2024 Volume 2, Issue 1
Big Data Analytics for Healthcare
ISSN: 2630-4767 (Online)

Screening and Verification of Hub Genes in Colorectal Cancer Using Bioinformatics

Shuhua Chen*, Rikui Wen, Huiqin Zhu, Rongzhang Xie
Department of Clinical Laboratory, Yunfu People's Hospital, Yunfu 527300, Guangdong Province, China

Corresponding author: Shuhua Chen, chenshuhua@163.com

Abstract:

Objective: To screen differentially expressed genes (DEGs) associated with colorectal cancer using bioinformatics analysis and verify their biological functions. *Methods:* The Department of Clinical Laboratory at Yunfu People's Hospital downloaded colorectal cancer microarray data (GSE21815, GSE31905, GSE35279) from the Gene Expression Omnibus (GEO). GEO2R was used to process the data and identify DEGs between colorectal cancer and normal colorectal tissues. Bioinformatics tools such as DAVID, STRING, and Cytoscape were utilized to construct a protein-protein interaction network for the DEGs, screen for hub genes, and analyze the biological functions of these hub genes using Gene Ontology (GO) and Kyoto Encyclopedia of Genes and Genomes (KEGG). Potential miRNAs regulating the hub genes were identified using MiRDB. Additionally, 30 colorectal cancer tissue samples and 30 normal colorectal tissue samples collected from August 2017 to August 2022 were used for validation through quantitative real-time PCR (qPCR). *Results:* The oxytocin receptor gene, matrix metalloproteinase-11 gene, mesenchymal-epithelial transition factor gene, matrix metalloproteinase-7 gene, kallikrein-8 gene, and kallikrein-10 gene were identified as key hub genes for the development and progression of colorectal cancer through bioinformatics analysis and protein-protein interaction network analysis. The expression levels of matrix metalloproteinase-11 gene (4.38 ± 1.58), mesenchymal-epithelial transition factor gene (2.69 ± 0.29), matrix metalloproteinase-7 gene (0.88 ± 0.14), kallikrein-8 gene (11.09 ± 3.90), and kallikrein-10 gene mRNA (7.88 ± 2.20) were significantly higher in colorectal cancer tissues compared to normal colorectal tissues, with statistically significant differences (t = 9.605, 25.339, 26.376, 9.541, 3.726; all $P < 0.001$). *Conclusion:* The abnormal expression of the matrix metalloproteinase-11 gene, mesenchymal-epithelial transition factor gene, matrix metalloproteinase-7 gene, kallikrein-8 gene, and kallikrein-10 gene in colorectal cancer tissues may be involved in the development of colorectal cancer, providing a basis for future basic research and clinical diagnosis and treatment of colorectal cancer.

Keywords:

Colorectal cancer
Bioinformatics
Hub genes

Online publication: June 28, 2024

1. Introduction

Colorectal cancer is one of the common malignancies of the digestive tract in clinical practice, ranking third in cancer mortality and posing a serious threat to human health. Therefore, early identification of biomarkers related to the diagnosis and prognosis of colorectal cancer is crucial [1]. Bioinformatics, which combines biology and informatics, plays an active role in elucidating the molecular mechanisms of diseases. Gene microarrays primarily involve the attachment of high-density DNA fragment arrays to materials such as glass or nylon using microarray technology, enabling the screening of valuable genes for further research and analysis. They are currently often used to collect disease expression profiling data [2]. Hub genes, selected through bioinformatics, are core genes and potential targets for clinical disease treatment [3]. The Department of Laboratory Medicine at Yunfu People's Hospital employed bioinformatics methods to screen for differentially expressed genes related to colorectal cancer. These genes were then subjected to biological pathway enrichment analysis, and a protein-protein interaction network was constructed to identify hub genes. From August 2017 to August 2022, 30 colorectal cancer tissue samples and 30 normal colorectal tissue samples were collected to further validate the expression of hub genes, providing a novel auxiliary tool for future clinical diagnosis, treatment, and prognosis evaluation of colorectal cancer. The details are reported below.

