Original Article

Screening and identification of hub gene and differential gene and mutation sequence analysis of related genes in colorectal cancer based on bioinformatics analysis

Xiaofei Zhang[1], Leichang Zhang[2], Jian Lai[3], Yan Lu[1], Jianming Ye[4], Ying Wang[1]

[1]Department of Anorectal Surgery, Ganzhou Traditional Chinese Medicine Hospital Affiliated to Jiangxi University of Traditional Chinese Medicine, Ganzhou, China; [2]Department of Anorectal Surgery, The First Affiliated Hospital of Jiangxi University of Traditional Chinese Medicine, Nanchang, China; [3]Department of Gastrointestinal Surgery, The First Affiliated Hospital of Gannan Medical College, Ganzhou, China; [4]Department of Oncology, The First Affiliated Hospital of Gannan Medical College, Ganzhou, China

Contributions: (I) Conception and design: X Zhang; (II) Administrative support: Y Wang; (III) Provision of study materials or patients: All authors; (IV) Collection and assembly of data: All authors; (V) Data analysis and interpretation: All authors; (VI) Manuscript writing: All authors; (VII) Final approval of manuscript: All authors.

Correspondence to: Leichang Zhang. Department of Anorectal Surgery, The First Affiliated Hospital of Jiangxi University of Traditional Chinese Medicine, Nanchang 330006, China. Email: acme32@163.com.

Background: At present, the research of genomics is in ascendency, and using bioinformatics analysis methods to systematically explore the pathogenic genes and their regulatory mechanisms will play a great role in promoting the research of cancer. This study was to search The Cancer Genome Atlas (TCGA) database and extract inflammation-related non-coding RNA to construct a prognosis model of colon cancer and search for new immunotherapeutic targets.

Methods: The transcriptome sequencing data and clinical data of 396 colon cancer patients were downloaded from TCGA database, and the inflammation-related non-coding RNA was obtained from the non-coding RNAs in Inflammation (ncRI) database. The prognostic model was constructed by univariate Cox regression, least absolute shrinkage and selection operator (LASSO) regression, and multivariate Cox regression, and the optimal grouping threshold of risk score was determined by X-Tile software. The patients were risk stratified to further explore the differences in immune cell infiltration and biological function between the high- and low-risk groups.

Results: The TCGA dataset of colon cancer was included to screen out 120 differentially expressed genes (DEGs) that overlapped in the 2 datasets, among which 29 genes were up-regulated and 91 genes were down-regulated. Kyoto Encyclopedia of Genes and Genomes (KEGG) pathway analysis of the above 120 DEGs showed that proximal tubule sodium bicarbonate recovery, nitrogen metabolism, pancreatic fluid secretion, and PPAR signaling pathways were closely related to the occurrence of colon cancer. The expression of copper death-related genes was significantly correlated with the correlation coefficient of colon cancer ($P<0.01$). Gene Ontology analysis showed that the DEGs were mainly enriched in messenger RNA processing, RNA splicing, small G protein-mediated signal transduction, adhesion junction, mitochondrial matrix, mitochondrial protein complex, chromatin binding, small G protein binding, and Ras G protein binding, among others. KEGG analysis showed that the DEGs were enriched in the following pathways: herpes simplex virus type 1 infection, pathways of neurodegenerative diseases, Huntington's disease, prion disease, Parkinson's disease, the Ras signaling pathway, and so on.

Conclusions: The key genes closely related to colon cancer were effectively screened by the bioinformatics method, which provided a theoretical basis for further study of its mechanism.

Keywords: Colon cancer; inflammation-related noncoding RNA; immune infiltration; prognostic model; bioinformatics analysis

Submitted Oct 14, 2022. Accepted for publication Dec 01, 2022.
doi: 10.21037/jgo-22-1131
View this article at: https://dx.doi.org/10.21037/jgo-22-1131

Introduction

Colon cancer is one of the most commonly diagnosed malignancies and a leading contributor to cancer-related mortality worldwide (1). At present, tumor-node-metastasis (TNM) staging is still the gold standard for prognostic evaluation of colon cancer (2-4). However, even among colon cancer patients with the same TNM stage and similar treatment strategies, clinical outcomes vary widely, suggesting that traditional clinicopathological staging does not provide sufficient evidence to determine risk stratification in colon cancer patients (5). Since the implementation of the Human Genome Project in 1990, great achievements have been made. The rapid development of high-throughput sequencing technology is a revolutionary change to traditional sequencing, which is faster, more accurate, and more sensitive than traditional sequencing (6). The development of sequencing data analysis methods has enabled it to also provide a basis for further exploration of the expression regulation relationship in human genomics (7).

