



# Identification of Potential Diagnostic and Prognostic Biomarkers for Colorectal Cancer Based on GEO and TCGA Databases

Zhenjiang Wang<sup>1†</sup>, Mingyi Guo<sup>1†</sup>, Xinbo Ai<sup>1</sup>, Jianbin Cheng<sup>1</sup>, Zaiwei Huang<sup>1</sup>, Xiaobin Li<sup>2\*</sup> and Yuping Chen<sup>1\*</sup>

<sup>1</sup> Department of Gastroenterology, Zhuhai People's Hospital (Zhuhai Hospital Affiliated With Jinan University), Zhuhai, China,

<sup>2</sup> Zhuhai Precision Medical Center, Zhuhai People's Hospital (Zhuhai Hospital Affiliated With Jinan University), Zhuhai, China

## OPEN ACCESS

### Edited by:

Zhichao Liu,  
National Center for Toxicological  
Research (FDA), United States

### Reviewed by:

Dan Li,  
University of Arkansas at Little Rock,  
United States  
Dongying Li,  
United States Food and Drug  
Administration, United States

### \*Correspondence:

Yuping Chen  
chenypmd@163.com  
Xiaobin Li  
li.xiaobin2009@163.com

<sup>†</sup>These authors have contributed  
equally to this work

### Specialty section:

This article was submitted to  
Computational Genomics,  
a section of the journal  
Frontiers in Genetics

**Received:** 04 September 2020

**Accepted:** 30 November 2020

**Published:** 14 January 2021

### Citation:

Wang Z, Guo M, Ai X, Cheng J,  
Huang Z, Li X and Chen Y (2021)  
Identification of Potential Diagnostic  
and Prognostic Biomarkers for  
Colorectal Cancer Based on GEO and  
TCGA Databases.  
*Front. Genet.* 11:602922.  
doi: 10.3389/fgene.2020.602922

Colorectal cancer (CRC) is one of the most common neoplastic diseases worldwide. With a high recurrence rate among all cancers, treatment of CRC only improved a little over the last two decades. The mortality and morbidity rates can be significantly lessened by earlier diagnosis and prompt treatment. Available biomarkers are not sensitive enough for the diagnosis of CRC, whereas the standard diagnostic method, endoscopy, is an invasive test and expensive. Hence, seeking the diagnostic and prognostic biomarkers of CRC is urgent and challenging. With that order, we screened the overlapped differentially expressed genes (DEGs) of GEO (GSE110223, GSE110224, GSE113513) and TCGA datasets. Subsequent protein–protein interaction network analysis recognized the hub genes among these DEGs. Further functional analyses including Gene Ontology and KEGG pathway analysis and gene set enrichment analysis were processed to investigate the role of these genes and potential underlying mechanisms in CRC. Kaplan–Meier analysis and Cox hazard ratio analysis were carried out to clarify the diagnostic and prognostic role of these genes. In conclusion, our present study demonstrated that CCNA2, MAD2L1, DLGAP5, AURKA, and RRM2 are all potential diagnostic biomarkers for CRC and may also be potential treatment targets for clinical implication in the future.

**Keywords:** colorectal cancer, diagnostic biomarker, prognostic biomarker, GEO, TCGA

## INTRODUCTION

Global Cancer Statistics 2018 indicates that colorectal cancer (CRC) accounts for ~10% of all diagnosed cancers and cancer-related deaths in the world each year (Bray et al., 2018). According to the data from China Cancer Registry Annual Report, the incidence and mortality of CRCs have been increasing in the past 10 years (Zheng et al., 2019). With the improvement of surgical methods and the launch of early tumor diagnosis and treatment, the current levels of diagnosis and treatment of CRC have been greatly improved. However, the prognosis of clinical CRC is still not optimistic. Many researches have shown that the occurrence and development of CRCs may be related to genetic, lifestyle, obesity, and environmental factors, while the exact etiology and the mechanism are still unclear (Bray et al., 2018). To further clarify the pathogenesis of CRCs and to improve the precision of treatment of CRCs, genetic research, study of tumor signaling pathways, and biological target therapy are continuing to deepen, which are gradually being applied in clinic. Meanwhile,

molecular stratification therapy and application of biomarkers to guide prognosis and treatment decisions are also increasing (Bogaert and Prenen, 2014).

As we all know, the occurrence, development, overall survival time, and recurrence and non-recurrence of tumors are not only related to the pathological type and clinical stage of the tumor but also closely related to the expression and pathway of tumor genes (Bogaert and Prenen, 2014). More and more studies suggested that there are many abnormally expressed genes in the gene expression of CRCs, relative to normal tissues, which are closely related to the proliferation, differentiation, apoptosis, metastasis, recurrence, and survival time of CRC (Lu et al., 2012; Liu et al., 2013; Gan et al., 2018; Branchi et al., 2019). The analysis of abnormally expressed genes has very important clinical significance for the targeted therapy, prognosis analysis, and recurrence risk prediction of CRC. Currently, there have been a lot of clinical researches on tumor recurrence genes and signaling pathways, and the gene recurrent model (GRM) has been established to make up for the traditional tumor classification and staging recurrence prediction, providing more genetic information and more accurate prediction data (Chen et al., 2019; Yang et al., 2020). For example, Sun et al. (2019) found that exosomal CPNE3 showed potential implications in CRC diagnosis and prognosis. Carcinoembryonic antigen (CEA) was a recommended prognostic marker in CRC for tumor diagnosis and monitoring response to therapy (Campos-da-Paz et al., 2018). Ahluwalia et al. (2019) identified a novel 4-gene prognostic signature that had clinical utility in colorectal cancer. However, there are few clinical studies about biomarkers and gene pathways which have no risk of recurrence and tumor survival time.

In this study, we obtained advanced colorectal cancer gene profiles (GSE110223, GSE110224, and GSE113513) from the Gene Expression Omnibus. Differentially expressed genes (DEG) were identified by comparing CRC tissues with non-cancerous gastric tissues using the GEO2R online analysis. Subsequently, the DEG were analyzed according to Gene Ontology (GO), KEGG pathway enrichment analysis, coexpression, and protein–protein interaction (PPI) analyses. We then performed the overall survival analysis for candidate genes. Finally, GEPIA and UALCAN online tools were performed to associate candidate genes with colorectal cancer overall survival (OS), disease-free survival (RFS), and pathological staging analysis through the Cancer Genome Atlas [TCGA] dataset and found that CCNA2, MAD2L1, DLGAP5, AURKA, and RRM2 could play an important role in colorectal cancer overall survival and disease-free survival, and these may be potential treatment targets for clinical implication in the future.

## MATERIALS AND METHODS

### Microarray Data

Studies from the GEO database were considered eligible satisfying the following criteria: (1) studies with CRC tissue samples, (2) studies containing information on technology and platform utilized for studies, and (3) studies including adjacent normal tissues as the control. Based on the above criteria, three datasets [GSE110223 (Vlachavas et al., 2019), GSE110224

(Vlachavas et al., 2019), GSE113513 (2018, unpublished)] are all downloaded from the GEO (Campos-da-Paz et al., 2018) database. The platform used by GSE110223 is [HG-U133A] Affymetrix Human Genome U133A Array, which includes 13 CRC tumor tissue samples and 13 normal tissue samples. The platform used by GSE110224 is [HG-U133\_Plus\_2] Affymetrix Human Genome U133 Plus 2.0 Array, which includes 17 CRC tumor tissue samples and 17 normal tissue samples. The platform used by GSE113513 is [PrimeView] Affymetrix Human Gene Expression Array, which includes 14 CRC tumor tissue samples and 14 normal tissue samples. A total of 44 CRC tumor tissue samples and 44 normal tissue samples were included in this study (Figure 1).

### Identification of DEGs

Using the GEOquery and limma R software packages in the Bioconductor project, the GEO2R web tool was used for identifying differential genes in selected datasets. We analyzed three GEO datasets online through GEO2R and selected the genes with  $P < 0.05$  and fold change (FC)  $> 1.5$  in the dataset as differentially expressed genes (DEGs). Then, we used FunRich\_3.1.3 (Pathan et al., 2017) to make a Venn diagram and extract the differentially expressed genes common to the three datasets.

