Identification of 10 Hub genes related to the progression of colorectal cancer by co-expression analysis

Jie Meng 1,2, Rui Su 2, Yun Liao 2, Yanyan Li 2, Ling Li Corresponding Author

1 Department of Pharmacy, Anhui University of Traditional Medicine, Hefei, China
2 Department of Pharmacy, Tongren Hospital, Shanghai Jiaotong University School of Medicine, Shanghai, China

Corresponding Author: Ling Li
Email address: LL2699@shtrhospital.com

Background Colorectal cancer (CRC) is the third most common cancer in the world. The present study is aimed at identifying Hub genes associated with the progression of CRC.

Method The data of the patients with CRC were obtained from the Gene Expression Omnibus (GEO) database and assessed by weighted gene co-expression network analysis (WGCNA), Gene Ontology (GO) and Kyoto Encyclopedia of Genes and Genomes (KEGG) enrichment analyses performed in R. by WGCNA, several hub genes that regulate the mechanism of tumorigenesis in CRC were identified. Differentially expressed genes in the data set GSE28000 and GSE42284 were used to construct a co-expression network for WGCNA. The yellow, black, and blue modules associated with CRC level were filtered. Combining the co-expression network and the PPI network, 15 candidate hub genes were screened. Results After validation using the TCGA-COAD dataset, a total of 10 Hub genes (MT1X, MT1G, MT2A, CXCL8, IL1B, CXCL5, CXCL11, IL10RA, GZMB, KIT) closely related to the progression of CRC were identified. The expressions of MT1G, CXCL8, IL1B, CXCL5, CXCL11 and GZMB in colorectal cancer tissues were higher than normal tissues (p-value <0.05). The expressions of MT1X, MT2A, IL10RA and KIT in colorectal cancer tissues were lower than normal tissues (p-value <0.05). Conclusions By combining with a series of methods including GO enrichment analysis, KEGG pathway analysis, PPI network analysis and gene co-expression network analysis, we identified 10 hub genes that were associated with the progression of CRC.
Identification of 10 Hub genes related to the progression of colorectal cancer
by co-expression analysis

Jie Meng¹², Rui Su², Yun Liao², Yanyan Li², Ling Li²
¹ Department of Pharmacy, Anhui University of Traditional Medicine, Hefei, China
² Department of Pharmacy, Tongren Hospital, Shanghai Jiaotong University School of Medicine, Shanghai, China

Corresponding author:
Ling Li²
Department of Pharmacy, Tongren Hospital, Shanghai Jiaotong University School of Medicine,
1111 Xianxia Road, Shanghai, 200336, China
Email address: LL2699@shtrhospital.com

Abstract

Background Colorectal cancer (CRC) is the third most common cancer in the world. The present study is aimed at identifying Hub genes associated with the progression of CRC. Method The data of the patients with CRC were obtained from the Gene Expression Omnibus (GEO) database and assessed by weighted gene co-expression network analysis (WGCNA), Gene Ontology (GO) and Kyoto Encyclopedia of Genes and Genomes (KEGG) enrichment analyses performed in R. by WGCNA, several hub genes that regulate the mechanism of tumorigenesis in CRC were identified. Differentially expressed genes in the data set GSE28000 and GSE42284 were used to construct a co-expression network for WGCNA. The yellow, black, and blue modules associated with CRC level were filtered. Combining the co-expression network and the PPI network, 15 candidate hub genes were screened. Results After validation using the TCGA-COAD dataset, a total of 10 Hub genes (MT1X, MT1G, MT2A, CXCL8, IL1B, CXCL5, CXCL11, IL10RA, GZMB, KIT) closely related to the progression of CRC were identified. The expressions of MT1G, CXCL8, IL1B, CXCL5, CXCL11 and GZMB in colorectal cancer tissues were higher than normal tissues (p-value <0.05). The expressions of MT1X, MT2A, IL10RA and KIT in colorectal cancer tissues were lower than normal tissues (p-value <0.05). Conclusions By combinating with a series of methods including GO enrichment analysis, KEGG pathway analysis, PPI network analysis and gene co-expression network analysis, we identified 10 hub genes that were associated with the progression of CRC.