Colon cancer is one of the most commonly diagnosed malignancies and a leading contributor to cancer-related mortality worldwide (8). Among the leading cancer cases and cancer types estimated by sex in 2021, the global incidence of colorectal cancer (CRC) was the third highest among women and the fourth highest among men. CRC is also one of the main causes of cancer death in China. The incidence of CRC in Chinese men and women ranks the fifth and fourth, respectively (9). The European Prospective Cancer Cohort Survey showed that women have a higher risk of proximal colon cancer (34%) than men (25%), and this is age-related, specifically, it increases with age (35% for those <60 years of age and up to 60% for those over 70 years of age). At present, surgical resection of colon cancer is still the main treatment method (10). For non-metastatic colon cancer, total tumor resection and lymph node dissection should be performed when the surgical margin is sufficient. Generally, the distal margin of the tumor should be greater than or equal to 5 cm. At least 12 lymph nodes should be collected and analyzed for proper lymph node staging. The American Joint Committee on Cancer (AJCC) staging manual has become a benchmark for classifying cancer patients, defining prognosis, and determining the best treatment. Although TNM staging is still the gold standard for evaluating the prognosis of colon cancer (11), some patients with colon cancer have significantly different clinical outcomes despite having the same TNM stage and using similar treatment strategies, suggesting that traditional clinicopathological staging does not provide sufficient evidence to determine risk stratification in patients with colon cancer. Over the past half century, changes in dietary patterns and overall diets have led to changes in the incidence and mortality of colon cancer. For example, milk and dietary fiber are considered protective factors, whereas high intake of meat and fat increases the risk, so colon cancer can be prevented through appropriate diet-related factors. Non-dietary factors currently thought to contribute to colon cancer include smoking, long-term use of non-steroidal anti-inflammatory drugs, and genetic predisposition (such as polyposis or Lynch syndrome). According to the origin of mutations, CRC can be classified as sporadic, hereditary, or familial. Hereditary cancers account for only 5% of all CRC cases and can be broadly divided into two groups: polyposis and non-polyposis (12-14).

Colon cancer is a gastrointestinal tumor with high morbidity and mortality. Data from the 2022 Cancer Report in the United States show that there are more than 130,000 new cases of colon cancer each year and more than 50,000 deaths (15). The prognosis of colon cancer is closely related to early diagnosis and treatment. Due to the lack of effective

Highlight box

Key findings

- The inflammation-related noncoding RNA model may predict survival for patients with colon cancer prognosis.

What is known and what is new?

- The pathogenic genes and their regulatory mechanisms will play a great role in promoting the research of cancer.
- These non-coding RNAs provide clues for further study of their functions.

What is the implication, and what should change now?

- The key genes closely related to colon cancer were effectively screened by the bioinformatics method, which provided a theoretical basis for further study of its mechanism.

early diagnostic methods, many patients are diagnosed in the late stage, most of whom cannot undergo radical surgery, which worsens the prognosis (16). Moreover, the efficacy of chemotherapy in advanced colon cancer patients is difficult to ensure due to the occurrence of drug resistance. Therefore, it is necessary to find more efficient early detection molecules and new therapeutic targets in order to establish a reliable new system of prevention, early detection, and treatment (17).

At present, the research of genomics is in ascendency, and using bioinformatics analysis methods to systematically explore the pathogenic genes and their regulatory mechanisms will play a great role in promoting the research of cancer. There were reports about the bioinformatics analysis of related genes in CRC (18,19). However, fewer studies focused on the prognosis model of colon cancer and new immunotherapeutic targets. Chronic inflammation can promote the occurrence and development of tumors by affecting the immune system. Chronic inflammatory stimulation will weaken the immune response ability of the body, causing the "peaceful coexistence" of the host and tumor, and promoting immune escape. The Cancer Genome Atlas (TCGA) project was launched in 2006. After more than 10 years, it has had a profound impact on global cancer research. Non-coding RNAs in Inflammation (ncRI) provides a manually curated database for experimentally validated non-coding RNAs in inflammatory disease. Based on this, by analyzing the microarray data set of colon cancer in TCGA database, this study used a variety of bioinformatics methods to screen and further analyze the hub genes regulating colon cancer. We present the following article in accordance with the TRIPOD reporting checklist (available at https://jgo.amegroups.com/article/view/10.21037/jgo-22-1131/rc).

Methods

Data acquisition

The total of 396 colon cancer transcriptome data and corresponding clinical data were obtained from TCGA, and samples with complete long non-coding RNA (lncRNA) expression data, microRNA (miRNA) expression data, and clinical data were collected after removing samples with follow-up time less than 30 days and missing survival status. The clinical data of the patients included age, gender, TNM stage, survival time (follow-up time), and survival status. A list of human inflammation-related non-coding RNAs was downloaded from the database. The study was conducted in accordance with the Declaration of Helsinki (as revised in 2013).

Data processing and analysis

According to the annotation information about lncRNAs and coding genes in the GENCODE database (https://www.gencodegenes.org/), lncRNAs in colon cancer transcriptome expression data were extracted. Different from previous studies, which extracted lncRNAs by calculating their association with genes, this study extracted experimentally validated non-coding RNAs associated with inflammation from the non-coding RNAs in Inflammation (ncRI) database. The identification of prognostic molecules in cancer is undoubtedly a very important issue in molecular oncology. It is very important to study the differentially expressed genes (DEGs) and prognosis. An increase in cancer-related genes tends to shorten the survival time of patients, and mutations in cancer-related genes will prolong patient survival time, but there have been reports indicating that screening for DEGs and prognosis has no impact on survival. In this study, 48 inflammation-related miRNAs and 23 inflammation-related lncRNAs were screened with median absolute deviation >0.5.

The survival package in R software (The R Foundation for Statistical Computing, Vienna, Austria) was used for univariate Cox regression analysis to initially screen the prognosis-related non-coding RNA, and then the prognostic model was gradually constructed by least absolute shrinkage and selection operator (LASSO) regression and multivariate Cox regression.

Differences in immune cells infiltrating analysis

To assess the relative abundance of tumor infiltrating immune cells in different risk groups, we used the Cell-type Identification by Estimating Relative Subsets of RNA Transcripts (CIBERSORT) algorithm, a newly proposed deconvolution algorithm that estimates the relative abundance of immune cells in sample tissues from the expression profile of the tissue based on a set of reference gene (547 genes) expression values. In this study, we used CIBERSORT to evaluate the abundance of 22 immune cell types in colon cancer tissues, including T cells, B cells, neutrophils, macrophages, natural killer cells, dendritic cells, plasma cells, and eosinophils, based on TCGA gene expression profile of colon cancer. The infiltration

abundance of immune cells was calculated between the high- and low-risk groups, and the difference of infiltration abundance between the two groups was compared. The difference of infiltration abundance between the two groups was considered statistically significant (P<0.05).

