Observational Study

Medicine

OPEN

RRP9 and DDX21 as new biomarkers of colorectal cancer

Xiaoqian Chi, MD[a,*], Ning Yang, MD[b], Yabin Liu, MD[c]

Abstract

Colorectal cancer originates from the epithelium of the large intestine and is a common malignant tumor in the gastrointestinal tract. However, the relationship between RRP9 and DDX21 and colorectal cancer (CRC) remains unclear. GSE134834, GSE206800, and GSE209892 profiles for CRC were downloaded from the gene expression omnibus database generated using GPL20115 and GPL23126. Differentially expressed genes (DEGs) were screened and weighted gene co-expression network analysis was performed. The construction and analysis of protein–protein interaction network. Functional enrichment analysis and gene set enrichment analysis were performed. Gene expression heat map was drawn and immune infiltration analysis was performed. Comparative toxicogenomics database analysis were performed to find the disease most related to the core gene. TargetScan was used to screen miRNAs regulating central DEGs. One thousand three hundred eighty DEGs were identified. According to gene ontology analysis, they were mainly concentrated in signal receptor activity regulation and metal titanase activity. Kyoto encyclopedia of gene and genome analysis showed that they mainly focused on IL17 signal pathway, PPAR signal pathway, protein digestion, and absorption, and the interaction of viral proteins with cytokines and cytokine receptors. The intersection of enrichment items and GOKEGG enrichment items of differentially expressed genes is mainly concentrated in PPAR signal pathway and the interaction of viral proteins with cytokines and cytokine receptors. The protein–protein interaction network obtained 16 core genes (MAD2L1, MELK, TPX2, UBE2C, RFC4, PLK1, RACGAP1, DKC1, DDX21, L Y AR, WDR3, RRP9, WDR43, NOLC1, BRIX1, and GTPBP4). Heat map of gene expression showed that core genes (TPX2, UBE2C, RFC4, PLK1, DKC1, LYAR, WDR3, NOLC1, and BRIX1) were not significantly differentially expressed between CRC and normal tissue samples. Core genes (MAD2L1, MELK, RACGAP1, RRP9, WDR43, DDX21, and GTPBP4) were highly expressed in CRC tissue samples and lowly expressed in normal tissue samples. Comparative toxicogenomics database analysis showed that 7 genes (MAD2L1, MELK, RACGAP1, RRP9, WDR43, DDX21, and GTPBP4) were related to necrosis, inflammation, tumor, precancerous symptoms, hemorrhage, and weightlessness. RRP9 and DDX21 are highly expressed in CRC. The higher the expression level of RRP9 and DDX21, the worse the prognosis.

Abbreviations: CRC = colorectal cancer, CTD = comparative toxicogenomics database, DEGs = differential epigenetic genes, GO = gene ontology, GSEA = gene set enrichment analysis, KEGG = Kyoto encyclopedia of gene and genome, PPI = protein–protein interaction, rRNA = ribosomal RNA, snoRNP = small nucleolar ribonucleoprotein, STRING = search tool for the retrieval of interacting genes, TOM = topological overlap matrix, WGCNA = weighted gene co-expression network analysis.

Keywords: biomarkers, colorectal cancer, DDX21, prognosis, RRP9

1. Introduction

Colorectal cancer (CRC) is one of the most common cancers in the world, with 1 million to 2 million new cases diagnosed each year. It has become the third most common cancer and the fourth leading cause of cancer-related death, with 700,000 deaths each year, second only to lung, liver, and stomach cancer.[1] Colorectal cancer is the second most common cancer in women (9.2%) and the third most common cancer in men (10%) by sex. From 1990 to 2012, the incidence of CRC increased by more than 200,000 new cases per year. Most CRC cases have been found in western countries (55%), but this trend is changing due to the rapid development in some countries in the past few years.[2–4] China has a low incidence of CRC, but its incidence has an increasing trend in many areas.

This research received no specific grant from any funding agency in the public, commercial, or not-for-profit sectors.

The authors have no conflicts of interest to disclose.

The datasets generated during and/or analyzed during the current study are available from the corresponding author on reasonable request.

This study was approved by the Fourth Affiliated Hospital of Hebei Medical University.

[a] Department of General Surgery, Beijing Rehabilitation Hospital, Capital Medical University, Beijing, China, [b] Department of General Surgery, The First Affiliated Hospital, Hebei Medical University, Shijiazhuang, China, [c] Department of General Surgery, The Fourth Affiliated Hospital, Hebei Medical University, Shijiazhuang, China.

** Correspondence: Xiaoqian Chi, Department of General Surgery, Beijing Rehabilitation Hospital, Capital Medical University, Beijing Shi Jingshan District Badachu West Xiazhuang, 100144 (e-mail: 18210034207@163.com).*

How to cite this article: Chi X, Yang N, Liu Y. RRP9 and DDX21 as new biomarkers of colorectal cancer. Medicine 2023;102:43(e34384).

Received: 11 March 2023 / Received in final form: 26 June 2023 / Accepted: 27 June 2023

http://dx.doi.org/10.1097/MD.0000000000034384

The disease mostly occurs in men over middle age, especially in the age of 40 to 70 years old. The ratio of male and female is about 2:1.[5] Colorectal cancer is a common malignant tumor in the gastrointestinal tract, the early symptoms are not obvious, with the increase of the degree of cancer, the symptoms such as change of defecation habit, hematochezia, diarrhea, alternation of diarrhea and constipation, local abdominal pain, and so on. Late stage shows anemia, weight loss, and other systemic symptoms. Mutations in specific genes can lead to the onset of CRC, just like other types of cancer. These mutations can be found in oncogenes, tumor suppressor genes, and genes related to DNA repair mechanism. According to the origin of the mutation, CRC can be divided into sporadic, hereditary, and familial.[6–8] Like other malignant tumors, the cause of this disease is still unknown and can occur in any part of the colon or rectum, but rectum and sigmoid colon are the most common, and the rest are found in cecum, ascending colon, descending colon, and transverse colon in turn. Most of the cancers were adenocarcinoma, and a few were squamous cell carcinoma and mucinous carcinoma. The disease may be related to genetic factors, chromosome abnormalities, gene fusion, and other factors. Therefore, it is particularly important to study the molecular mechanism of CRC.

As an important part of the development of life science, bioinformatics has been at the forefront of life science and technology research. In recent years, China's biotechnology has developed by leaps and bounds, and bioinformation resources have also grown explosively. Bioinformatics reveals the biological significance represented by big data, which is a bridge between data and clinic. Represented by the analysis and reporting of gene detection data, bioinformatics plays an important role in tumor treatment.[9,10]However, the relationship between RRP9, DDX21, and CRC is not clear.

Therefore, this paper intends to use bioinformatics technology to mine the core genes between CRC and normal tissues, and carry out enrichment analysis and pathway analysis. Using common data sets to verify the significant role of RRP9 and DDX21 in CRC.

2. Methods

2.1. Colorectal cancer data set

In this study, CRC data sets GSE134834, GSE206800, and GSE209892 configuration files were downloaded from the gene expression omnibus data library (http://www.ncbi.nlm.nih.gov/geo/) generated by GPL20115 and GPL23126. GSE134834 includes 5 CRC and 5 normal tissue samples, GSE206800 includes 11 CRC and 4 normal tissue samples, and GSE209892 includes 6 CRC and 6 normal tissue samples to identify differentially expressed genes (DEGs) in CRC.

2.2. De-batch processing

For merging and debatching of multiple datasets, we first use R software package to merge datasets GSE134834, GSE206800, and GSE209892. For the combination of multiple data sets, we first use the R software package in Silico Merging [DOD:10.1186/1471-2105-13335] to merge the data sets to get the merge matrix. Furthermore, we use the remove Batch Effect function of the R software package limma (version 3.42.2,) to remove the batch effect, and finally obtain the matrix after removing the batch effect, and apply it to the follow-up analysis.

2.3. Screening of DEGs screening of differentially expressed genes (DEGs)

R packet "limma" is used for probe summation background correction of merging matrices of GSE134834, GSE206800, and GSE209892. The Benjamini-Hochberg method is used to adjust the original P value. Use false discovery rate (FDR) to calculate fold change. The cutoff standard of DEG is $P < .05$. And make a visual representation of the volcano map.