Keywords: Colorectal cancer, Hub genes, Progression, Co-expression network analysis

Introduction

Colorectal cancer (CRC) is the third most common cancer in the world and considered as the second leading cause of cancer-associated deaths (Brody 2015). Due to lack of early specific disease symptoms it is often identified in advanced stages leading to poor prognosis (Simon...
Multiple biomarkers that can help in improving the diagnosis and treatment monitoring have been identified for CRC (Lech et al. 2016; Okugawa et al. 2015). It has been reported that THBS2 can serve as a prognostic biomarker and also the expression of THBS2 is significantly associated with lymphatic invasion and TNM staging of CRC patients (Wang et al. 2016). The overexpression of CHD4 induced microsatellite instability-high (MSI-H) colorectal cell (CRC) radio-resistance is also reported. The knockdowns of CHD4 enhances radio-sensitivity in microsatellite stabilization (MSS) especially in CRC (Wang et al. 2019). Few studies have shown that the immune system also plays an important role in the development of CRC (Becht et al. 2016; Yin et al. 2017). As can be seen that the mechanism of CRC is complicated and multifaceted in nature, it requires further exploration of mechanisms for the occurrence and development of CRC.

Advances in the sequencing technologies have provided excellent tools and platforms for cancer research including CRC (Srivastava et al. 2014). By correlating the clinical data with molecular mechanisms, new biomarkers for diagnosis, treatment, and prognosis can be restored. Microarray could be used to probe the key biomarkers and provide a better understanding of the molecular mechanisms involved in CRC. Until now, clinically applicable biomarkers are still lacking. Therefore, exploring novel and effective molecular biomarkers to elucidate effective therapeutic targets for CRC patients is still imperative. In this study, we focused on the different expression pattern between the CRC tumor tissues and matched normal tissues. To discover the hub genes and key pathways associated with the initiation and progression of CRC, we applied differential gene expression analysis and functional enrichment analysis. Recently it was demonstrated that the genetic characteristics of the ER and PR pathways can serve as a new marker for CRC prognosis and management (Liu 2016). Weighted Correlation Network Analysis (WGCNA) is an R-package for weighted correlation network analysis and can be used as a data exploration tool for genetic screening (sorting) to find clusters (modules) of highly related genes. It has been widely used to find hub genes in various cancers. For example, studies using WGCNA and UBE2S to identify 11 gene co-expression clusters from large-scale breast cancer data suggest poor prognosis in breast cancer (Clarke et al. 2013). In order to explore the progression of CRC, we have used this algorithm to identify hub genes associated with clinical features.

In the current study, by combining GO enrichment analysis, KEGG pathway analysis, PPI network analysis, gene co-expression network analysis and other bioinformatic methods, we aim to identify the potential key hub genes and modules involved in CRC initiation and progression. In conclusion, we identified 10 hub genes that participated in several cancer-relevant pathways and their abnormal expression are correlated with the clinical progression and prognosis of CRC people by overall survival analysis.

Materials and Method
Data download and processing
GSE28000 and GSE42284 were obtained from NCBI Gene Expression Omnibus (GEO).
The GSE28000 has a total of 115 samples, the platform was Agilent-014850 Whole Human Genome Microarray 4x44K G4112F. GSE42284 was consist of 188 diseased samples, the platform was Agilent Homo sapiens 37K DiscoverPrint_19742. We used batch normalization to correct the two-chip data and all data are normalized.

Screening of differentially expressed genes (DEGs)
Difference analysis of 34 normal and 269 CRC samples was performed using the limma package, \(|\log FC | > 1, p\)-value < 0.05 was defined as DEGs.

GO enrichment analysis and KEGG pathway analysis
A total of 694 upregulated and 1271 downregulated genes were analyzed by GO (Gene Ontology) enrichment and KEGG (Kyoto Encyclopedia of Genes and Genomes) pathway using the R package cluster profiler. \(P\)-value < 0.05 was defined as a meaningful enrichment analysis result. The KEGG pathway was analyzed by R package with a threshold \(p\)-value < 0.05. GO and KEGG pathway analysis was used to predict potential functions.

