

Received: 2018.03.21  
Accepted: 2018.04.10  
Published: 2018.08.31

# Identification and Interaction Analysis of Molecular Markers in Colorectal Cancer by Integrated Bioinformatics Analysis

Authors' Contribution:  
Study Design A  
Data Collection B  
Statistical Analysis C  
Data Interpretation D  
Manuscript Preparation E  
Literature Search F  
Funds Collection G

ABCDG 1,2,3 **Bin Han\***  
BF 2 **Dan Feng\***  
CD 3 **Xin Yu**  
BD 1 **Yuanyuan Zhang**  
DF 1,2 **Yuanqi Liu**  
AE 1 **Liming Zhou**

1 Department of Pharmacology, West China School of Preclinical and Forensic Medicine, Sichuan University, Chengdu, Sichuan, P.R. China  
2 Department of Pharmacy, Affiliated Hospital of North Sichuan Medical College, Nanchong, Sichuan, P.R. China  
3 Health Service Center of Southeast Community, Nanchong, Sichuan, P.R. China

\* These authors contributed equally to this work

**Corresponding Author:** Liming Zhou, e-mail: [zhoulm108@163.com](mailto:zhoulm108@163.com)

**Source of support:** This work was supported by the Health and Family Planning Commission, Sichuan Province (16PJ136)

---

**Background:** Colorectal cancer (CRC) is an extremely common gastrointestinal malignancy.





**Material/Methods:** Three mRNA and 2 microRNA microarray datasets were downloaded from the Gene Expression Omnibus (GEO) database. Differentially expressed genes (DEGs) and microRNAs (DEMs) were obtained. The Database for Annotation, Visualization, and Integrated Discovery (DAVID) program was utilized to perform gene ontology (GO) enrichment analysis and Kyoto Encyclopedia of Genes and Genomes (KEGG) pathway analysis. Protein-protein interaction (PPI) network analysis was performed with the Search Tool for the Retrieval of Interacting Genes (STRING) and Cytoscape and Molecular Complex Detection (MCODE). Kaplan-Meier curves were plotted to determine overall survival (OS) estimates. DEMs targets were predicted by miRWalk. Quantitative reverse transcription polymerase chain reaction (QRT-PCR) was utilized to detect the expression of genes and microRNAs.

**Results:** A total of 264 DEGs and 8 DEMs were obtained. GO analysis revealed that the DEGs were enriched in terms of cell structure, digestion, receptor binding, and extracellular material (ECM). KEGG pathway analysis showed that the DEGs were enriched in ECM interaction and mineral absorption. Additionally, a PPI network consisting of 181 nodes and 450 edges was established. Three modules with 38 high-degree hubs were extracted from the PPI network and found to be involved in pathways such as chemokine signaling. Five DEGs located in the network of DEM-DEG pairs were associated with the overall survival of CRC patients. Furthermore, hsa-miR-551b was demonstrated to be significantly down-regulated in CRC tissues.

**Conclusions:** The key biomarkers could provide new clues for CRC.

**MeSH Keywords:** **Biological Markers • Colorectal Neoplasms • High-Throughput Screening Assays**

**Full-text PDF:** <https://www.medscimonit.com/abstract/index/idArt/910106>

 2832  4  5  42



## Background

Colorectal cancer (CRC) is the third most commonly diagnosed cancer in the world [1]. In China, CRC is the 4th most common cancer, with the 5th highest mortality rate [2]. CRC carries an extremely poor prognosis, particularly due to its strong resistance to chemotherapy, early vascular and lymphatic invasion, and its high rates of distant metastasis and disease recurrence.

MicroRNAs (mRNAs), which are long noncoding RNA (lncRNA) with specific expression profiles in tumor tissues or peripheral blood have been developed as diagnostic and therapeutic biomarkers for CRC [3,4]. Nevertheless, effective diagnostic methods for CRC are still lacking, especially for the early stages of the disease. It is imperative that newer methods to detect CRC-related genetic alterations are invented and diagnostic markers specific for the early stages of this debilitating disease are discovered.

Microarray analysis has been applied to investigate the processes of CRC as it is an effective tool to detect general genetic alterations in the study of oncology [5,6]. Based on the microarray data, bioinformatics analysis has enabled the identification of differentially expressed genes (DEGs), differentially expressed miRNAs (DEMs), signaling pathways, biological processes, and molecular functions [7].

In this study, multiple mRNA and microRNA datasets (3 and 2, respectively) were downloaded from the Gene Expression Omnibus database (GEO DataSets) to minimize false-positive results. Bioinformatics methods enabled reliable DEGs and DEMs between CRC tissues and normal tissues to be obtained, while functional and pathway analyses were carried out to categorize the DEGs. Key mRNAs and microRNAs were selected as potential candidate biomarkers for CRC via protein-protein interaction (PPI) network establishment, overall survival (OS) analysis, and mRNA-microRNA interaction pairing. One of the DEMs was rarely reported in the context of CRC, and we sought to preliminarily verify its expression in CRC tissues.

## Material and Methods

### Microarray data

mRNA microarray datasets GSE81582 [8], GSE44076 [9], and GSE44861 [10] and microRNA datasets GSE41655 [11] and GSE18392 [12] were downloaded from the Gene Expression Omnibus database (GEO DataSets). A total of 385 CRC samples and 206 normal control samples were collected from these datasets.

### Identification of DEGs and DEMs

To obtain differentially expressed genes (DEGs) and microRNAs (DEMs) between colorectal tumor and normal tissue samples, tumor samples and normal samples were divided into 2 groups. After GEO2R (<http://www.ncbi.nlm.nih.gov/geo/geo2r/>) analysis [13], results including adjusted P values (adj. P. Val) and log FC were provided. Cut-off criterion was set as adj. P. Val <0.01 and |log FC| >1. A list of candidate DEGs and DEMs was obtained via the above methods.

### Functional and pathway enrichment of DEGs analysis

Gene ontology (GO) analysis and Kyoto Encyclopedia of Genes and Genomes (KEGG) pathway analysis are both integrated in the Database for Annotation, Visualization and Integrated Discovery (DAVID, <http://david.abcc.ncifcrf.gov/>) program. Therefore, DAVID is capable of providing comprehensive annotations for functional and pathway interpretations [14]. In this experiment, DEGs were uploaded onto DAVID in order to perform related GO and KEGG pathway enrichment analyses. The cut-off criterion was set as  $P < 0.05$ .

### Visualized PPI network establishment and modules selection

Search Tool for the Retrieval of Interacting Genes (STRING, <http://string.embl.de/>) is a database of known and predicted protein-protein interactions [15]. All candidate DEGs were entered into the STRING website, with a confidence score of  $\geq 0.4$  set as the cut-off criterion for PPI network construction. A simple tabular text of PPI created by STRING was visualized using Cytoscape (version 3.6.0, <http://www.cytoscape.org/>). Significant modules in the visible PPI network were screened using the Molecular Complex Detection (MCODE) plugin. Degree cut-off=2, node score cut-off=0.2, k-core=2, and max depth=100 were set as the cut-off criterion. Three highest-degree modules were extracted, and the potential mechanisms of each module were analyzed with DAVID. A degree of  $\geq 10$  was set as the filter criterion. Eligible hub genes were selected as the potential key genes and biomarkers.