2.4. Weighted gene co-expression network analysis (WGCNA)

First, we use the de-batch and post-merge matrices of GSE134834, GSE206800, and GSE209892 to calculate the median absolute deviation of each gene, eliminate the top 50% of the smallest genes in median absolute deviation, remove the outlier genes and samples by the good Samples Genes method of R software package WGCNA, and then use WGCNA to construct scale-free co-expression network. Specifically, first, Pearson correlation matrix and average linkage method are performed on all paired genes, and then. Construct a weighted adjacency matrix (Pearson correlation between C_mn = Gene_m and Gene_n) using the power function Abelimn = | C_mn | ^ β. Adjacency between A_mn = Genem and Genen). β is a soft threshold parameter, which can emphasize the strong correlation between genes and weaken the influence of weak correlation and negative correlation. After selecting the power of 10, the adjacency is transformed into a topological overlap matrix (TOM), which can measure the network connectivity of a gene. The network connectivity is defined as the sum of its adjacency with all other genes, which is used for the network gene ratio, and the corresponding dissimilarity degree (1-TOM) is calculated. In order to classify the genes with similar expression profiles into gene modules, the average linkage hierarchical clustering is carried out according to the dissimilarity measure based on TOM, and the minimum size (genome) of the gene tree is 30. Set the sensitivity to: 3. In order to further analyze the module, we calculate the difference of the feature genes of the module, select a cut line for the module tree, and merge some modules. It is worth noting that the gray module is considered to be a collection of genes that cannot be assigned to any module.

2.5. Construction and analysis of protein–protein interaction (PPI) network

The Search Tool for the Retrieval of Interacting Genes (STRING, https://cn.string-db.org/) aims to collect, score and integrate all publicly available sources of PPI information, and to supplement these sources by calculating predictions. In this study, the list of differential genes was input into STRING database to construct a PPI network for predicting core genes (confidence > 0.4). Cytoscape software (https://cytoscape.org. Version 3.7.1)[11] can provide biologists with biological network analysis and two-dimensional (2D) visualization. In this study, the PPI network formed by STRING database is visualized and core genes are predicted by Cytoscape software. First of all, we import the PPI network into the cytoscape software, find the module with the best correlation through MCODE, and calculate the ten genes with the best correlation through 3 algorithms (MCC, MNC, and DMNC). After visualization, we derive the list of core genes.

2.6. Functional enrichment analysis

Gene ontology (GO) and Kyoto encyclopedia of gene and genome (KEGG) analysis are computational methods for evaluating the function and biological pathways of genetics. In this study, the list of differential genes screened by Wayne map was input into KEGG rest API (https://www.kegg.jp/kegg/rest/keggapi.html) obtained the latest KEGG pathway gene annotation, which was used as the background, the genes were mapped to the background set, and the R software package clusterProfiler

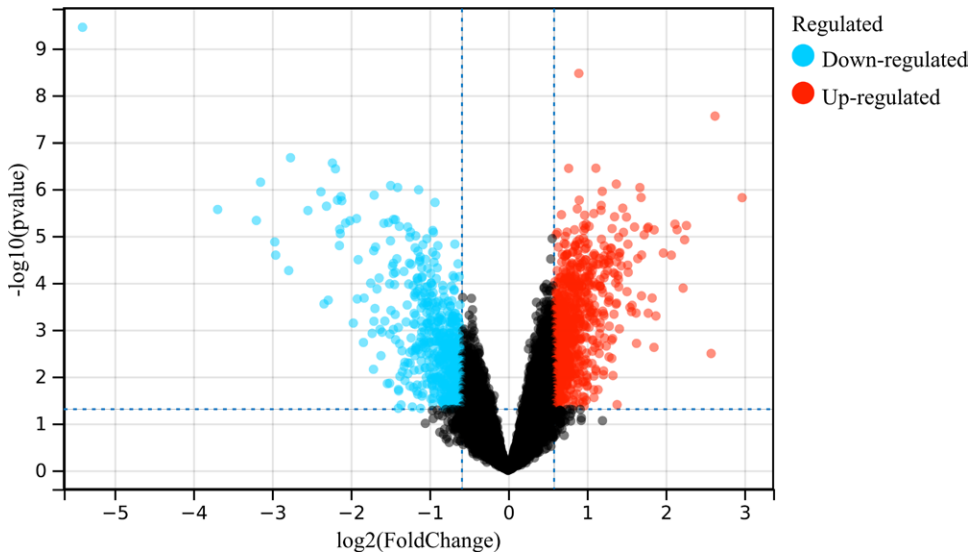

Figure 1. Analysis of differentially expressed genes. 1380 DEGs were identified. DEGs = differentially expressed genes.

(version 3.14.3) was used for enrichment analysis to obtain the results of gene set enrichment. The GO annotation of genes in R software package org.Hs.e.g..db (version 3.1.0) was also used as a background to map genes into the background set. The minimum gene set was set to 5, and the maximum gene set was set to 5000 value of <.05and a FDR of <.25. It was considered to be a statistically significant measure.

In addition, Metascape database can provide comprehensive gene list annotation and analysis resources, and can be visually exported. We used Metascape (http://metascape.org/gp/index.html) database to analyze the functional enrichment of the above differential gene list and derive it.

2.7. GSEA

For gene set enrichment analysis (GSEA), we obtained the GSEA software (version 3.0) from GSEA (DOI:10.1073/pnas.0506580102, http://software.broadinstitute.org/gsea/index.jsp). The samples were divided into 2 groups according to CRC and normal tissue, and from Molecular Signatures Database (DOI:10.1093/bioinformatics/btr260) downloaded the c2.cp.kegg.v7.4.symbols.gmt subset. In order to evaluate the related pathways and molecular mechanisms, based on gene expression profile and phenotypic grouping, the minimum gene set is 5, the maximum gene set is 5000, and a thousand resampling times, *P*-value of <.05 and FDR of <.25 is considered to be statistically significant. The whole genome was analyzed by GO and KEGG. Developed by GSEA.

2.8. Gene expression heat map

We use R-packet heatmap to map the expression of core genes found by 2 algorithms in PPI network in GSE134834, GSE206800, and GSE209892, and to visualize the difference of core gene expression between CRC and normal tissue samples.

2.9. Immune infiltration analysis

CIBERSORT (http://CIBERSORT.stanford.edu/) is a very common method for calculating immune cell infiltration. The LM22 gene file is used to define 22 immune cell subsets. We applied the integrated bioinformatics method, used CIBERSORT software package to analyze the de-batch merging matrix of GSE134834, GSE206800, and GSE20989, and used the principle of linear support vector regression to deconvolution the expression matrix of immune cell subtypes to estimate the abundance of immune cells. At the same time, the samples with sufficient confidence were selected by using confidence $P < .05$ as the truncation standard.

2.10. CTD analysis

Comparative toxicogenomics database (CTD, http://ctdbase.org/) integrates a large number of chemical substances, genes, functional phenotypes and disease interaction data, which provides great convenience for the study of disease-related environmental exposure factors and drug potential mechanism. We input the core gene into the CTD website, find the disease most related to the core gene, and use Excel to draw the radar map of the differential expression of each gene.

2.11. miRNA

TargetScan (www.targetscan.org) is an online database for predicting and analyzing miRNA and target genes. In our study, TargetScan was used to screen the miRNA that regulates central DEG.

3. Results

3.1. Analysis of differentially expressed genes

In this study, according to the set cutoff value and the de-batch merging matrix of GSE134834, GSE206800, and GSE209892, a total of 1380 DEGs were identified (Fig. 1).

3.2. Functional enrichment analysis

1.3.2.. DEGs. We carried out GO and KEGG analysis of these differentially expressed genes. According to GO analysis, they were mainly concentrated in signal receptor activity regulation and metal titanase activity. In KEGG analysis, they mainly focused on IL17 signal pathway, PPAR signal pathway, protein digestion and absorption, and the interaction of viral proteins with cytokines and cytokine receptors (Figure 2A, B, E, and F).

