Genomic Expression Profiling of Colorectal Cancer in Silico

Feifei Liu

Peking University People's Hospital

Yu Wang

Peking University People's Hospital

Wenxue Li

Peking University People's Hospital

Diancheng Li

Peking University People's Hospital

Yuwei Xin

Peking University People's Hospital

Fang Liu

Peking University People's Hospital

Jiaan Zhu (✉ zhujiaan@pkuph.edu.cn)

Peking University People's Hospital <https://orcid.org/0000-0001-8700-639X>

Research article

Keywords: colorectal cancer, bioinformatic, differentially expressed genes, GEO, protein-protein interaction

Posted Date: March 15th, 2021

DOI: <https://doi.org/10.21203/rs.3.rs-310514/v1>

License: This work is licensed under a Creative Commons Attribution 4.0 International License.

[Read Full License](#)

Abstract

Background: Colorectal cancer (CRC) is one of the most common malignancies of the digestive system; the progression and prognosis of which are affected by a complicated network of genes and pathways. The aim of this study was to identify potential hub genes associated with the progression and prognosis of colorectal cancer (CRC).

Methods: We obtained gene expression profiles from GEO database to search differentially expressed genes (DEGs) between CRC tissues and normal tissue. Subsequently, we conducted a functional enrichment analysis, generated a protein–protein interaction (PPI) network to identify the hub genes, and analyzed the expression validation of the hub genes. Kaplan–Meier plotter survival analysis tool was performed to evaluate the prognostic value of hub genes expression in CRC patients.

Results: A total of 370 samples, involving CRC and normal tissues were enrolled in this article. 283 differentially expressed genes (DEGs), including 62 upregulated genes and 221 downregulated genes between CRC and normal tissues were selected. We finally filtered out 6 hub genes, including *INSL5*, *MTIM*, *GCG*, *SPP1*, *HSD11B2*, and *MAOB*. In the database of TCGA-COAD, the mRNA expression of *INSL5*, *MT1M*, *HSD11B2*, *MAOB* in tumor is lower than that in normal; the mRNA expression of *SPP1* in tumor is higher than that in normal. In the HPA database, the expression of *INSL5*, *GCG*, *HSD11B2*, *MAOB* in tumor is lower than that in normal tissues; the expression of *SPP1* in the tumor is higher than that in normal tissues. Survival analysis revealed that *INSL5*, *GCG*, *SPP1* and *MT1M* may serve as prognostic biomarkers in CRC.

Conclusions: We screened out six hub genes to predict the occurrence and prognosis of patients with CRC using bioinformatics methods, which may provide new targets and ideas for diagnosis, prognosis and individualized treatment for CRC.

Background

Colorectal cancer (CRC) is one of the primary malignant tumors of the digestive tract[1]. It is estimated that there are approximately 1.4 million new CRC diagnoses each year and approximately 694,000 people died of CRC in 2012[2]. The historical changes in risk factors, and advances in early detection and treatment, the incidence and mortality of CRC are decreasing over the past 2 decades, but the situation is still not optimistic[3]. Although studies have found that genetic mutations, diet, inflammation, and gut microbiota can play roles in the occurrence and development of CRC, the pathogenesis of CRC remains unclear [4]. There has been a significant development of the molecular understanding of colorectal carcinogenesis in the last 30 years [5]. Increasing evidences have shown that CRC is a polygenic disease, and these genes may be oncogenes, tumor suppressor genes, mismatch repair genes, and cell cycle regulating genes [6]. In the past three decades, the *APC*, *KRAS*, and *TP53* genes have been considered to be important tumor suppressor genes or oncogenes [7]. Wnt/ β -catenin, RAS-extracellular signal-regulated kinase (ERK), p38 α -mitogen-activated protein kinase (MAPK), Hippo, and Serrated pathway may play

critical roles in the development of CRC [8–12]. These genes and pathways can serve as diagnostic biomarkers and therapeutic targets, and also be used to evaluate the clinicopathological features of CRC.

Currently, stool-based tests (eg, guaiac-based fecal occult blood testing [gFOBT], immunochemical-based fecal occult blood testing [FIT], stool DNA [sDNA] testing), endoscopy (eg, flexible sigmoidoscopy [SIG], colonoscopy), imaging (eg, double-contrast barium enema, computed tomographic colonography [CTC]) and pathological biopsy are effective screening and diagnostic methods for CRC [13, 14]. Despite strong evidence that CRC screening reduces the incidence and mortality effectively, CRC screening rates are still low worldwide. Approximately 31% of US adults had never been screened for CRC in 2018 [15], Japan's target population of CRC screenings that had been done was less than 50% in 2013, and the target population's participation rate in CRC screening was more lower in China [16]. Therefore, the treatment of CRC is still very important. In the treatment of cancer, precision medicine is the current trend [17]. With the deepening of research, doctors' therapeutic decisions are being developed from "one gene, one drug" to "multi-gene, multi-drug" [18]. Although many aberrant genes and biomarkers have already been identified in CRC, there are still many challenges to overcome in order to establish a personalized therapy, ranging from the early diagnosis and treatment to the determination of prognosis factors. Therefore, it is urgent and imperative to search for reliable biomarkers for diagnosis and therapy.

At present, bioinformatics methods have been applied as very important tools to predict cancer-related genes [19]. In our study, we screened the differentially expressed genes (DEGs) between CRC and normal tissues in five datasets from Gene Expression Omnibus (GEO, <https://www.ncbi.nlm.nih.gov/geo/>) [20]. Subsequently, Gene Ontology (GO) functional annotation analysis and Kyoto Encyclopedia of Genes and Genomes (KEGG) pathways analysis were performed for the screened DEGs. Then, we established protein–protein interaction (PPI) network of the DEGs and selected the hub genes related to CRC. Moreover, expression validation and survival analysis of these hub genes were analyzed using GEPIA platform and Kaplan–Meier plotter. This study would provide reliable molecular biomarkers for screening and diagnosis, prognosis, as well as novel therapeutic targets for CRC.

Methods

Gene expression microarray data.

High-throughput gene expression profiles of normal colorectal tissues and CRC tissues were extracted from the GEO database. The independent datasets of GSE28000 [21], GSE33113 [22], GSE41328 [23], GSE47063 [24], and GSE74602 were selected which included 86 CRC samples and 40 normal tissues, 90 CRC tissues and 6 normal tissues, 5 CRC samples and 5 normal tissues, 31 CRC samples and 47 normal tissues, and 30 CRC samples and 30 normal tissues, respectively.

Data Processing of DEGs.

