



## OPEN ACCESS

EDITED BY  
Jialiang Yang,  
Geneis (Beijing) Co. Ltd., China

REVIEWED BY  
Junlin Xu,  
Hunan University, China  
Lijuan Zhu,  
Shijiazhuang Tiedao University, China

\*CORRESPONDENCE  
Sheng Zheng,  
zs19860909@163.com

SPECIALTY SECTION  
This article was submitted to RNA,  
a section of the journal  
Frontiers in Genetics

RECEIVED 12 August 2022  
ACCEPTED 08 September 2022  
PUBLISHED 27 September 2022

CITATION  
Leng X, Yang J, Liu T, Zhao C, Cao Z, Li C,  
Sun J and Zheng S (2022), A  
bioinformatics framework to identify the  
biomarkers and potential drugs for the  
treatment of colorectal cancer.  
*Front. Genet.* 13:1017539.  
doi: 10.3389/fgene.2022.1017539

COPYRIGHT  
© 2022 Leng, Yang, Liu, Zhao, Cao, Li,  
Sun and Zheng. This is an open-access  
article distributed under the terms of the  
[Creative Commons Attribution License  
\(CC BY\)](https://creativecommons.org/licenses/by/4.0/). 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.

# A bioinformatics framework to identify the biomarkers and potential drugs for the treatment of colorectal cancer

Xiaogang Leng, Jianxiu Yang, Tie Liu, Chunbo Zhao,  
Zhongzheng Cao, Chengren Li, Junxi Sun and Sheng Zheng\*

Department of Colorectal and Anal Surgery, Weifang People's Hospital, Weifang, China

Colorectal cancer (CRC), a common malignant tumor, is one of the main causes of death in cancer patients in the world. Therefore, it is critical to understand the molecular mechanism of CRC and identify its diagnostic and prognostic biomarkers. The purpose of this study is to reveal the genes involved in the development of CRC and to predict drug candidates that may help treat CRC through bioinformatics analyses. Two independent CRC gene expression datasets including The Cancer Genome Atlas (TCGA) database and GSE104836 were used in this study. Differentially expressed genes (DEGs) were analyzed separately on the two datasets, and intersected for further analyses. 249 drug candidates for CRC were identified according to the intersected DEGs and the Crowd Extracted Expression of Differential Signatures (CREEDS) database. In addition, hub genes were analyzed using Cytoscape according to the DEGs, and survival analysis results showed that one of the hub genes, *TIMP1* was related to the prognosis of CRC patients. Thus, we further focused on drugs that could reverse the expression level of *TIMP1*. Eight potential drugs with documentary evidence and two new drugs that could reverse the expression of *TIMP1* were found among the 249 drugs. In conclusion, we successfully identified potential biomarkers for CRC and achieved drug repurposing using bioinformatics methods. Further exploration is needed to understand the molecular mechanisms of these identified genes and drugs/small molecules in the occurrence, development and treatment of CRC.

## KEYWORDS

colorectal cancer, differentially expressed gene, hub gene, survival analysis, *TIMP1*, drug repurposing

## Introduction

Colorectal cancer (CRC) is the most common subtype in gastrointestinal cancers, and its early symptoms are unobvious, which results in a high mortality rate. The continuous rise of new cases and deaths of CRC will lead to a significant increase in the economic burden globally (Rogler, 2014; Arnold et al., 2017; Hong et al., 2021; Liu et al., 2021). As

the second leading cause of cancer death worldwide (Zhao et al., 2020; Sung et al., 2021), CRC has become a major global public health concern. Studies have shown that the clinical tumor stage at diagnosis affects the prognosis of patients. The 5-years relative survival rate of patients with stage I was 90%, while that of patients with stage IV was only 10% (Siegel et al., 2012; O'Connell et al., 2004; Yang et al., 2022). Currently, various diagnostic strategies for CRC include both invasive and non-invasive methods. Invasive methods rely on endoscopy and imaging. Imaging tests such as nuclear magnetic resonance (NMR) and computed tomography (CT) can be used to diagnose severe focal lesions, but both tests are expensive (Grassetto et al., 2012; Swiderska et al., 2014). Hence, there is an urgent need for alternative, cheap and easy-to-measure screening methods. Despite recent advances in treatment and multidisciplinary care, CRC patients continue to suffer from serious adverse reactions, which can impair prognosis and reduce survival (McQuade et al., 2017; Kong et al., 2020). The developing drugs with low toxicity, especially drug repositioning (Liu et al., 2020; Meng et al., 2022) is of great significance for improving the clinical treatment and reducing adverse reactions.