### Survival analysis of DEGs

Survival analysis is a key step in identifying the potential of hub genes as key genes and biomarkers. The Kaplan-Meier plotter online tool in R2: Genomics Analysis and Visualization Platform (<http://r2.amc.nl>) is a tool capable of assessing the effect of genes on survival [16]. Patients with CRC were first categorized into 2 groups depending on the degree of gene expression (high or low). The overall survival curves were then plotted with their respective significances analyzed to determine the effect of hub genes. The cut-off criterion was set as  $P < 0.05$ .

## Prediction of miRNA targets

DEM targets were predicted with the help of the miRWalk database. MiRWalk is a database which not only documents miRNA binding sites within the complete sequence of a gene, but also combines this information with a comparison of binding sites resulting from other existing miRNA-target prediction programs [17]. A total of 9 established miRNA-target prediction programs (RNA22, miRWalk, miRanda, miRDB, TargetScan, RNAhybrid, PITA, PICTAR, and Diana-microT) are available on miRWalk. Targets of low abundance miRNAs (e.g., hsa-miR-30a\*) are genes predicted by at least 3 programs, while targets of high abundance miRNAs (e.g., hsa-miR-30a) must be genes predicted by 5 or more programs.

## Patients and samples

To determine the expression of key genes or microRNAs in CRC tissues, tumor and surrounding normal tissues samples (>5 cm away from the tumor) were collected from 16 patients with CRC who underwent radical surgery in the Department of General Surgery, West China Hospital, Sichuan University, between May 2010 and September 2011. The 16 CRC patients were 10 males and 6 females ages 30–72 years and the median age was 52.8 years. The number of patients in stage I, stage II, stage III, and stage IV were 2, 4, 6, and 4, respectively. The pathological diagnosis was performed by 2 pathologists independently. The excised tumor and adjacent tissues were stored in liquid nitrogen until further analysis. All tissue samples were collected from consenting individuals according to protocols approved by the Ethics Review Board at Sichuan University.

## Quantitative reverse transcription polymerase chain reaction (QRT-PCR)

Total RNA was extracted from colorectal tumor tissues and their adjacent non-tumor tissues by using the E.Z.N.A.® Total DNA/RNA/Protein Kit (Omega Bio-Tekinc, Norcross, GA). The RNA extracted was subjected to reverse transcription using the Revert Aid First-Strand cDNA Synthesis Kit (Thermo scientific Inc., MA, US). QRT-PCR master-mix was prepared using SYBR® Premix Ex TaqTMII (Tli RnaseH Plus) (Takara Bio Inc., Japan) and real-time PCR analysis was performed with the CFX96TM Real-Time PCR Detection System. All the experiments were carried out according to the manufacturer's instructions. The primer sequences for real-time PCR were as follows: miR-551b (forward: 5'-ACACTCAGCTGGCTGAAACCAAGTATGGGTCG-3'; reverse: 5'-TGGTGTCTGCAGTCG-3'); U6 (forward: 5'-CTCGCTTCGGCAGCACA-3'; reverse: 5'-AACGCTTACGAATTTGCGT-3'). The relative expression of miR-551b was normalized against U6 and analyzed with the  $2^{-\Delta\Delta Ct}$  method [18].

## Statistical analysis

SPSS software (version 20.0, SPSS, Inc., Chicago, IL, USA) was used for statistical analysis. Data are expressed as mean  $\pm$  standard error. The comparison between 2 groups was performed by using the *t* test. A *p* value of <0.05 was considered to be statistically significant.