The online analysis tool – GEO2R (<https://www.ncbi.nlm.nih.gov/geo/geo2r/>) was used to search DEGs between CRC samples and normal samples. The adjusted *p* values were utilized to decrease the false

positive rate using Benjamini and Hoch-berg false discovery rate method by default. The adjust p value < 0.01 and $|\log FC| \geq 1$ were adjusted as the cutoff criterion.

Analysis of enriched GO terms and KEGG pathways.

GO analysis is commonly performed for functional enrichment analysis; gene function can be classified into biological process (BP), molecular function (MF) and cellular component (CC) terms. KEGG is widely used for exploring the advanced functions and mechanisms involved in the biological system at the molecular level. In this study, GO terms and KEGG pathways analysis was performed with the Database for Annotation, Visualization and Integrated Discovery (DAVID) platform (v6.8; <https://david.ncifcrf.gov/>) [25]. $P < 0.05$ and gene counts ≥ 10 were considered statistically significant.

Construction of PPI network and identification of module's seeds.

The Search Tool for the Retrieval of Interacting Genes (STRING) database (<http://string-db.org/>) [26] was performed for PPI network construction. The minimum required interaction score was set to 0.4 as default. Then, Cytoscape software v3.7.1 (<http://www.cytoscape.org/index.html>) [27] was used for the analysis and visualization of the PPI network. In addition, the remarkable module of PPI network was selected via Molecular Complex Detection(MCODE) app (node score cutoff = 0.2, degree cutoff = 2, max. depth = 100, and k-score = 2). Nodes with higher degree of connectivity tend to be more essential in maintaining the stability of the entire network. The CytoHubba plugin was used to calculate the degree of each protein node. In our study, hub genes were selected based on a node degree ≥ 4 .

Expression and survival analysis of hub genes

We checked the expression of hub genes corresponding to key proteins in TCGA database (NIH) through GEPIA platform (<http://gepia.cancer-pku.cn/>), which is the largest oncogene chip database currently in the world, constituting a conveniently integrated data mining platform[28]. The expression of hub genes was compared between CRC tissues and normal tissues in HPA database. Student's unpaired t-test was used in determining the statistical significance of the calculated differential expression. The fold change was defined as 2 and $P < 0.01$ was considered statistically significant.

To investigate the potential prognostic role of hub genes, Kaplan-Meier (KM) survival curves was performed to illuminate correlations between expression of hub genes and the overall survival of CRC patients in the TCGA database. $P < 0.01$ was considered statistically significant.

Results

Identification of DEGs

In the present study, we analyzed gene expression datasets of 5 datasets and extracted DEGs on the basis of the cut-off criteria. The present study covered 242 CRC and 128 normal tissues in total. The

results showed that 1573, 3081, 1431, 2505, and 1675 DEGs between CRC and normal tissues respectively. There were 532 (1041) (Fig. 1A), 857 (2224) (Fig. 1B), 649 (782) (Fig. 1C), 1185 (1320) (Fig. 1D), 674 (1001) (Fig. 1E) up- (down-) regulated genes among these DEGs (Table 1). There were a total of 62 (221) overlapping up- (down-) regulated genes in these datasets, and the number of overlapping DEGs is 283 (Fig. 1F and G).

Table 1
Up- and downregulated differentially expressed genes in GSE datasets

Datasets ID	Up-regulated gene number	Down-regulated gene number	Total number
GSE28000	532	1041	1573
GSE33113	857	2224	3081
GSE41328	649	782	1431
GSE47063	1185	1320	2505
GSE74602	674	1001	1675

Enriched GO terms and KEGG pathway analysis of the overlapping DEGs

85 enriched GO terms and 15 KEGG pathways were obtained from the overlapping DEGs using DAVID. The numbers of biological process (BP), cellular component (CC) and molecular function (MF) among these enriched GO terms are 49, 16, and 20, respectively. Table 2 showed the enriched GO terms of the overlapping DEGs based on the counts of genes ≥ 10 . These terms were integral component of membrane, plasma membrane, extracellular exosome, integral component of plasma membrane, extracellular space, extracellular region, zinc ion binding, endoplasmic reticulum, calcium ion binding, proteolysis, cell surface, cell proliferation, negative regulation of cell proliferation, apical plasma membrane, proteinaceous extracellular matrix, receptor binding, ion transmembrane transport, and actin binding.

Table 2
Enriched GO terms of the overlapping DEGs based on the counts of genes ≥ 10 .

Category	Term	Count	Ratio (%)	P Value	Genes
CC	GO:0016021 integral component of membrane	101	35.69	0.00	AQP8, TSPAN7, LPAR1, VIPR1, GPAT3, AGTR1, ADTRP, FAP, CD46, SLC25A23, SLC51B, SLC51A, SLC4A4, HIGD1A, PDPN, SLC04A1, CA12, SMIM14, SLC22A23, TNFRSF17, METTL7A, CWH43, BACE2, HHLA2, HSD11B2, SLC30A10, DCUN1D5, EMP1, FXYD1, ABCA8, CAV1, GCNT3, GCNT2, PAQR8, PPAT, BCHE, NDRG2, HRASLS2, MAOA, MAOB, MET, SLC6A14, HILPDA, ABCG2, SEMA6A, C1ORF115, SEMA6D, DIO2, SFRP1, ATP2A3, MS4A12, NXPE1, CLDN8, LRRC19, ANPEP, SLC26A2, FRMD3, ST6GALNAC6, DSEL, CEACAM7, TMEM100, SLC22A5, CEACAM1, TRPM6, PIGZ, PTGER4, BTNL3, HEPACAM2, CLDN1, LRMP, CA4, HSD17B2, ABHD3, GPM6B, TMEM35A, PLPP1, ITM2C, ITM2A, TMEM37, FMO5, ETFDH, ETNK1, UGT2A3, SCNN1B, CNNM4, SLC12A2, OLR1, SI, NPY1R, ATP1A2, CYP4F12, GDPD3, CLDN23, Tmprss3, RNF43, SLC17A4, SLC16A9, CDH19, MEP1A, MEP1B, LRP8
CC	GO:0005886 plasma membrane	84	29.68	0.00	CLDN8, AQP8, GNA11, LPAR1, VIPR1, SLC26A2, AZGP1, ST6GALNAC6, AGTR1, ADTRP, ANK3, CD46, FAP, TGFBI, CEACAM7, MGLL, SLC22A5, TMEM100, SLC51B, SLC51A, FAM129A, SLC4A4, CEACAM1, AHNK, GNG7, SGK1, TRPM6, CLCA1, PTGER4, CLCA4, SLC04A1, PDPN, CA12, HTR4, CHP2, TNFRSF17, ARRDC4, PIGR, EPB41L3, SLC26A3, PLCE1, CEMIP, CA4, CA2, SLC30A10, EMP1, FXYD1, ABCA8, CAV1, FHL1, TRIB3, ABHD3, PAQR8, GPM6B, C2ORF88, PLPP1, ITM2C, ITM2A, PDE6A, CD177, ETNK1, PLCD1, SCNN1B, AXIN2, TESC, OLR1, SLC12A2, SI, MET, SLC6A14, ITGA2, NPY1R, ATP1A2, CLDN23, ABCG2, GCG, RNF43, SEMA6D, DIO2, SFRP1, SLC16A9, CDH19, MEP1B, LRP8