Screening and identification of differential genes

We used the GGPLOT2, LimMA, and PheATMap in R software (version: 4.0.2) and other software packages to process 3 datasets. The selection criteria for DEGs were as follows: false discovery rate (FDR), and P<0.05. Fold change (FC) represented the difference multiple of DEGs. The intersection of the selected DEGs from the 3 datasets was taken, and the VennDiagram software package was used to make the VennDiagram, and the total DEGs were obtained.

Construct a protein-protein interaction (PPI) network and screen core genes

A PPI network was constructed using the Search Tool for the Retrieval of Interacting Genes/Proteins (STRING; https://string-db.org/) online database for DEGs. The PPI network was then imported into Cytoscape software (version: 3.7.2; https://cytoscape.org/), and the CytoHubba plug-in was used to screen the core genes. The top 10 genes were selected for each analysis method, and all genes were sorted according to their occurrence times in the 12 analysis methods. Finally, the top 10 genes with the most occurrence times were selected as the core genes, and the PPI network of core genes was constructed.

Prediction of relevant transcription factors

The transcription factors of each lncRNA were predicted in the CONSITE database (http://consite.genereg.net/) and their common transcription factors were screened out by intersection. The regulatory mechanism of transcription factors in colon cancer was queried through PubMed. The transcription factors of each miRNA were predicted and analyzed by CONSITE database, and the common transcription factors of the above miRNAs were screened out.

Gene enrichment analysis

The top 200 genes with the greatest correlation with the expression of related genes in colon cancer tissues were screened from Gene Expression Profiling Interactive Analysis 2 (GEPIA2) online tool, and they were imported into the WebGestalt online tool for KEGG pathway enrichment analysis. With genomic protein-coding genes as the reference set and FDR <0.05 as the criterion for judging significantly enriched pathways, the results were presented in bar charts.

Survival analysis

FPKM-UQ expression data and clinical prognosis information of colon cancer-related genes were extracted from TCGA dataset. The patients were divided into two groups: a high-risk group and a low-risk group. Kaplan-Meier (KM) analysis and log-rank test were used to analyze the overall survival (OS) between the two groups.

Statistical analysis

All statistical analyses were performed using R software. Kaplan-Meier analysis was used for survival curve analysis, and log-rank method was used to compare the survival differences among the groups. Cox proportional hazards regression model was used to analyze prognostic factors. Survival curves were plotted by R software package "SurvMiner", and a level of significance α=0.05 was considered significant.

Results

Construction of prognostic model

The prognostic non-coding RNAs were further screened by LASSO Cox regression with 10-fold cross-validation and 1,000 iterations. Finally, 5 non-coding RNAs (OIP5-AS1, NKILA, SNHG16, HSA-miR-9-2, HSA-Mir-3651) obtained by multivariate Cox regression analysis were used to construct the prognostic model. Risk score = (OIP5-AS1 ×1.997669) + (NKILA ×0.550143) – (SNHG16 ×1.66744) + (HSA-Mir-9-2 ×0.803013) + (HSA-Mir-3651 ×0.728344). Compared with tumor stage, the model had better predictive performance. X-tile software (Yale School of Medicine, New Haven, CT, USA) was used to determine the optimal grouping threshold of risk score was 5.2, and the samples were divided into high- and low-risk groups. The survival rate of patients in the high risk group was significantly lower than that in the low risk group (P<0.001) (*Figure 1*).

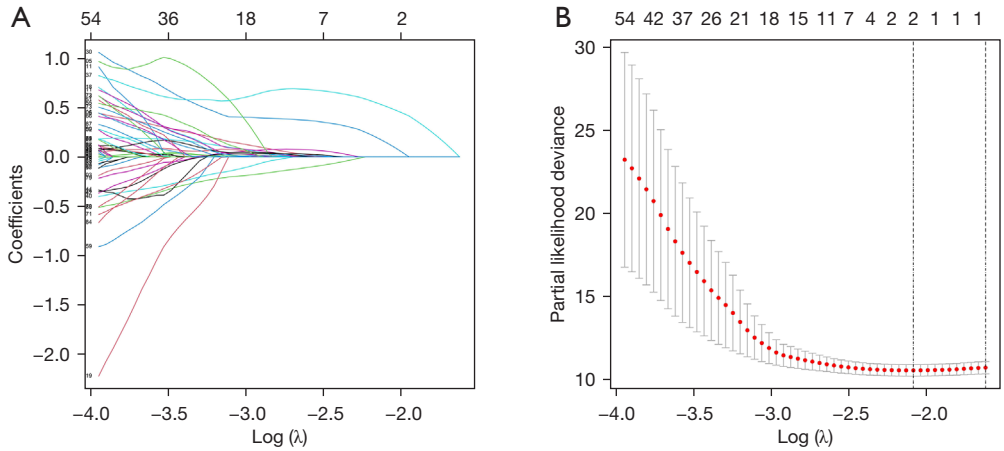

Figure 1 Construction of a prognosis model for colon cancer patients with LASSO. LASSO, least absolute shrinkage and selection operator.