Gene co-expression network analysis
The gene co-expression network was constructed using the WGCNA package. The top 25% of the genes from the variance plot were screened for constructing a weighted co-expression network. The network module was segmented using a dynamic cut tree algorithm. In order to test the stability of each identified module randomly divided training and test set were generated using the preservation function module stability in the WGCNA package. Correlation between modules and clinical features was assessed by Pearson correlation testing to search for key modules. Clinical information included gender, age, race, stage, location for WGCNA analysis.

Hub gene screening
Genes in the overlapping region were selected after combining the DEGs and the module genes. These were obtained from the string database. The PPI network was constructed and a higher degree node in the network was selected as the hub genes.

Data set verification
In order to verify if the gene expression significantly associated with colorectal cancer, we validated 15 candidate hub genes using TCGA-COAD data in GEPIA database.

Hub gene survival analysis
Kaplan Meier - Plotter database (http://kmplot.com/analysis/) in colorectal cancer data set were used for survival analysis.

Results
Acquisition of microarray data
A total of 303 samples and 18588 genes (Supplementary Table S1) in GSE28000 and
GSE42284 were corrected by batch normalization. File annotation information from Agilent-014850 Whole Human Genome Microarray 4x44K G4112F (Feature Number version) and Agilent Homo sapiens 37K DiscoverPrint_19742 platforms is shown in (Supplementary Table S2).

Differentially expressed genes (DEGs) screening

The expression matrix data from 34 normal and 269 cancer samples were compared and analyzed using the limma package. By employing $p$-value < 0.05 and $|\log FC| \geq 1$ as critical criteria, a total of 1216 DEGs were obtained (Fig. 1A and Supplementary Table S3). DEGs are shown in the volcano and the heat map (Fig. 1A and 1B).

GO enrichment analysis and KEGG pathway analysis results

A total of 694 up-regulated and 1271 down-regulated genes were analyzed by GO (Gene Ontology) enrichment using the R package cluster profiler. $P$-value < 0.05 was defined as a meaningful enrichment analysis result (Fig. 2A and Supplementary Table S4). As shown in Fig. 2A, 1965 genes were significantly enriched in receptor ligand activity, G protein-coupled receptor binding and cytokine activity. KEGG (Kyoto Encyclopedia of Genes and Genomes) pathway enrichment analysis results were shown in Fig. 2B and Supplementary Table S5, which were significantly enriched in cytokine-cytokine receptor interaction.

Gene co-expression network analysis results

GSE28000 and GSE42284 consisted of a total of 303 samples and 18588 genes. The top 25% of the genes in the variance plot were screened to construct a weighted co-expression network. In order to ensure the reliability of the network structure 42 outliers were removed (Fig. 3A). A total of 11 modules (Fig. 3C and Fig. 3D) were obtained by selecting an appropriate soft threshold power =9 (Fig. 3B) according to the scale-free network. Further, to evaluate the stability of each identified module, the training (Supplementary Table S6) and test set (Supplementary Table S7) were randomly divided and the module stability was calculated using the module preservation function (permutations = 200) in the WGCNA package. The clinical information related to GSE26712 is shown in Supplementary Table S8. The yellow, black, and blue modules that are significantly associated with clinical features were selected as candidate key modules (Fig. 3E).

Screening of hub genes

The yellow, black, and blue modules were respectively intersected with the differential genes, and the genes in the intersection region are shown in Fig. 4A-C. The results of gene enrichment analysis of the intersection of DEGs and yellow, black and blue modules are shown in Fig. 5A-F.
Survival analysis of hub genes

PPI network analysis was performed on the genes from the intersection regions, and the interacting proteins with confidence >0.7 was selected. The yellow module had 27 nodes and 33 edges, the black module had 19 nodes and 73 edges, and the blue module had 18 nodes, 20 sides (Fig. 6A-C).

The degree of each node in the network based on Cytoscape was calculated. The top 5 nodes in the degree ranking in the module network were selected as candidate hub genes. A total of 15 candidate hub genes were identified. The top 5 nodes in the yellow module were MT1H, MT1X, MT1E, MT1G, MT2A, and the top 5 nodes in the black module were CXCL8, CSF2, IL1B, CXCL5, IL1A, respectively. The top 5 nodes in the blue module were TLR3, CXCL11, IL10RA, GZMB, and KIT.