Category	Term	Count	Ratio (%)	P Value	Genes
CC	GO:0070062 extracellular exosome	77	27.21	0.00	NAAA, GNA11, PTGS1, SELENBP1, ANPEP, CXCL12, SLC26A2, CMBL, CKB, AZGP1, ACTG2, HSPH1, SMPDL3A, CD46, TGFB1, GUCA2A, SLC22A5, FAM129A, SLC4A4, GUCA2B, CEACAM1, AHNAK, GNG7, MB, ACAA2, CLCA4, PIGR, METTL7A, NEBL, SERPINB5, AKR1B10, PGM1, SCIN, CA4, CA2, CA1, UGP2, C7, GCNT3, UGDH, PLPP1, ITM2C, ITM2A, KRT24, TIMP1, PBLD, SIAE, GLIPR2, HSPA2, CD177, ENTPD5, PLCD1, NDRG2, SCNN1B, DPT, SPP1, CPNE8, BCAS1, CES2, OLR1, SLC12A2, SI, MAOB, SLC6A14, EPHX2, S100A11, GDDP3, FUCA1, APPL2, PCK1, TST, SFRP1, MEP1A, NXPE4, FABP1, PSAT1, GDF15
CC	GO:0005887 integral component of plasma membrane	42	14.84	0.00	FXD1, CAV1, AQP8, TSPAN7, ANPEP, LPAR1, PLPP1, VIPR1, SLC26A2, BEST2, AGTR1, CD46, SLC22A5, SCNN1B, SLC4A4, CEACAM1, CLCA1, OLR1, SLC12A2, CLCA4, PTGER4, SLC04A1, PDPN, PTPRH, SLC22A23, MET, SLC6A14, HTR4, PIGR, NPY1R, SLC26A3, SEMA6A, RNF43, VSIG2, SEMA6D, SLC17A4, ATP2A3, CLDN1, MEP1A, LRMP, MEP1B, SCARA5
CC	GO:0005615 extracellular space	40	14.13	0.00	CXCL1, EDN3, CTHRC1, CXCL2, CXCL8, ANPEP, SELENBP1, MMP3, CXCL12, GREM2, ABI3BP, TIMP1, CKB, ACTG2, AZGP1, SIAE, SMPDL3A, FAP, TGFB1, SPP1, DPT, CES2, BMP2, CLCA1, S100A11, HILPDA, PIGR, C2ORF40, GCG, TST, LAMA1, SFRP1, SRPX2, SERPINB5, MEP1A, MEP1B, PYY, CA2, GDF15, SST
CC	GO:0005576 extracellular region	38	13.43	0.01	CXCL1, VIP, C7, EDN3, INSL5, CXCL2, CXCL8, MMP3, CXCL12, GREM2, DNASE1L3, MMP1, TIMP1, HSPH1, AZGP1, GLIPR2, BCHE, TGFB1, GUCA2A, FAM150B, NMU, SPP1, BMP2, DHRS11, MET, MMP12, MMP11, GCG, LAMA1, SCGN, SFRP1, CEMIP, LRP8, PYY, GDF15, SST, CHGB, NXPE1
MF	GO:0008270 zinc ion binding	27	9.54	0.03	FHL1, NR3C2, ADH1C, ADH1B, ANPEP, ADH1A, MMP3, ZEF1, MMP1, TCEA3, SMPDL3A, MT1E, MT1G, MT1F, MT1M, CA12, LDB3, MMP12, NEBL, MMP11, RNF43, MEP1A, MEP1B, CA4, CA2, CA1, KLF4
CC	GO:0005783 endoplasmic reticulum	20	7.07	0.05	CLDN8, CES2, CAV1, PIGZ, MAMDC2, SMIM14, METTL7A, GPAT3, DNASE1L3, TMPRSS3, FMO5, SRPX, ANK3, ATP2A3, BACE2, CEMIP, ENTPD5, PDE9A, TMEM100, AHCYL2

Category	Term	Count	Ratio (%)	P Value	Genes
MF	GO:0005509 calcium ion binding	19	6.71	0.03	TESC, PPP2R3A, NKD2, CHP2, S100A11, MMP3, DNASE1L3, ZZEF1, MMP12, MMP1, MMP11, SCGN, SLC25A23, VSNL1, CDH19, SCIN, LRP8, PLCD1, CASQ2
BP	GO:0006508 proteolysis	16	5.65	0.01	CLCA1, OLR1, CLCA4, ANPEP, MMP3, TMPRSS3, MMP12, MMP1, MMP11, SFRP1, CD46, BACE2, FAP, MEP1A, MEP1B, LRP8
CC	GO:0009986 cell surface	16	5.65	0.02	BMP2, MET, ITGA2, HILPDA, LPAR1, SRPX, HSPA2, SRPX2, SFRP1, ANK3, CD46, BACE2, FAP, CA4, SCARA5, CEACAM1
BP	GO:0008283 cell proliferation	15	5.30	0.00	CXCL1, GCG, PLCE1, PDPN, TGFBI, MET, ENTPD5, TNFRSF17, ITGA2, DACH1, PYY, AXIN2, APPL2, EMP1, CITED2
BP	GO:0008285 negative regulation of cell proliferation	15	5.30	0.00	CXCL1, TESC, BMP2, S100A11, CXCL8, SOX4, PLPP1, AZGP1, SFRP1, BCHE, SCIN, AXIN2, SST, KLF4, DPT
CC	GO:0016324 apical plasma membrane	14	4.95	0.00	TRPM6, CAV1, CLCA4, SLC12A2, SI, CYP4F12, SLC26A3, SORBS2, SLC17A4, CLDN1, CA4, SLC22A5, SCNN1B, CEACAM1
CC	GO:0005578 proteinaceous extracellular matrix	12	4.24	0.00	CTHRC1, LAMA1, SFRP1, MAMDC2, ZG16, TGFBI, MMP3, MMP12, MMP1, MMP11, TIMP1, DPT
MF	GO:0005102 receptor binding	11	3.89	0.04	CXCL1, GCG, EDN3, LAMA1, CAV1, BMP2, SRPX2, EPHX2, HILPDA, CXCL12, NMU
BP	GO:0034220 ion transmembrane transport	10	3.53	0.01	FXYD1, SGK1, CLCA1, SGK2, CLCA4, ATP2A3, AQP8, ATP1A2, SCNN1B, CASQ2
MF	GO:0003779 actin binding	10	3.53	0.03	EPB41L3, CGN, ENC1, SCIN, SNTB1, ANLN, LMOD1, CNN1, CEACAM1, MYOT