3.2.2. GSEA. In addition, we carried out GSEA of the whole genome in order to find out the possible enrichment items in

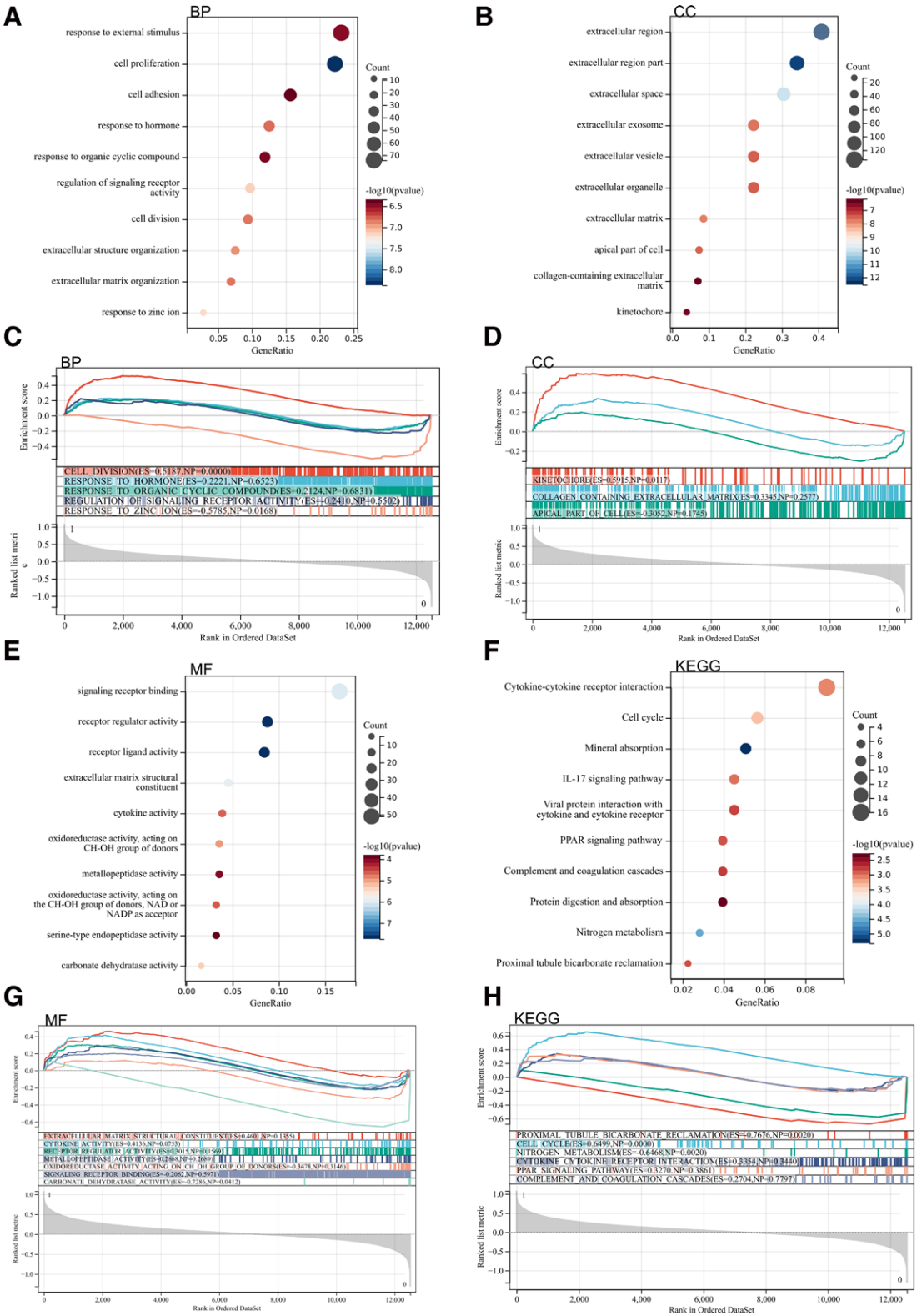

Figure 2. Functional enrichment analysis. (A, B, E, and F) of DEGs. (C, D, G, and H) GSEA. DEGs = differentially expressed genes, GSEA = gene set enrichment analysis.

non-differentially expressed genes and verify the results of differentially expressed genes. The intersection of enrichment items and GOKEGG enrichment items of differentially expressed genes is shown in the figure, which is mainly concentrated in PPAR signal pathway and the interaction of viral proteins with cytokines and cytokine receptors (Fig. 2C, D, G, and H).

3.3. Metascape enrichment analysis

In the metascape enrichment project, PIDFOXM1 pathway and gastric cancer network can be seen in the GO enrichment project (Fig. 3A). At the same time, we also output the enrichment network colored by enrichment term and *P*-value (Figs. 3B, C and 4), which visually shows the correlation and confidence of each enrichment project.

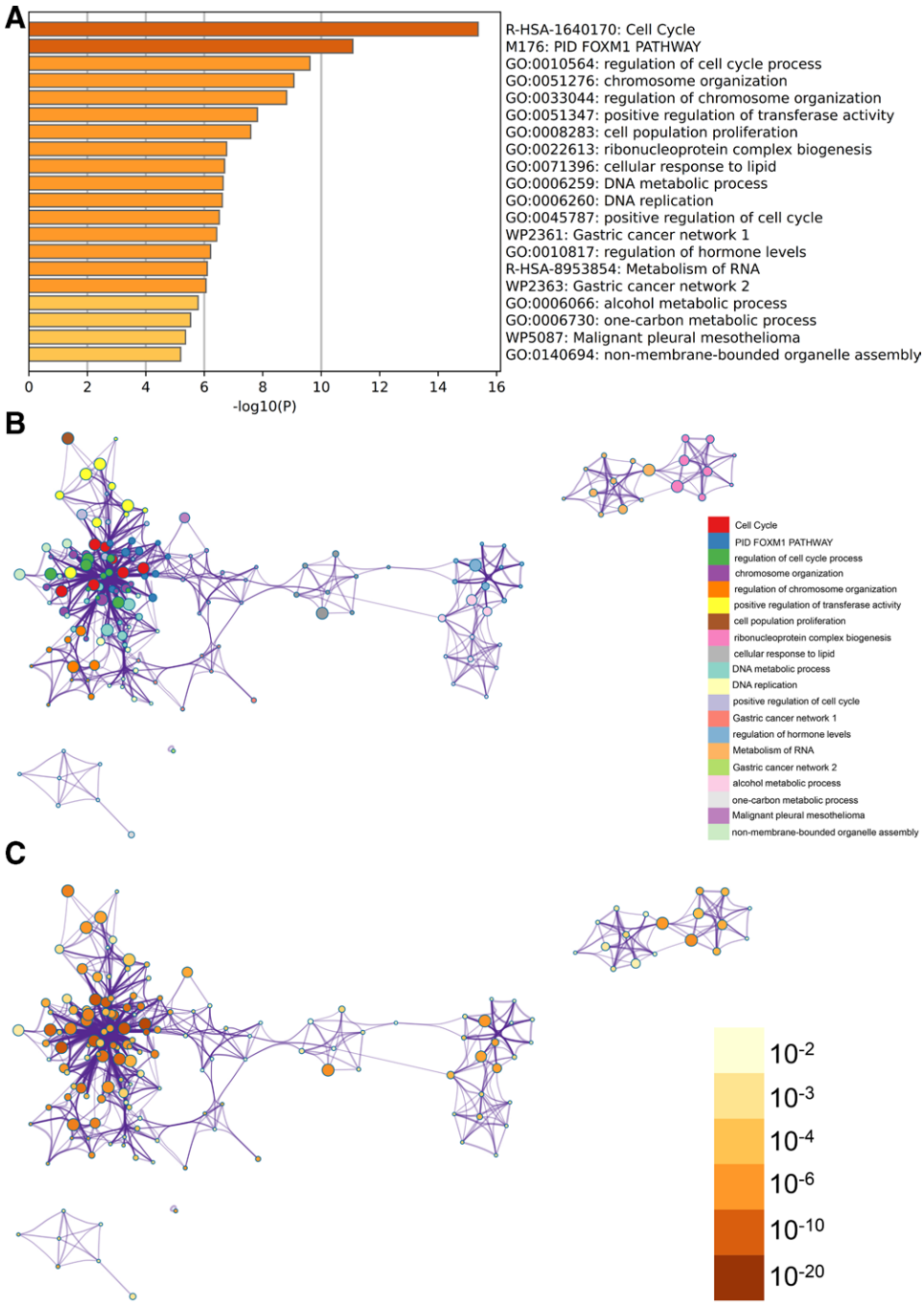

Figure 3. Metascape enrichment analysis. (A) PIDFOXM1 pathway and gastric cancer network can be seen in the GO enrichment project (B) enrichment networks colored by enrichment terms (C) enrichment networks colored by *P* values. GO = gene ontology.

3.4. WGCNA

The selection of soft threshold power is an important step in WGCNA. The network topology is analyzed to determine the soft threshold power. The soft threshold power in the WGCNA analysis is set to 9, which is the lowest power of the scale-free topology fitting index of 0.9 (Fig. 5A and B). The hierarchical clustering tree of all genes was constructed, and 3 important modules were generated (Fig. 5C). Then analyze the interaction between these modules (Fig. 5D). The module-phenotypic correlation heat map (Fig. 6A) and the GS–MM correlation scatter map of related hub genes (Fig. 6B–G) were generated. We calculated the correlation between module feature vector and gene expression to obtain MM. According to the cutoff criteria (| MM | > 0.8), 8 genes with high connectivity were identified as hub genes in the clinically significant module.

We also draw the Wayne diagram of the differential genes screened by WGCNA and DEGs and take the intersection to create and analyze the PPI network (Fig. 5E).

3.5. Construction and analysis of PPI network

DEGs's PPI network was constructed from STRING online database and analyzed by Cytoscape software (Fig. 7A). The core gene cluster was obtained (Fig. 7B and F). Three different algorithms were used to identify the central gene (Fig. 7C–E, G–I), and the Wayne graph was used to merge (Fig. 7J and K)).