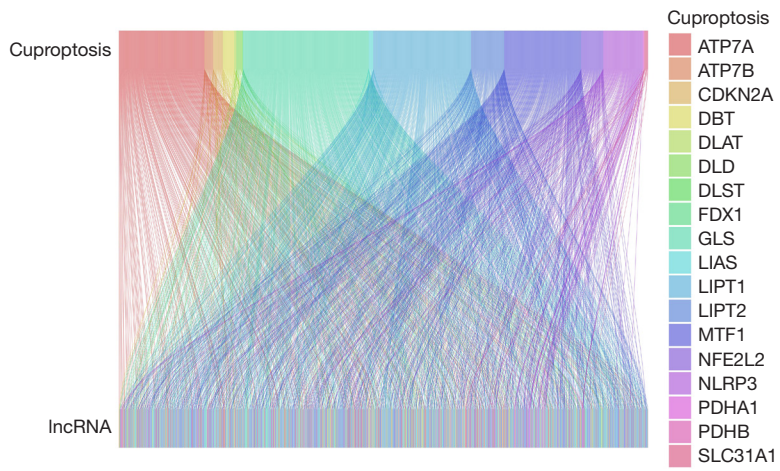

Figure 2 Sankey diagram analysis results.

Sankey diagram analysis

The Sankey diagram is the energy distribution diagram. The width of extended branches in the figure corresponds to the size of data flow, which can be used to show the distribution trend of high and low expression of a gene for different copper death mutations and patient survival in a tumor sample. We obtained raw counts and corresponding clinical information for all colon cancer RNA sequencing data (*Figure 2*).

Correlation heat map analysis

The horizontal and vertical coordinates represent genes, and different colors represent correlation coefficients (blue represents positive correlation, red represents negative correlation). The darker the color, the stronger the correlation between the two variables, *P<0.05, **P<0.01, and the asterisk represents the importance. In Spearman correlation analysis between gene and gene expression, we found that the expression of copper death-related genes was significantly correlated with the correlation coefficient of colon cancer (P<0.01) (*Figure 3*).

Gene Ontology (GO) functional enrichment analysis

GO analysis showed that the DEGs were mainly enriched in messenger RNA (mRNA) processing, RNA splicing, small G protein-mediated signal transduction, adhesion junction, mitochondrial matrix, mitochondrial protein

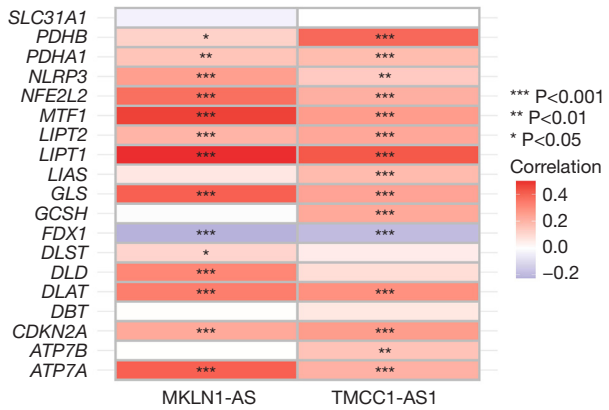

Figure 3 Risk correlation analysis of mutated genes.

complex, chromatin binding, small G protein binding, and Ras G protein binding, among others (*Figure 4*).

KEGG functional enrichment analysis

KEGG analysis showed that the DEGs were enriched in the following pathways: herpes simplex virus type 1 infection, pathways of neurodegenerative diseases, Huntington's disease, prion disease, Parkinson's disease, the Ras signaling pathway, and so on (*Figure 5*).

Genetic characterization of variable shear mutations in colon cancer

We visually analyzed the mutation data, transcriptome data, and clinical data of colon cancer copper death-related genes. We found that the mutation rate of colon cancer copper death-related genes in colon cancer patients of the two batches of samples was 82.68% and 79.12%, which suggested that primary colon cancer patients with high expression had a certain risk of metastasis. Missense mutations were the most prominent among the 536 colon cancer samples. Copper death-related mutations were associated with mononucleotide variants, mainly concentrated in nonsense mutations, insertional mutations, and deletion mutations (*Figure 6*).

Discussion

Colon cancer is one of the most prevalent and deadly cancers worldwide. Due to the heterogeneity of cancer, not all patients benefit from surgical resection, chemoradiotherapy, or chemotherapy (20). Therefore, it is important to develop personalized therapies for colon cancer. The rapid development of modern gene chip technology and bioinformatics analysis makes it possible to identify and study more novel cancer markers. Many researchers (21-24) have begun to search for important DEGs by data mining of public databases. Then, the functional enrichment analysis and experimental verification of target genes are used to further explore the secrets of life science. With the popularization of next-generation sequencing technology, it has been found that most of the transcription products do not have the function of coding protein. The biological function of these non-coding RNAs, especially in cancer, has attracted mounting attention. After small interfering RNAs (siRNAs) and microRNAs, lncRNAs are another class of transcripts that are commonly associated with human diseases (25-28).