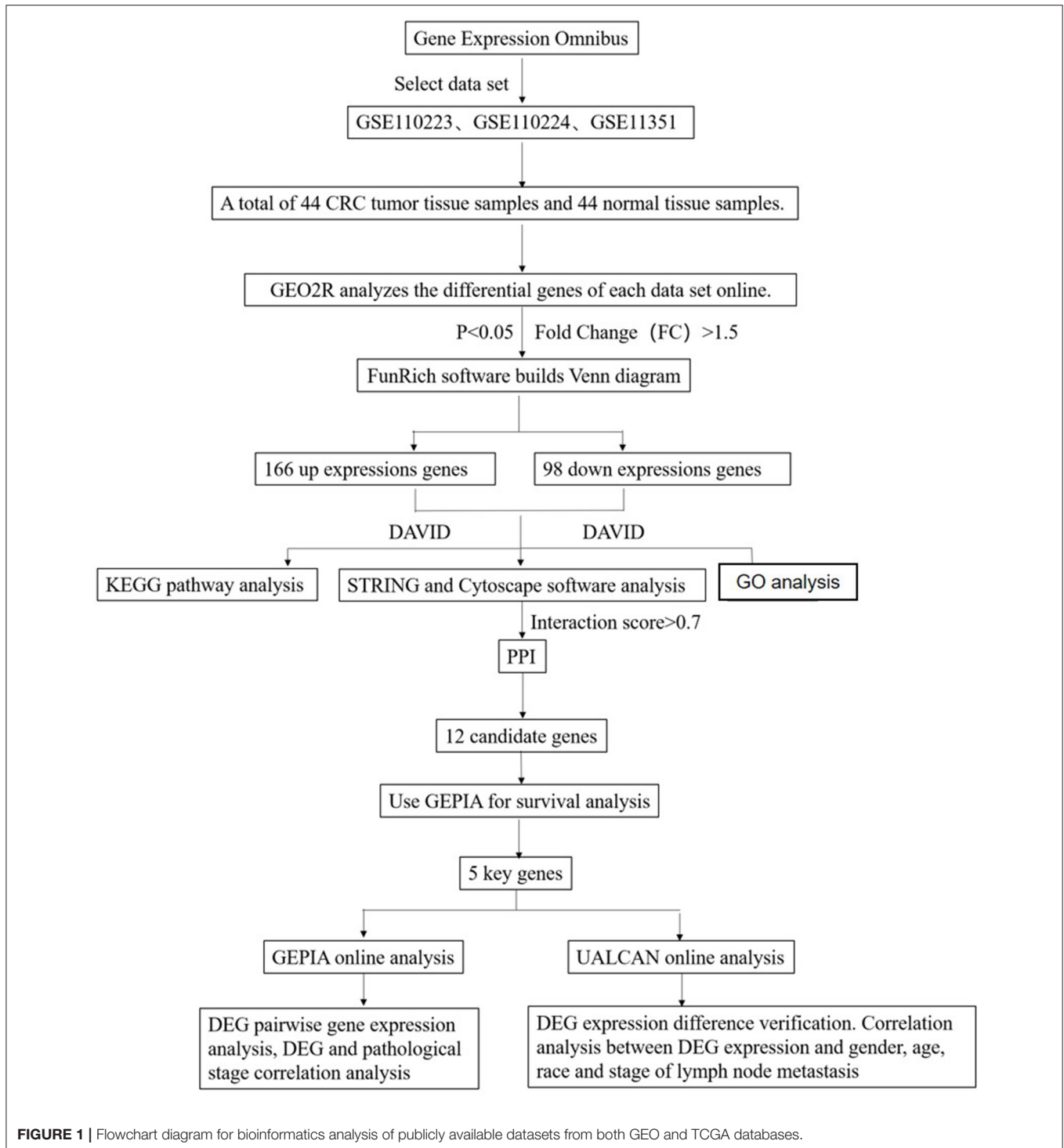
### GO and KEGG Pathway Analysis

In this study, the GO term ([www.geneontology.org](http://www.geneontology.org)) and Kyoto Encyclopedia of Genes and Genomes (KEGG, [www.genome.jp](http://www.genome.jp)) approach were identified and analyzed by using DAVID v6.8 (<https://david-d.ncifcrf.gov/summary.jsp>) (Huang et al., 2009a,b). The identifier and species were selected as “official\_gene\_symbol” and “Homo Sapiens,” respectively. The enrichment of  $P < 0.05$  was set as the critical standard for significant enrichment and by using the ggplot2 package (version 3.3.1) (Wickham, 2009) and the R language (version 3.6.3, <http://www.r-project.org/>) to visualize the analysis results of the DAVID.

### Protein–Protein Interaction Network Analysis

STRING (version 11.0), covering 24,584,628 proteins from 5,090 organisms, and integrating known and predicted interactions between more than 932,000,000 proteins from multiple organisms including Homo sapiens (Szklarczyk et al., 2019), was used to conduct PPI network analysis on DEGs. The “Multiple proteins” button was selected, and the species were selected as “Homo Sapiens.” When the  $P$ -value was  $< 0.05$ , the network interaction relationship is considered to be statistically significant, with interaction score  $> 0.7$  considered to be a high-confidence interaction relationship.

The Cytoscape software (version 3.6.0) (Shannon et al., 2003) was used to visualize the PPI network, and the plugin CytoHubba (Chin et al., 2014) was applied to identify the central node genes in the PPI network, and then the central node gene as a candidate DEG for the following analysis.



## The Expressions and Survival Analysis of Candidate DEGs

GEPIA (<http://gepia.cancer-pku.cn/>) is a database that uses standard processing methods to analyze the RNA sequencing expression data of 9,736 tumors and 8,587 normal samples from the TCGA and GTEx projects (Tang et al., 2017). GEPIA provides

various functions such as tumor/normal differential expression analysis, analysis according to cancer type or pathological stage, survival analysis, correlation analysis, etc. By using the GEPIA, we performed Kaplan–Meier survival analysis on the relative expression of candidate DEGs in CRC patients and the overall survival time and disease-free survival time, with hazard ratio

(HR) and corresponding 95% confidence interval. DEG related to the overall survival time and disease-free survival time was used as the study purpose DEG and conduct data analysis. Multiple-gene comparison and principal component analysis for the biomarker candidates were also conducted using the GEPIA.

UALCAN (<http://ualcan.path.uab.edu/analysis.html>) is an online database that uses TCGA transcriptome and clinical patient data to enable researchers to analyze the differential

expression of tumor tissue and normal tissue, tumor stage, lymph node metastasis, and other related clinical parameters (Chandrashekar et al., 2017). We validated DEGs by using the UALCAN database, reanalyzed their expression differences in CRC tissue samples and normal tissue samples, and performed correlation analysis between DEG and gender, age, race, and stage of lymph node metastasis.