Data set verification

Similar results were obtained for the hub genes screened by the TCGA-COAD data validation in Fig. 7A-J, suggesting reliable findings of the present study. A total of 15 candidate hub genes were verified, among these 10 genes: MT1X, MT1G, MT2A, CXCL8, IL1B, CXCL5, CXCL11, IL10RA, GZMB, KIT were significantly different between the normal group and the cancer group.

Survival analysis

Survival analysis of the hub genes selected above was performed in Fig. 8A and Fig. 8B. Findings suggest that two genes were significantly associated with the prognosis of patients (p-value <0.05), which is IL10RA and KIT. With the increase of IL10RA and KIT expression, the total survival time was significantly prolonged.

Discussion

Over the past few decades, the mortality rates associated with CRC has increased due to the higher incidence in the young population (Connell et al. 2017; The Lancet 2017). Long term research in colorectal cancer has led to many substantial advances in the diagnostic and therapeutic techniques. For example, effective prognostic biomarkers for CRC are CEA levels, circulating tumor DNA (ctDNA), MS instability and certain genetic characteristics (Duffy 2015). Analysis can be used to diagnose, identify, and track tumor-specific changes associated with disease progression and to guide treatment decisions (Osumi et al. 2019). miRNAs can also be used as diagnostic and prognostic biomarkers for assessing tumor development, progression, invasion, metastasis, and reaction to chemotherapeutic drugs (Shirafkan et al. 2018). It can also help in the early identification of differences between responder and non-responder individuals (Ballester et al. 2016). Advances in bioinformatics and genetics have led to the development of biomarkers and genetic models that can help to select responders and to assess prognosis, and
thereby rationalizing, individualizing, and improving the prognosis.

In this study, we identified Hub genes associated with colorectal cancers and established a gene co-expression network analysis. Initially, most of the response genes demonstrated a tendency to separate, with only a few being significantly concentrated. Expression changes occurring in one gene can interact with the linked genes and can affect the downstream biological functions. Finally, through survival analysis, we screened 10 genetic models including MT1X, MT1G, MT2A, CXCL8, IL1B, CXCL5, CXCL11, IL10RA, GZMB, KIT. Abnormal expression of these genes may influence the survival time of patients. The expression of MT1G, CXCL8, IL1B, CXCL5, CXCL11, GZMB in colorectal cancer tissues is higher than normal tissues. In contrast the expression of MT1X, MT2A, IL10RA, KIT was lower in cancer tissues as compared to normal tissues.