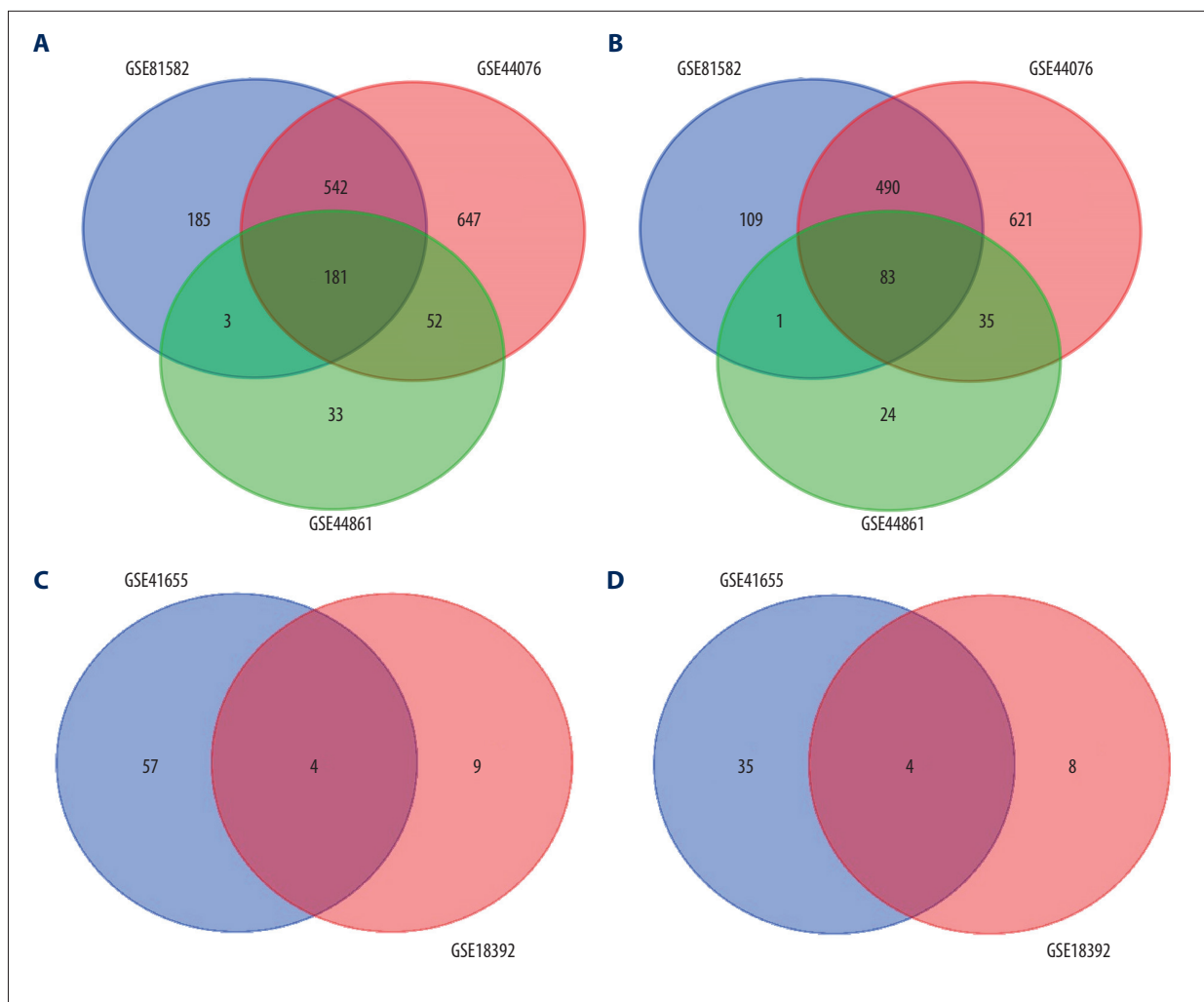
## Results

### Identification of DEGs and DEMs

A total of 264 DEGs expressed in CRC tissues were extracted from the GSE81582, GSE44076, and GSE44861 datasets. Expressions of 181 genes were down-regulated (Figure 1A) and expressions of 83 genes were up-regulated (Figure 1B) in comparison to normal controls. Eight DEMs in CRC tissues were identified in the GSE41655 and GSE18392 datasets. Compared to samples from normal controls, the relative expressions of 4 microRNAs (hsa-miR-378\*, hsa-miR-375, hsa-miR-30a\*, and hsa-miR-551b) were down-regulated (Figure 1C) while the remaining 4 microRNAs (hsa-miR-224, hsa-miR-584, hsa-miR-183, and hsa-miR-135b) were up-regulated (Figure 1D).

### Functional and pathway enrichment analysis

DAVID was used to perform functional and pathway enrichment analysis for the identified DEGs. GO analysis results showed that for biological processes (BP), up-regulated DEGs were significantly enriched in extracellular matrix and structure organization, cell migration, motility, and localization. Down-regulated DEGs were significantly enriched in digestion, homeostasis, and cellular response to zinc ion and bicarbonate transport (Table 1). Up-regulated DEGs were significantly enriched in cell components (CC), including proteinaceous extracellular matrices, banded collagen fibrils, and extracellular spaces and matrices. Down-regulated DEGs were significantly enriched in CCs, including extracellular regions, exosomes, vesicles, and organelles (Table 1). GO molecular function (MF) showed that the up-regulated DEGs were significantly enriched in extracellular matrix structural constituents, including regions of CXCR chemokine receptor binding, cytokine activity, receptor binding, and glycosaminoglycan binding. Down-regulated DEGs were significantly enriched in carbonate dehydratase, phospholipase, oxidoreductase activity, zinc ion binding, and lipase activity (Table 1). Moreover, 21 KEGG pathways were over-represented in the up-regulated DEGs, including ECM-receptor interaction, focal adhesion, and the PI3K-Akt signaling pathway, while the down-regulated DEGs were significantly enriched in 15 KEGG pathways, including mineral absorption, nitrogen metabolism, and sulfur metabolism (Table 1).



**Figure 1.** Differentially expressed genes and microRNAs in microarray datasets. Down-regulated (**A**) and up-regulated (**B**) differentially expressed genes in GSE81582, GSE44076, and GSE44861 datasets; down-regulated (**C**) and up-regulated (**D**) differentially expressed microRNAs in GSE41655 and GSE18392 datasets.

### PPI network construction and modules selection

By using the STRING database, the PPI network of DEGs was established and consisted of 181 nodes and 450 edges. There were 65 up-regulated and 116 down-regulated genes contained in the 181 nodes (Figure 2A). The PPI network was observed by Cytoscape, and the cut-off criterion of hub gene selection was set as degrees  $\geq 10$ . A total of 23 genes were selected for key biomarker identification. They consisted of 13 up-regulated genes (*THY1*, *MET*, *MMP7*, *CXCL1*, *CCND1*, *SPP1*, *CD44*, *COL1A1*, *COL1A2*, *TIMP1*, *MMP1*, *MYC*, *SERPINE1*) and 10 down-regulated genes (*BMP2*, *PHLPP2*, *ACACB*, *LPAR1*, *GNA11*, *GCG*, *CXCL12*, *SST*, *PYY*, *BCL2*). The 3 most significant modules, which included a total of 38 nodes and 120 edges, were extracted from the PPI network by MCODE (Figure 2B–2D). Genes in these modules were mainly associated with chemokine signaling, extracellular matrix, and mineral absorption (Tables 2–4); 12 KEGG

pathways were represented by these nodes, which comprised of chemokine signaling pathways, the TNF signaling pathway, and the PI3K-Akt signaling pathway (Tables 2–4).

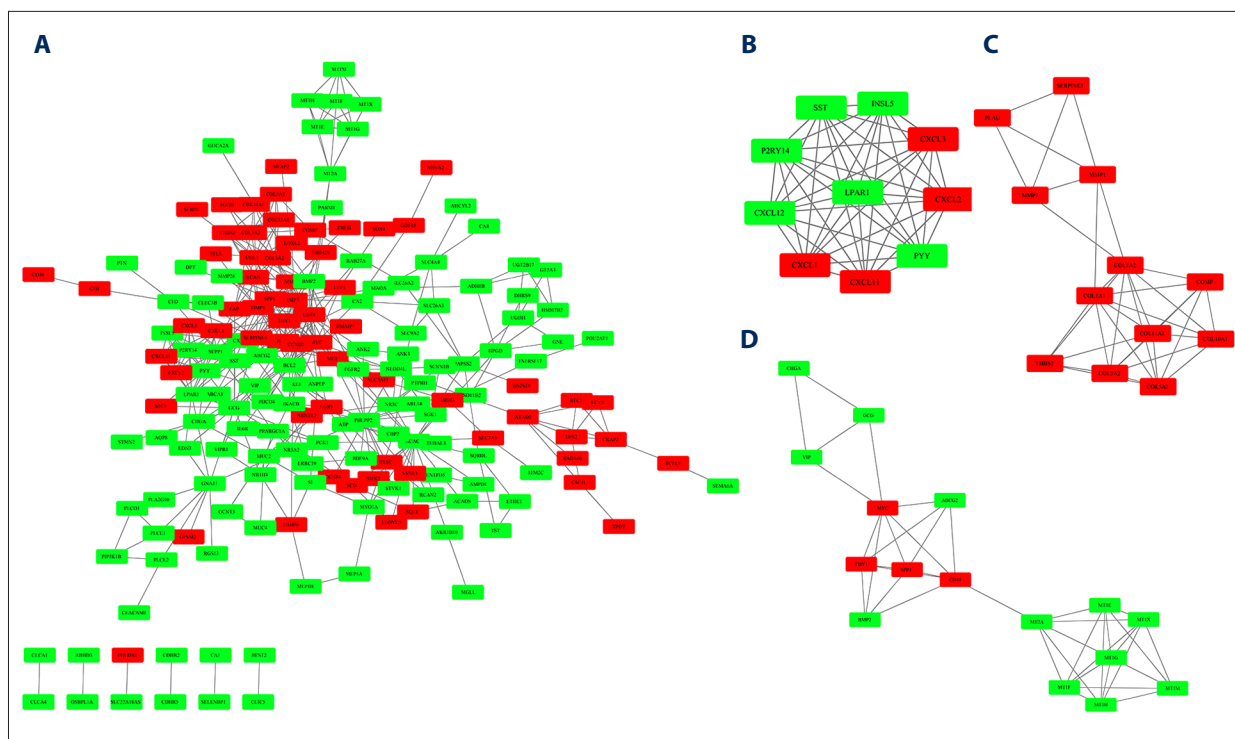
### The survival analysis

The prognostic values of the 23 selected key biomarker genes were assessed using the Kaplan-Meier online tool in R2. Results showed that elevated expressions of 4 genes (*TIMP1*, *SERPINE1*, *CCND1*, *COL1A2*) and low expression of the *SST* gene in CRC patients were associated with worse overall survival rates in CRC patients (Figure 3). Therefore, these 5 DEGs were considered to be potential key biomarkers of CRC.

