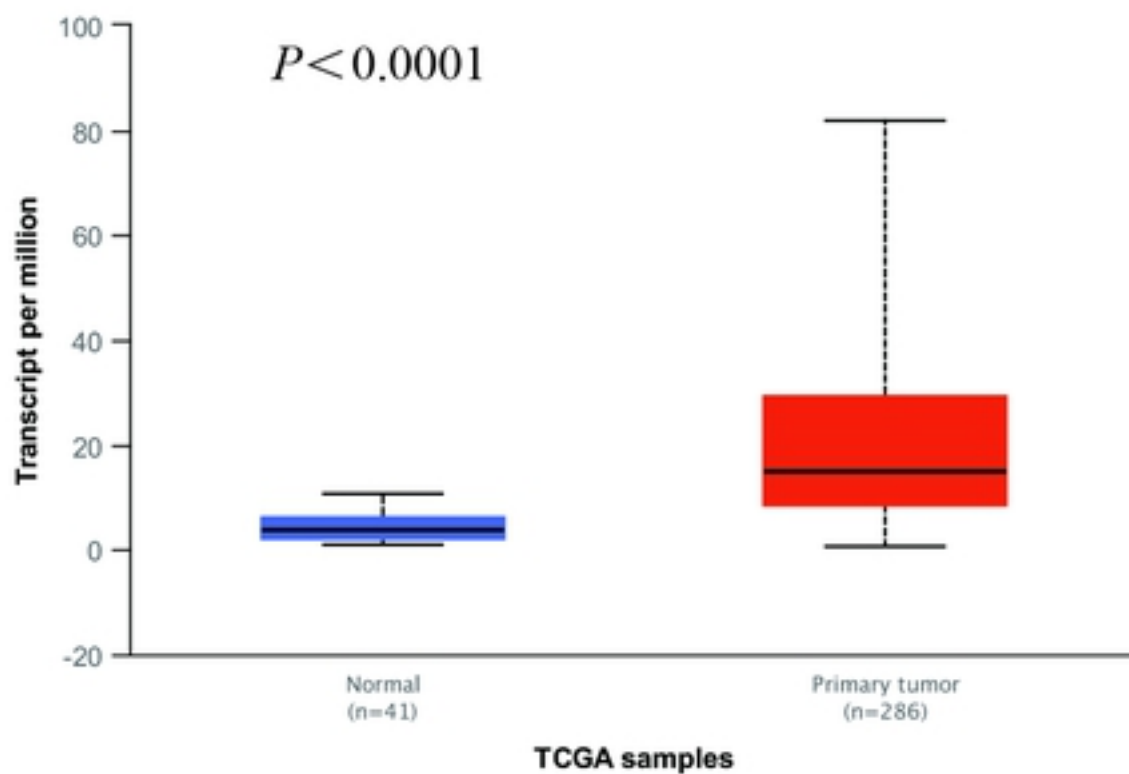
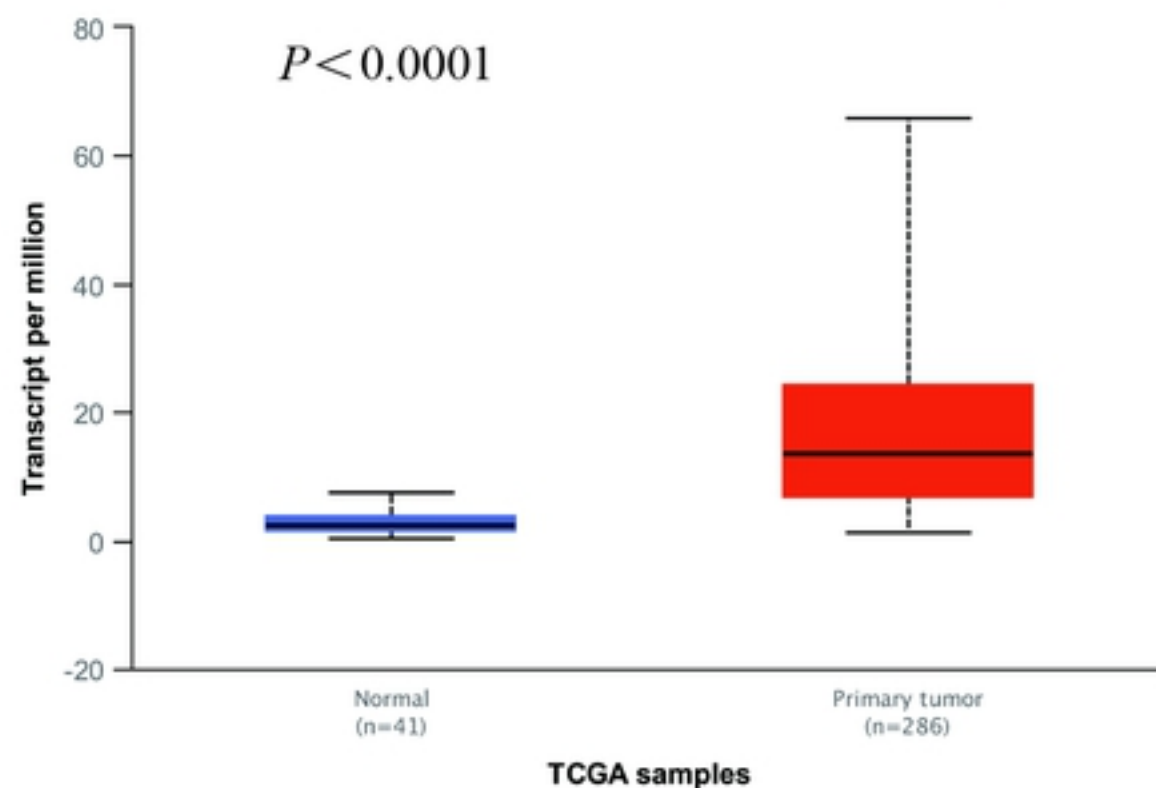