The researchers have found that these 10 Hub genes are involved in cytokine-cytokine receptor interactions and in the receptor-ligand activity. Previous findings suggest that MT1X induces cell cycle arrest and apoptosis by inactivating NF-κB signaling in HCC (Liu et al. 2018). MT1G promotes new tumor suppressor activity in CRC tumor differentiation and indicates that MT and zinc signaling as new participants in colorectal differentiation (Arriaga et al. 2017). MT1G promotes methylation and tumor aggressiveness in prostate cancers and could serve as a marker for locally advanced disease (Henrique et al. 2005). The genetic polymorphisms in MT2A (rs10636 and rs28366003) increases the risk of breast cancer in Chinese Han population (Liu et al. 2017). In another study, overexpression of CXCL8 induced cell proliferation, migration, and invasion of colon cancer LoVo cells, and CXCL8 induced EMT via the PI3K / AKT / NF-κB signaling axis was reported (Shen et al. 2017). Many articles have shown that IL1B SNPs may be involved in the pathogenesis of NSCLC and thyroid cancer in Chinese population and can also be used as a new prognostic genetic biomarker for non-small cell lung cancer (Li et al. 2019; Li et al. 2015; Perez-Ramirez et al. 2017). Tumor-derived CXCL5 can promote human colorectal cancer metastasis by activating ERK / Elk-1 / Snail and AKT /GSK3β/β-catenin pathway, accelerate osteosarcoma growth and can serve as a biomarker for non-small cell carcinoma, etc. (Roca et al. 2018; Wu et al. 2017; Zhao et al. 2017). CXCL11 is overexpressed in CRC tissues and cell lines, inhibiting CXCL11 to significantly affect the CRC cell migration, invasion, and EMT in vitro. In addition, down-regulation of CXCL11 also reduces CRC growth and metastasis in vivo (Gao et al. 2018). IL10 is a key anti-inflammatory cytokine inhibiting the pro-inflammatory responses of innate and adaptive immune cells. Spontaneous intestinal inflammation in IL10 and IL10R-deficient mice is reported in the reference. Also patients with deleterious mutations in IL10, IL10RA or IL10RB develop severe enterocolitis in the first few months of life (Shouval et al. 2014). Moreover, the expression of IL10 in CRC tissues was significantly higher than the healthy intestinal endothelial cells. The correlation between the expression of IL10RA and the proliferation index or clinical stage of CRC confirms the importance of IL10RA in the pathogenesis of CRC (Zadka et al. 2018). In NK cell-based anticancer therapy, activation of autophagy in hypoxic cells operates through selective degradation of the pro-apoptotic NK-derived serine protease GZMB/granzyme B and by blocking the NK-mediated target cell apoptosis. The autophagy targeting cancer cells promotes...
tumor regression by promoting the elimination of NK cells (Viry et al. 2014). Subpopulations of naive and memory B cells also express GZMB in mammary gland draining lymph nodes (Arabpour et al. 2019). Oncogenic signaling of Kit tyrosine kinase selectively occurs in the Golgi apparatus in the gastrointestinal stromal tumors and causing inhibition of the carcinogenesis by blocking the secretory transport of M-COPA in the gastrointestinal stromal tumors (Obata et al. 2018; Obata et al. 2017). Multiple reports have identified genes directly or indirectly involved in different biological pathways associated with CRC. The 10 genes identified in this study were further validated using a separate data set to verify if the 10 gene model can significantly affect the prognosis of colorectal cancer.

In practice, this 10-gene model can be used to roughly predict the prognosis. In order to further elucidate the relationship between DEG and clinical outcomes in patients with colorectal cancer additional mathematical analysis and modeling independent of standard clinical and pathological criteria needs to be conducted. Also, further prognostic gene models and cluster analysis of TNM (tumor, lymph node, and metastasis) are needed to assess the independent nature of this model. At the same time, the identified genetic model requires further validation using qPCR and other laboratory methods.

Conclusions

In summary, we screened differentially expressed genes based on GEO's gene expression profiling. GO enrichment analysis and KEGG pathway analysis were performed on these genes. The top 25% of the genes were screened for constructing a weighted co-expression network. Finally, hub genes were screened, and survival curves were generated. A separate data set was used to verify the validity of the 10 gene model. The hub genes identified in this study can help to predict postoperative treatment and prognosis of CRC patients.

Acknowledgments

The authors thank the reviewers for their helpful comments on our report.

Conflicts of Interest

The authors declare that there are no conflicts of interest.

Data Availability Statement

The following information was supplied regarding data availability:

The code to generate the random data set used in the study is provided in a Supplemental File.

Funding

This research did not receive any specific grant from funding agencies in the public, commercial, or not-for-profit sectors.

ORCID
References

Arabpour M, Rasolmali R, Talei AR, Mehdipour F, and Ghaderi A. 2019. Granzyme B production by activated B cells derived from breast cancer-draining lymph nodes. *Mol Immunol* 114:172-178. 10.1016/j.molimm.2019.07.019

Arriaga JM, Bravo AI, Mordoh J, and Bianchini M. 2017. Metallothionein 1G promotes the differentiation of HT-29 human colorectal cancer cells. *Oncol Rep* 37:2633-2651. 10.3892/or.2017.5547

Ballester V, Rashtak S, and Boardman L. 2016. Clinical and molecular features of young-onset colorectal cancer. *World J Gastroenterol* 22:1736-1744. 10.3748/wjg.v22.i5.1736