Inflammation involves leukocyte recruitment, cytokine and chemokine production, and oxidative stress; chronic inflammation promotes tumor transformation (29). Malignancy and inflammation are closely related to each other, which in turn can affect tumor cell survival, metastasis, and angiogenesis, such as in chronic hepatitis, chronic gastritis, inflammatory bowel disease, and chronic pancreatitis (30). Cancer can also induce local or systemic inflammation, which is mediated by activation of transcription factors and production of major inflammatory cytokines (31). Inflammation associated with cancer affects cell proliferation, cell survival, angiogenesis, tumor cell migration, invasion, metastasis, and suppression of adaptive immunity. Oip5-as1 is a well-known tumor-related lncRNA, which plays a complex cellular mechanism in the development of malignant tumors. For example, not only does it affect mitosis by inhibiting the expression of cyclin G-associated kinases, but OIP5-AS1 is associated with the development of many malignancies and regulates cell proliferation and apoptosis in many cancers, including lung adenocarcinoma, breast cancer, glioma, and hepatoblastoma. SNG16 was initially identified as an oncogenic lncRNA in neuroblastoma (32). The overexpression of SNHG16 is related to the clinical and pathological features of cancer patients, and regulates cell proliferation and apoptosis. SNHG16 expression is commonly elevated in a variety of cancers, including acute lymphoblastic leukemia, bladder cancer, breast cancer, cervical cancer, and ovarian cancer, gastric cancer, and retinoblastoma. However, the opposite has been reported for SNHG16 expression in CRC and hepatocellular carcinoma. Conversely, analyses have suggested that up-regulation of SNHG16 is a common

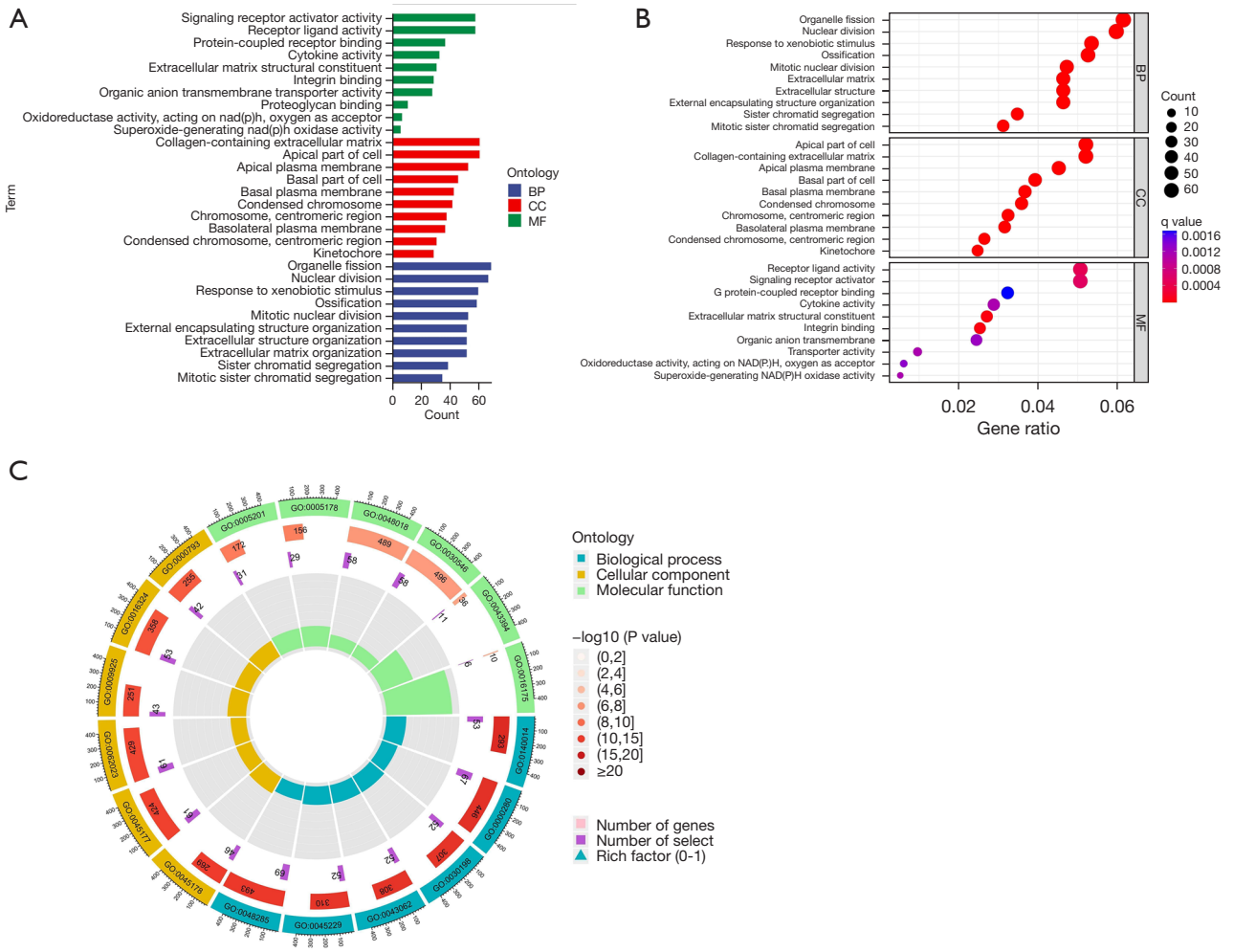

Figure 4 GO enrichment analysis of copper death-related genes in colon cancer. GO, Gene Ontology; BP, Biological process; CC, cellular component; MF, molecular function.

event in the colorectal area and affects the expression of genes involved in lipid metabolism (33). SNHG16 is located in the cytoplasm and is associated with polyribosomes. It is regulated by the Wnt signaling pathway in CRC. Silencing SNHG16 can increase cell apoptosis and inhibit cell migration. Such differences may be further caused by differences in SNHG16 quantification, isotype detection, RNA quality, percentage of cancer cells in cancer tissues, and the use of preoperative chemotherapy. Therefore, basic experiments are still needed to verify the role of SNHG16 in colon cancer. NKILA is a long noncoding RNA that interacts with NF-kB. It regulates the sensitivity of T cells to cell death by inhibiting NF-kB activity, and it can inhibit the malignant development of esophageal cancer by blocking NF-kB signaling. Epigenetic inactivation of miRNA genes is a common and pervasive phenomenon in human tumors (34).