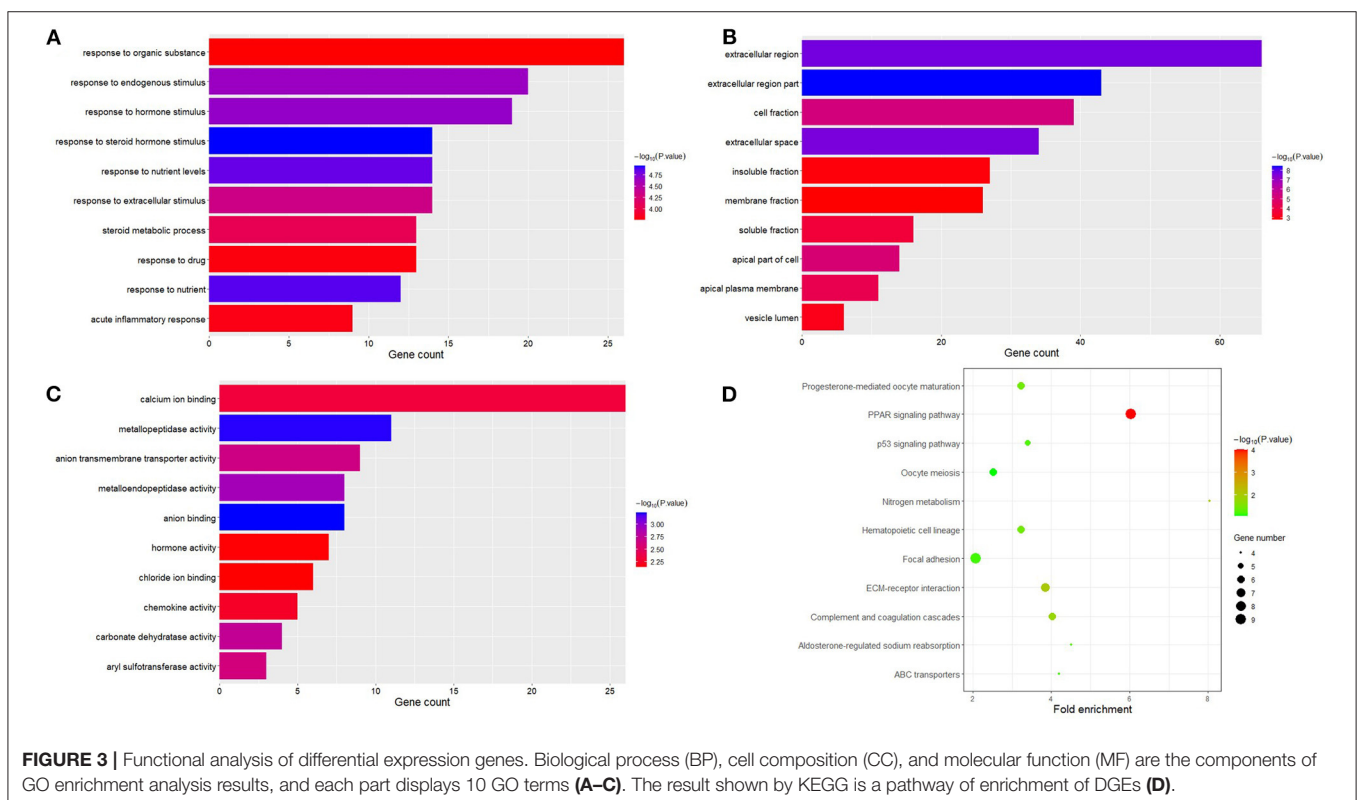
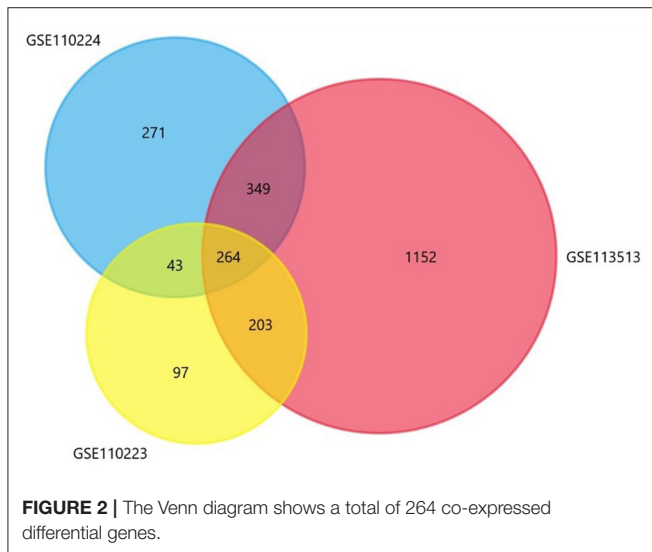
## RESULTS

### Identification of DEGs

This study included three gene sets (GSE110223, GSE110224, GSE113513), of which GSE110223 included 13 tumor samples and 13 normal samples; GSE110224 included 17 tumor samples and 17 normal samples; and GSE113513 included 14 tumor samples and 14 normal samples. In all included datasets, compared with normal samples, there are 264 significantly different genes in all datasets (Figure 2), including 166 up-expression genes and 98 down-expression genes.

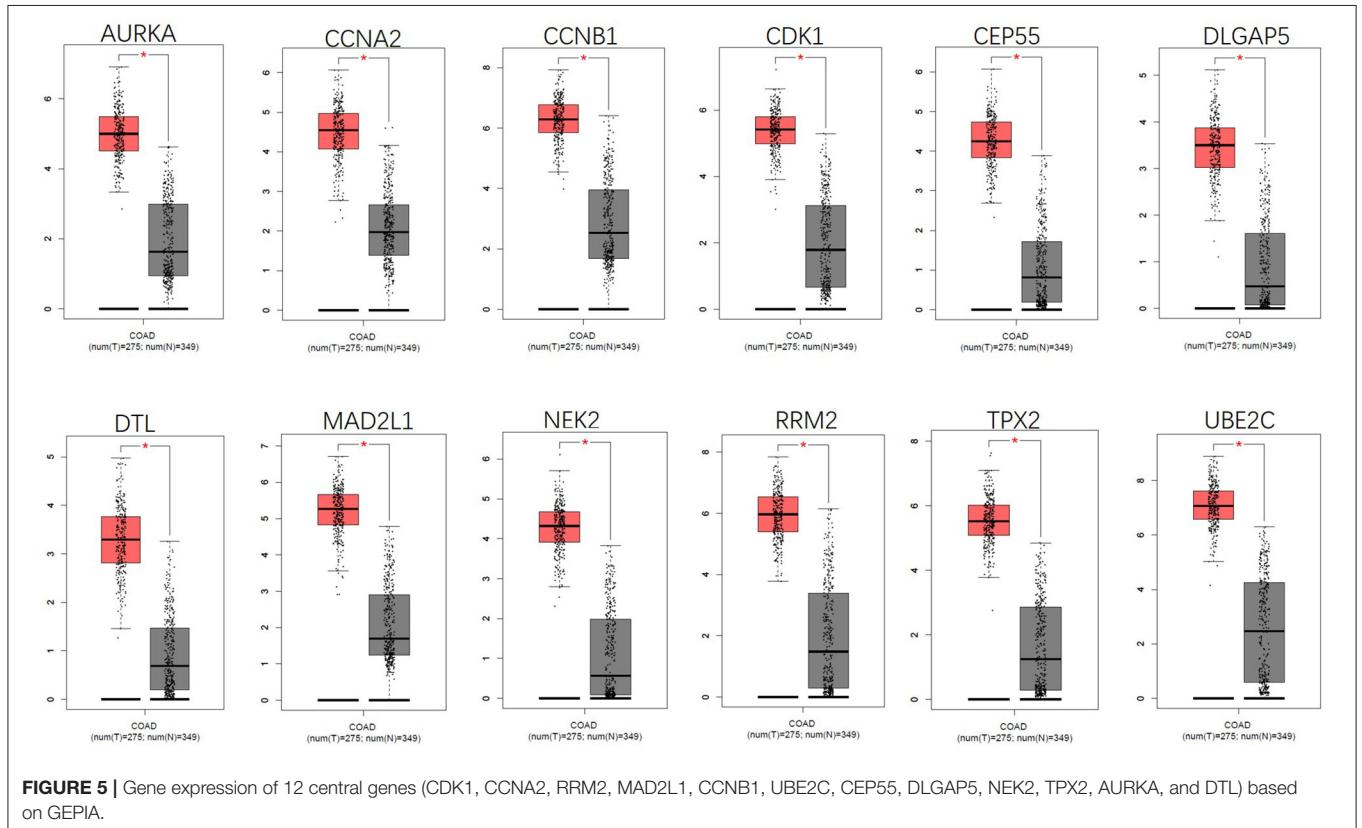
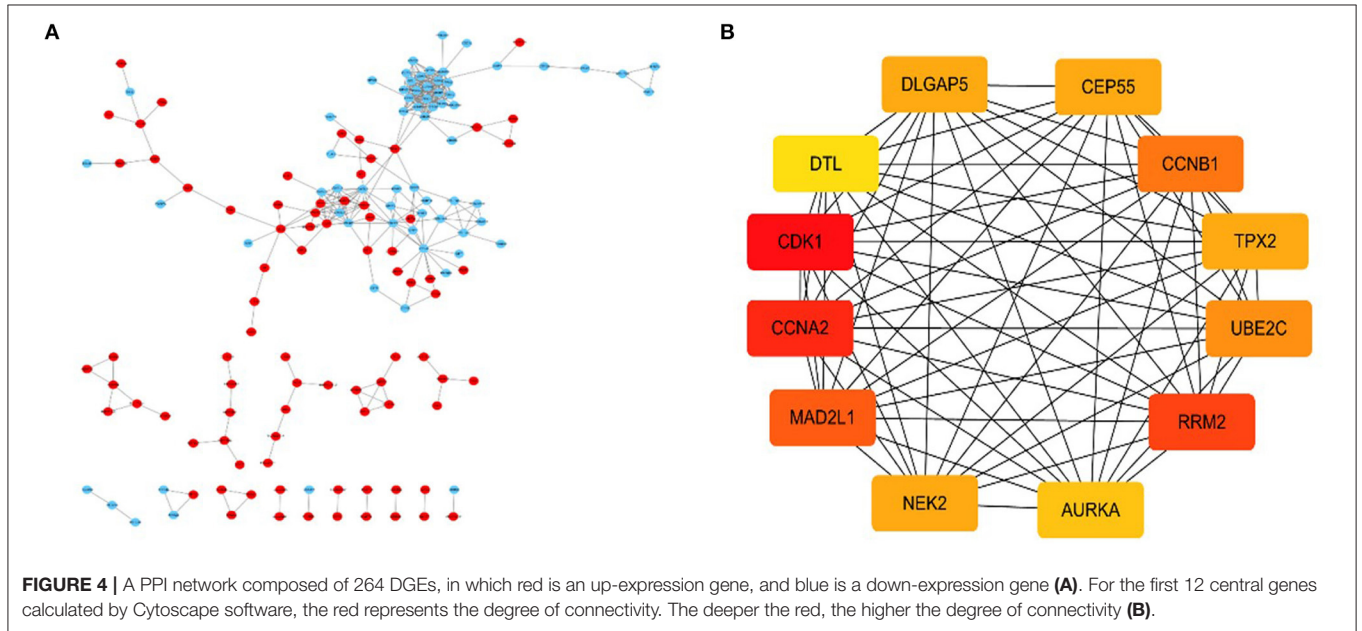
### Enrichment Analysis of DEGs