Figure5

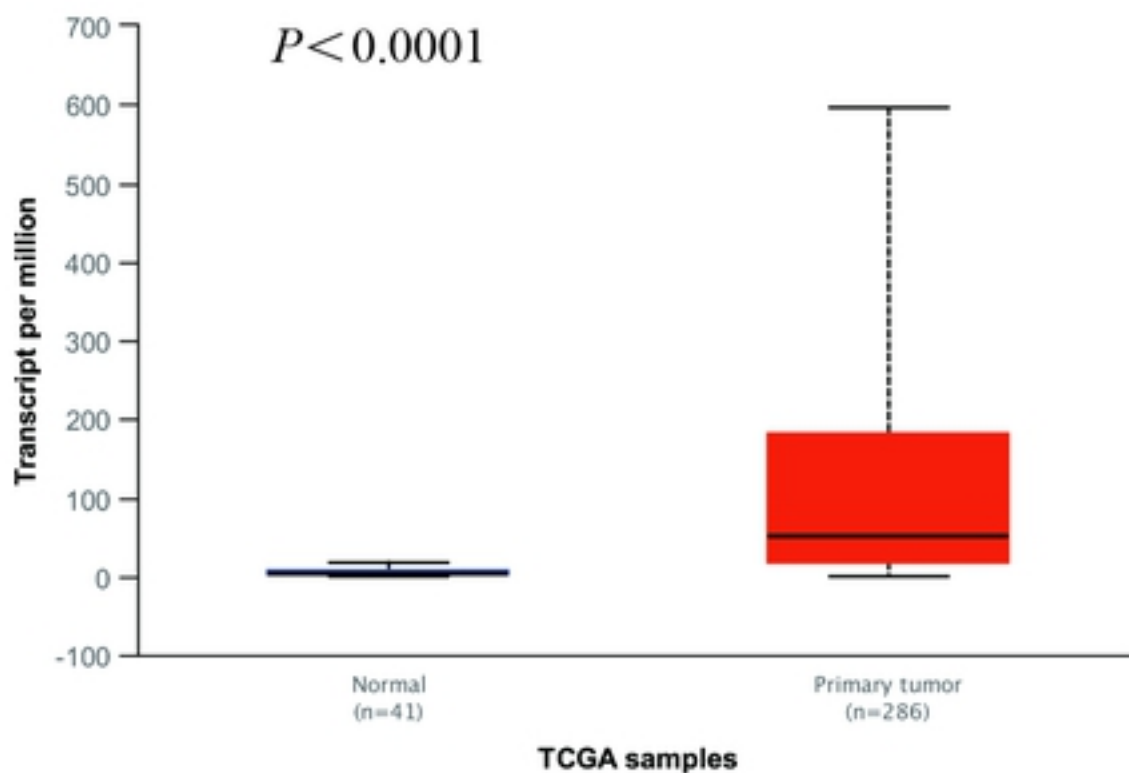
Expression of CXCL2 in COAD based on Sample types