Aberrant methylation of miRNA genes can inhibit tumor function. The presence of DNA methylation in miRNA genes can also serve as a novel biomarker. Survival analysis and univariate and multivariate analysis showed that the risk score of the model was an independent prognostic factor for colon cancer patients, and compared with the prognostic prediction effect of TNM stage, the risk score of the model had a higher predictive effect. According to the risk score, the patients were divided into high- and low-risk groups. The infiltration abundance of M2 macrophages and resting dendritic cells in the high-risk group was significantly lower than that in the low-risk group. The M2 macrophages produce many factors that contribute to tumor progression

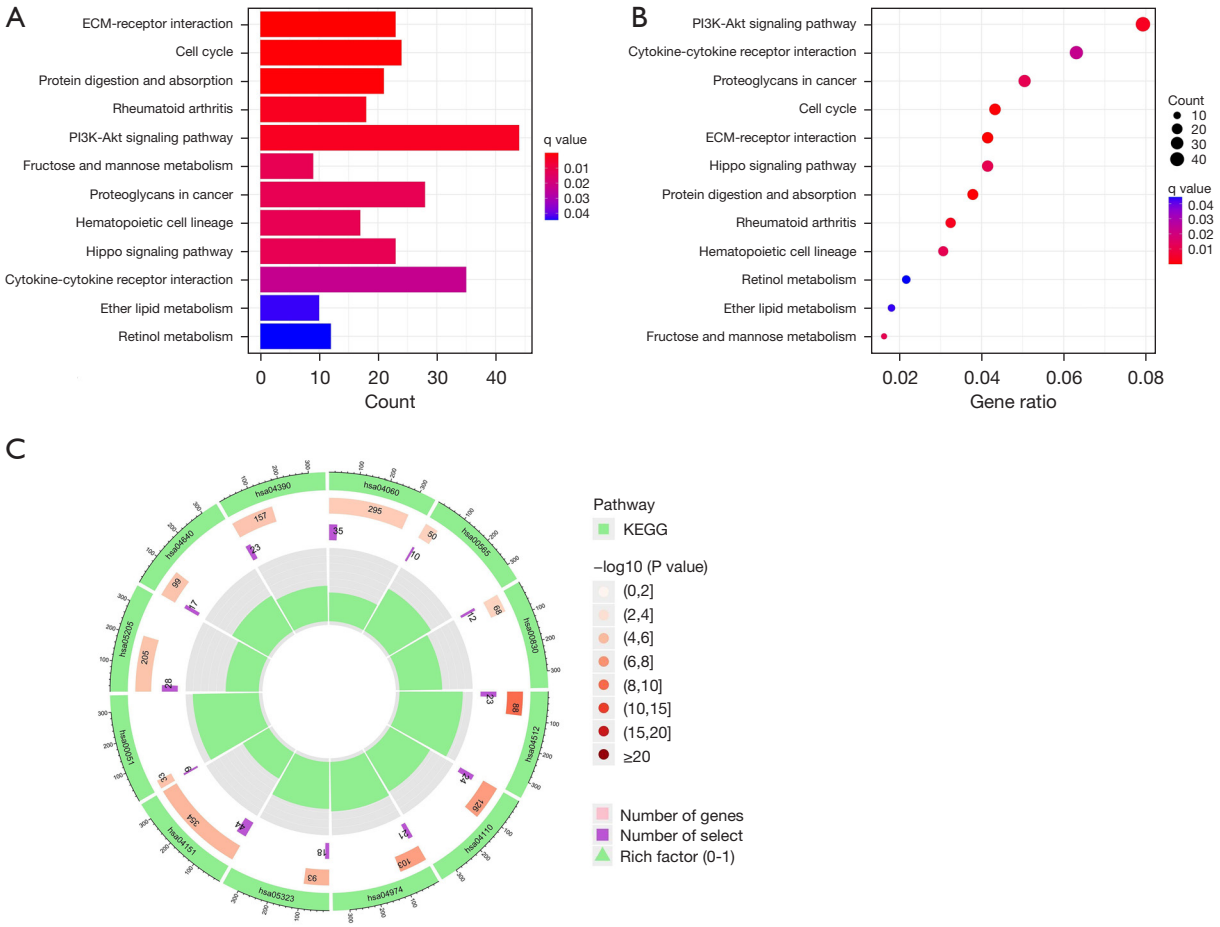

Figure 5 KEGG enrichment analysis of copper death-related genes in colon cancer. ECM, extracellular matrix; KEGG, Kyoto Encyclopedia of Genes and Genomes.