The obtained DEGs were analyzed for functional enrichment by using DAVID. GO enrichment analysis mainly predicts the function of target genes through three aspects: biological process (BP), cell composition (CC), and molecular function (MF). By using DAVID, we found that DEG is mainly enriched in BP, including response to steroid hormone stimulus, response to nutrient, response to nutrient levels, response to hormone stimulus, response to endogenous stimulus, response to organic substance, response to extracellular stimulus, steroid metabolic process, response to drug, response to nutrient, acute inflammatory response.



response to extracellular stimulus, steroid metabolic process, acute inflammatory response, response to drug, response to organic substance, etc. (Figure 3A). CC analysis includes extracellular region part, extracellular region, extracellular

space, cell fraction, apical part of the cell, apical plasma membrane, soluble fraction, vesicle lumen, insoluble fraction, and membrane fraction (Figure 3B). MF analysis includes anion binding, metallopeptidase activity, metalloendopeptidase



activity, carbonate dehydratase activity, anion transmembrane transporter activity, aryl sulfotransferase activity, calcium ion binding, chemokine activity, hormone activity, and chloride ion binding (Figure 3C).

At the same time, the analysis of the KEGG pathway shows that 264 DEGs are mainly enriched in eight pathways, namely, PPAR signaling pathway, ECM-receptor interaction, nitrogen metabolism, complement and coagulation cascades, hematopoietic cell lineage, progesterone-mediated oocyte maturation, aldosterone-regulated sodium reabsorption, p53 signaling pathway, focal adhesion, ABC transporters, and oocyte meiosis (Figure 3D).

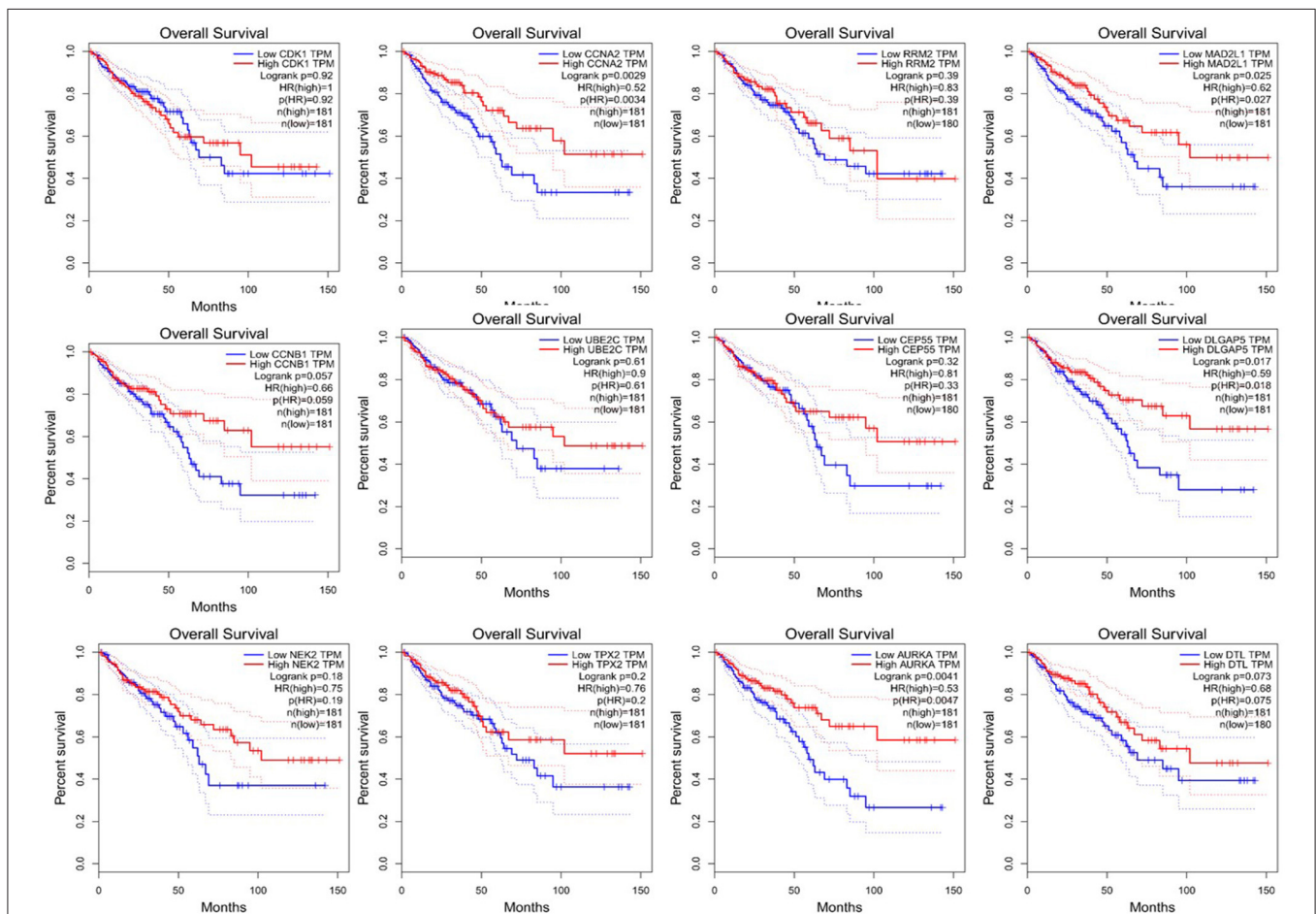
## PPI Network to Identify Central Genes

By using the STRING database and Cytoscape 3.6.0 software, a PPI network was constructed for the 264 DGEs obtained and the central genes were determined. The PPI network constructed by STRING (Figure 4A) has a total of 262 nodes and 802 edges, and an interaction score  $>0.7$  is considered a high-confidence interaction relationship. Using Cytoscape 3.6.0 software, we identified the top 12 genes with the most