2. Subjects and methods

2.1. Study subjects and microarray data

Thirty tissue samples from patients with colorectal cancer treated at our hospital from August 2017 to August 2022 were collected as the abnormal group, and another 30 non-cancerous colonic tissue samples were collected as the normal group. Colorectal cancer microarray data, including GSE21815, GSE31905, and GSE35279, were downloaded from the Gene Expression Omnibus (GEO) of the National Center for Biotechnology Information. Specifically, GSE21815 comprised 12 colorectal cancer tissues and nine normal colorectal control tissues; GSE31905 included 55 colorectal cancer tissues and seven normal colorectal tissues; GSE35279 encompassed 74 cancer tissues and five normal tissues.

2.2. Methods

2.2.1. Data preprocessing

The tidyverse software was used to process the acquired microarray data, naming the gene files for each sample and organizing the gene expression levels into a unified expression matrix. Sample annotation content was extracted from the metadata database to match the expression matrix. Finally, the biomaRt package in R was utilized to annotate standard gene symbols, resulting in a standardized gene expression matrix.

2.2.2. Differential gene screening

The edgeR and limma packages were applied to analyze differentially expressed genes from the preprocessed gene expression matrix. The GEO analysis tool "analyze with geo2r" was then used to divide the data into colorectal cancer tissue and normal tissue groups. Effective genes were selected based on the criteria of $|\log FC| > 2$ and $P < 0.01$ [4]. Differential genes from the GSE21815, GSE31905, and GSE35279 datasets were identified, and a Venn diagram was constructed to find common genes among these datasets.

2.2.3. lncRNA extraction

Relevant information on the screened differential gene lncRNAs was collected and organized from the Gencode database. The lncRNA ensemble IDs and their expression levels were extracted using R software.

2.2.4. Biological function enrichment analysis

The bioinformatics tool DAVID was employed to screen for differentially expressed genes (DEGs) related to colorectal cancer that differed from normal colorectal tissues. Functional enrichment analysis was conducted using Gene Ontology (GO), including cellular components, molecular functions, and biological processes. Additionally, the Kyoto Encyclopedia of Genes and Genomes (KEGG) pathway enrichment analysis was performed.

2.2.5. Protein-protein interaction (PPI) network

Analysis PPI analysis was performed on the DEGs involved in the enrichment analysis using the STRING

database. The PPI network was constructed using Cytoscape. Specifically, network centrality parameters, such as degree centrality, were calculated through network analysis. The MCODE plugin was used to extract key subnetworks, obtain relatively concentrated expression subsets, and identify potential regulatory genes in colorectal cancer. Finally, ClueGO and CluePedia were utilized to visualize the KEGG pathways and construct a gene interaction network between biological pathways. The top 10 differentially expressed genes interacting with the pathways were extracted as hub genes.

2.2.6. Validation of hub gene expression in colorectal cancer tissue samples

Quantitative real-time PCR (qPCR) was used to measure the expression levels of hub genes in cancerous and normal colorectal tissues. The specific steps are as follows: (1) Total RNA extraction: A tissue sample weighing 0.3–0.5 g was ground into a powder, placed in a 1.5 mL centrifuge tube, and total RNA was extracted according to the instructions provided by the RNA extraction kit. (2) Reverse transcription reaction: The reverse transcription reaction was performed following the operating procedures specified in the reverse transcription kit. A total of 200 ng of RNA was used, 1 μL of specific reverse transcription primer and 12 μL of sterile water were added, and incubation was carried out for 5 minutes. Subsequently, Ribolock RNase inhibitor and RevertAid M-MuLV reverse transcriptase were sequentially added. (3) RT-PCR reaction detection: The reaction system was set up using the SYBR green method with the following conditions: 95°C for 10 minutes, followed by 95°C for 15 seconds, 60°C for 1 minute, for 40 cycles. U6 was used as the internal reference gene, and the gene expression level was calculated as $2^{-\Delta\Delta Ct}$. Normal colorectal tissue served as the control to determine the hub expression level in colorectal cancer tissue. $\Delta\Delta Ct = (Ct_{\text{target gene}} - Ct_{U6})$ colorectal cancer tissue – $(Ct_{\text{target gene}} - Ct_{U6})$ normal colorectal tissue.

2.3. Statistical methods

SPSS 26.0 statistical software was used for data analysis. Measurement data (mRNA expression levels) followed a normal distribution and were described as mean ± standard deviation (SD). The *t*-test was performed, and $P < 0.05$ was considered statistically significant.