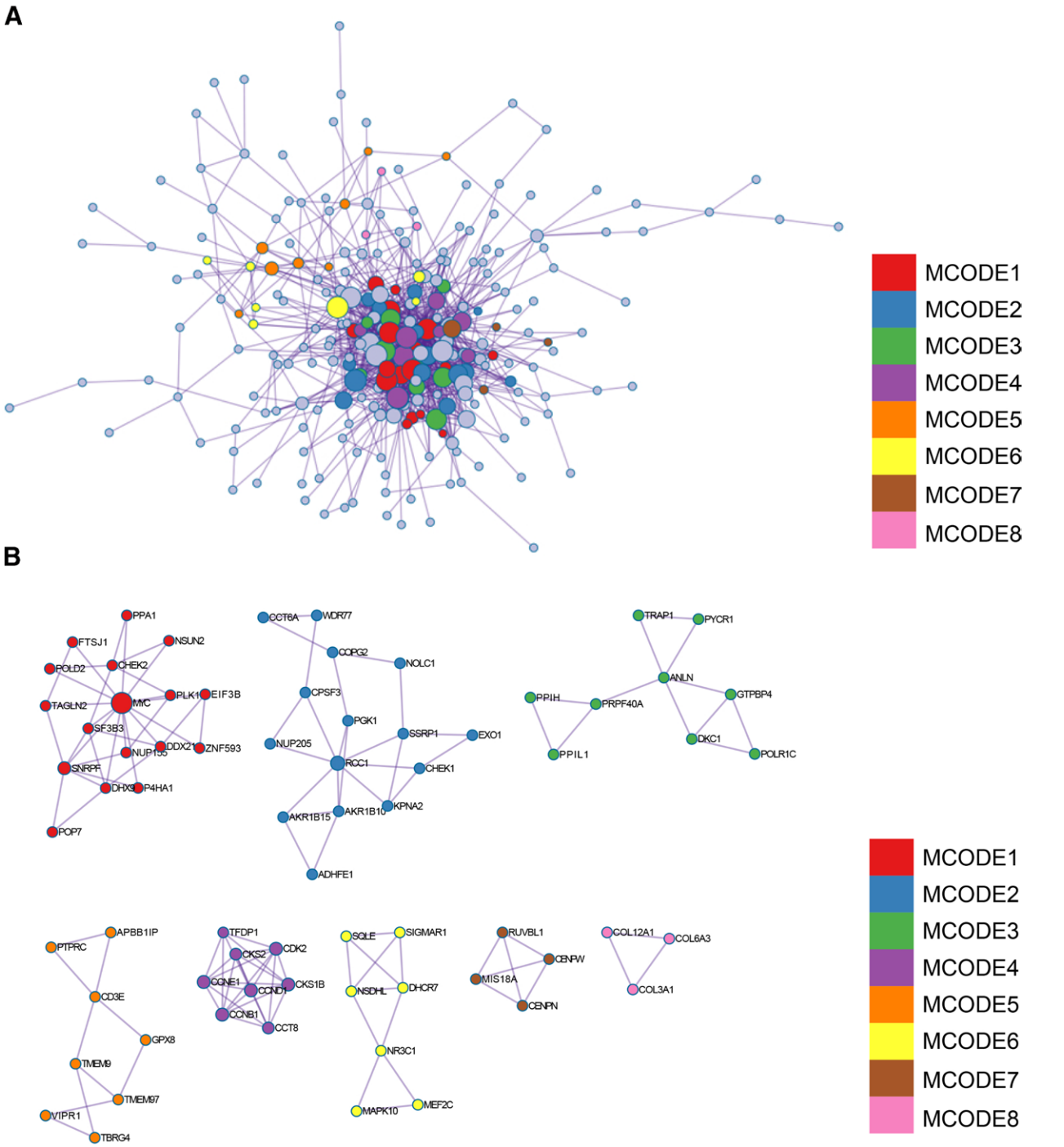

Figure 4. Metascape enrichment analysis.

Sixteen core genes (MAD2L1, MELK, TPX2, UBE2C, RFC4, PLK1, RACGAP1, DKC1, DDX21, L Y AR, WDR3, RRP9, WDR43, NOLC1, BRIX1, and GTPBP4) were obtained.

At the same time, we also use the metascape website to output the PPI network, and identify the core module to verify the PPI network results in STRING. Among them, MAD2L1, MELK, RACGAP1, RRP9, WDR43, DDX21, and GTPBP4 were identified as core genes.

3.6. Gene expression heat map

We visualized the heat map of the expression of core genes in the samples (Fig. 8). We found that there was no significant difference in the expression of core genes (TPX2, UBE2C, RFC4, PLK1, DKC1, LYAR, WDR3, NOLC1, and BRIX1) between CRC and normal tissues. Core genes (MAD2L1, MELK, RACGAP1, RRP9, WDR43, DDX21, and GTPBP4) are highly expressed in CRC tissue samples and low in normal tissue samples.

3.7. Immune infiltration analysis

We used CIBERSORT software package to analyze the de-batch merging matrix of GSE134834, GSE206800, and GSE209892. Under 95% confidence, we obtained the proportion of immune cells of the whole gene expression matrix (Fig. 9A) and the heat

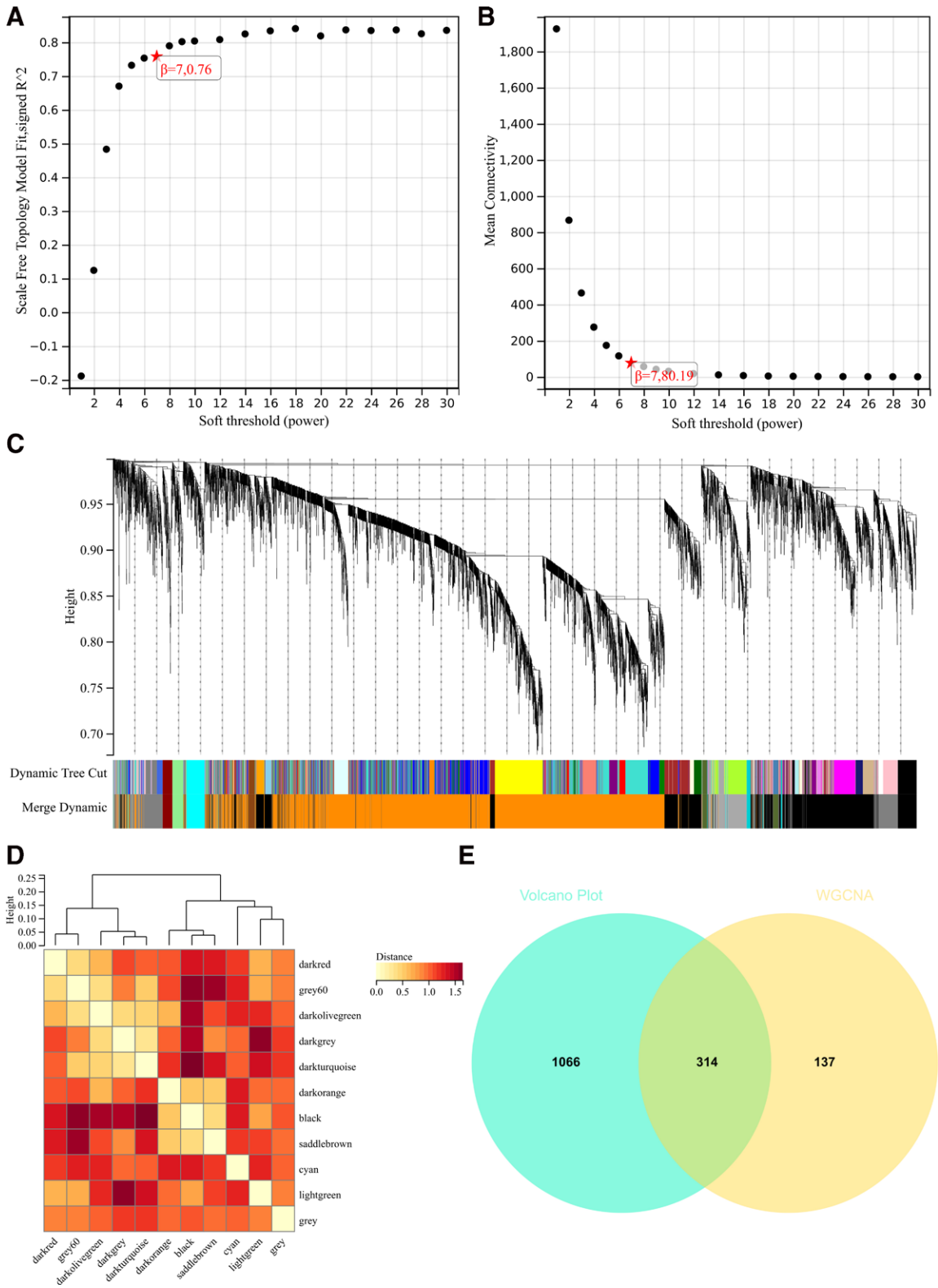

Figure 5. WGCNA. (A) β = 7, 0.76. (B) β = 7, 80.19. (C) A hierarchical clustering tree of all genes is constructed and important modules are generated. (D) The interaction between these modules. (E) Wayne diagram of the differential genes screened by WGCNA and DEGs and take the intersection to create and analyze the protein–protein interaction network. DEGs = differentially expressed genes, WGCNA = weighted gene co-expression network analysis.

map of immune cell expression in the data set (Fig. 9B). We also analyzed the correlation of infiltrating immune cells and obtained the co-expression pattern among immune cell components (Fig. 9C).