Expression of SERPINE1 in COAD based on Sample types



Expression of SPP1 in COAD based on Sample types



Expression of TIMP1 in COAD based on Sample types

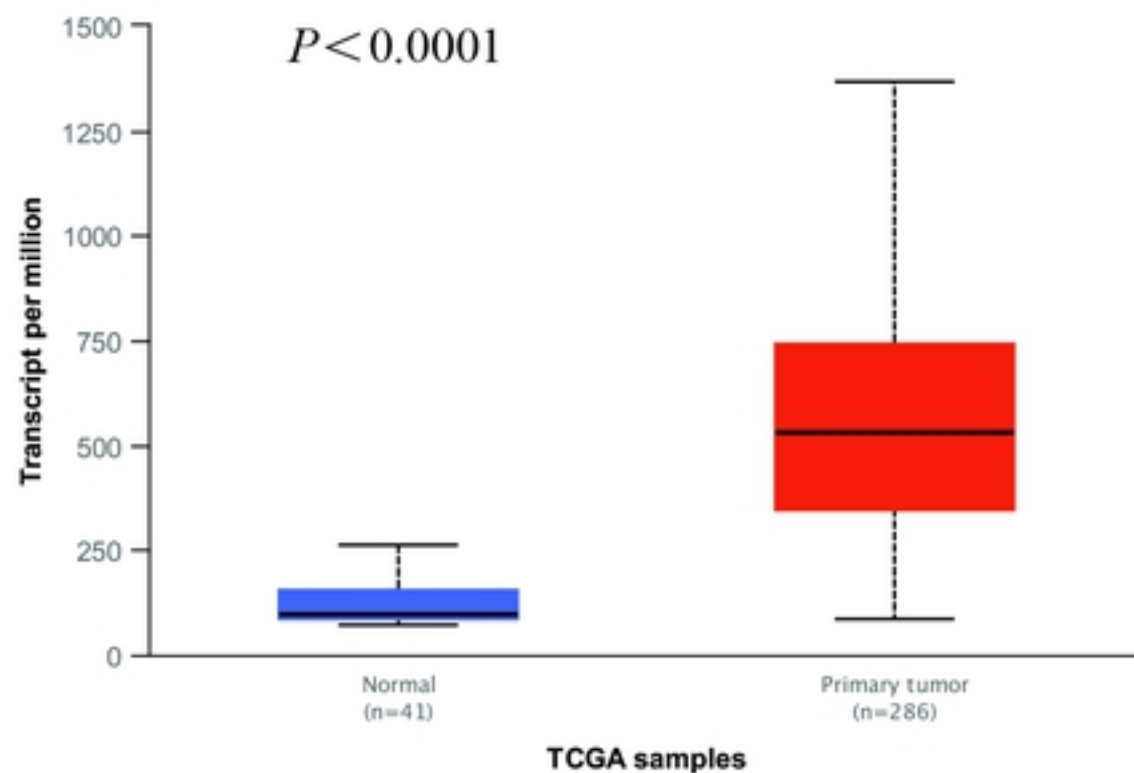


Figure6