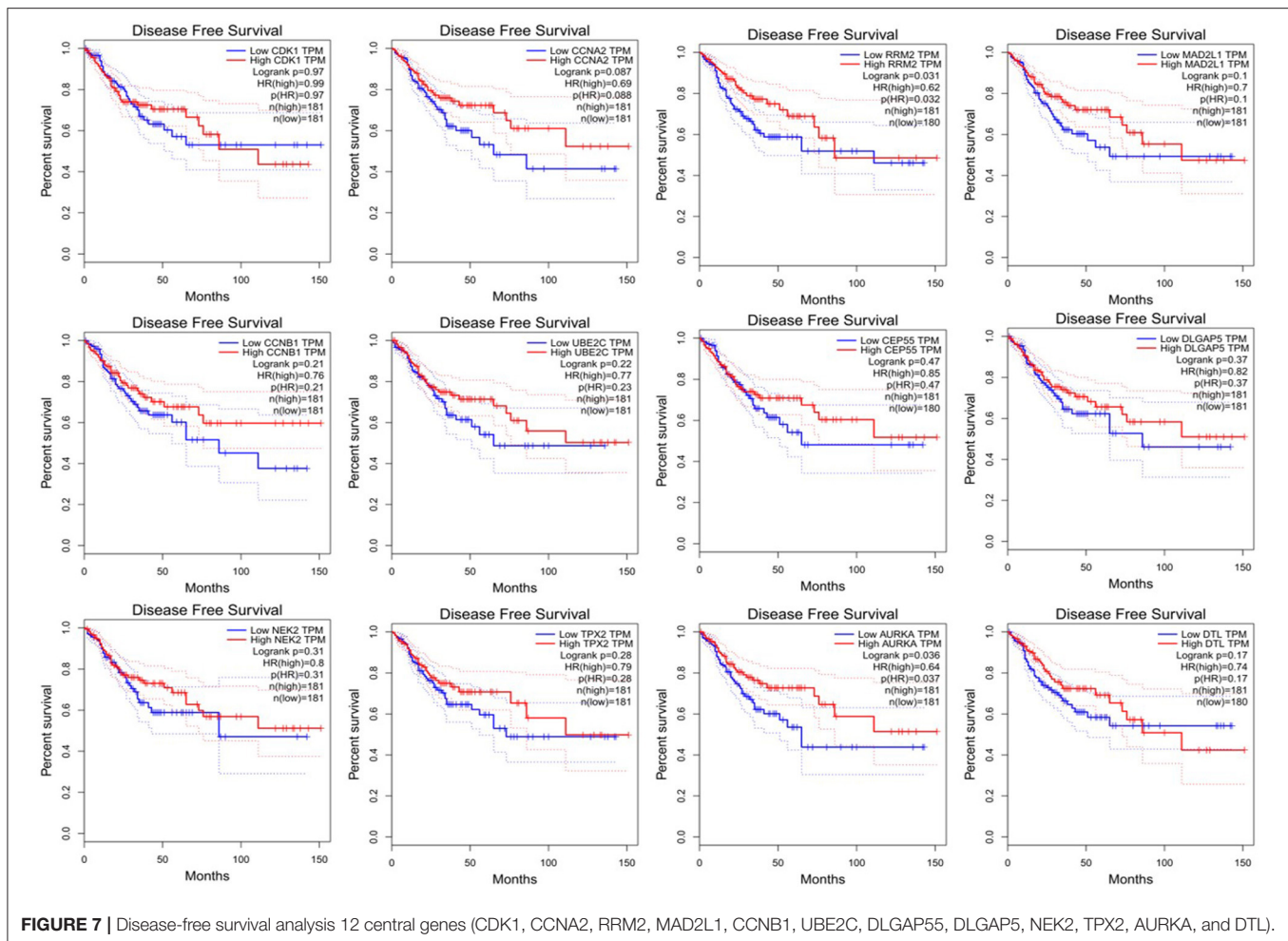
connectedness (Figure 4B). The most connected gene is CDK1, followed by CCNA2, RRM2, MAD2L1, CCNB1, UBE2C, CEP55, DLGAP5, NEK2, TPX2, AURKA, and DTL. These 12 genes can form a module. Using GEPIA, gene expression profiles of the 12 central genes between CRC tumor samples and normal samples were displayed in Figure 5.

## Overall Survival Analysis and Disease-Free Survival Analysis

Since CRC is mainly adenocarcinoma, which can account for 90%, we used the GEPIA database to analyze the overall survival and disease-free survival of colorectal adenocarcinoma on 12 central genes. We found that among the 12 central genes, CCNA2, MAD2L1, DLGAP5, and AURKA were associated with the overall survival of colorectal adenocarcinoma ( $P < 0.05$ ) (Figure 6), and RRM2 and AURKA were associated with disease-free survival of colorectal adenocarcinoma ( $P < 0.05$ ) (Figure 7). Therefore, this study will focus on the five genes CCNA2, MAD2L1, DLGAP5, AURKA, and RRM2.



**FIGURE 6 |** Overall survival analysis of 12 central genes (CDK1, CCNA2, RRM2, MAD2L1, CCNB1, UBE2C, CEP55, DLGAP5, NEK2, TPX2, AURKA, and DTL) based on GEPIA.



## Correlation Analysis Based on the GEPIA

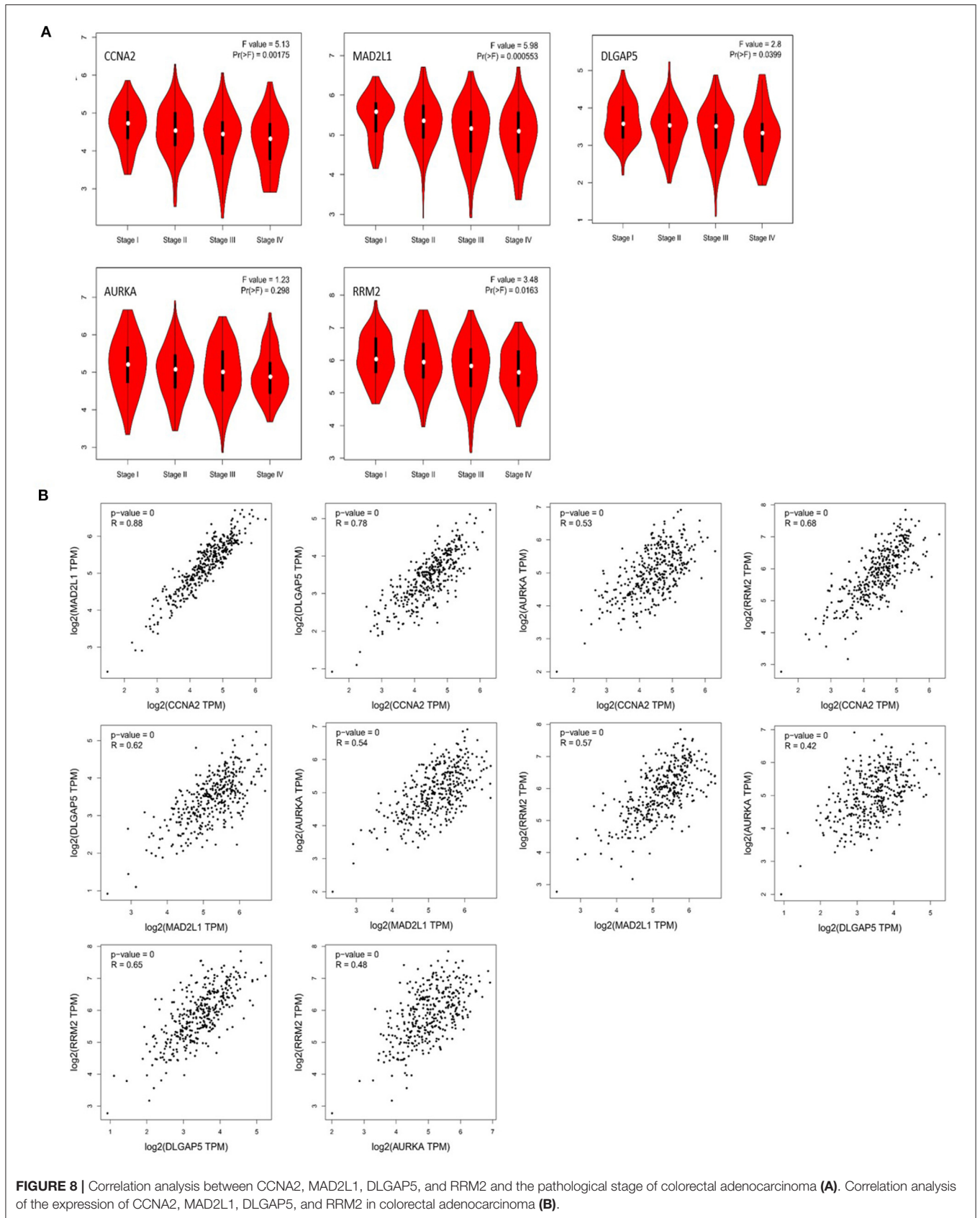
Through the analysis of the correlation between these five genes and the pathological staging of colorectal adenocarcinoma based on GEPIA, we found that CCNA2, MAD2L1, DLGAP5, and RRM2 are all significantly related to the pathological stage of COAD and READ ( $P < 0.05$ ), while AURKA is associated with the colorectal gland, with no significant correlation between cancer pathological staging ( $P > 0.05$ ) (Figure 8A). The correlation analysis of the expression of these five genes in colorectal adenocarcinoma showed that CCNA2 was highly correlated with MAD2L1 ( $P < 0.001$ ,  $R = 0.88$ ), and it was also correlated with DLGAP5 ( $P < 0.001$ ,  $R = 0.78$ ), AURKA ( $P < 0.001$ ,  $R = 0.53$ ), and RRM2 ( $P < 0.001$ ,  $R = 0.68$ ). Moderate positive correlations between MAD2L1 and DLGAP5 ( $P < 0.001$ ,  $R = 0.62$ ), MAD2L1 and AURKA ( $P < 0.001$ ,  $R = 0.54$ ), and MAD2L1 and RRM2 ( $P < 0.001$ ,  $R = 0.57$ ) were observed. DLGAP5 and AURKA ( $P < 0.001$ ,  $R = 0.42$ ) had low expression correlation, and there was a moderate positive correlation between DLGAP5 and RRM2 ( $P < 0.001$ ,  $R = 0.65$ ). AURKA and RRM2 ( $P < 0.001$ ,  $R = 0.48$ ) have a low expression correlation (Figure 8B).