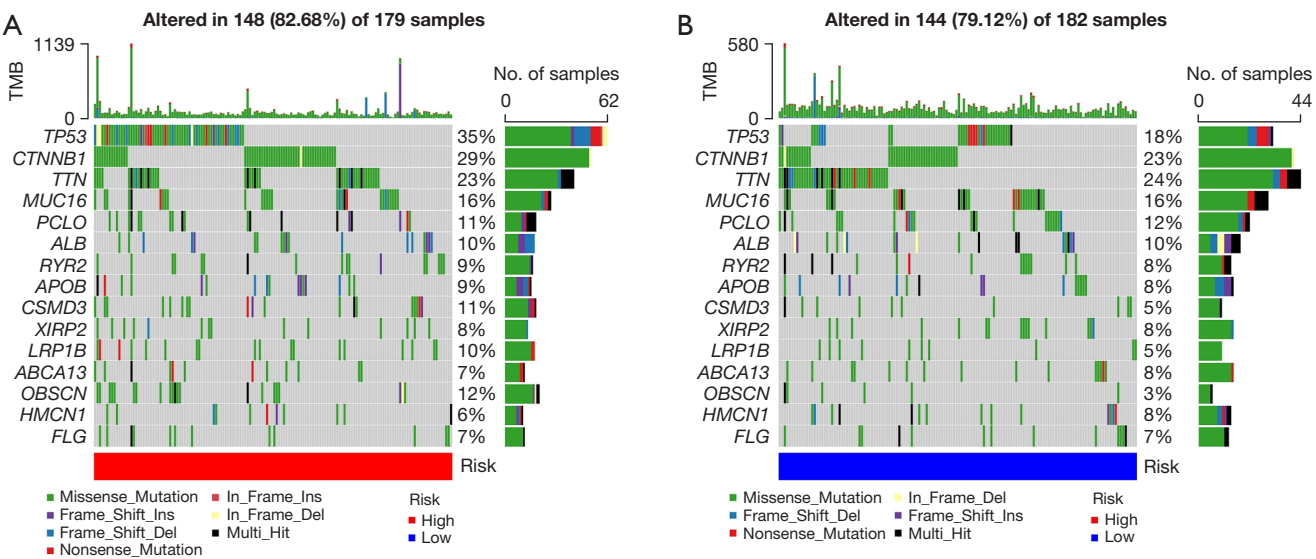

Figure 6 Genetic mutation distribution characteristics. TMB, tumor mutation burden.

and stimulate tumor growth and stromal remodeling. The different roles and prognosis of macrophages in different cancers may involve changes in the balance between M1 and M2 phenotypes, as well as the interaction of the tumor microenvironment in different cancer sites (tumor center *vs.* tumor edge). Dendritic cells are considered the strongest antigen presenting cells, stimulating naïve resting T cells and initiating primary immune responses, enhancing the immunogenicity of tumor antigens and stimulating specific cytotoxic T cells to inhibit tumorigenesis (35).

Limitations

There were some limitations to the current study. TCGA data were used, which made this study retrospective. The prognostic model was not validated in an independent sample group and lacked external validation. More samples are needed to verify the function of inflammation-related non-coding genes in the prognostic model, such as real-time polymerase chain reaction (RT-PCR), gene knockout, and other experimental methods to further explore their biological functions. The analysis of immune infiltration in this study was based on transcriptomic data of whole tissue samples, but the difference of infiltration in different tumor sites (tumor center *vs.* tumor edge) also affected the prognosis of patients. Finally, experimental data was further needed to validate the representative hub gene.

Conclusions

To sum up, our inflammation-related noncoding RNA model may predict survival for patients with colon cancer prognosis. At the same time, these non-coding RNAs provide clues for further study of their functions, and are expected to become new molecular markers for the diagnosis or treatment of colon cancer.

Acknowledgments

Funding: The study was supported by the National Natural Science Foundation of China (No. 82260938) and the Jiangxi Natural Science Foundation (No. 2020BABL206135).

Footnote

Reporting Checklist: The authors have completed the TRIPOD reporting checklist. Available at https://jgo.amegroups.com/article/view/10.21037/jgo-22-1131/rc

Conflicts of Interest: All authors have completed the ICMJE uniform disclosure form (available at https://jgo.amegroups.com/article/view/10.21037/jgo-22-1131/coif). The authors have no conflicts of interest to declare.

Ethical Statement: The authors are accountable for all aspects of the work in ensuring that questions related to the accuracy or integrity of any part of the work are appropriately investigated and resolved. The study was conducted in accordance with the Declaration of Helsinki (as revised in 2013).

References

1. González-Ramírez M, Ballaré C, Mugianesi F, et al. Differential contribution to gene expression prediction of histone modifications at enhancers or promoters. PLoS Comput Biol 2021;17:e1009368.
2. Fernandopulle MS, Prestil R, Grunseich C, et al. Transcription Factor-Mediated Differentiation of Human iPSCs into Neurons. Curr Protoc Cell Biol 2018;79:e51.
3. Waldman BS, Schwarz D, Wadsworth MH 2nd, et al. Identification of a Master Regulator of Differentiation in Toxoplasma. Cell 2020;180:359-372.e16.
4. Tran N, Broun A, Ge K. Lysine Demethylase KDM6A in Differentiation, Development, and Cancer. Mol Cell Biol 2020;40:e00341-20.
5. Ullah R, Naz A, Akram HS, et al. Transcriptomic analysis reveals differential gene expression, alternative splicing, and novel exons during mouse trophoblast stem cell differentiation. Stem Cell Res Ther 2020;11:342.
6. Wu L, Zheng Y, Liu J, et al. Comprehensive evaluation of the efficacy and safety of LPV/r drugs in the