Becht E, de Reynies A, Giraldo NA, Pilati C, Buttard B, Lacroix L, Selves J, Sautes-Fridman C, Laurent-Puig P, and Fridman WH. 2016. Immune and Stromal Classification of Colorectal Cancer Is Associated with Molecular Subtypes and Relevant for Precision Immunotherapy. *Clin Cancer Res* 22:4057-4066. 10.1158/1078-0432.CCR-15-2879

Brody H. 2015. Colorectal cancer. *Nature* 521:S1. 10.1038/521S1a

Clarke C, Madden SF, Doolan P, Aherne ST, Joyce H, O'Driscoll L, Gallagher WM, Hennessy BT, Moriarty M, Crown J, Kennedy S, and Clynes M. 2013. Correlating transcriptional networks to breast cancer survival: a large-scale coexpression analysis. *Carcinogenesis* 34:2300-2308. 10.1093/carcin/bgt208

Connell LC, Mota JM, Braghiroli MI, and Hoff PM. 2017. The Rising Incidence of Younger Patients With Colorectal Cancer: Questions About Screening, Biology, and Treatment. *Curr Treat Options Oncol* 18:23. 10.1007/s11864-017-0463-3

Duffy MJ. 2015. Personalized treatment for patients with colorectal cancer: role of biomarkers. *Biomark Med* 9:337-347. 10.2217/bmm.15.3

Gao YJ, Liu L, Li S, Yuan GF, Li L, Zhu HY, and Cao GY. 2018. Down-regulation of CXCL11 inhibits colorectal cancer cell growth and epithelial-mesenchymal transition. *Onco Targets Ther* 11:7333-7343. 10.2147/OTT.S167872

Henrique R, Jeronimo C, Hoque MO, Nomoto S, Carvalho AL, Costa VL, Oliveira J, Teixeira MR, Lopes C, and Sidransky D. 2005. MT1G hypermethylation is associated with higher tumor stage in prostate cancer. *Cancer Epidemiol Biomarkers Prev* 14:1274-1278. 10.1158/1055-9965.EPI-04-0659

Lech G, Slotwinski R, Slodkowski M, and Krasnodebski IW. 2016. Colorectal cancer tumour markers and biomarkers: Recent therapeutic advances. *World J Gastroenterol* 22:1745-1755. 10.3748/wjg.v22.i5.1745

Li H, Duan N, Zhang Q, and Shao Y. 2019. IL1A & IL1B genetic polymorphisms are risk factors for thyroid cancer in a Chinese Han population. *Int Immunopharmacol* 76:105869. 10.1016/j.intimp.2019.105869

Li Y, Zhao W, Zhao Z, Wu J, Chen L, Ma Y, Li Q, Lu D, Jin L, and Wang J. 2015. IL1B gene polymorphisms, age and the risk of non-small cell lung cancer in a Chinese population. *Lung Cancer* 89:232-237. 10.1016/j.lungcan.2015.06.009
Liu D. 2016. Gene signatures of estrogen and progesterone receptor pathways predict the prognosis of colorectal cancer. *FEBS J* 283:3115-3133. 10.1111/febs.13798

Liu D, Wang M, Tian T, Wang XJ, Kang HF, Jin TB, Zhang SQ, Guan HT, Yang PT, Liu K, Liu XH, Xu P, Zheng Y, and Dai ZJ. 2017. Genetic polymorphisms (rs10636 and rs28366003) in metallothionein 2A increase breast cancer risk in Chinese Han population. *Aging (Albany NY)* 9:547-555. 10.18632/aging.101177

Liu Z, Ye Q, Wu L, Gao F, Xie H, Zhou L, Zheng S, and Xu X. 2018. Metallothionein 1 family profiling identifies MT1X as a tumor suppressor involved in the progression and metastatic capacity of hepatocellular carcinoma. *Mol Carcinog* 57:1435-1444. 10.1002/mc.22846

Obata Y, Horikawa K, Shiina I, Takahashi T, Murata T, Tasaki Y, Suzuki K, Yonekura K, Esumi H, Nishida T, and Abe R. 2018. Oncogenic Kit signalling on the Golgi is suppressed by blocking secretory trafficking with M-COPA in gastrointestinal stromal tumours. *Cancer Lett* 415:1-10. 10.1016/j.canlet.2017.11.032