3.8. CTD analysis

In this study, we entered the hub gene list into the CTD website to find diseases related to core genes, improving the understanding of the association between genes and diseases. Seven genes

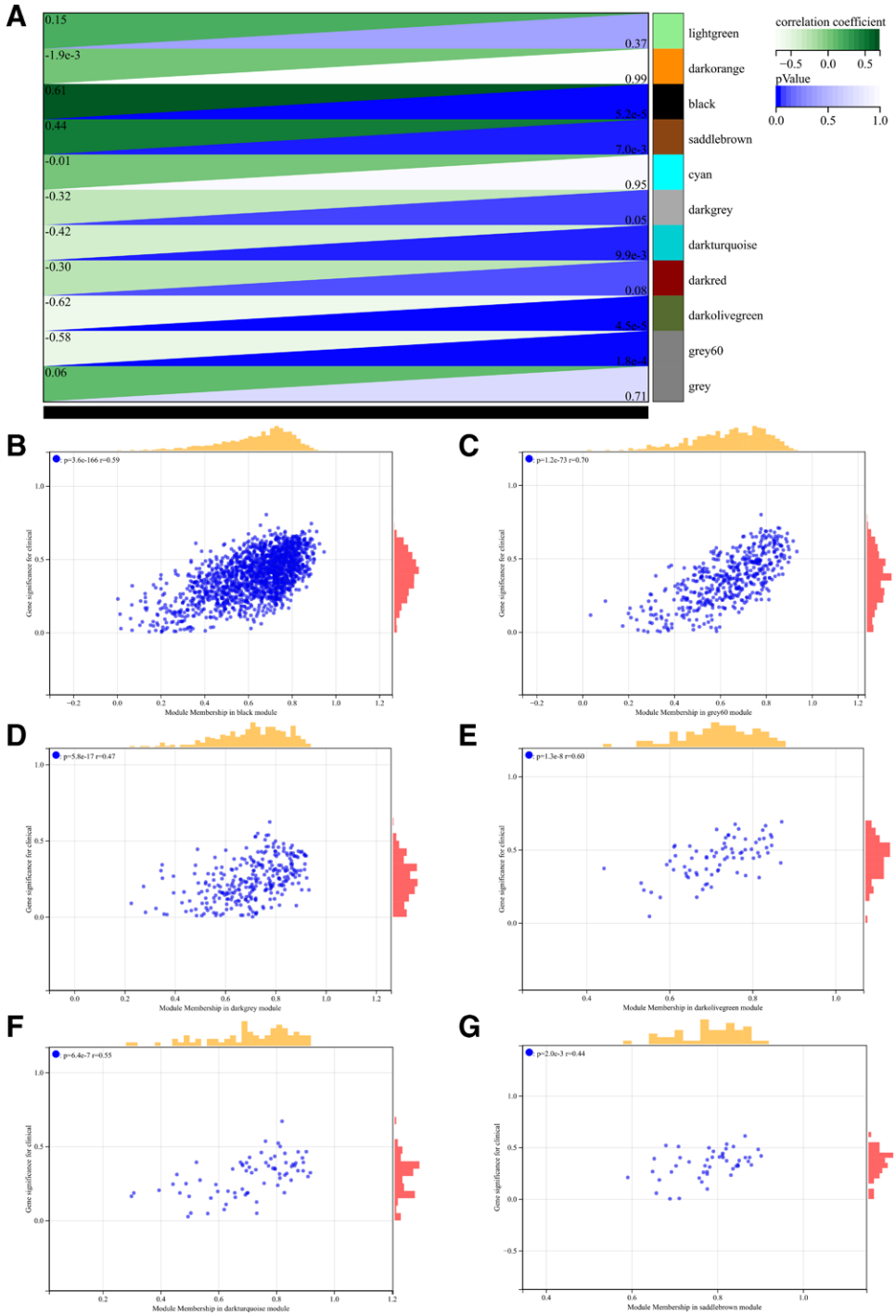

Figure 6. WGCNA. (A) The module–phenotypic correlation heat map. (B–G) GS-MM correlation scatter map of related hub genes. WGCNA = weighted gene co-expression network analysis.

(MAD2L1, MELK, RACGAP1, RRP9, WDR43, DDX21, and GTPBP4) were found to be associated with necrosis, inflammation, tumor, precancerous symptoms, bleeding, and weightlessness (Fig. 10).

3.9. miRNA prediction and functional annotation related to hub gene

In this study, we input the hub gene list into targetscan to find the relevant miRNA to improve the understanding of gene expression regulation (Table 1). We found that the related miRNA of MAD2L1 gene is hsa-miR-6088, hsa-miR-4770, hsa-miR-143-3ptermelk gene related miRNA is hsamiR-802;RACGAP1 gene related miRNA is hsa-miR-19a-3p, the related miRNA of hsa-miR-19b-3p;RRP9 gene is hsa-miR-374c-5p, the related miRNA of hsa-miR-655-3pten WDR43 gene is hsa-miR-223-3pten miRNA is hsa-miR-218-5p. The related miRNA of GTPBP4 gene is hsa-miR-330-3p.

4. Discussion

CRC is the third most common cancer in the world. The highest incidence of CRC mainly occurs in high-income countries such as the United States, Japan and South Korea.[12] CRC accounts