According to gene counts, Table 3 showed the most significantly enriched KEGG pathways of the overlapping DEGs. These DEGs were enriched in metabolic pathways, pancreatic secretion, mineral absorption, drug metabolism-cytochrome P450, bile secretion, proximal tubule bicarbonate reclamation, tyrosine metabolism, aldosterone-regulated sodium reabsorption, fatty acid degradation, retinol metabolism, glycolysis / gluconeogenesis, nitrogen metabolism, galactose metabolism, starch and sucrose metabolism, and pentose and glucuronate interconversions.

Table 3
The most significantly enriched KEGG pathways of the overlapping DEGs.

Term	Count	Ratio (%)	P Value	Genes
hsa01100:Metabolic pathways	35	12.36749	0.02	GCNT3, GCNT2, HSD17B2, PTGS1, ADH1C, UGDH, ADH1B, ADH1A, ANPEP, PLPP1, GPAT3, PPAT, CMBL, CKB, ST6GALNAC6, ETNK1, MGLL, PLCD1, UGT2A3, PAPSS2, ACAA2, ACADS, MAOA, MAOB, SI, EPHX2, PCK1, GBA3, TST, PLCE1, AKR1B10, PGM1, AHCYL2, PSAT1, UGP2
hsa04972:Pancreatic secretion	8	2.826855	0.00	SLC26A3, CLCA1, SLC12A2, CLCA4, ATP2A3, ATP1A2, CA2, SLC4A4
hsa04978:Mineral absorption	7	2.473498	0.00	SLC26A3, TRPM6, MT1M, MT1E, ATP1A2, MT1G, MT1F
hsa00982:Drug metabolism - cytochrome P450	7	2.473498	0.00	FM05, MAOA, MAOB, ADH1C, ADH1B, ADH1A, UGT2A3
hsa04976:Bile secretion	7	2.473498	0.00	AQP8, SLC51B, ATP1A2, CA2, SLC51A, SLC4A4, ABCG2
hsa04964:Proximal tubule bicarbonate reclamation	5	1.766784	0.00	CA4, ATP1A2, CA2, SLC4A4, PCK1
hsa00350:Tyrosine metabolism	5	1.766784	0.00	MAOA, MAOB, ADH1C, ADH1B, ADH1A
hsa04960:Aldosterone-regulated sodium reabsorption	5	1.766784	0.01	SGK1, NR3C2, HSD11B2, ATP1A2, SCNN1B
hsa00071:Fatty acid degradation	5	1.766784	0.01	ACAA2, ACADS, ADH1C, ADH1B, ADH1A
hsa00830:Retinol metabolism	5	1.766784	0.04	ADH1C, ADH1B, ADH1A, UGT2A3, RETSAT
hsa00010:Glycolysis / Gluconeogenesis	5	1.766784	0.04	PGM1, ADH1C, ADH1B, ADH1A, PCK1
hsa00910:Nitrogen metabolism	4	1.413428	0.00	CA12, CA4, CA2, CA1
hsa00052:Galactose metabolism	4	1.413428	0.02	AKR1B10, SI, PGM1, UGP2
hsa00500:Starch and sucrose metabolism	4	1.413428	0.03	GBA3, SI, PGM1, UGP2
hsa00040:Pentose and glucuronate interconversions	4	1.413428	0.03	AKR1B10, UGDH, UGT2A3, UGP2

PPI network analysis and module identification

Protein interactions among the DEGs were performed using STRING tools. A total of 166 nodes and 322 edges were identified (Fig. 2). These nodes and edges represented the overlapping DEGs involved in the PPI and the interaction between these DEGs, respectively. Table 4 displayed the degrees of all nodes. We obtained 11 Modules and 10 seeds in the PPI network based on MCODE app (Table 5). The modules with a score of at least 4 were selected as hub genes (Fig. 3). These modules might play important roles in CRC.

Table 4
Degree value of all nodes.

Node name	Degree	Node name	Degree	Node name	Degree	Node name	Degree
IL8	23	HSPH1	4	CNN1	2	CLCA1	1
PHLPP2	22	MAOB	4	DACH1	2	CLCA4	1
GNG7	19	PDE6A	4	DIO2	2	CMBL	1
GCG	18	PGM1	4	ENTPD5	2	CTHRC1	1
CXCL12	16	SCNN1B	4	FABP1	2	DPT	1
SST	16	SGK2	4	FXYD1	2	FHL1	1
ITGA2	15	ACAA2	3	KLF4	2	FOSB	1
NMU	15	ACADS	3	MGLL	2	FUCA1	1
AGTR1	14	AXIN2	3	MMP11	2	GCNT3	1
LPAR1	14	BCHE	3	PCK1	2	GDF15	1
PYY	14	CEACAM1	3	PDK4	2	GLIPR2	1
CXCL1	13	CITED2	3	PLCD1	2	GREM2	1
SPP1	13	CLDN1	3	PLCE1	2	GTF3A	1
GNA11	11	CLDN8	3	PLCL2	2	HSD17B2	1
NPY1R	11	ETV4	3	POU2AF1	2	HSPB8	1
CXCL2	10	FAP	3	PPAP2A	2	LAMA1	1
INSL5	10	LDB3	3	SCGN	2	MEP1A	1
MMP1	10	LMOD1	3	SFRP1	2	MEP1B	1
BMP2	9	LRRC19	3	SLC16A9	2	MYOT	1
TIMP1	9	MB	3	SULT1A2	2	NAAA	1
ACTG2	8	MT1E	3	TESC	2	NDRG2	1
CAV1	8	MT1F	3	TNFRSF17	2	NXPE1	1
VIP	8	MT1G	3	TUBAL3	2	NXPE4	1
ADH1B	7	MT1M	3	UGP2	2	PEX26	1
CA2	7	OLR1	3	VSNL1	2	PIGR	1
MMP3	7	PDE9A	3	ZNF575	2	PTPRH	1
PTGER4	7	PTGS1	3	ABCA8	1	RAD54B	1