**Table 1.** Gene ontology and KEGG pathway analysis of up-regulated and down-regulated genes in colorectal cancer.

Expression	Category	Term/Description	Count	P value
Up-regulated	GOTERM_BP_FAT	GO: 0030198~extracellular matrix organization	18	2.97E-13
	GOTERM_BP_FAT	GO: 0043062~extracellular structure organization	18	3.12E-13
	GOTERM_BP_FAT	GO: 0016477~cell migration	27	1.99E-11
	GOTERM_BP_FAT	GO: 0048870~cell motility	27	2.58E-10
	GOTERM_BP_FAT	GO: 0051674~localization of cell	27	2.58E-10
	GOTERM_CC_FAT	GO: 0005578~proteinaceous extracellular matrix	16	8.11E-10
	GOTERM_CC_FAT	GO: 0005615~extracellular space	28	4.09E-09
	GOTERM_CC_FAT	GO: 0031012~extracellular matrix	17	2.02E-08
	GOTERM_CC_FAT	GO: 0044420~extracellular matrix component	10	2.95E-08
	GOTERM_CC_FAT	GO: 0098643~banded collagen fibril	5	3.88E-07
	GOTERM_MF_FAT	GO: 0005201~extracellular matrix structural constituent	8	1.76E-07
	GOTERM_MF_FAT	GO: 0045236~CXCR chemokine receptor binding	4	7.96E-05
	GOTERM_MF_FAT	GO: 0005125~cytokine activity	8	1.27E-04
	GOTERM_MF_FAT	GO: 0005102~receptor binding	19	2.79E-04
	GOTERM_MF_FAT	GO: 0005539~glycosaminoglycan binding	7	6.30E-04
	KEGG_PATHWAY	hsa04512: ECM-receptor interaction	9	6.72E-08
	KEGG_PATHWAY	hsa04510: Focal adhesion	10	5.45E-06
	KEGG_PATHWAY	hsa04151: PI3K-Akt signaling pathway	11	5.62E-05
	KEGG_PATHWAY	hsa04974: Protein digestion and absorption	6	2.44E-04
	KEGG_PATHWAY	hsa05146: Amoebiasis	5	0.004809
Down-regulated	GOTERM_BP_FAT	GO: 0007586~digestion	14	9.98E-09
	GOTERM_BP_FAT	GO: 0048878~chemical homeostasis	33	1.34E-08
	GOTERM_BP_FAT	GO: 0071294~cellular response to zinc ion	7	2.27E-08
	GOTERM_BP_FAT	GO: 0015701~bicarbonate transport	8	2.52E-07
	GOTERM_BP_FAT	GO: 0042592~homeostatic process	40	2.53E-07
	GOTERM_CC_FAT	GO: 0044421~extracellular region part	74	1.32E-08
	GOTERM_CC_FAT	GO: 0005576~extracellular region	82	3.03E-08
	GOTERM_CC_FAT	GO: 0070062~extracellular exosome	59	4.84E-08
	GOTERM_CC_FAT	GO: 1903561~extracellular vesicle	59	5.81E-08
	GOTERM_CC_FAT	GO: 0043230~extracellular organelle	59	5.88E-08
	GOTERM_MF_FAT	GO: 0004089~carbonate dehydratase activity	5	8.96E-06
	GOTERM_MF_FAT	GO: 0004620~phospholipase activity	8	7.82E-05
	GOTERM_MF_FAT	GO: 0016616~oxidoreductase activity	8	1.74E-04
	GOTERM_MF_FAT	GO: 0008270~zinc ion binding	26	2.41E-04
	GOTERM_MF_FAT	GO: 0016298~lipase activity	8	3.03E-04
	KEGG_PATHWAY	hsa04978: Mineral absorption	8	1.78E-06
	KEGG_PATHWAY	hsa00910: Nitrogen metabolism	5	5.63E-05
	KEGG_PATHWAY	hsa00920: Sulfur metabolism	4	2.40E-04
	KEGG_PATHWAY	hsa04972: Pancreatic secretion	7	0.001269
	KEGG_PATHWAY	hsa04960: Aldosterone-regulated sodium reabsorption	5	0.001563

If there were more than 5 terms enriched in this category, top 5 terms were selected according to *P* value. BP – biological process; CC – cellular component; MF – molecular function; GO – gene ontology; KEGG – Kyoto Encyclopedia of Genes and Genomes.



**Figure 2.** Protein–protein interaction (PPI) network and the top 3 modules from the PPI network. (A) The protein-protein interaction network of differentially expressed genes; (B) module 1; (C) module 2; (D) module 3.

**Table 2.** Functional and pathway enrichment analysis of module 1.

Category	Term/description	Count	P value
GOTERM_BP_FAT	GO: 0050921~positive regulation of chemotaxis	6	2.30E-09
GOTERM_BP_FAT	GO: 0007186~G-protein coupled receptor signaling pathway	9	8.90E-09
GOTERM_BP_FAT	GO: 0050920~regulation of chemotaxis	6	2.00E-08
GOTERM_BP_FAT	GO: 0070098~chemokine-mediated signaling pathway	5	6.76E-08
GOTERM_BP_FAT	GO: 0002690~positive regulation of leukocyte chemotaxis	5	6.76E-08
GOTERM_CC_FAT	GO: 0005615~extracellular space	7	6.15E-05
GOTERM_CC_FAT	GO: 0005576~extracellular region	8	0.006196
GOTERM_CC_FAT	GO: 0044421~extracellular region part	7	0.013994
GOTERM_MF_FAT	GO: 0045236~CXCR chemokine receptor binding	5	1.25E-10
GOTERM_MF_FAT	GO: 0001664~G-protein coupled receptor binding	7	1.59E-09
GOTERM_MF_FAT	GO: 0008009~chemokine activity	5	1.10E-08
GOTERM_MF_FAT	GO: 0042379~chemokine receptor binding	5	3.09E-08
GOTERM_MF_FAT	GO: 0005102~receptor binding	8	2.01E-06
KEGG_PATHWAY	hsa04062: Chemokine signaling pathway	5	1.67E-05
KEGG_PATHWAY	hsa05134: Legionellosis	3	0.001228
KEGG_PATHWAY	hsa05132: Salmonella infection	3	0.002879
KEGG_PATHWAY	hsa04668: TNF signaling pathway	3	0.004656

If there were more than 5 terms enriched in this category, top 5 terms were selected according to P value. BP – biological process; CC – cellular component; MF – molecular function; GO – gene ontology; KEGG – Kyoto Encyclopedia of Genes and Genomes.