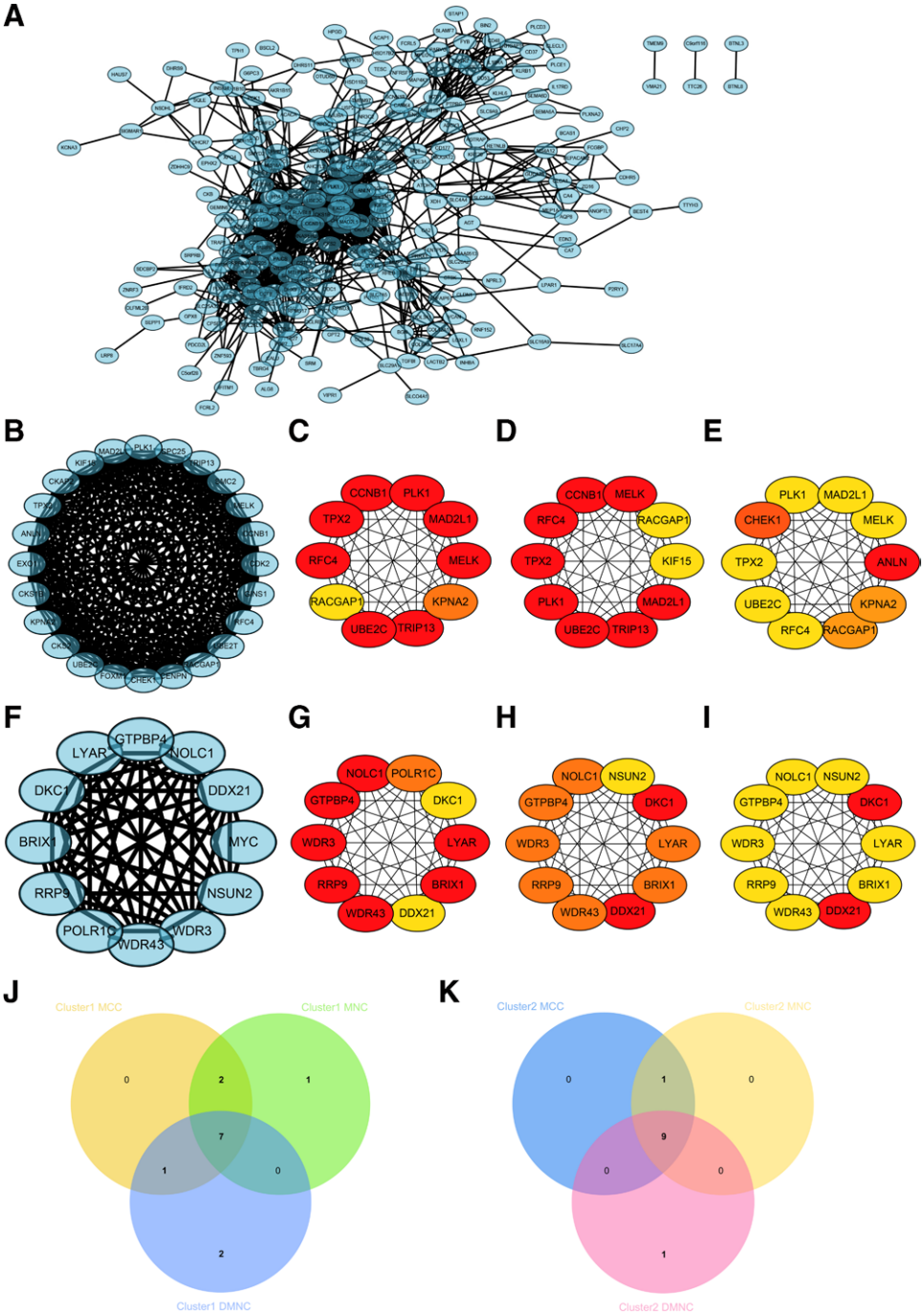

Figure 7. Construction and analysis of protein–protein interaction (PPI) Network. (A) PPI network. (B and F) The core gene cluster. (C–E, G–I) Three different algorithms were used to identify the central gene (J, K) Wayne graph was used to merge.

for 8% of all cancer deaths and is the fourth most common cause of cancer death. However, the trends of morbidity and mortality of CRC are different between developed and developing countries. Although CRC morbidity and mortality have declined in the United States and most European countries, a sharp increase in CRC morbidity and mortality has been observed in low- and middle-income countries such as Brazil and China.[13,14] Colorectal cancer is the fourth most common cancer in men and the fifth most common cancer in women in China. From 1991 to 2005, its morbidity and mortality increased by 37% and 71% respectively.[15,16] High morbidity and mortality make CRC a serious public health issue. In the past few decades, with the progress of CRC diagnostic technology and treatment strategies, CRC patients have been well treated. However, the high recurrence rate and low survival rate of CRC are still serious problems affecting the quality of life of patients with CRC. Therefore, there is still an urgent need to develop new and effective treatments for patients with CRC, especially metastatic CRC, to explore the molecular mechanism of CRC, and the study of targeted drugs is very important. The main result of this study is the high expression of RRP9 and DDX21 in CRC. The higher the expression of RRP9 and DDX21, the worse the prognosis.

Ribosomal biogenesis is a highly complex process that occurs in the nucleolus and requires the coordination and cooperation of small nucleolar ribonucleoprotein (snoRNP) and ribosomal proteins.[17] SnoRNA is mainly involved in the posttranscriptional modification and maturation of ribosomal RNA (rRNA),

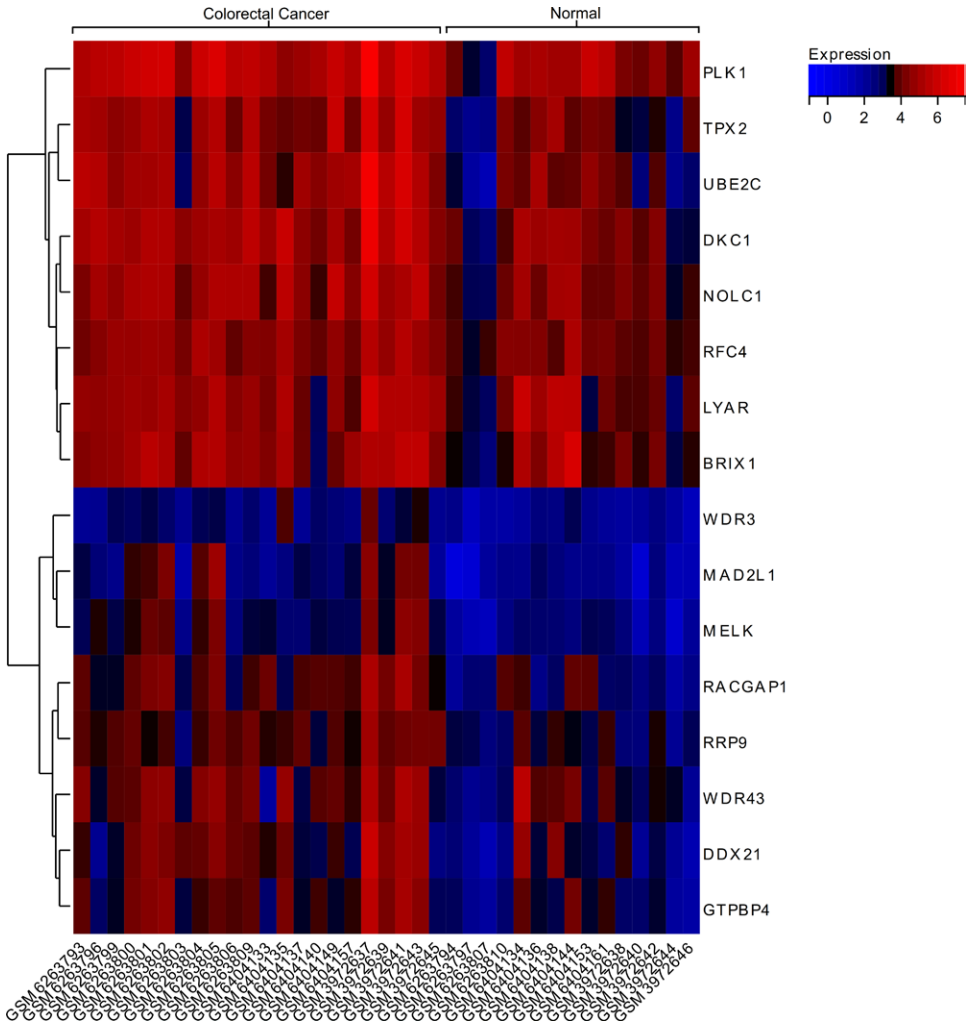

Figure 8. Gene expression heat map.

small nuclear RNA, and other cellular RNA.[18] Since cell growth requires the production of new ribosomes, it makes sense for cancer cells to use the mechanisms involved in ribosome biogenesis to support their accelerated growth.[19] In fact, recent studies have suggested the potential carcinogenic role of snoRNA in a variety of cancers, including breast cancer, CRC, hepatocellular carcinoma, and pancreatic ductal adenocarcinoma. RRP9, also known as U3-55K, is a U3snoRNA-associated protein. RRP9 is identified as the core subunit of the U3-snoRNP complex and participates in pre-rRNA processing.[20] Studies have shown that Smurf1 interacts with RRP9. RRP9 is the core component of U3snoRNP complex and participates in pre-rRNA processing. Diacylation modification tests in vivo and in vitro showed that RRP9 was coupled with Nedd8. RRP9Neddylation was catalyzed by Smurf1 and removed by NEDP1 adenosinase. It was confirmed that Lys221 was the main Nadierization site in RRP9. Lack of RRP9Neddylation inhibits pre-rRNA processing and leads to downregulation of ribosomal biogenesis. Therefore, functional studies have shown that ectopic expression of RRP9 promotes tumor cell proliferation, colony formation and cell migration, while non-Neddylation RRP9 and K221R mutants do not. In addition, the increased expression of RRP9 and Smurf1 is associated with cancer progression in human CRC.[21–23] These results suggest that Smurf1 plays a multifaceted role in pre-rRNA processing by catalyzing RRP9Neddylation, and provides new clues for the carcinogenic role of RRP9. Zhang Zhi qi et al found that the expression of U3snoRNA-related protein RRP9/U3-55K was significantly increased in human pancreatic tissue and PC cell line. RRP9 activates the AKT signal pathway by interacting with the DNA binding domain of IGF2BP1. Treatment of RRP9 overexpressed PC cells with AKT inhibitors MK-2206 and gemcitabine could significantly inhibit tumor proliferation. RRP9 is identified as a new target and may be beneficial to the treatment of PC. RRP9 plays an indispensable role in promoting chemical resistance to gemcitabine. Therefore, targeting RRP9 may provide a potential therapeutic strategy.[24] The above literature review is consistent with our results. The higher the expression of RRP9 in CRC, the higher the RRP9, the worse the prognosis. Based on the above literature analysis and our research results, we speculate that RRP9 may play an important role in the occurrence and development of CRC.