Node name	Degree	Node name	Degree	Node name	Degree	Node name	Degree
SGK1	7	RERGL	3	AHCYL2	1	S100A11	1
ADH1A	6	RETSAT	3	AKR1B10	1	SCIN	1
EDN3	6	SDPR	3	ANK3	1	SELENBP1	1
HTR4	6	SLC26A3	3	ANLN	1	SI	1
SLC4A4	6	SOX9	3	ANPEP	1	SIAE	1
HSPA2	5	TRIB3	3	AQP8	1	SLC12A2	1
MAOA	5	UGDH	3	ARRDC4	1	SLC17A4	1
MET	5	AGPAT9	2	ATP1A2	1	SLC22A5	1
NR3C2	5	ATP2A3	2	BEST2	1	SLC26A2	1
PAPSS2	5	CA1	2	CA4	1	SLC30A10	1
UGT2A3	5	CASQ2	2	CD177	1	SLC51A	1
VIPR1	5	CGN	2	CD46	1	SLC51B	1
ABCG2	4	CHP2	2	CES2	1	SPIB	1
ETFDH	4	CKB	2	CHGB	1		
HSD11B2	4	CLDN23	2	CKAP2	1		

Table 5
11 Modules and 10 seeds in the PPI network based on MCODE app.

Cluster	Score	Nodes	Edges	Node IDs
1	11	11	55	CXCL2, INSL5 , SST, GNG7, PYY, NPY1R, CXCL12, NMU, IL8, LPAR1, CXCL1
2	5	5	10	GCG , VIPR1, VIP, PTGER4, HTR4
3	4	4	6	SPP1 , MMP3, TIMP1, MMP1
4	4	4	6	MT1M , MT1F, MT1E, MT1G
5	4	4	6	HSD11B2 , NR3C2, SCNN1B, SGK1
6	4	4	6	ADH1A, MAOB , ADH1B, MAOA
7	3.33	4	5	CLDN8, CLDN1, CLDN23, CGN
8	3	3	3	AGTR1, EDN3, GNA11
9	3	3	3	HSPA2, HSPH1, PAPSS2
10	3	3	3	ETFDH, ACADS, ACAA2
11	2.67	4	4	CHP2, PHLPP2, TESC, ATP2A3

Expression and survival analysis of the 6 selected hub genes

GEPIA was utilized to compare the expression levels of the hub genes between CRC tumor samples and adjacent normal samples from the TCGA database. We selected the seeds of the modules with a score of at least 4 (*INSL5*, *GCG*, *SPP1*, *HSD11B2*, *MT1M* and *MAOB*) for further analysis. In the database of TCGA-COAD, the mRNA expression of *INSL5* in tumor is lower than that in normal (Fig. 4B, $p < 0.05$); the mRNA expression of *SPP1* in tumor is higher than that in normal (Fig. 4C, $p < 0.05$); the mRNA expression of *MT1M* in tumor is lower than that in normal (Fig. 4D, $p < 0.05$); the mRNA expression of *HSD11B2* in tumor is lower than that in normal (Fig. 4E, $p < 0.05$); and the mRNA expression of *MAOB* in tumor is lower than that in normal (Fig. 4F, $p < 0.05$).

In the HPA database, the expression of *INSL5* in tumor is lower than that in normal (Fig. 5A, B); the expression of *GCG* in tumor is lower than that in normal (Fig. 5C and D), the expression of *SPP1* in the tumor is higher than that in normal (Fig. 5E and F); the expression of *HSD11B2* in tumor is lower than that in normal (Fig. 5G and H); and the expression of *MAOB* in tumor is lower than that in normal (Fig. 5I and J). Unfortunately, there was no *MT1M* related immunohistochemistry data of patients with or without CRC in the HPA database.

To investigate the prognostic values of the six potential hub genes, we used Kaplan–Meier plotter bioinformatics analysis platform. Through survival analysis in the HPA database, high level expression of

INSL5 (Fig. 6A), GCG (Fig. 6B), and MT1M (Fig. 6D) and low level expression of SPP1 (Fig. 6C) were associated with better prognosis patients with CRC ($P < 0.05$). However, the expression of HSD11B2 (Fig. 6E) and MAOB (Fig. 6F) were not significantly associated with prognosis in patients with CRC ($P > 0.05$).

Discussion

CRC is the second largest malignant tumor of the digestive system, and the third most lethal malignancy in the world [29]. The 5-year survival rate of patients with CRC at its early stage is 90%, while that of patients at later stages was less than 12% [30] [31]. Currently, the treatment of colorectal cancer is still mainly surgical resection, chemotherapy and physical therapy. However, due to the insufficient diagnostic methods for early-onset CRC, the vast majority of patients have been diagnosed at a late stage; and due to gene mutation and changes in immune microenvironment in the process of tumor development, chemotherapy cannot achieve targeted efficacy, resulting in a high mortality rate [32]. Therefore, it is of great significance to identify the key marker genes and to understand the molecular mechanisms of CRC for its early diagnosis, effective treatment strategy development and prognosis evaluation. As an important discipline of biological science, integrated bioinformatics analysis can easily access gene expression profile information, which is quite useful for exploring various molecular mechanisms [32], and is considered as an efficient method for systematically predicting disease-related genes.

In the present study, we screened the common differentially expressed genes (DEGs) in 5 GSE datasets via GEO2R, and obtained a total of 62 (221) overlapping up- (down-) regulated genes in these datasets, and 283 overlapping DEGs. To obtain an in-depth understanding of these DEGs, we performed the GO function and KEGG pathway analysis of these DEGs. The results of GO and KEGG pathway enrichment analysis suggested that these DEGs were significantly enriched in different cancer-related functions and pathways. 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. After the above series of operations, 6 hub genes, including INSL5, MT1M, GCG, SPP1, HSD11B2, and MAOB were screened out.

In addition, we tested the expression of these 6 genes at the mRNA as well as protein levels and found that the mRNA expression of INSL5, MT1M, HSD11B2, MAOB in CRC tissues were significantly lower than that in normal tissues, while SPP1 in tumor is higher than that in normal. Through survival analysis in the HPA database, the high level expression of INSL5, GCG, and MT1M and the low level expression of SPP1 were associated with better prognosis patients with CRC.

INSL5 (insulin-like 5) was a member of the insulin/relaxin superfamily. It can activate the G-protein-coupled receptor relaxin/insulin-like family peptide receptor 4 (RXFP4). Mashima et al [33] have shown that INSL5 is expressed in enteroendocrine cells (EECs) along the colorectum with a gradient increase toward the rectum. INSL5-RXFP4 signaling may play a role in an autocrine/paracrine fashion in the

colorectum with effects other than those on cell proliferation, inflammation and mucosal healing, and INSL5 might represent a unique marker of colorectal EECs and NETs.