3. Results

3.1. Differential gene screening results

The GSE 21815 (937), 31905 (1470), and 35279 (1458) datasets included 141 colorectal cancer tissue samples and 21 normal colorectal tissue samples. 937, 1470, and 1458 DEGs were identified in the three datasets, respectively. After drawing a Venn diagram, 105 common up-regulated genes and 140 common down-regulated genes were obtained.

3.2. Core gene verification

Based on PPI network analysis, 245 common differential genes were imported into the STRING database to generate a PPI network diagram. Visual analysis of the PPI network revealed the top 10 hub genes with the highest interaction degree: oxytocin receptor (*OXTR*), matrix metalloproteinase-11 (*MMP11*), erythropoietin-producing hepatoma-B2 (*EPHB2*), mesenchymal-epithelial transition factor (*MET*), transforming growth factor-β1 (*TGF-β1*), stanniocalcin-2 (*STC2*), matrix metalloproteinase-7 (*MMP7*), kallikrein-8 (*KLK8*), kallikrein-10 (*KLK10*), and keratin-23 (*KRT23*). The mRNA expression levels of *MMP11*, *c-MET*, *MMP7*, *KLK8*, and *KLK10* in colorectal cancer tissues were significantly higher than those in normal colorectal tissues, with statistically significant differences ($P < 0.05$ for all). However, there were no statistically significant differences in the mRNA expression levels of *OXTR*, *EPHB2*, *TGF-β1*, *STC2*, and *KRT23* between colorectal cancer tissues and normal colorectal tissues ($P > 0.05$ for all). **Table 1** shows the details.

4. Discussion and conclusion

Currently, numerous basic and clinical studies have revealed the etiology and pathogenesis of colorectal cancer development [5,6]. However, the prevalence and mortality rates have remained high in recent years. Some scholars believe that this is mainly due to most studies focusing on single genetic events in cohort studies, leading to biased results [7]. In this study, we selected three

Table 1. Comparison of mRNA expression levels of 10 hub genes between colorectal cancer tissues and normal colorectal tissues (mean ± SD)

Group	OXTR	MMP11	EPHB2	MET	TGF-β1	STC2	MMP7	KLK8	KLK10	KRT23
Colorectal cancer tissue (n = 30)	0.87 ± 0.09	4.38 ± 1.58	2.99 ± 0.51	2.69 ± 0.29	1.40 ± 0.31	1.43 ± 0.23	0.88 ± 0.14	11.09 ± 3.90	7.88 ± 2.20	0.96 ± 0.15
Normal colorectal tissue (n = 30)	0.79 ± 0.12	1.57 ± 0.27	2.85 ± 0.38	1.01 ± 0.21	1.24 ± 0.36	1.37 ± 0.21	0.17 ± 0.04	4.00 ± 1.18	5.96 ± 1.76	0.98 ± 0.15
t-value	1.661	9.605	1.213	25.339	1.856	1.024	26.376	9.541	3.726	0.545
P-value	0.102	< 0.001	0.230	< 0.001	0.068	0.310	< 0.001	< 0.001	< 0.001	0.588

Abbreviations: *OXTR*: Oxytocin receptor gene, *MMP11*: Matrix metalloproteinase-11 gene, *EPHB2*: Ephrin type-B receptor 2 gene, *MET*: Mesenchymal-epithelial transition factor gene, *TGF-β1*: Transforming growth factor beta 1 gene, *STC2*: Stanniocalcin-2 gene, *MMP7*: Matrix metalloproteinase 7 gene, *KLK8*: Kallikrein-8 gene, *KLK10*: Kallikrein-10 gene, *KRT23*: Keratin-23 gene.