Obata Y, Horikawa K, Takahashi T, Akieda Y, Tsujimoto M, Fletcher JA, Esumi H, Nishida T, and Abe R. 2017. Oncogenic signaling by Kit tyrosine kinase occurs selectively on the Golgi apparatus in gastrointestinal stromal tumors. *Oncogene* 36:3661-3672. 10.1038/onc.2016.519

Okugawa Y, Grady WM, and Goel A. 2015. Epigenetic Alterations in Colorectal Cancer: Emerging Biomarkers. *Gastroenterology* 149:1204-1225 e1212. 10.1053/j.gastro.2015.07.011

Osumi H, Shinozaki E, Yamaguchi K, and Zembutsu H. 2019. Clinical utility of circulating tumor DNA for colorectal cancer. *Cancer Sci* 110:1148-1155. 10.1111/cas.13972

Perez-Ramirez C, Canadas-Garre M, Alnatsha A, Molina MA, Robles AI, Villar E, Delgado JR, Faus-Dader MJ, and Calleja-Hernandez MA. 2017. Interleukins as new prognostic genetic biomarkers in non-small cell lung cancer. *Surg Oncol* 26:278-285. 10.1016/j.suronc.2017.05.004

Roca H, Jones JD, Purica MC, Weidner S, Koh AJ, Kuo R, Wilkinson JE, Wang Y, Daignault-Newton S, Pienta KJ, Morgan TM, Keller ET, Nor JE, Shea LD, and McCauley LK. 2018. Apoptosis-induced CXCL5 accelerates inflammation and growth of prostate tumor metastases in bone. *J Clin Invest* 128:248-266. 10.1172/JCI92466

Shen T, Yang Z, Cheng X, Xiao Y, Yu K, Cai X, Xia C, and Li Y. 2017. CXCL8 induces epithelial-mesenchymal transition in colon cancer cells via the PI3K/Akt/NF-kappaB signaling pathway. *Oncol Rep* 37:2095-2100. 10.3892/or.2017.5453

Shirafkan N, Mansoori B, Mohammadi A, Shomali N, Ghasbi M, and Baradaran B. 2018. MicroRNAs as novel biomarkers for colorectal cancer: New outlooks. *Biomed Pharmacother* 97:1319-1330. 10.1016/j.biopha.2017.11.046

Shouval DS, Ouahed J, Biswas A, Goettel JA, Horwitz BH, Klein C, Muise AM, and Snapper SB. 2014. Interleukin 10 receptor signaling: master regulator of intestinal mucosal homeostasis in mice and humans. *Adv Immunol* 122:177-210. 10.1016/B978-0-12-800267-4.00005-5
Simon K. 2016. Colorectal cancer development and advances in screening. Clin Interv Aging 11:967-976. 10.2147/CIA.S109285

Srivastava P, Mangal M, and Agarwal SM. 2014. Understanding the transcriptional regulation of cervix cancer using microarray gene expression data and promoter sequence analysis of a curated gene set. Gene 535:233-238. 10.1016/j.gene.2013.11.028

The Lancet O. 2017. Colorectal cancer: a disease of the young? Lancet Oncol 18:413. 10.1016/S1470-2045(17)30202-4

Viry E, Baginska J, Berchem G, Noman MZ, Medves S, Chouaib S, and Janji B. 2014. Autophagic degradation of GZMB/granzyme B: a new mechanism of hypoxic tumor cell escape from natural killer cell-mediated lysis. Autophagy 10:173-175. 10.4161/auto.26924

Wang HC, Chou CL, Yang CC, Huang WL, Hsu YC, Luo CW, Chen TJ, Li CF, and Pan MR. 2019. Over-Expression of CHD4 Is an Independent Biomarker of Poor Prognosis in Patients with Rectal Cancers Receiving Concurrent Chemoradiotherapy. Int J Mol Sci 20:1-14. 10.3390/ijms20174087

Wang X, Zhang L, Li H, Sun W, Zhang H, and Lai M. 2016. THBS2 is a Potential Prognostic Biomarker in Colorectal Cancer. Sci Rep 6:33366. 10.1038/srep33366