## Verification of the Differential Expression of CCNA2, MAD2L1, DLGAP5, AURKA, and RRM2 and the Analysis of Related Clinical Parameters

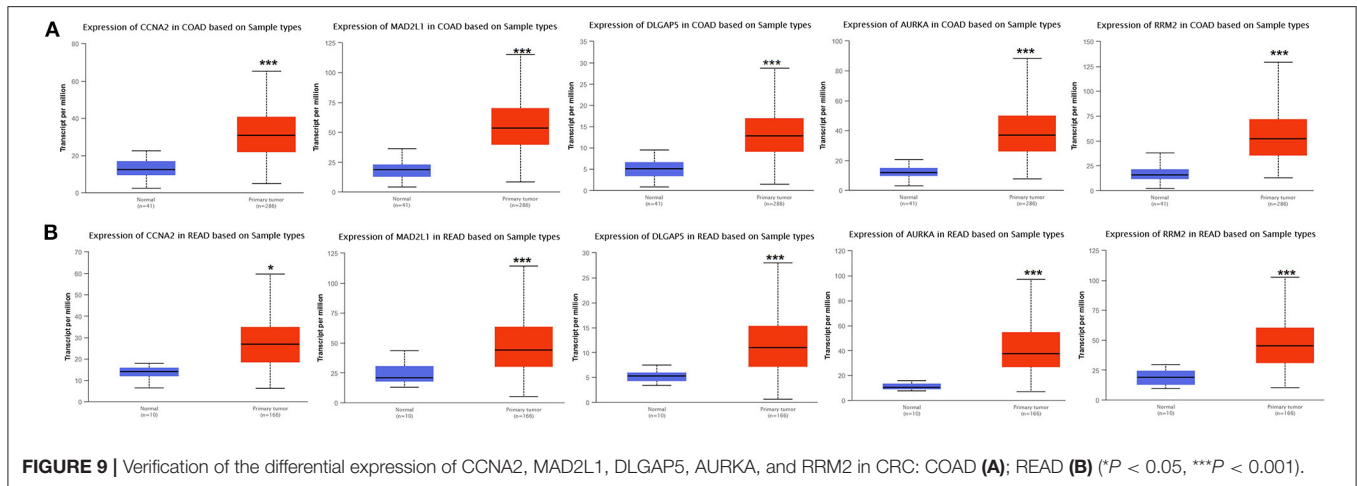
Analysis for gene differential expression through the UALCAN software (<http://ualcan.path.uab.edu/analysis.html>) indicated that there were significant differences in expression of the five genes in the normal group and the tumor group, both in colon adenocarcinoma (COAD) and rectal adenocarcinoma (READ) (Figure 9).

For COAD, taking the normal samples as the reference group, the expression of these five genes all significantly increased, whether in males or in females (Figure S1A), various age groups (Figure S2A), various races (Figure S3A), and various lymph node metastasis stages (Figure S4A).

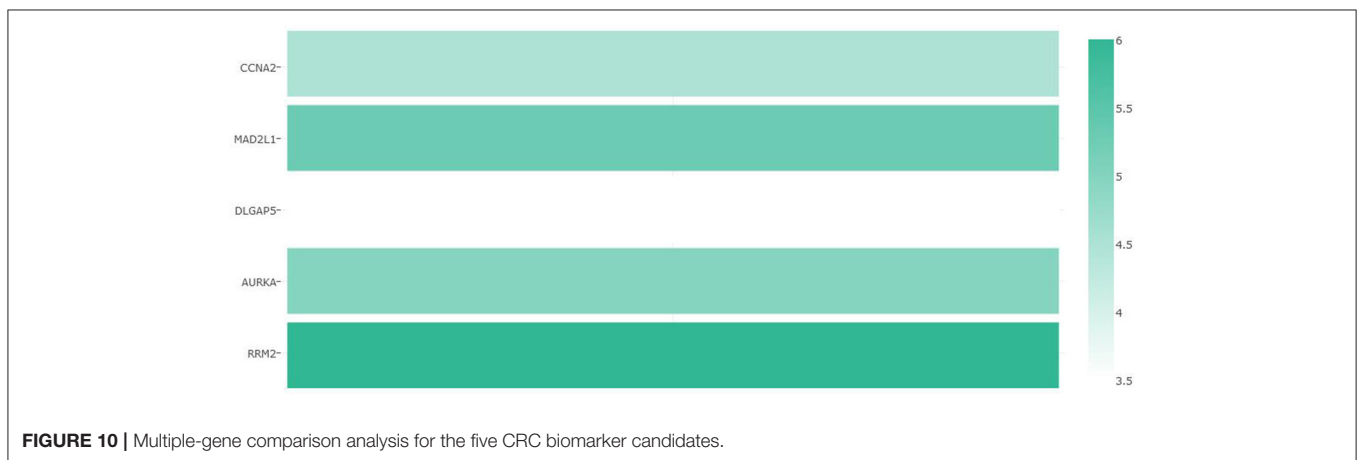
For READ, taking the normal samples as the reference group, in terms of gender (Figure S1B), the expression of these five genes was significantly increased in males or females. In terms of age (Figure S2B), CCNA2, MAD2L1, and AURKA were 21–40. The expression of DLGAP5 was not obvious in the age groups







**FIGURE 9** | Verification of the differential expression of CCNA2, MAD2L1, DLGAP5, AURKA, and RRM2 in CRC: COAD (A); READ (B) (\* $P < 0.05$ , \*\*\* $P < 0.001$ ).



**FIGURE 10** | Multiple-gene comparison analysis for the five CRC biomarker candidates.

of 41–60, 61–80, and 81–100. DLGAP5 was not significantly expressed in the groups of 21–40 and 81–100. Both the 60-year-old group and the 61–80-year-old group were significantly expressed, and RRM2 was significantly expressed in all age groups. In terms of race (Figure S3B), as the sample number of Asian patients was only one case, the Asian patient sample group was not tested. By comparison, the expression of these five genes was significantly increased in both the African-American patient group and the Caucasian patient group. In terms of lymph node metastasis (Figure S4B), the expression of CCNA2 was significantly increased in N1 and N2, but not in N3. MAD2L1, DLGAP5, AURKA, and RRM2 were significantly increased in the metastasis stage of each lymph node.

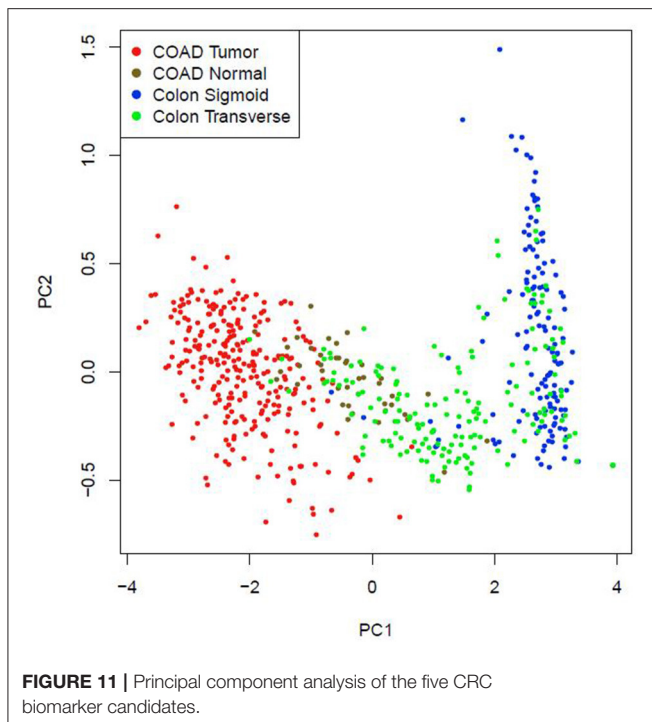
### Possibility of a Five-Gene Biomarker in Diagnosis of CRC

Multiple-gene comparison analysis for the five CRC biomarker candidates was conducted using the GEPIA, with the only tumor data (COAD). Among the five genes, RRM2 had the highest expression level, followed by MAD2L1, AURKA, CCNA2, and DLGAP5 (Figure 10). Principal component analysis of the five genes was performed with TCGA tumor data, TCGA normal

data, and GTEx data (both colon-sigmoid and colon-transverse); we found that the five genes could effectively distinguish between CRC samples and normal samples (Figure 11), indicating the possibility of a five-gene biomarker in diagnosis of CRC.