databases, GSE 21815, GSE 31905, and GSE 35279, to conduct an in-depth bioinformatics analysis. We identified 245 DEGs associated with colorectal cancer. Functional enrichment analysis using GO was performed on the upregulated and downregulated genes. The results showed that: (1) The cellular components (CC) of upregulated DEGs involve the extracellular matrix, extracellular region, and plasma membrane, while the downregulated genes involve the extracellular region, extracellular space, and intercellular adhesion molecules. (2) The molecular functions of upregulated genes mainly include sequence-specific DNA binding, RNA polymerase II activity, and growth factor activity, while the downregulated genes are primarily involved in hormone activity and zinc ion binding. (3) The biological processes (BP) of upregulated genes encompass cell proliferation, apoptosis, drug response, and proteolysis, whereas the downregulated genes are involved in proteolysis and transmembrane transport of small molecules. KEGG pathway enrichment analysis indicated that the upregulated DEGs are mainly enriched in the transforming growth factor-β (TGF-β) signaling pathway and the Wnt signaling pathway, while the downregulated DEGs are enriched in the PPAR signaling pathway and nitrogen metabolism pathway. These findings suggest that upregulated genes, primarily located in the nucleus, cell membrane, and extracellular matrix, participate in cell migration, proliferation, and apoptosis of colorectal cancer cells by regulating DNA replication and cell cycle processes. On the other hand, downregulated genes, predominantly found in the extracellular region, are involved in metabolic processes that contribute to cancer development and progression.

In this study, we constructed a PPI network for the differentially expressed genes and identified 10 hub genes. Validation using clinical tissue samples revealed that the mRNA expression levels of matrix metalloproteinase-11 (*MMP11*) (4.38 ± 1.58), mesenchymal-epithelial transition factor (*MET*) (2.69 ± 0.29), matrix metalloproteinase-7 (*MMP7*) (0.878 ± 0.143), kallikrein-8 (*KLK8*) (11.09 ± 3.90), and kallikrein-10 (*KLK10*) (7.88 ± 2.20) were significantly higher in colorectal cancer tissues compared to normal colorectal tissues ($P < 0.05$ for all). Matrix metalloproteinases (MMPs) play a role in dissolving the extracellular matrix and basement membrane proteases. Under normal conditions, their expression levels are low in the body. However, in pathological changes, especially during cancer cell proliferation, differentiation, and migration, their expression levels increase significantly [8]. MMP11, a member of the MMP family, can induce tumor progression by remodeling the extracellular matrix and promote tumor cell survival by inhibiting apoptosis [9]. MMP7, the smallest secreted protein among MMPs, can dissolve the extracellular matrix, hydrolyze proteoglycans and collagen substrates, and promote tumor cell proliferation and differentiation [10]. Previous studies have reported high expression levels of *MMP11* (4.25 ± 1.19) and *MMP7* (0.886 ± 0.125) in various cancer tissues [11]. According to Li *et al.* [12], *MMP7* expression can indicate the occurrence and progression of colorectal cancer. Similarly, Wen *et al.* [13] found that detecting *MMP11* can aid in diagnosing

the development of colorectal cancer. *C-Met*, a proto-oncogene, can phosphorylate signaling pathways, promote cell movement, and lead to epithelial cell dispersion, endothelial cell migration, and tumor cell invasion and migration. Other studies have reported that *C-Met* can induce tumor angiogenesis [14]. Cao *et al.* [15] demonstrated a positive correlation between *C-Met* expression levels in colorectal cancer tissues and cancer cell invasion and metastasis, providing guidance for prognosis. It has been established that kallikreins (KLKs) are closely related to the development of malignant tumors. Their mechanism involves encoding hK proteins that dissolve extracellular matrix proteins, facilitating cancer cell migration and tumor angiogenesis [16]. Relevant literature reports that KLK8 can promote the proliferation and metastasis of colorectal cancer cells, and its overexpression can inhibit tumor cell apoptosis. Conversely, knocking down *KLK8* promotes cancer cell apoptosis [17]. Another study found abnormal expression of *KLK10* in colorectal cancer tissues, positively correlated with tumor pathological staging and the occurrence of liver metastasis [18]. Thus, the high expression of *MMP11*, *c-MET*, *MMP7*, *KLK8*, and *KLK10* can promote the proliferation, differentiation, invasion, and migration of colorectal cancer cells.

In summary, this study utilized bioinformatics methods to screen and validate a set of genes closely related to the development of colorectal cancer. These genes have the potential to serve as early predictors of colorectal cancer, providing a basis for future basic research and clinical diagnosis and treatment of the disease.

Funding

2021 Yunfu Medical and Health Science and Technology Plan Project (20210310)

Disclosure statement

The authors declare no conflict of interest.