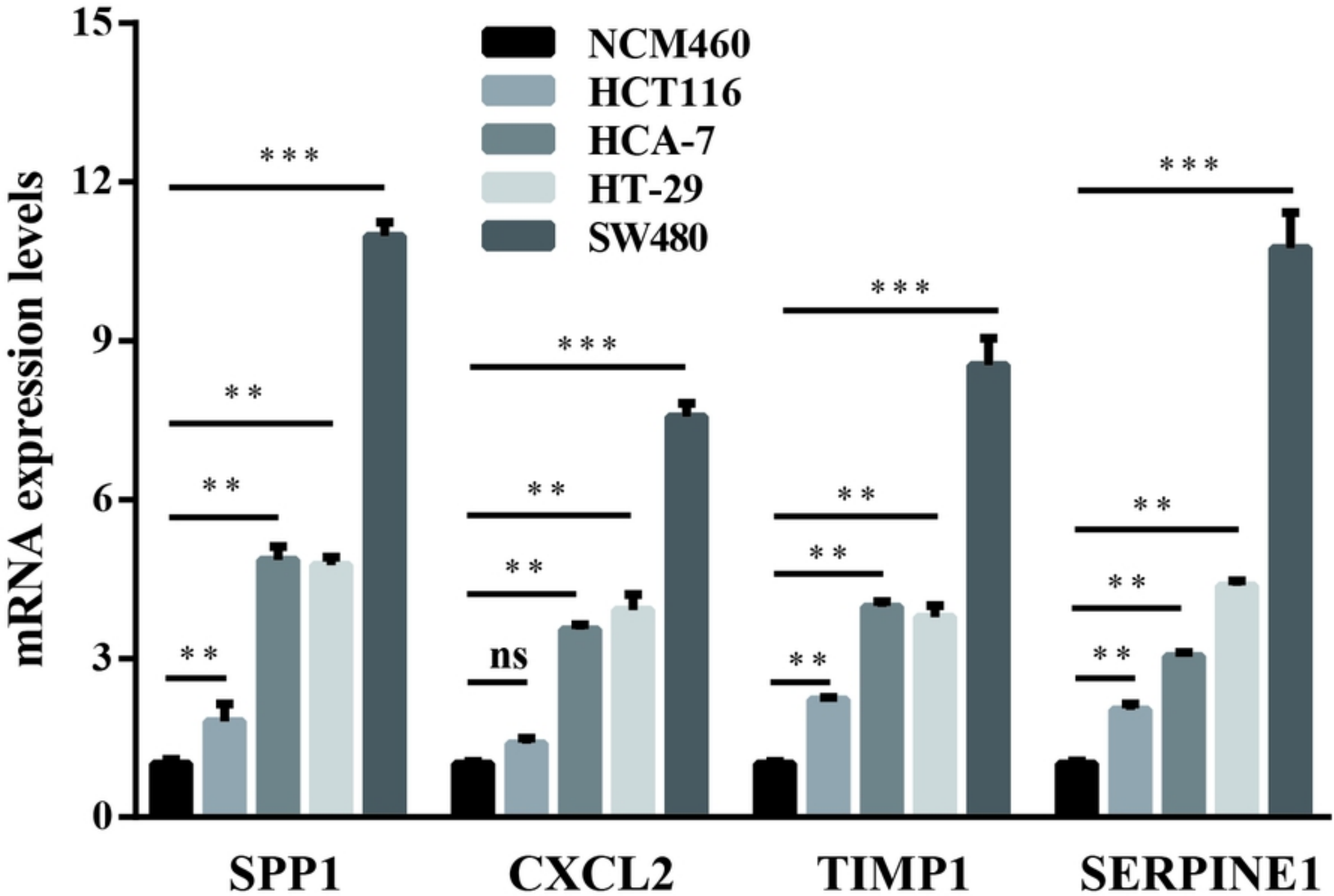


Figure7



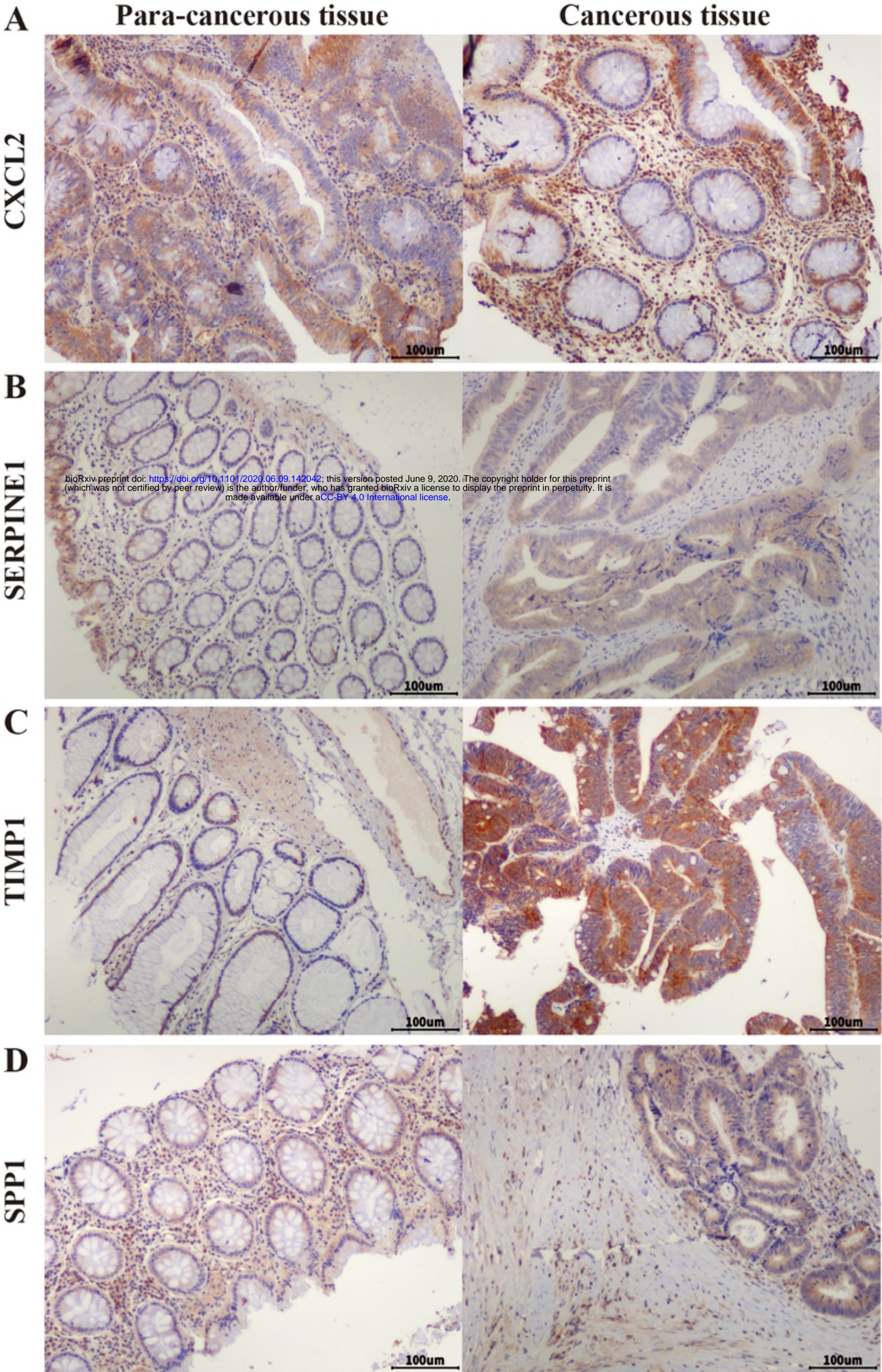


Figure8