**Table 3.** Functional and pathway enrichment analysis of module 2.

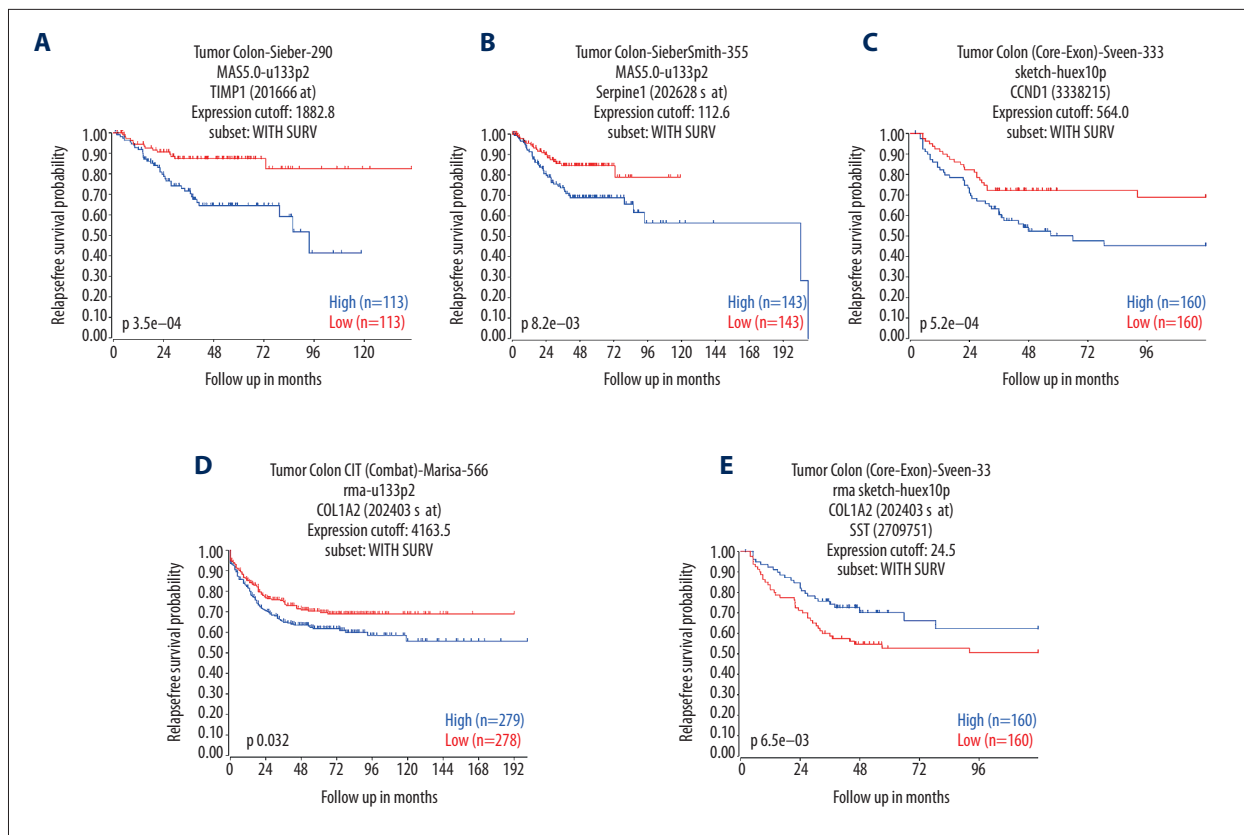
Category	Term/description	Count	P value
GOTERM_BP_FAT	GO: 0030574~collagen catabolic process	8	3.22E-15
GOTERM_BP_FAT	GO: 0044243~multicellular organism catabolic process	8	6.81E-15
GOTERM_BP_FAT	GO: 0030198~extracellular matrix organization	10	2.51E-14
GOTERM_BP_FAT	GO: 0043062~extracellular structure organization	10	2.58E-14
GOTERM_BP_FAT	GO: 0032963~collagen metabolic process	8	1.46E-13
GOTERM_CC_FAT	GO: 0031012~extracellular matrix	11	4.09E-14
GOTERM_CC_FAT	GO: 0005578~proteinaceous extracellular matrix	10	1.72E-13
GOTERM_CC_FAT	GO: 0005581~collagen trimer	7	4.64E-11
GOTERM_CC_FAT	GO: 0005583~fibrillar collagen trimer	5	8.78E-11
GOTERM_CC_FAT	GO: 0098643~banded collagen fibril	5	8.78E-11
GOTERM_MF_FAT	GO: 0005201~extracellular matrix structural constituent	6	1.66E-09
GOTERM_MF_FAT	GO: 0048407~platelet-derived growth factor binding	3	2.52E-05
GOTERM_MF_FAT	GO: 0005198~structural molecule activity	6	1.16E-04
GOTERM_MF_FAT	GO: 0008201~heparin binding	4	1.68E-04
GOTERM_MF_FAT	GO: 0046872~metal ion binding	10	2.33E-04
KEGG_PATHWAY	hsa04512: ECM-receptor interaction	7	1.47E-09
KEGG_PATHWAY	hsa04974: Protein digestion and absorption	6	1.30E-07
KEGG_PATHWAY	hsa04510: Focal adhesion	7	2.66E-07
KEGG_PATHWAY	hsa04151: PI3K-Akt signaling pathway	7	5.54E-06
KEGG_PATHWAY	hsa05146: Amoebiasis	5	1.59E-05

If there were more than 5 terms enriched in this category, top 5 terms were selected according to *P* value. BP – biological process; CC – cellular component; MF – molecular function; GO – gene ontology; KEGG – Kyoto Encyclopedia of Genes and Genomes.

**Table 4.** Functional and pathway enrichment analysis of module 3.

Category	Term/description	Count	P value
GOTERM_BP_FAT	GO: 0071294~cellular response to zinc ion	7	4.56E-15
GOTERM_BP_FAT	GO: 0010043~response to zinc ion	7	4.28E-12
GOTERM_BP_FAT	GO: 0045926~negative regulation of growth	8	6.91E-10
GOTERM_BP_FAT	GO: 0071276~cellular response to cadmium ion	5	1.01E-09
GOTERM_BP_FAT	GO: 1990267~response to transition metal nanoparticle	7	1.05E-09
GOTERM_CC_FAT	GO: 0048471~perinuclear region of cytoplasm	9	6.57E-08
GOTERM_MF_FAT	GO: 0008270~zinc ion binding	7	3.24E-04
GOTERM_MF_FAT	GO: 0046914~transition metal ion binding	7	9.38E-04
GOTERM_MF_FAT	GO: 0005102~receptor binding	5	0.036725
KEGG_PATHWAY	hsa04978: Mineral absorption	7	1.03E-10

If there were more than 5 terms enriched in this category, top 5 terms were selected according to *P* value. BP – biological process; CC – cellular component; MF – molecular function; GO – gene ontology; KEGG – Kyoto Encyclopedia of Genes and Genomes.



**Figure 3.** Prognostic value of 5 genes in patients with colorectal cancer. Prognostic value: (A) TIMP1; (B) SERPINE1; (C) CCND1; (D) COL1A2; (E) SST.