Wu K, Yu S, Liu Q, Bai X, Zheng X, and Wu K. 2017. The clinical significance of CXCL5 in non-small cell lung cancer. Onco Targets Ther 10:5561-5573. 10.2147/OTT.S148772

Yin Y, Yao S, Hu Y, Feng Y, Li M, Bian Z, Zhang J, Qin Y, Qi X, Zhou L, Fei B, Zou J, Hua D, and Huang Z. 2017. The Immune-microenvironment Confers Chemoresistance of Colorectal Cancer through Macrophage-Derived IL6. Clin Cancer Res 23:7375-7387. 10.1158/1078-0432.CCR-17-1283

Zadka L, Kulus MJ, Kurnol K, Piotrowska A, Glatzel-Plucinska N, Jurek T, Czuba M, Nowak A, Chabowski M, Janczak D, and Dziegieł P. 2018. The expression of IL10RA in colorectal cancer and its correlation with the proliferation index and the clinical stage of the disease. Cytokine 110:116-125. 10.1016/j.cytto.2018.04.030

Zhao J, Ou B, Han D, Wang P, Zong Y, Zhu C, Liu D, Zheng M, Sun J, Feng H, and Lu A. 2017. Tumor-derived CXCL5 promotes human colorectal cancer metastasis through activation of the ERK/Elk-1/Snail and AKT/GSK3beta/beta-catenin pathways. Mol Cancer 16:70. 10.1186/s12943-017-0629-4
Figure 1

DEGs in GSE28000 and GSE42284.

(A) Volcanic map of gene expression values between colorectal cancer tissues and normal tissues. Vermilion is the up-regulated gene and blue is the down-regulated gene. (B) Differential gene heat map.
Figure 2

Enrichment analysis results.

(A) GO enrichment analysis results of DEGs. (B) Results of the first 12 KEGG pathways in enrichment analysis of DEGs.
Figure 3

WGCNA.

(A) Sample cluster analysis. (B) The top panel represents gene tree and the bottom a gene module with different colors. (C) relationship between the two preservation statistics. (D) Module preservation as a function of module quality. (E) Correlation between modules and features. The upper number in each cell is the correlation coefficient between the clinical features and each module, and the lower number is the corresponding *p*-value. Module size and preservation scores are shown in the x and y axes, respectively. Module numbers are shown next to the circles. Modules with Z summary scores > 10 (above the red dotted line) are considered highly preserved, Z summary scores between 2 and 10 (between the blue and red dotted lines) are weak to moderately preserved, and Z summary scores < 2 (below the blue dotted line) are not preserved.
Figure 4

DEGs and module gene Venn diagram.

(A) DEGs and yellow module gene Venn diagram. (B) DEGs and black module gene Venn diagram. (C) DEGs and blue module gene Venn diagram.
Manuscript to be reviewed
Figure 5

The results of gene enrichment analysis of the intersection of DEGs and yellow, black and blue modules.

(A) Gene enrichment analysis results of the intersection region of DEGs and yellow modules.  
(B) Gene enrichment analysis results of the intersection region of DEGs and black modules  
(C) Gene enrichment analysis results of the intersection region of DEGs and blue modules.  
(D) Yellow module GO enrichment analysis results. (E) Black module GO enrichment analysis results. (F) Blue Yellow module GO enrichment analysis results.
Figure 6

PPI network analysis results.

(A) Black module. (B) Blue module. (C) Yellow module.
Figure 7

Hub genes screened by TCGA-COAD data validation.

The expression level of hub genes in colorectal cancer tissues. The expression level of hub genes in colorectal cancer tissues were higher than those in normal tissues ($p$-value < 0.05). (A) CXCL5. (B) CXCL8. (C) CXCL11. (D) GZMB. (E) IL1B. (F) MT1G. The expression level of DEGs in colorectal cancer tissues were lower than normal tissues (indicates $p$-value < 0.05). (G) MT2A. (H) IL10RA. (I) KIT. (J) MT1X.
Figure 8

The expression level of two hub genes was significantly associated with prognosis ($p$-value<0.05).

The expression of the two hub genes selected increased and the overall survival time was significantly prolonged. (A) IL10RA. (B) KIT.