References

[1] Wang X, Liu Z, Gong H, et al., 2022, Advances in the Study of Biomarkers for Colorectal Cancer. Journal of Colorectal & Anal Surgery, 28(1): 34–38.

[2] China Anti-Cancer Association, 2021, Targeted Therapy Committee. Expert Consensus on Molecular Testing for Colorectal Cancer Using High-Throughput Sequencing in China. Chinese Journal of Clinical Oncology, 26(3): 253–264.

[3] Fan X, Shao C, Cui L, 2019, Prediction and Validation of Hub Genes in Colorectal Cancer Based on PPI Network Analysis. Chinese Journal of Medical Library and Information Science, 28(6): 17–25.

[4] Yin X, Xue X, Zhao H, 2021, Bioinformatics Analysis for Screening Key Genes and Related Pathways in Gallbladder Cancer. Chinese Journal of Cancer Surgery, 13(1): 12–19.

[5] Chen S, Ma T, Zhang Z, et al., 2022, Mechanism of Long Non-Coding RNA LINC01419 in Regulating Proliferation, Apoptosis, Migration, and Invasion of Colorectal Cancer Cells. Anhui Medical and Pharmaceutical Journal, 26(3): 572–577, Cover 3.

[6] Zhao L, Han T, Li K, et al., 2023, Differential Analysis of Genes and Proteins Related to Carcinogenesis in Colon and Rectal Cancer. Beijing Medicine, 45(4): 284–288.

[7] Gong H, Huang X, Chen H, et al., 2022, Research Progress on Epigenetic Biomarkers for Colorectal Cancer. Journal of Colorectal & Anal Surgery, 28(1): 14–18.

[8] Yi M, Lü S, Du Q, et al., 2022, Value of Combined Detection of MMP-11 and Macrophage Inhibitory Cytokine-1 in the

Diagnosis and Prognosis Assessment of Colorectal Cancer. Cancer Progress, 20(5): 505–507, 512.

[9] Shi C, Cheng J, Liu H, et al., 2023, Expression and Clinical Value of SPOCK1, MMP11, and MMP9 in Colorectal Cancer. Chongqing Medicine, 52(1): 102–107.

[10] Chen W, 2021, Value of Detecting MMP-7 Expression by Immunohistochemistry in Assessing Safe Surgical Margins for Rectal Cancer. Heilongjiang Medicine, 34(3): 708–709.

[11] Luo S, Li T, 2022, Application Value of Serum MMP-11 and Angiopoietin-2 in Diagnosing Colorectal Cancer. Basic and Clinical Oncology, 35(3): 198–200.

[12] Li J, Chen W, Li Z, et al., 2022, Expression of MMP7, CD44v7, Survivin, and PPARγ in Serum of Colorectal Cancer Patients and Its Clinical Significance. Practical Oncology Journal, 37(7): 1068–1071.

[13] Wen L, Wang Q, Mai Q, et al., 2023, Expression of Five Genes in Stool mRNA and Their Clinical Significance in Colorectal Cancer Patients. Journal of Clinical Digestive Diseases, 35(4): 301–306.

[14] Fan G, Liang X, Gao B, et al., 2021, Relationship Between c-MET, CXCR4 Proteins, and Microvessel Density in Colorectal Cancer with Liver Metastasis. Tumor Research and Clinical, 33(11): 830–833.

[15] Cao M, Huai J, Yang X, 2023, Application Value of MACC1, c-Met, and ERCC1 Expression in Pathological Assessment of Colorectal Cancer. Contemporary Medicine Symposium, 21(9): 5–9.

[16] Li X, Zheng J, Yao R, et al., 2016, Research Progress on Prognostic Value of Human Tissue Kallikrein in Colorectal Cancer Patients. Shandong Medicine, 56(34): 99–102.

[17] Hua Q, Shen X, Xu P, 2021, Mechanism of KLK8 in Regulating Colorectal Cancer Cell Apoptosis via EGF. Chinese Cancer Journal, 31(5): 383–389.

[18] Guo P, He N, Xia Y, 2021, Data Mining Analysis of Human Tissue Kallikrein Family in Colorectal Cancer Expression and Prognosis. Cancer Progress, 19(2): 149–153, 177.