**miRNA-mRNA pairs**

Predicted targets of the 8 DEMs were obtained with miRWalk. Results revealed 14 common target genes of the 4 down-regulated DEMs (Figure 4A), while the other 4 up-regulated DEMs were found to have a total of 68 common target genes (Figure 4B). *PAG1* was the only predicted common target of all 8 DEMs (Figure 4C). According to the negative regulatory mechanism between microRNA and mRNA and based on the expression trend of DEGs and DEMs in CRC, we found that *SERPINE1* was the predicted target of hsa-miR-378\* and hsa-miR-30a\*, *COL1A2* was the predicted target of hsa-miR-30a\*, and *CCND1* was the predicted target of hsa-miR-551b (Figure 4C).

**Relative expression of hsa-miR-551b in CRC tissues**

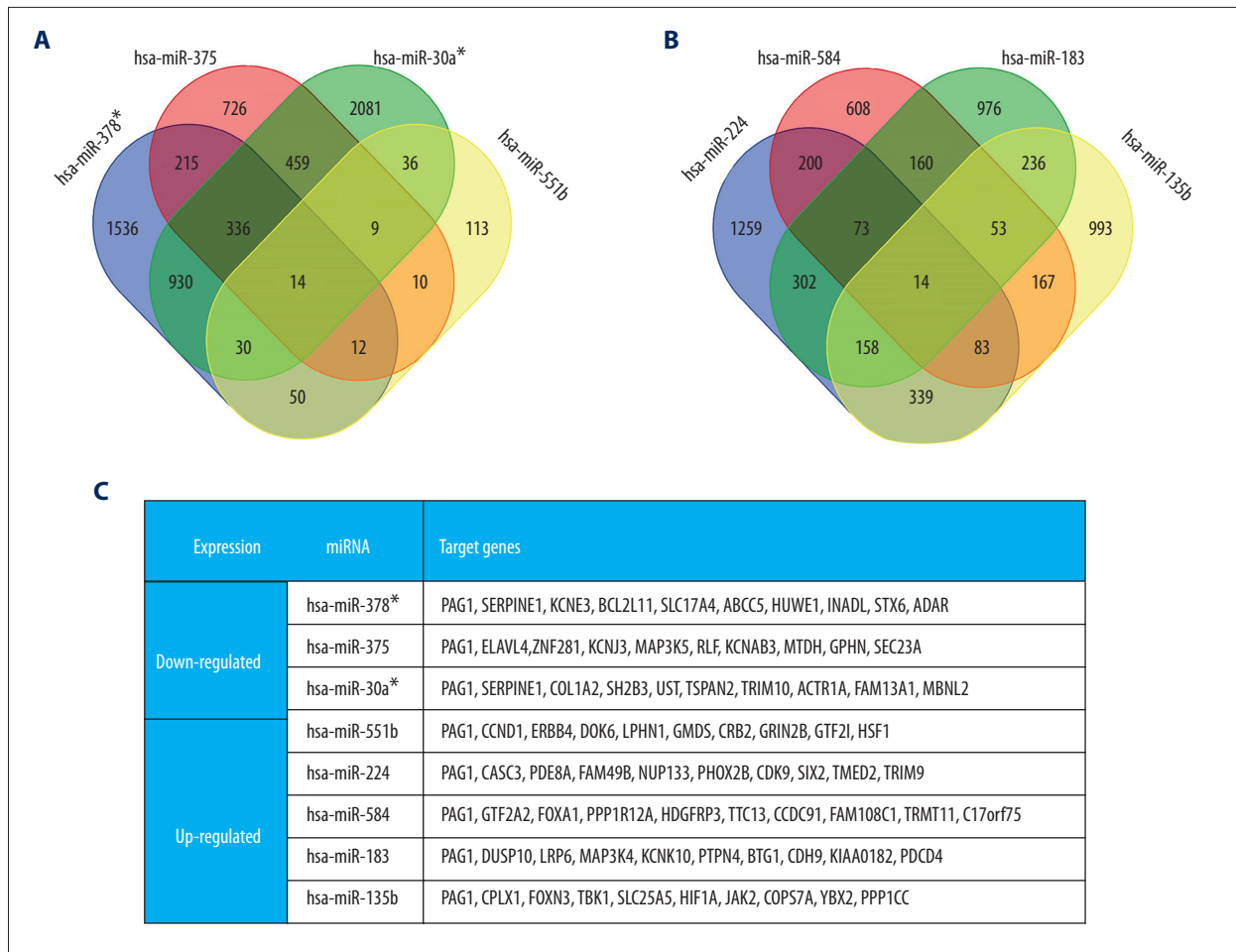
To the best of our knowledge, the expression and function of hsa-miR-551b in CRC has not been previously reported. Therefore, we examined the relative expression of hsa-miR-551b in 16 freshly-frozen CRC tissues via real-time PCR. Compared to the normal tissues, the expression of hsa-miR-551b was significantly down-regulated in CRC tissues and also displayed a decreasing expression trend in bioinformatics analysis (Figure 5). The result highlights the possible function of hsa-miR-551b as an anti-oncogene in CRC.

**Discussion**

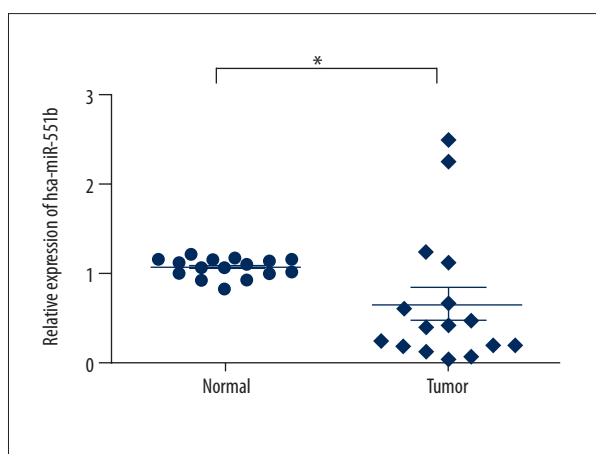
Colorectal cancer is an extremely debilitating gastrointestinal malignancy [1]. In China, CRC remains a huge threat to human life due to its high incidence and mortality rates [2]. Early and accurate detection of the disease based on expression analysis of molecular biomarkers has been proven to be an effective way to improve survival rates in patients with CRC [19,20]. Recent advances in microarray technology have succeeded in elucidating major genetic events that contribute to CRC [7]. As a result, more clues are discerned regarding potentially diagnostic, therapeutic, and prognostic biomarkers that are evident during the progression of CRC [19,21–23].

In the present study, a total of 385 CRC samples and 206 normal control samples were collected from 5 datasets. A total of 264 DEGs and 8 DEMs were identified from this analysis. Out of all the DEGs, up-regulated DEGs were found to be enriched in extracellular matrix structures, cell migration, and localization, while down-regulated DEGs were mainly enriched in digestive and homeostatic mechanisms. The 3 most significant cluster modules were extracted from the visualized PPI network of DEGs. Meanwhile, 23 high-degree genes were selected from the PPI network for key biomarker identification.