Recently, Shi-Bing Li's work found INSL5 was elevated in nasopharyngeal carcinoma (NPC) tumor tissue and the plasma of NPC patients[34]. INSL5 acts on its own receptors through an autocrine pathway to promote tumor growth. Mechanistically, INSL5 enhanced phosphorylation and nuclear translocation of STAT5 and promoted glycolytic gene expression, leading to induced glycolysis in cancer cells.

Monoamine oxidases (MAOs) including MAOA and MAOB are enzymes located on the outer membranes of mitochondria, which are responsible for catalyzing monoamine oxidation. Recently, increased level of MAOs was shown in several cancer types. Elevation of MAOB activity in cells resulted in induction of oxidative stress in association with mitochondrial dysfunction. Yi-Chieh Yang et al [35] found that MAOB is highly expressed in CRC tissues compared to normal colorectal tissues, and its overexpression was significantly associated with a higher recurrence rate and poorer prognosis. However, in this study, we found that the expression of MAOB were not significantly associated with prognosis, which might be owing to the different samples used for the research; thus, this point deserves further investigation.

As a major member of the metallothionein family, MT1M is a cysteine-rich metal-binding protein that is induced by inflammation and protects the cell against carcinogens. Emerging evidence has shown that MT1M is related to pancreatic ductal adenocarcinoma [36], breast cancer [37], renal cell carcinoma [38] and colorectal carcinoma [39]. In colorectal cancer, MT1M was a highly expressed isoform in tumor epithelia. It is reported that low MT1M expression is an independent prognosticator for relatively poor survival of patients. Heya Na's study first found that MT1M regulated the expression of lncRNA STEAP3-AS1 and that high levels of STEAP3-AS1 were associated with poor overall survival in colon cancer patients [40].

The protein encoded by the glucagon (GCG) gene is a preproprotein which is cleaved into four distinct mature peptides. These peptides are involved in maintaining nutrient homeostasis, and are regulators of cell proliferation, differentiation, and apoptosis [41]. GCG, a remarkable decrease was found in the gene expression already in the adenoma stage. Previous study showed that activity of GCG possibly interferes with tumor progression[42]. The current study showed that GCG expression was significantly down regulated in CRC tissues. Therefore, we inferred that GCG might play a role in suppressing the development of CRC. Further molecular studies are necessary to elucidate the regulatory mechanisms of GCG in CRC progression.

Type-2 11 β -hydroxysteroid dehydrogenase (HSD11B2) is a key enzyme which converts cortisol to inactive cortisone and is involved in tumor progression and metastasis. Several studies have shown that the promotion of tumor progression and metastasis by HSD11B2 resulted from its physiological function of inactivating glucocorticoids (GC). Lu Qi et al[43] found that the expression of the HSD11B2 was markedly decreased in colorectal cancer tissues. Jin Chen et al[44] performed a series of in vivo and in vitro assays, and found the ectopic expression of HSD11B2 significantly promoted the migration, invasion and

metastasis of colorectal cancer (CRC) cells via the Fgfbp1-AKT pathway. Therefore, targeting HSD11B2 may be a novel treatment strategy for inhibiting the metastasis of CRC.

SPP1, also known as osteopontin, is a phosphoprotein secreted by malignant cells. SPP1 is involved in multiple physiological and pathological processes. Recent studies have reported that SPP1 is significantly associated with cell growth, adherence and invasion in tumourigenesis and metastasis, and is over-expressed in breast, colorectal, liver, and lung cancers[45–47]. In addition, high levels of SPP1 gene expression were associated with a poor prognosis for these cancers. Cheng et al. [47] found that the expression level of SPP1 in CRC tissues was higher than that in normal tissue, which was closely correlated to tumor TNM stage and aggressiveness. A study by Choe et al [48] confirmed the higher SPP1 expression along with shorter survival times in CRC patients.

While providing insight into the prediction of cancer-related genes and biomarkers of CRC, our study still had several limitations. Firstly, bioinformatics was used to analyze the related genes and biomarkers of CRC, but there was no further validation of its function and mechanism by clinical samples or cytological experiments, therefore, further experiments in cell, animal models and clinical samples are required to validate the present results, and these critical issues will be further investigated in our future work. Secondly, the disparity between platforms and analytical tools and algorithms poses challenge for bioinformatics analysis, as results for the same data using different bioinformatics tools may vary.

Conclusions

In summary, our integrated bioinformatics analysis screened out a total of 283 DEGs and 6 hub genes, which might play crucial roles in the occurrence, development, and prognosis of CRC. In order to get more accurate correlation results, further experiments are needed to validate the results of the prediction. Anyway, this study provided new diagnostic biomarkers and therapeutic targets, as well as promising prognostic indicator for CRC patients.

Abbreviations

CRC: colorectal cancer; INSL5: insulin-like 5; GCG: glucagon; SPP1: secreted phosphoprotein 1; MT1M: metallothionein 1; HSD11B2: type-2 11 β -hydroxysteroid dehydrogenase; MAOB: Monoamine oxidases B; DEGs: Differentially expressed genes; KEGG: Kyoto Encyclopedia of Genes and Genomes; GO: gene ontology; DAVID: Database for Annotation, Visualization and Integrated Discovery; PPI: protein–protein interaction; MCODE: Molecular Complex Detection; TCGA: The Cancer Genome Atlas; GEPIA: Gene Expression Profiling Interactive Analysis; GO: Gene Ontology.

Declarations

Acknowledgments

Not applicable.

Funding

Not applicable.

Authors' contributions

JAZ participated in conceiving and designing the study. FFL and YW drafted the manuscript and analyzed the data. DCL and YWX handled the picture and article format. WXL and FL reviewed the data. All authors have read and approved the manuscript.

Ethics approval and consent to participate

Ethical approval for this study was obtained from the Ethics Committee of Peking University People's Hospital.

Availability of data and materials

The datasets used and/or analyzed during the current study are available from GEO database and TCGA database.

Consent for publication

Not applicable.

Competing interests

The authors declare that they have no competing interests.