DExD-box helicase 21 (DDX21) protein, characterized by a conserved Asp-Glu-Ala-Asp (DEAD) motif, is a RNA helicase involved in various cellular processes, including RNA secondary structural changes, translation initiation, nuclear and mitochondrial splicing, and ribosome and splice assembly.[25] DDX21 performs multiple functions in ribosomal biogenesis, including transcription and ribosomal RNA processing.[26] DDX21 can effectively unlock the R ring (a three-stranded nucleic acid complex composed of RNA: DNA heteroduplex) and prevent R-loop-mediated RNA polymerase arrest, while DDX21 deficiency leads to cell R-ring accumulation and DNA damage. In the past decade, DDX21 has attracted much attention because of its potential role in tumorigenesis. A study found that DDX21 promotes the proliferation of gastric cancer by regulating cell cycle.[27] Another recent study found that

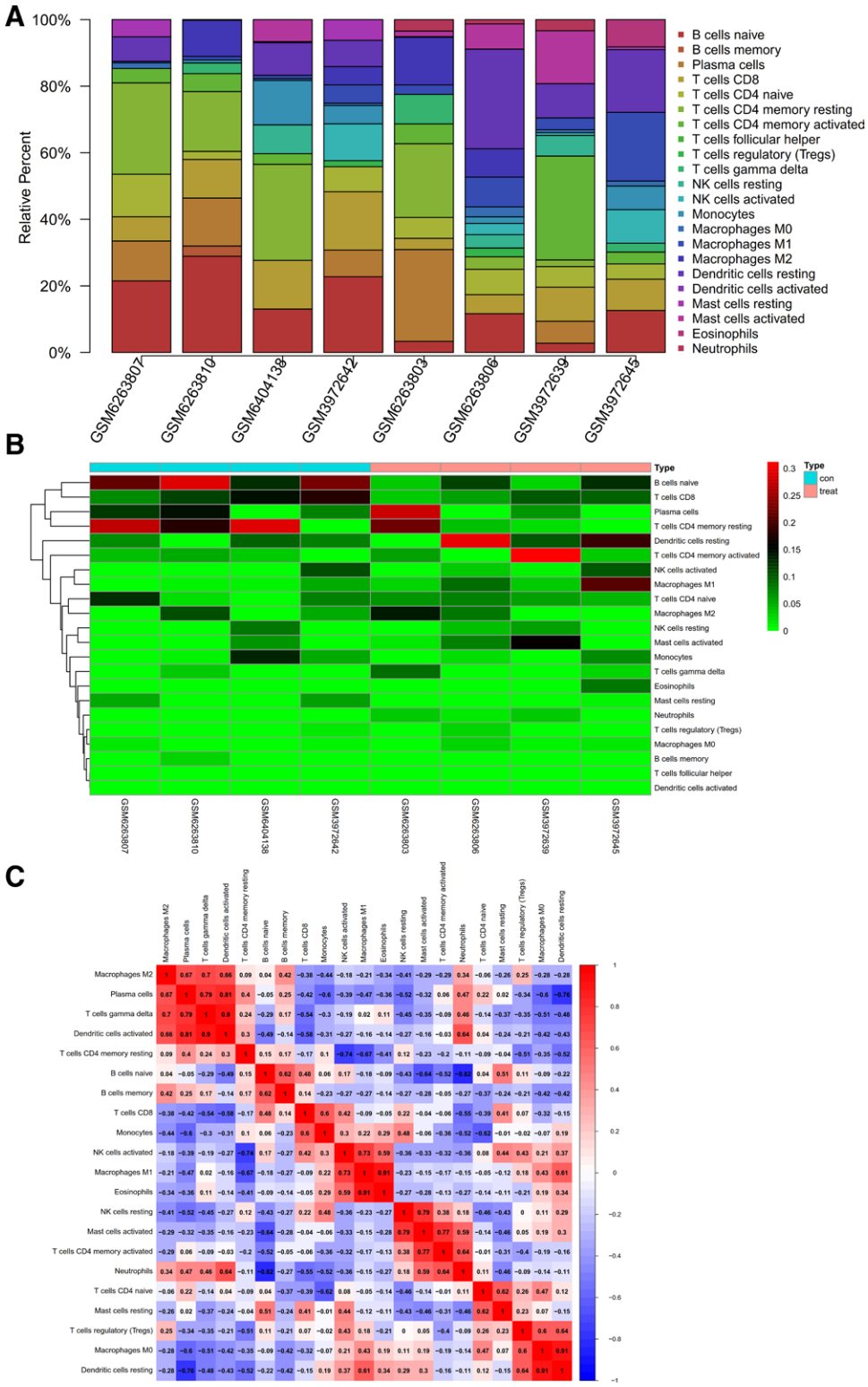

Figure 9. Immune infiltration analysis. (A) The proportion of immune cells of the whole gene expression matrix. (B) The heat map of immune cell expression in the data set. (C) The co-expression pattern among immune cell components.

downregulation of long noncoding RNA (lncRNA) HCP5 can limit the proliferation, migration and invasion of gastric cancer cells by regulating the expression of DDX21.[28] Tanaka et al's unbiased proteomic analysis of two independent CRC cohorts using mass spectrometry showed that DDX21 protein was significantly upregulated in cancer compared with benign mucosa. In addition, the expression of DDX21 protein was examined in the verification group of 710 patients, including 619 with early CRC and 91 with advanced CRC. DDX21 was mainly detected in tumor nuclei and highly expressed in some mitotic cells. The expression level of DDX21 was associated with non-mucous histology of early cancer. Survival analysis showed that high level of DDX21 protein was associated with longer survival of patients with early CRC.[29] The research of Zheng Yanzhu et al

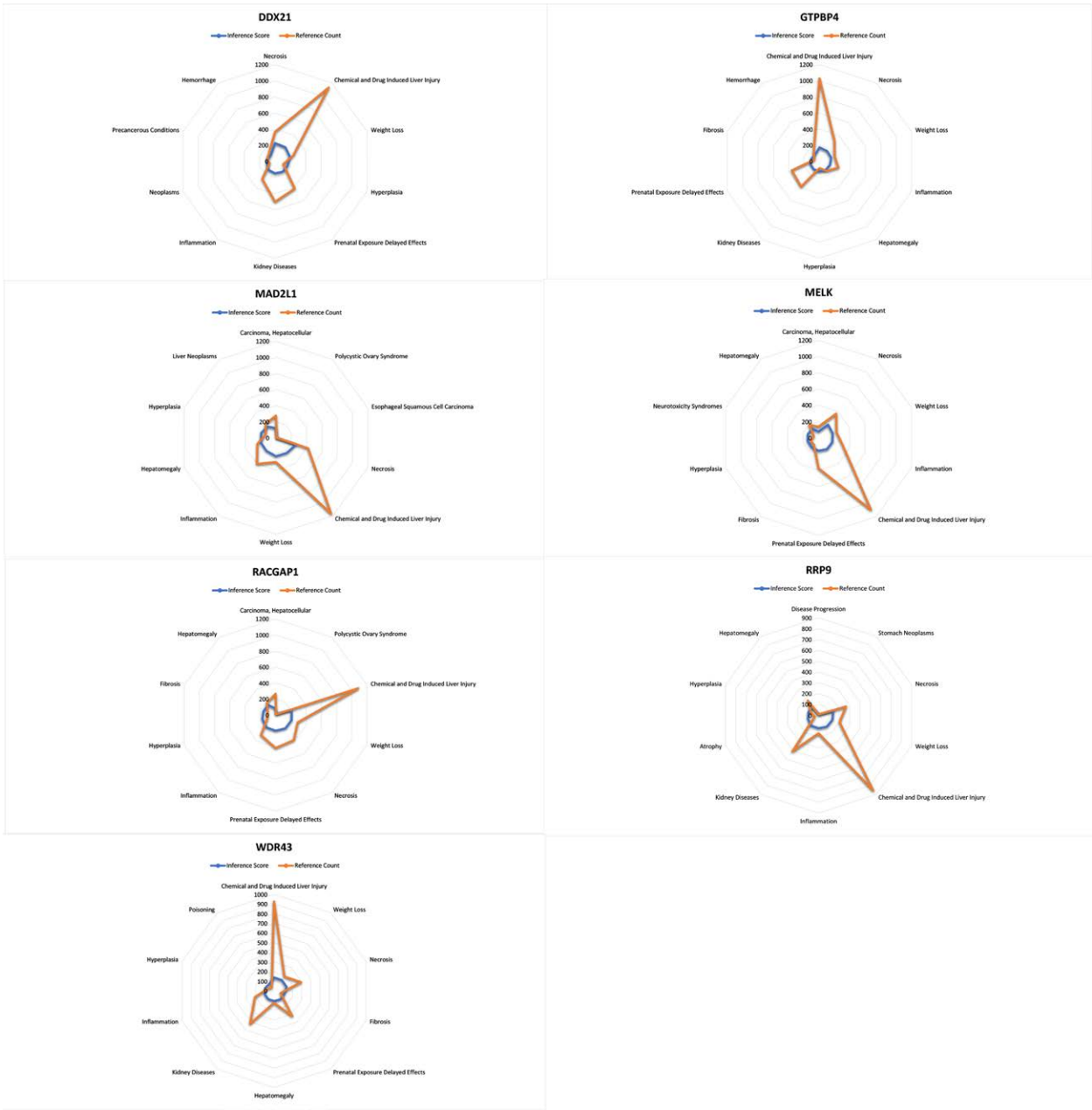

Figure 10. CTD analysis. Seven genes (MAD2L1, MELK, RACGAP1, RRP9, WDR43, DDX21, and GTPBP4) were found to be associated with necrosis, inflammation, tumor, precancerous symptoms, bleeding and weightlessness. CTD = comparative toxicogenomics database.