**Figure 4.** Differentially expressed microRNAs (DEMs) in colorectal cancer and their targets. (A) Targets of down-regulated DEMs; (B) targets of up-regulated DEMs; (C) list of partial targets.



**Figure 5.** Expression levels of hsa-miR-551b in colorectal cancer and normal tissues. Data were pooled from 3 independent experiments, \*  $P < 0.05$  vs. normal tissues.

Survival analysis of these genes showed that 4 up-regulated DEGs (*TIMP1*, *CCND1*, *SERPINE1*, *COL1A2*) and 1 down-regulated DEG (*SST*) were significantly related to worse overall survival rates of patients with CRC.

*TIMP1* (TIMP metalloproteinase inhibitor 1) functions in tissue remodeling, tumor angiogenesis, and tumor cell invasion and metastasis by inhibiting the activity of metalloproteinases [24,25]. It is widely overexpressed in multiple tumors, including CRC [23]. Aberrations of *CCND1* (cyclin D1) contribute to excessive cell proliferation and cancer occurrence [26]. Compared with the adjacent normal tissues, the expression of *CCND1* was significantly up-regulated in CRC tumor samples [27]. It has been associated with the poor prognosis of different types of tumors by promoting tumor invasion and metastasis [28]. Based on previous studies, *TIMP1* and *CCND1* are generally considered to be oncogenes [27,29]. Their oncogenic roles are reinforced by the present study, as we demonstrated that an over-expression of both genes was significantly related to worse overall survival of patients with CRC.

*SERPINE1* (serpin family E member 1), also named *PAI-1*, can be combined with *uPA* to form an inactive complex [30]. Silencing of *PAI-1* suppressed the progression and occurrence of liver metastasis in patients with CRC [31]. Interestingly, previous studies have not found *PAI-1* to be an independent unfavorable prognostic factor for overall 5- and 10-year survival of patients with CRC, despite significant plasma levels of *PAI-1* [29]. *COL1A2* (collagen type I alpha 2 chain) has previously been frequently reported by many bioinformatic analyses for CRC [32,33], and has been suggested to play an important role in the disease. In the present study, we found that *COL1A2* and *SERPINE1* were both included in MCODE cluster 2, a gene cluster associated with collagen and ECM. These results indicate that aberrations of these 2 genes disrupt the normal collagen physiology. The abnormal expression of *COL1A2* and *SERPINE1* may be considered as a typical feature in CRC development.

*SST* (somatostatin) inhibits the release of numerous secondary hormones by binding to G-protein-coupled somatostatin receptors [34,35]. Somatostatin signaling contributes to the quiescence of colon cancer stem cells via somatostatin receptor type 1 (SSTR1) [36]. Reduced production of *SST* can promote uncontrolled cell proliferation in CRC [37]. In the present study, our data show that *SST* is a hub node in MCODE cluster 1, with its low expression significantly correlated with worse overall survival rates of patients with CRC. Taken together, our results suggest that *TIMP1*, *SERPINE1*, *CCND1*, *COL1A2*, and *SST* may be potentially valuable biomarkers, as each is intricately involved in the pathogenesis of CRC by affecting different physiological processes.

Mounting evidence indicates that abnormal expression of miRNAs may be involved in the pathogenesis of CRC [38,39]. Some miRNAs have been proven to be appropriate biomarkers because of their specific expression profiles [40]. In our study, 8 DEMs from CRC and normal samples were identified by GEO2R analysis. The 4 down-regulated and 4 up-regulated DEMs had 14 and 68 common predicted target genes, respectively. These miRNA-mRNA pairs form a large gene signal network in CRC. As the predicted targets of some DEMs, *SERPINE1*, *COL1A2*,

and *CCND1* are also involved in the network. In our study, we uncovered a previously unreported DEM involved in CRC: hsa-miR-551b. However, low hsa-miR-551b levels have been associated with EMT, metastasis, and a poor prognosis in gastric cancer patients [41]. It has also been investigated as a candidate diagnostic biomarker for prostate cancer [42]. Our results show that the relative expression of hsa-miR-551b in CRC tissues were significantly down-regulated in real-time PCR analysis, as well as in bioinformatics analysis. Interestingly, hsa-miR-551b has previously been predicted to be an upstream regulator of *CCND1*. Additional studies are needed to further clarify the role of hsa-miR-551b in the CRC gene network.

In summary, our study provides a comprehensive bioinformatics analysis of CRC. Key biomarkers and molecular mechanisms underlying CRC progression were investigated. Although our conclusions would benefit from additional molecular experiments and clinical practice validation, we are hopeful that our results provide new diagnosis, therapeutic, and prognostic clues for the management of CRC patients.

## Conclusions

Taken together, our results show the functional and pathway enrichment of DEGs in CRC, established a visualized protein-protein interaction network, and demonstrated that 5 key genes are associated with the overall survival of CRC patients. We also identified 8 DEMs that form a gene signal network with miRNA-mRNA pairing. Furthermore, we reported for the first time that hsa-miR-551b is down-regulated in CRC tissues, as shown by our bioinformatics analysis. Our study identified potential biomarkers and we are currently performing clinical evaluation of their validity. Results of the present study may provide new diagnostic, therapeutic, and prognostic clues for the management of CRC patients.

## Conflict of interest

None.

## References:

- Brenner H, Kloor M, Pox CP: Colorectal cancer. *Lancet*, 2014; 383: 1490–502
- Chen W, Zheng R, Baade PD et al: Cancer statistics in China, 2015. *Cancer J Clin*, 2016; 66: 115–32
- Rokavec M, Horst D, Hermeking H: Cellular model of colon cancer progression reveals signatures of mRNAs, miRNA, lncRNAs, and epigenetic modifications associated with metastasis. *Cancer Res*, 2017; 77: 1854–67
- Slaby O: Non-coding RNAs as biomarkers for colorectal cancer screening and early detection. *Adv Exp Med Biol*, 2016; 937: 153–70
- Chen Y, Yu X, Xu Y, Shen H: Identification of dysregulated lncRNAs profiling and metastasis-associated lncRNAs in colorectal cancer by genome-wide analysis. *Cancer Med*, 2017; 6: 2321–30
- Zong S, Li W, Li H et al: Identification of hypoxia-regulated angiogenic genes in colorectal cancer. *Biochem Biophys Res Commun*, 2017; 493: 461–67
- Slattery ML, Herrick JS, Mullany LE et al: The co-regulatory networks of tumor suppressor genes, oncogenes, and miRNAs in colorectal cancer. *Genes Chromosomes Cancer*, 2017; 56: 769–87
- Sayagues JM, Corchete LA, Gutierrez ML et al: Genomic characterization of liver metastases from colorectal cancer patients. *Oncotarget*, 2016; 7: 72908–22
- Sole X, Crous-Bou M, Cordero D et al: Discovery and validation of new potential biomarkers for early detection of colon cancer. *PLoS One*, 2014; 9: e106748
- Ryan BM, Zanetti KA, Robles AI et al: Germline variation in *NCF4*, an innate immunity gene, is associated with an increased risk of colorectal cancer. *Int J Cancer*, 2014; 134: 1399–407