References

1. Siegel RL, Miller KD, Jemal A: Cancer Statistics, 2017. *CA: a cancer journal for clinicians* 2017, 67(1):7-30.
2. Jenkins MA, Win AK, Templeton AS, Angelakos MS, Buchanan DD, Cotterchio M, Figueiredo JC, Thibodeau SN, Baron JA, Potter JD *et al*: Cohort Profile: The Colon Cancer Family Registry Cohort (CCFRC). *International Journal of Epidemiology* 2018, 47(2):387-388i.
3. Siegel RL, Miller KD, Fedewa SA, Ahnen DJ, Meester RGS, Barzi A, Jemal A: Colorectal cancer statistics, 2017. *CA: a cancer journal for clinicians* 2017, 67(3):177-193.
4. Li SKH, Martin A: Mismatch Repair and Colon Cancer: Mechanisms and Therapies Explored. *Trends in molecular medicine* 2016, 22(4):274-289.
5. Obuch JC, Ahnen DJ: Colorectal Cancer: Genetics is Changing Everything. *Gastroenterology clinics of North America* 2016, 45(3):459-476.
6. Nikolouzakis TK, Vassilopoulou L, Fragkiadaki P, Mariolis Sapsakos T, Papadakis GZ, Spandidos DA, Tsatsakis AM, Tsiaoussis J: Improving diagnosis, prognosis and prediction by using biomarkers in

- CRC patients (Review). *Oncology reports* 2018, 39(6):2455-2472.
7. Fearon ER: Molecular genetics of colorectal cancer. *Annual review of pathology* 2011, 6:479-507.
 8. Bahrami A, Amerizadeh F, ShahidSales S, Khazaei M, Ghayour-Mobarhan M, Sadeghnia HR, Maftouh M, Hassanian SM, Avan A: Therapeutic Potential of Targeting Wnt/beta-Catenin Pathway in Treatment of Colorectal Cancer: Rational and Progress. *Journal of cellular biochemistry* 2017, 118(8):1979-1983.
 9. Jeong WJ, Ro EJ, Choi KY: Interaction between Wnt/beta-catenin and RAS-ERK pathways and an anti-cancer strategy via degradations of beta-catenin and RAS by targeting the Wnt/beta-catenin pathway. *NPJ precision oncology* 2018, 2(1):5.
 10. Grossi V, Peserico A, Tezil T, Simone C: p38alpha MAPK pathway: a key factor in colorectal cancer therapy and chemoresistance. *World journal of gastroenterology* 2014, 20(29):9744-9758.
 11. Wierzbicki PM, Rybarczyk A: The Hippo pathway in colorectal cancer. *Folia histochemica et cytobiologica* 2015, 53(2):105-119.
 12. Yamane L, Scapulatempo-Neto C, Reis RM, Guimaraes DP: Serrated pathway in colorectal carcinogenesis. *World journal of gastroenterology* 2014, 20(10):2634-2640.
 13. Lin JS, Piper MA, Perdue LA, Rutter CM, Webber EM, O'Connor E, Smith N, Whitlock EP: Screening for Colorectal Cancer: Updated Evidence Report and Systematic Review for the US Preventive Services Task Force. *Jama* 2016, 315(23):2576-2594.
 14. Bressenot A, Cahn V, Danese S, Peyrin-Biroulet L: Microscopic features of colorectal neoplasia in inflammatory bowel diseases. *World journal of gastroenterology* 2014, 20(12):3164-3172.
 15. Joseph DA, King JB, Dowling NF, Thomas CC, Richardson LC: Vital Signs: Colorectal Cancer Screening Test Use - United States, 2018. *MMWR Morbidity and mortality weekly report* 2020, 69(10):253-259.
 16. Sano Y, Byeon JS, Li XB, Wong MC, Chiu HM, Rerknimitr R, Utsumi T, Hattori S, Sano W, Iwatate M *et al*: Colorectal cancer screening of the general population in East Asia. *Digestive endoscopy : official journal of the Japan Gastroenterological Endoscopy Society* 2016, 28(3):243-249.
 17. Friedman AA, Letai A, Fisher DE, Flaherty KT: Precision medicine for cancer with next-generation functional diagnostics. *Nature reviews Cancer* 2015, 15(12):747-756.
 18. Dienstmann R, Vermeulen L, Guinney J, Kopetz S, Tejpar S, Tabernero J: Consensus molecular subtypes and the evolution of precision medicine in colorectal cancer. *Nature reviews Cancer* 2017, 17(2):79-92.
 19. Fang L, Wang Y, Chen X: Gene expression profiling and bioinformatics analysis of hereditary gingival fibromatosis. *Biomedical reports* 2018, 8(2):133-137.
 20. Barrett T, Wilhite SE, Ledoux P, Evangelista C, Kim IF, Tomashevsky M, Marshall KA, Phillippy KH, Sherman PM, Holko M *et al*: NCBI GEO: archive for functional genomics data sets—update. *Nucleic acids research* 2013, 41(Database issue):D991-995.

21. Jovov B, Araujo-Perez F, Sigel CS, Stratford JK, McCoy AN, Yeh JJ, Keku T: Differential gene expression between African American and European American colorectal cancer patients. *PloS one* 2012, 7(1):e30168.
22. de Sousa EMF, Colak S, Buikhuisen J, Koster J, Cameron K, de Jong JH, Tuynman JB, Prasetyanti PR, Fessler E, van den Bergh SP *et al*: Methylation of cancer-stem-cell-associated Wnt target genes predicts poor prognosis in colorectal cancer patients. *Cell stem cell* 2011, 9(5):476-485.
23. Lin G, He X, Ji H, Shi L, Davis RW, Zhong S: Reproducibility Probability Score—incorporating measurement variability across laboratories for gene selection. *Nature biotechnology* 2006, 24(12):1476-1477.
24. Uribe-Lewis S, Stark R, Carroll T, Dunning MJ, Bachman M, Ito Y, Stojic L, Halim S, Vowler SL, Lynch AG *et al*: 5-hydroxymethylcytosine marks promoters in colon that resist DNA hypermethylation in cancer. *Genome biology* 2015, 16(1):69.
25. Huang da W, Sherman BT, Lempicki RA: Systematic and integrative analysis of large gene lists using DAVID bioinformatics resources. *Nature protocols* 2009, 4(1):44-57.
26. Szklarczyk D, Morris JH, Cook H, Kuhn M, Wyder S, Simonovic M, Santos A, Doncheva NT, Roth A, Bork P *et al*: The STRING database in 2017: quality-controlled protein-protein association networks, made broadly accessible. *Nucleic acids research* 2017, 45(D1):D362-d368.
27. Kohl M, Wiese S, Warscheid B: Cytoscape: software for visualization and analysis of biological networks. *Methods in molecular biology (Clifton, NJ)* 2011, 696:291-303.
28. Tang Z, Li C, Kang B, Gao G, Li C, Zhang Z: GEPIA: a web server for cancer and normal gene expression profiling and interactive analyses. *Nucleic acids research* 2017, 45(W1):W98-w102.
29. McGuire S: World Cancer Report 2014. Geneva, Switzerland: World Health Organization, International Agency for Research on Cancer, WHO Press, 2015. *Advances in nutrition (Bethesda, Md)* 2016, 7(2):418-419.
30. Brenner H, Kloor M, Pox CP: Colorectal cancer. *Lancet (London, England)* 2014, 383(9927):1490-1502.
31. 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):557-569.
32. 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.
33. Mashima H, Ohno H, Yamada Y, Sakai T, Ohnishi H: INSL5 may be a unique marker of colorectal endocrine cells and neuroendocrine tumors. *Biochemical and biophysical research communications* 2013, 432(4):586-592.
34. Li SB, Liu YY, Yuan L, Ji MF, Zhang A, Li HY, Tang LQ, Fang SG, Zhang H, Xing S *et al*: Autocrine INSL5 promotes tumor progression and glycolysis via activation of STAT5 signaling. *EMBO molecular medicine* 2020, 12(9):e12050.
35. Yang YC, Chien MH, Lai TC, Su CY, Jan YH, Hsiao M, Chen CL: Monoamine Oxidase B Expression Correlates with a Poor Prognosis in Colorectal Cancer Patients and Is Significantly Associated with