## DISCUSSION

CRC is one of the most common cancers and carries a major global health burden. Globally, among all cancers, CRC ranked third in incidence and second in mortality in 2018 (Bray et al., 2018). There is no specific clinical symptom of colorectal cancer at the early stage, but when it was found, it had become in the middle and late stages. Surgery and radiotherapy and chemotherapy at the perioperative period or adjuvant chemotherapy is still the first choice for colorectal cancers (Schmoll et al., 2012). In those with metastatic disease, the treatment repertoire has been extended to include biologically targeted agents, including monoclonal antibodies targeting EGFR, such as cetuximab or panitumumab (Tabernero et al., 2015). As a result of improved treatment options, the overall



survival (OS) of patients with metastatic CRC has increased from  $\sim 1$  year in the era of 5-fluorouracil (5-FU) therapy alone, to  $\sim 3$  years with currently available therapies (Cremolini et al., 2015). However, recurrence and overall survival are still the challenge for the treatment of colorectal cancer in clinic. Therefore, it is crucial to identify new markers that can predict CRC recurrence, overall survival (OS), and disease-free survival (RFS), subsequently separating patients into high- or low-risk groups for enhanced efficacy in further treatment. There are many clinical studies related to tumor recurrence. In this study, we focus on the genes about overall survival (OS) and disease-free survival (DFS) in CRC. Three datasets (GSE110223, GSE110224, GSE113513) from the GEO database were introduced into the analysis; 264 DEGs were overlapped in all the three datasets including 166 upregulated DEGs and 98 downregulated DEGs. Even though the three datasets had the similar number of samples, the number of DEGs of each dataset varied greatly. GSE113513 had the largest number of DEGs, largely more than those of other two datasets.

We found that CCNA2, MAD2L1, DLGAP5, AURKA, and RRM2 were closely related to the prognosis of CRC. Among them, we found that CCNA2, MAD2L1, DLGAP5, and AURKA were related to the overall survival (OS) of colorectal tumors, while RRM2 and AURKA had the relation between disease-free survival (DFS) and colorectal cancer. CCNA2, MAD2L1, DLGAP5, and RRM2 were all significantly related to the pathological stages of colorectal cancer and were also closely related to the stage of lymph node metastasis. There was a study that revealed that the expression of CCNA2 in CRC tissues is higher than that in normal tissues. The knockdown

of CCNA2 could significantly suppress CRC cell growth by impairing cell cycle progression and inducing cell apoptosis (Gan et al., 2018). Our study showed that the abnormal expression of CCNA2 was also significantly related to the overall survival time, pathological stage of the tumor, and lymph node metastasis. There were articles that showed that CCNA2 is an important sign to judge the poor prognosis of the tumor, as it also highly expressed in pancreatic cancer, breast cancer, lung cancer, and other tumors (Gao et al., 2014; Peng et al., 2018; Brcic et al., 2019). MAD2L1, DLGAP5, and AURKA are the key genes for spindle assembly. When these genes are abnormally expressed, they will cause chromosome mismatch and other genetic problems during mitosis (Ooi et al., 2012). What is worse is that the unstable gene expression will eventually lead to cancer (Wassmann and Benezra, 2001; Weaver and Cleveland, 2005). Clinical studies had shown that DLGAP5 was related to the invasion and migration of CRC (Branchi et al., 2019); besides, DLGAP5 expression was also related to overall survival and lymph node metastasis but had no correlation with disease-free survival. It is an important measure of poor prognosis. Previous studies had verified that MAD2L1, DLGAP5, and AURKA were highly expressed in CRC (Chuang et al., 2016; Branchi et al., 2019; Ding et al., 2020). In our study, we found that these abnormally expressed genes not only induce the occurrence and development of tumors but also are significantly related to the overall survival, pathological staging of tumors, and tumor lymph node metastasis. When it comes to the expression of RRM2, studies showed that it related to the depth of invasion, degree of differentiation, disease-free survival (RFS), and metastasis of CRC (Lu et al., 2012; Liu et al., 2013). Our study showed that RRM2 was associated with disease-free survival of colorectal adenocarcinoma and was an important target gene predicted after tumor treatment.

In summary, it can be seen that CCNA2, MAD2L1, DLGAP5, RRM2, and AURKA are significantly related to the overall survival prognosis, disease-free survival, pathological stage, and lymph node metastasis stage of CRC, which are also important indicators for the evaluation of the prognosis of CRC and the evaluation of further treatment. Principal component analysis of the five genes indicated that they could effectively distinguish between CRC samples and normal samples. In order to increase the accuracy of diagnosing CRC, we suggested that the five biomarker candidates as a five-gene biomarker for diagnosis of CRC.

Once we understand the site where abnormal gene expression induces tumors, more targeted drugs will be applied. Most importantly, it provides a theoretical basis for future gene-level treatment of tumors and can achieve more precise targeted therapy. It is also important to guide the development of genetic kits and the non-invasive diagnosis of colorectal tumors.

## DATA AVAILABILITY STATEMENT

The datasets presented in this study can be found in online repositories. The names of the repository/repositories and accession number(s) can be found in the article/**Supplementary Material**.

## ETHICS STATEMENT

Ethical review and approval was not required for the study on human participants in accordance with the local legislation and institutional requirements. Written informed consent for participation was not required for this study in accordance with the national legislation and the institutional requirements. Written informed consent was not obtained from the individual(s) for the publication of any potentially identifiable images or data included in this article.

## AUTHOR CONTRIBUTIONS

YC and XL designed and directed the research. ZW and MG conducted data collection, data statistics, and article writing