treatment of SARS and MERS to provide potential treatment options for COVID-19. Aging (Albany NY) 2021;13:10833-52.
7. Kim DS, Risca VI, Reynolds DL, et al. The dynamic, combinatorial cis-regulatory lexicon of epidermal differentiation. Nat Genet 2021;53:1564-76.
8. Uribe-Lewis S, Carroll T, Menon S, et al. 5-hydroxymethylcytosine and gene activity in mouse intestinal differentiation. Sci Rep 2020;10:546.
9. Xie J, Yin Y, Yang F, et al. Differential Network Analysis Reveals Regulatory Patterns in Neural Stem Cell Fate Decision. Interdiscip Sci 2021;13:91-102.
10. Chisolm DA, Weinmann AS. Metabolites, genome organization, and cellular differentiation gene programs. Curr Opin Immunol 2018;51:62-7.
11. Yang S, Kam WR, Liu Y, et al. Comparative influence of differentiation and proliferation on gene expression in human meibomian gland epithelial cells. Exp Eye Res 2021;205:108452.
12. da Silva RA, Fuhler GM, Janmaat VT, et al. HOXA cluster gene expression during osteoblast differentiation involves epigenetic control. Bone 2019;125:74-86.
13. Ahmed M, Min DS, Kim DR. Curated gene expression dataset of differentiating 3T3-L1 adipocytes under pharmacological and genetic perturbations. Adipocyte 2020;9:600-8.
14. Wu L, Zhong Y, Wu D, et al. Immunomodulatory Factor TIM3 of Cytolytic Active Genes Affected the Survival and Prognosis of Lung Adenocarcinoma Patients by Multi-Omics Analysis. Biomedicines 2022;10:2248.
15. Li J, Ma X, Chakravarti D, et al. Genetic and biological hallmarks of colorectal cancer. Genes Dev 2021;35:787-820.
16. Heinimann K. Hereditary Colorectal Cancer: Clinics, Diagnostics and Management. Ther Umsch 2018;75:601-6.
17. Baidoun F, Elshiwy K, Elkeraie Y, et al. Colorectal Cancer Epidemiology: Recent Trends and Impact on Outcomes. Curr Drug Targets 2021;22:998-1009.
18. Ebadfardzadeh J, Kazemi M, Aghazadeh A, et al. Employing bioinformatics analysis to identify hub genes and microRNAs involved in colorectal cancer. Med Oncol 2021;38:114.
19. Wang Y, Huang L, Shan N, et al. Establishing a three-miRNA signature as a prognostic model for colorectal cancer through bioinformatics analysis. Aging (Albany NY) 2021;13:19894-907.
20. Fan A, Wang B, Wang X, et al. Immunotherapy in colorectal cancer: current achievements and future perspective. Int J Biol Sci 2021;17:3837-49.
21. Smith KM, Desai J. Nivolumab for the treatment of colorectal cancer. Expert Rev Anticancer Ther 2018;18:611-8.
22. Saus E, Iraola-Guzmán S, Willis JR, et al. Microbiome and colorectal cancer: Roles in carcinogenesis and clinical potential. Mol Aspects Med 2019;69:93-106.
23. Wu L, Zheng Y, Ruan X, et al. Long-chain noncoding ribonucleic acids affect the survival and prognosis of patients with esophageal adenocarcinoma through the autophagy pathway: construction of a prognostic model. Anticancer Drugs 2022;33:e590-603.
24. Siegel RL, Jakubowski CD, Fedewa SA, et al. Colorectal Cancer in the Young: Epidemiology, Prevention, Management. Am Soc Clin Oncol Educ Book 2020;40:1-14.
25. Picard E, Verschoor CP, Ma GW, et al. Relationships Between Immune Landscapes, Genetic Subtypes and Responses to Immunotherapy in Colorectal Cancer. Front Immunol 2020;11:369.
26. Dai Z, Zhang J, Wu Q, et al. The role of microbiota in the development of colorectal cancer. Int J Cancer 2019;145:2032-41.
27. Patel SG, Karlitz JJ, Yen T, et al. The rising tide of early-onset colorectal cancer: a comprehensive review of epidemiology, clinical features, biology, risk factors, prevention, and early detection. Lancet Gastroenterol Hepatol 2022;7:262-74.
28. IJsselsteijn ME, Sanz-Pamplona R, Hermitte F, et al. Colorectal cancer: A paradigmatic model for cancer immunology and immunotherapy. Mol Aspects Med 2019;69:123-9.
29. Messersmith WA. NCCN Guidelines Updates: Management of Metastatic Colorectal Cancer. J Natl Compr Canc Netw 2019;17:599-601.
30. Jin K, Ren C, Liu Y, et al. An update on colorectal cancer microenvironment, epigenetic and immunotherapy. Int Immunopharmacol 2020;89:107041.
31. Zhang X, Hu F, Li G, et al. Human colorectal cancer-derived mesenchymal stem cells promote colorectal cancer progression through IL-6/JAK2/STAT3 signaling. Cell Death Dis 2018;9:25.
32. Yu H, Hemminki K. Genetic epidemiology of colorectal cancer and associated cancers. Mutagenesis 2020;35:207-19.
33. Mitsala A, Tsalikidis C, Pitiakoudis M, et al. Artificial Intelligence in Colorectal Cancer Screening, Diagnosis and Treatment. A New Era. Curr Oncol 2021;28:1581-607.
34. Wu L, Zhong Y, Yu X, et al. Selective poly adenylation

predicts the efficacy of immunotherapy in patients with lung adenocarcinoma by multiple omics research. Anticancer Drugs 2022;33:943-59.
35. Bhullar DS, Barriuso J, Mullamitha S, et al. Biomarker concordance between primary colorectal cancer and its metastases. EBioMedicine 2019;40:363-74.

(English Language Editor: J. Jones)

Cite this article as: Zhang X, Zhang L, Lai J, Lu Y, Ye J, Wang Y. Screening and identification of hub gene and differential gene and mutation sequence analysis of related genes in colorectal cancer based on bioinformatics analysis. J Gastrointest Oncol 2022;13(6):3056-3066. doi: 10.21037/jgo-22-1131