- Epithelial-to-Mesenchymal Transition-Related Gene Signatures. *International journal of molecular sciences* 2020, 21(8).
36. Wu Y, Wei J, Ming Y, Chen Z, Yu J, Mao R, Chen H, Zhou G, Fan Y: Orchestrating a biomarker panel with lncRNAs and mRNAs for predicting survival in pancreatic ductal adenocarcinoma. *Journal of cellular biochemistry* 2018, 119(9):7696-7706.
37. Ma X, Liu C, Xu X, Liu L, Gao C, Zhuang J, Li H, Feng F, Zhou C, Liu Z *et al*: Biomarker expression analysis in different age groups revealed age was a risk factor for breast cancer. *Journal of cellular physiology* 2020, 235(5):4268-4278.
38. Maleckaite R, Zalimas A, Bakavicius A, Jankevicius F, Jarmalaite S, Daniunaite K: DNA methylation of metallothionein genes is associated with the clinical features of renal cell carcinoma. *Oncology reports* 2019, 41(6):3535-3544.
39. Arriaga JM, Levy EM, Bravo AI, Bayo SM, Amat M, Aris M, Hanois A, Bruno L, Roberti MP, Loria FS *et al*: Metallothionein expression in colorectal cancer: relevance of different isoforms for tumor progression and patient survival. *Human pathology* 2012, 43(2):197-208.
40. Na H, Li X, Zhang X, Xu Y, Sun Y, Cui J, Chen Z, Shi X, Ren S, Zuo Y: lncRNA STEAP3-AS1 Modulates Cell Cycle Progression via Affecting CDKN1C Expression through STEAP3 in Colon Cancer. *Molecular therapy Nucleic acids* 2020, 21:480-491.
41. Drucker DJ: Glucagon-like peptides: regulators of cell proliferation, differentiation, and apoptosis. *Molecular endocrinology (Baltimore, Md)* 2003, 17(2):161-171.
42. Spisák S, Kalmár A, Galamb O, Wichmann B, Sipos F, Péterfia B, Csabai I, Kovalszky I, Semsey S, Tulassay Z *et al*: Genome-wide screening of genes regulated by DNA methylation in colon cancer development. *PloS one* 2012, 7(10):e46215.
43. Qi L, Zhang Y, Song F, Ding Y: Chinese herbal medicine promote tissue differentiation in colorectal cancer by activating HSD11B2. *Archives of biochemistry and biophysics* 2020, 695:108644.
44. Chen J, Liu QM, Du PC, Ning D, Mo J, Zhu HD, Wang C, Ge QY, Cheng Q, Zhang XW *et al*: Type-2 11 β -hydroxysteroid dehydrogenase promotes the metastasis of colorectal cancer via the Fgfbp1-AKT pathway. *American journal of cancer research* 2020, 10(2):662-673.
45. Zhang Y, Du W, Chen Z, Xiang C: Upregulation of PD-L1 by SPP1 mediates macrophage polarization and facilitates immune escape in lung adenocarcinoma. *Experimental cell research* 2017, 359(2):449-457.
46. Insua-Rodríguez J, Pein M, Hongu T, Meier J, Descot A, Lowy CM, De Braekeleer E, Sinn HP, Spaich S, Sütterlin M *et al*: Stress signaling in breast cancer cells induces matrix components that promote chemoresistant metastasis. *EMBO molecular medicine* 2018, 10(10).
47. Cheng Y, Wen G, Sun Y, Shen Y, Zeng Y, Du M, Zhu G, Wang G, Meng X: Osteopontin Promotes Colorectal Cancer Cell Invasion and the Stem Cell-Like Properties through the PI3K-AKT-GSK/3 β - β /Catenin Pathway. *Medical science monitor : international medical journal of experimental and clinical research* 2019, 25:3014-3025.

48. 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):1833-1840.

Figures

Figure 1

Gene expression datasets and overlapping up- (down-) regulated genes in these datasets. (A-E), Volcano plots of DEGs in GSE28000, GSE33113, GSE41328, GSE47063, GSE74602; (F), 62 overlapping up-regulated genes; (G), 221 overlapping down-regulated genes.

Figure 2

Protein-protein interaction network of the overlapping DEGs.

Figure 3

Top 6 module analysis by MCODE in the Cytoscape software A, INSL5; B, GCG; C, SPP1; D, MT1M; E, HSD11B2; F, MAOB Each node indicates a protein module; the edges represent proteins interaction.

Figure 4

The expression levels of hub genes in COAD based on sample types. The abscissa of each figure represents the type of colorectal tissue, and the ordinate represents the gene expression signal value in the expression profile chip. Red indicate CRC tissues and green indicate normal tissues. (A) INSL5, (B) GCG, (C) SPP1, (D) MT1M, (E) HSD11B2 and (F) MAOB.

Figure 5

Expression of hub genes in CRC tissues. Expression of (A, B) INSLS (C, D) GCG, (E, F) SPP1, (G, H) HSD11B2, (I, J) MAOB. Magnification, $\times 100$

Figure 6

Kaplan-Meier overall survival of the top 6 hub genes. (A) INSL5, (B) GCG, (C) SPP1, (D) MT1M, (E) HSD11B2 and (F) MAOB. Red lines represent high and blue lines low expression genes. The x-axis shows the survival time of CRC patients and the y-axis the survival probability.