## REFERENCES

- Ahluwalia, P., Mondal, A. K., Bloomer, C., Fulzele, S., Jones, K., Ananth, S., et al. (2019). Identification and clinical validation of a novel 4 gene-signature with prognostic utility in colorectal Cancer. *Int. J. Mol. Sci.* 20:3818. doi: 10.3390/ijms20153818
- Bogaert, J., and Prenen, H. (2014). Molecular genetics of colorectal cancer. *Ann. Gastroenterol.* 27, 9–14. doi: 10.1111/j.1749-6632.1995.tb12114.x
- Branchi, V., García, S. A., Radhakrishnan, P., Gyorffy, B., Hissa, B., Schneider, M., et al. (2019). Prognostic value of DLGAP5 in colorectal cancer. *Int. J. Colorectal Dis.* 34, 1455–1465. doi: 10.1007/s00384-019-03339-6
- Bray, F., Ferlay, J., Soerjomataram, I., Siegel, R. L., Torre, L. A., and Jemal, A. (2018). Global cancer statistics 2018: GLOBOCAN estimates of incidence and mortality worldwide for 36 cancers in 185 countries. *CA Cancer J. Clin.* 68, 394–424. doi: 10.3322/caac.21492
- Brcic, L., Heidinger, M., Sever, A. Z., Zacharias, M., Jakopovic, M., Fediuk, M., et al. (2019). Prognostic value of cyclin A2 and B1 expression in lung carcinoids. *Pathology* 51, 481–486. doi: 10.1016/j.pathol.2019.03.011
- Campos-da-Paz, M., Dórea, J. G., Galdino, A. S., Lacava, Z. G. M., and de Fatima Menezes Almeida Santos, M. (2018). Carcinoembryonic Antigen (CEA) and hepatic metastasis in colorectal cancer: update on biomarker for clinical and biotechnological approaches. *Recent Pat. Biotechnol.* 12, 269–279. doi: 10.2174/1872208312666180731104244
- Chandrashekar, D. S., Bashel, B., Balasubramanya, S. A. H., Creighton, C. J., Ponce-Rodriguez, I., Chakravarthi, B., et al. (2017). UALCAN: a portal for facilitating tumor subgroup gene expression and survival analyses. *Neoplasia* 19, 649–658. doi: 10.1016/j.neo.2017.05.002
- Chen, L., Lu, D., Sun, K., Xu, Y., Hu, P., Li, X., et al. (2019). Identification of biomarkers associated with diagnosis and prognosis of colorectal cancer patients based on integrated bioinformatics analysis. *Gene* 692, 119–125. doi: 10.1016/j.gene.2019.01.001
- Chin, C. H., Chen, S. H., Wu, H. H., Ho, C. W., Ko, M. T., and Lin, C. Y. (2014). cytoHubba: identifying hub objects and sub-networks from complex interactome. *BMC Syst. Biol.* 8(Suppl. 4):S11. doi: 10.1186/1752-0509-8-S4-S11
- Chuang, T. P., Wang, J. Y., Jao, S. W., Wu, C. C., Chen, J. H., Hsiao, K. H., et al. (2016). Over-expression of AURKA, SKA3, and DSN1 contributes to colorectal adenoma to carcinoma progression. *Oncotarget* 7, 45803–45818. doi: 10.18632/oncotarget.9960
- Cremolini, C., Schirripa, M., Antoniotti, C., Moretto, R., Salvatore, L., Masi, G., et al. (2015). First-line chemotherapy for mCRC—a review and evidence-based algorithm. *Nat. Rev. Clin. Oncol.* 12, 607–619. doi: 10.1038/nrclinonc.2015.129
- Ding, X., Duan, H., and Luo, H. (2020). Identification of Core gene expression signature and key pathways in colorectal cancer. *Front. Genet.* 11:45. doi: 10.3389/fgene.2020.00045
- Gan, Y., Li, Y., Li, T., Shu, G., and Yin, G. (2018). CCNA2 acts as a novel biomarker in regulating the growth and apoptosis of colorectal cancer. *Cancer Manag. Res.* 10, 5113–5124. doi: 10.2147/CMAR.S176833
- Gao, T., Han, Y., Yu, L., Ao, S., Li, Z., and Ji, J. (2014). CCNA2 is a prognostic biomarker for ER+breast cancer and tamoxifen resistance. *PLoS One* 9:e91771. doi: 10.1371/journal.pone.0091771
- Huang da, W., Sherman, B. T., and Lempicki, R. A. (2009a). Systematic and integrative analysis of large gene lists using DAVID bioinformatics resources. *Nat. Protoc.* 4, 44–57. doi: 10.1038/nprot.2008.211
- Huang da, W., Sherman, B. T., and Lempicki, R. A. (2009b). Bioinformatics enrichment tools: paths toward the comprehensive functional analysis of large gene lists. *Nucleic Acids Res.* 37, 1–13. doi: 10.1093/nar/gkn923
- Liu, X., Zhang, H., Lai, L., Wang, X., Loera, S., Xue, L., et al. (2013). Ribonucleotide reductase small subunit M2 serves as a prognostic biomarker and predicts poor survival of colorectal cancers. *Clin. Sci. (Lond.)* 124, 567–578. doi: 10.1042/CS20120240
- Lu, A. G., Feng, H., Wang, P. X., Han, D. P., Chen, X. H., and Zheng, M. H. (2012). Emerging roles of the ribonucleotide reductase M2 in colorectal cancer and ultraviolet-induced DNA damage repair. *World J. Gastroenterol.* 18, 4704–4713. doi: 10.3748/wjg.v18.i34.4704
- Ooi, W. F., Re, A., Sidarovich, V., Canella, V., Arseni, N., Adami, V., et al. (2012). Segmental chromosome aberrations converge on overexpression of mitotic spindle regulatory genes in high-risk neuroblastoma. *Genes Chromosomes Cancer* 51, 545–556. doi: 10.1002/gcc.21940
- Pathan, M., Keerthikumar, S., Chisanga, D., Alessandro, R., Ang, C. S., Askenase, P., et al. (2017). A novel community driven software for functional enrichment analysis of extracellular vesicles data. *J. Extracell. Vesicles* 6:1321455. doi: 10.1080/20013078.2017.1321455
- Peng, X., Pan, K., Zhao, W., Zhang, J., Yuan, S., Wen, X., et al. (2018). NPTX1 inhibits colon cancer cell proliferation through down-regulating cyclin A2 and CDK2 expression. *Cell Biol. Int.* 42, 589–597. doi: 10.1002/cbin.10935
- Schmoll, H. J., Van Cutsem, E., Stein, A., Valentini, V., Glimelius, B., Haustermans, K., et al. (2012). ESMO consensus guidelines for management of patients with colon and rectal cancer. a personalized approach to clinical decision making. *Ann. Oncol. Off. J. Eur. Soc. Med. Oncol.* 23, 2479–2516. doi: 10.1093/annonc/mds236
- Shannon, P., Markiel, A., Ozier, O., Baliga, N. S., Wang, J. T., Ramage, D., et al. (2003). Cytoscape: a software environment for integrated models of biomolecular interaction networks. *Genome Res.* 13, 2498–2504. doi: 10.1101/gr.1239303
- Sun, B., Li, Y., Zhou, Y., Ng, T. K., Zhao, C., Gan, Q., et al. (2019). Circulating exosomal CPNE3 as a diagnostic and prognostic biomarker for colorectal cancer. *J. Cell. Physiol.* 234, 1416–1425. doi: 10.1002/jcp.26936
- Szklarczyk, D., Gable, A. L., Lyon, D., Junge, A., Wyder, S., Huerta-Cepas, J., et al. (2019). STRING v11: protein-protein association networks with increased coverage, supporting functional discovery in genome-wide experimental datasets. *Nucleic Acids Res.* 47, D607–D613. doi: 10.1093/nar/gky1131
- Taberero, J., Yoshino, T., Cohn, A. L., Obermannova, R., Bodoky, G., Garcia-Carbonero, R., et al. (2015). Ramucirumab versus placebo in combination with second-line FOLFIRI in patients with metastatic colorectal carcinoma that

- progressed during or after first-line therapy with bevacizumab, oxaliplatin, and a fluoropyrimidine (RAISE): a randomised, double-blind, multicentre, phase 3 study. *Lancet Oncol.* 16, 499–508. doi: 10.1016/S1470-2045(15)70127-0
- Tang, Z., Li, C., Kang, B., Gao, G., Li, C., and Zhang, Z. (2017). GEPIA: a web server for cancer and normal gene expression profiling and interactive analyses. *Nucleic Acids Res.* 45, W98–W102. doi: 10.1093/nar/gkx247
- Vlachavas, E. I., Pilalis, E., Papadodima, O., Koczan, D., Willis, S., Klippel, S., et al. (2019). Radiogenomic analysis of F-18-fluorodeoxyglucose positron emission tomography and gene expression data elucidates the epidemiological complexity of colorectal cancer landscape. *Comput. Struct. Biotechnol. J.* 17, 177–185. doi: 10.1016/j.csbj.2019.01.007
- Wassmann, K., and Benezra, R. (2001). Mitotic checkpoints: from yeast to cancer. *Curr. Opin. Genet. Dev.* 11, 83–90. doi: 10.1016/S0959-437X(00)00161-1
- Weaver, B. A., and Cleveland, D. W. (2005). Decoding the links between mitosis, cancer, and chemotherapy: the mitotic checkpoint, adaptation, and cell death. *Cancer Cell* 8, 7–12. doi: 10.1016/j.ccr.2005.06.011
- Wickham, H. (2009). *Ggplot2: Elegant Graphics for Data Analysis*. New York, NY: Springer Publishing Company, Incorporated.
- Yang, W. J., Wang, H. B., Wang, W. D., Bai, P. Y., Lu, H. X., Sun, C. H., et al. (2020). A network-based predictive gene expression signature for recurrence risks in stage II colorectal cancer. *Cancer Med.* 9, 179–193. doi: 10.1002/cam4.2642
- Zheng, R. S., Sun, K. X., Zhang, S. W., Zeng, H. M., Zou, X. N., Chen, R., et al. (2019). [Report of cancer epidemiology in China, 2015]. *Zhonghua Zhong Liu Za Zhi* 41, 19–28. doi: 10.3760/cma.j.issn.0253-3766.2019.01.005

**Conflict of Interest:** The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Copyright © 2021 Wang, Guo, Ai, Cheng, Huang, Li and Chen. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner(s) are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.