Table 1

A summary of miRNAs that regulate hub genes.

	Gene	MIRNA		
1	MAD2L1	hsa-miR-6088	hsa-miR-4770	hsa-miR-143-3p
2	MELK	hsa-miR-802		
3	RACGAP1	hsa-miR-19a-3p	hsa-miR-19b-3p	
4	RRP9	hsa-miR-374c-5p	hsa-miR-655-3p	
5	WDR43	hsa-miR-223-3p		
6	DDX21	hsa-miR-218-5p		
7	GTPBP4	hsa-miR-330-3p		

showed that the expression level of DDX21 in CRC tissue was significantly higher than that in normal tissue.[30] Hu Ankang et al also found that DDX21 gene was significantly expressed in 13 cancers including CRC.[31] The above literature review is consistent with our results. The higher the expression of DDX21 in CRC, the higher the DDX21, the worse the prognosis. Based on the above literature analysis and our research results, we speculate that DDX21 may play an important role in the occurrence and development of CRC.

Although this paper has carried out rigorous bioinformatics analysis, there are still some shortcomings. In this study, no animal experiments of gene overexpression or knockout were carried out to further verify its function. Therefore, in the future research, we should make an in-depth exploration in this aspect.

5. Conclusion

Signal proteins and pathways are usually attractive targets for cancer therapy. We have identified RRP9 and DDX21 as new biomarkers

of CRC. Further analysis of its function and its relationship with signal pathway may reveal the potential value of RRP9 and DDX21 in the diagnosis, prognosis and treatment of CRC.

Author contributions

Conceptualization: Xiaoqian Chi.
Data curation: Chunbo Kang, Ning Yang, Yabin Liu.
Methodology: Jie Zhang, Xiaoqian Chi, Ning Yang.
Software: Shiyang Hou, Ning Yang, Yabin Liu.
Writing – original draft: Xiaoqian Chi, Yabin Liu.
Writing – review & editing: Xiaoqian Chi.

References

[1] Mármol I, Sánchez-de-Diego C, Pradilla Dieste A, et al. Colorectal carcinoma: a general overview and future perspectives in colorectal cancer. Int J Mol Sci. 2017;18:197.
[2] Mathey MD, Pennella CL, Zubizarreta P. Colorectal carcinoma in children and adolescents. Arch Argent Pediatr. 2021;119:e487–98.
[3] Montalban-Arques A, Scharl M. Intestinalmicrobiota and colorectal carcinoma: implications for pathogenesis, diagnosis, and therapy. EBioMedicine. 2019;48:648–55.
[4] Beech C, Hechtman JF. Molecular approach to colorectal carcinoma: current evidence and clinical application. Surg Pathol Clin. 2021;14:429–41.
[5] Li N, Lu B, Luo C, et al. Incidence, mortality, survival, risk factor and screening of colorectal cancer: a comparison among China, Europe, and northern America. Cancer Lett. 2021;522:255–68.
[6] Alghandour R, Saleh GA, Shokeir FA, et al. Metastatic colorectal carcinoma initially diagnosed by bone marrow biopsy: a case report and literature review. J Egypt Natl Canc Inst. 2020;32:30.
[7] Bhalla A, Zulfiqar M, Bluth MH. Molecular diagnostics in colorectal carcinoma: advances and applications for 2018. Clin Lab Med. 2018;38:311–42.
[8] Bülow S, Svendsen LB, Mellemgaard A. Metachronous colorectal carcinoma. Br J Surg. 1990;77:502–5.
[9] Goh J, Goh C, Lim QW, et al. Transcriptomics indicate nuclear division and cell adhesion not recapitulated in MCF7 and MCF10A compared to luminal A breast tumours. Sci Rep. 2022;12:20902.
[10] Hartsough EJ, Weiss MB, Heilman SA, et al. CADM1 is a TWIST1-regulated suppressor of invasion and survival. Cell Death Dis. 2019;10:281.
[11] Reimand J, Isserlin R, Voisin V, et al. Pathway enrichment analysis and visualization of omics data using g:Profiler, GSEA, Cytoscape and EnrichmentMap. Nat Protoc. 2019;14:482–517.
[12] Cao W, Chen HD, Yu YW, et al. Changing profiles of cancer burden worldwide and in China: a secondary analysis of the global cancer statistics 2020. Chin Med J (Engl). 2021;134:783–91.
[13] Zhang Y, Chen Z, Li J. The current status of treatment for colorectal cancer in China: a systematic review. Medicine (Baltim). 2017;96:e8242.
[14] Niu L, Gao C, Li Y. Identification of potential core genes in colorectal carcinoma and key genes in colorectal cancer liver metastasis using bioinformatics analysis. Sci Rep. 2021;11:23938.
[15] Shi JF, Wang L, Ran JC, et al. Clinical characteristics, medical service utilization, and expenditure for colorectal cancer in China, 2005 to 2014: overall design and results from a multicenter retrospective epidemiologic survey. Cancer. 2021;127:1880–93.
[16] Yoshino T, Arnold D, Taniguchi H, et al. Pan-Asian adapted ESMO consensus guidelines for the management of patients with metastatic colorectal cancer: a JSMO-ESMO initiative endorsed by CSCO, KACO, MOS, SSO and TOS. Ann Oncol. 2018;29:44–70.
[17] Lafontaine DL. Noncoding RNAs in eukaryotic ribosome biogenesis and function. Nat Struct Mol Biol. 2015;22:11–9.
[18] Liang J, Wen J, Huang Z, et al. Small nucleolar RNAs: insight into their function in cancer. Front Oncol. 2019;9:587.
[19] Pecoraro A, Pagano M, Russo G, et al. Ribosome biogenesis and cancer: overview on ribosomal proteins. Int J Mol Sci. 2021;22:5496.
[20] Beltrame M, Tollervey D. Base pairing between U3 and the pre-ribosomal RNA is required for 18S rRNA synthesis. EMBO J. 1995;14:4350–6.
[21] Du MG, Liu F, Chang Y, et al. Neddylation modification of the U3 snoRNA-binding protein RRP9 by Smurf1 promotes tumorigenesis. J Biol Chem. 2021;297:101307.
[22] Zhou L, Jia L. Targeting protein neddylation for cancer therapy. Adv Exp Med Biol. 2020;1217:297–315.
[23] Xie P, Zhang M, He S, et al. The covalent modifier Nedd8 is critical for the activation of Smurf1 ubiquitin ligase in tumorigenesis. Nat Commun. 2014;5:3733.
[24] Zhang Z, Yu H, Yao W, et al. RRP9 promotes gemcitabine resistance in pancreatic cancer via activating AKT signaling pathway. Cell Commun Signal. 2022;20:188.
[25] Marcaida MJ, Kauzlaric A, Duperrex A, et al. The human RNA helicase DDX21 presents a dimerization interface necessary for helicase activity. iScience. 2020;23:101811.
[26] Ju Q, Li X, Zhang H, et al. NFE2L2 is a potential prognostic biomarker and is correlated with immune infiltration in brain lower grade glioma: a pan-cancer analysis. Oxid Med Cell Longev. 2020;2020:3580719.
[27] Cao J, Wu N, Han Y, et al. DDX21 promotes gastric cancer proliferation by regulating cell cycle. Biochem Biophys Res Commun. 2018;505:1189–94.
[28] Wang K, Yu X, Tao B, et al. Downregulationof lncRNA HCP5 has inhibitory effects on gastric cancer cells by regulating DDX21 expression. Cytotechnology. 2021;73:1–11.
[29] Tanaka A, Wang JY, Shia J, et al. DEAD-box RNA helicase protein DDX21 as a prognosis marker for early stage colorectal cancer with microsatellite instability. Sci Rep. 2020;10:22085.
[30] Jung Y, Lee S, Choi HS, et al. Clinical validation of colorectal cancer biomarkers identified from bioinformatics analysis of public expression data. Clin Cancer Res. 2011;17:700–9.
[31] Hu A, Wang Y, Tian J, et al. Pan-cancer analysis reveals DDX21 as a potential biomarker for the prognosis of multiple tumor types. Front Oncol. 2022;12:947054.