11. Sun M, Song H, Wang S et al: Integrated analysis identifies microRNA-195 as a suppressor of Hippo-YAP pathway in colorectal cancer. *J Hematol Oncol*, 2017; 10: 79
12. Sarver AL, French AJ, Borralho PM et al: Human colon cancer profiles show differential microRNA expression depending on mismatch repair status and are characteristic of undifferentiated proliferative states. *BMC Cancer*, 2009; 9: 401
13. Barrett T, Wilhite SE, Ledoux P et al: NCBI GEO: Archive for functional genomics data sets – update. *Nucleic Acids Res*, 2013; 41: D991–95
14. Huang da W, Sherman BT, Lempicki RA: Systematic and integrative analysis of large gene lists using DAVID bioinformatics resources. *Nat Protoc*, 2009; 4: 44–57
15. Szklarczyk D, Franceschini A, Wyder S et al: STRING v10: Protein-protein interaction networks, integrated over the tree of life. *Nucleic Acids Res*, 2015; 43: D447–52
16. Xie Y, Zhu S, Zhong M et al: Inhibition of aurora kinase A induces necroptosis in pancreatic carcinoma. *Gastroenterology*, 2017; 153: 1429–43e5
17. Dweep H, Sticht C, Pandey P, Gretz N: miRWalk – database: Prediction of possible miRNA binding sites by “walking” the genes of three genomes. *J Biomed Inform*, 2011; 44: 839–47
18. Varnholt H, Drebber U, Schulze F et al: MicroRNA gene expression profile of hepatitis C virus-associated hepatocellular carcinoma. *Hepatology*, 2008; 47: 1223–32
19. Mahasneh A, Al-Shaheri F, Jamal E: Molecular biomarkers for an early diagnosis, effective treatment and prognosis of colorectal cancer: Current updates. *Exp Mol Pathol*, 2017; 102: 475–83
20. Seleznick MJ. Tumor markers. *Prim Care*, 1992; 19: 715–26
21. Sun X, Yuan W, Hao F et al: Promoter methylation of RASSF1A indicates prognosis for patients with stage II and III colorectal cancer treated with oxaliplatin-based chemotherapy. *Med Sci Monit*, 2017; 23: 5389–95
22. Liu LG, Yan XB, Xie RT et al: Stromal expression of vimentin predicts the clinical outcome of stage II colorectal cancer for high-risk patients. *Med Sci Monit*, 2017; 23: 2897–905
23. Lorenc Z, Waniczek D, Lorenc-Podgorska K et al: Profile of expression of genes encoding matrix metalloproteinase 9 (MMP9), matrix metalloproteinase 28 (MMP28) and TIMP metalloproteinase inhibitor 1 (TIMP1) in colorectal cancer: Assessment of the role in diagnosis and prognostication. *Med Sci Monit*, 2017; 23: 1305–11
24. Stanciu AE, Zamfir-Chiru-Anton A, Stanciu MM et al: Imbalance between matrix metalloproteinases and tissue inhibitors of metalloproteinases promotes invasion and metastasis of head and neck squamous cell carcinoma. *Clin Lab*, 2017; 63: 1613–20
25. Kwiatkowska E, Domanski L, Bober J et al: Urinary metalloproteinases-9 and -2 and their inhibitors TIMP-1 and TIMP-2 are markers of early and long-term graft function after renal transplantation. *Kidney Blood Press Res*, 2016; 41: 288–97
26. Zhou W, Xu S, Ying Y et al: Resveratrol suppresses growth and migration of myelodysplastic cells by inhibiting the expression of elevated cyclin D1 (CCND1). *DNA Cell Biol*, 2017; 36(11): 966–75
27. Li Y, Wei J, Xu C et al: Prognostic significance of cyclin D1 expression in colorectal cancer: A meta-analysis of observational studies. *PLoS One*, 2014; 9: e94508
28. Zhang B, Liu W, Li L et al: KAI1/CD82 and cyclin D1 as biomarkers of invasion, metastasis and prognosis of laryngeal squamous cell carcinoma. *Int J Clin Exp Pathol*, 2013; 6: 1060–67
29. Gershtein ES, Korotkova EA, Prorokov VV, Kushlinskii NE: Tumor associated proteases – prognostic markers of colorectal cancer. *Klin Lab Diagn*, 2013; 43–47, 5–10
30. Dupont DM, Madsen JB, Kristensen T et al: Biochemical properties of plasminogen activator inhibitor-1. *Front Biosci (Landmark Ed)*, 2009; 14: 1337–61
31. Chen H, Peng H, Liu W et al: Silencing of plasminogen activator inhibitor-1 suppresses colorectal cancer progression and liver metastasis. *Surgery*, 2015; 158: 1704–13
32. Shen X, Yue M, Meng F et al: Microarray analysis of differentially-expressed genes and linker genes associated with the molecular mechanism of colorectal cancer. *Oncol Lett*, 2016; 12: 3250–58
33. Rodia MT, Ugolini G, Mattei G et al: Systematic large-scale meta-analysis identifies a panel of two mRNAs as blood biomarkers for colorectal cancer detection. *Oncotarget*, 2016; 7: 30295–306
34. Minye HM, Fabritius AL, Vesa J, Peltonen L: Data on characterizing the gene expression patterns of neuronal ceroid lipofuscinosis genes: CLN1, CLN2, CLN3, CLN5 and their association to interneuron and neurotransmission markers: Parvalbumin and Somatostatin. *Data Brief*, 2016; 8: 741–49
35. Orciani M, Caffarini M, Sorgentoni G et al: Effects of somatostatin and its analogues on progenitor mesenchymal cells isolated from human pituitary adenomas. *Pituitary*, 2017; 20: 251–60
36. Modarai SR, Opdenaker LM, Viswanathan V et al: Somatostatin signaling via SSTR1 contributes to the quiescence of colon cancer stem cells. *BMC Cancer*, 2016; 16: 941
37. Leiszter K, Sipos F, Galamb O et al: Promoter hypermethylation-related reduced somatostatin production promotes uncontrolled cell proliferation in colorectal cancer. *PLoS One*, 2015; 10: e0118332
38. Nagy ZB, Bartak BK, Kalmar A et al: Comparison of circulating miRNAs expression alterations in matched tissue and plasma samples during colorectal cancer progression. *Pathol Oncol Res*, 2017 [Epub ahead of print]
39. Schneiderova M, Naccarati A, Pardini B et al: MicroRNA-binding site polymorphisms in genes involved in colorectal cancer etiopathogenesis and their impact on disease prognosis. *Mutagenesis*, 2017; 32: 533–42
40. Yuan Z, Baker K, Redman MW et al: Dynamic plasma microRNAs are biomarkers for prognosis and early detection of recurrence in colorectal cancer. *Br J Cancer*, 2017; 117: 1202–10
41. Song G, Zhang H, Chen C et al: miR-551b regulates epithelial-mesenchymal transition and metastasis of gastric cancer by inhibiting ERBB4 expression. *Oncotarget*, 2017; 8: 45725–35
42. Haldrup C, Kosaka N, Ochiya T et al: Profiling of circulating microRNAs for prostate cancer biomarker discovery. *Drug Deliv Transl Res*, 2014; 4: 19–30