The improvement of molecular biology technology provides opportunities to develop more curative effect and enhance the outcomes of CRC. With the progress of high-throughput sequencing technology, gene expression profiling methods, such as RNA sequencing (RNA-seq), have been applied to scientific research and become a hot field of gene expression research (Saito et al., 2018; Deshiere et al., 2019; Zhang et al., 2021). The molecular mechanism of CRC holds the key to the prognosis and treatment response of patients, and is of great potential for the clinical practice (De Sousa et al., 2013; Sadanandam et al., 2013; Nguyen and Duong, 2018; Cheng et al., 2020; Cheng et al., 2021; Liu et al., 2022). Therefore, understanding of the molecular mechanism in the occurrence and development of CRC will help to develop novel therapies to optimize the treatment response throughout the disease course. In recent years, a large number of relevant CRC sequencing data have been generated, archived, and stored in public databases (Guo et al., 2017). Researches combining high-throughput sequencing data and bioinformatics analysis has gradually become a hot spot (Alves Martins et al., 2019; Zhao et al., 2019). Here, bioinformatics analysis of RNA-seq data of CRC patients may provide insights for drug repositioning for the treatment of CRC.

In this study, bioinformatics analysis was used to identify biomarkers of CRC and potential drugs that can improve the outcomes of CRC patients. Specifically, based on the TCGA data set and GSE104836 data set, we compared the transcriptome data of tumor samples and normal samples to identify differentially expressed genes (DEGs) on the two independent datasets. The DEGs were intersected for further analysis. Then these DEGs were further explored to detect the enriched GO terms and KEGG pathways. From those DEGs, latent drugs that can

improve the prognosis of patients from the Crowd Extracted Expression of Differential Signatures (CREEDS) were also predicted. In addition, the hub genes in the protein-protein interaction (PPI) network were discovered according to the DEGs and survival analysis was carried out on these hub genes. Finally, drug candidates could reverse hub genes were also predicted by CREEDS and validated by literatures.

## Materials and methods

### Data collection

RNA-seq data of CRC patients were downloaded from the Cancer Genome Atlas (TCGA) database (<https://portal.gdc.cancer.gov/>) and the Gene Expression Omnibus (GEO) database (<https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE104836>). Meanwhile, the associated clinical information of 478 tumor samples and 41 normal samples from TCGA, and 10 patients and 10 healthy controls from the GSE104836 dataset was obtained.

### Differentially expressed gene analysis

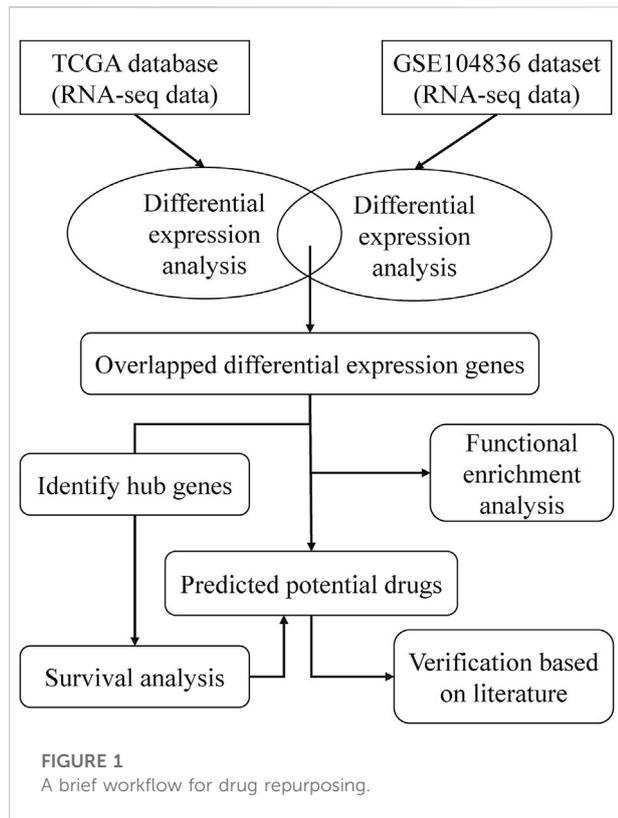
DESeq2 is a R package that can identify DEGs from raw count data. It uses the contraction estimation of discrete and the fold change of the gene expression to improve the stability and interpretability of the estimation, which makes the more quantitative analysis focus on intensity (Love et al., 2014). DEGs in CRC tumor samples and normal samples were detected using DESeq2 package with the criteria of  $p$ -value < 0.001 and  $\log_2$  |fold change|  $\geq 2$ .

### Functional and pathway enrichment analysis of DEGs

After DEG analysis of the TCGA dataset and GSE104836 dataset, overlapping DEGs were screened, and then enrichment analysis of KEGG pathway and GO (The Gene Ontology, 2019) including biological process (BP), cellular component (CC), and molecular function (MF) were carried out to reveal the altered biological characteristics of CRC. The R packages "clusterProfiler" and "ggplot" were used to visualize the results of the enrichment analysis.

### PPI network and hub genes analysis

The online database STRING (<http://string-db.org>) was used to develop a PPI network of DEGs, and the minimum required interaction score was 0.7. The Cytoscape software was used to visualize the PPI network and to analyze the structural properties



of the constructed network. The cytoHubba plug-in was used to identify hub genes in the PPI network.

## Potential drug identification

The CREEDS database consists of gene expression characteristics induced by single drug perturbation, which can be used to identify the relationship between genes, diseases, and drugs. To identify potential drugs for the treatment of CRC, we used the CREEDS database to find drugs that can reverse the DEGs. Specifically, for each drug in the CREEDS database, we calculated the  $p$ -value of the overlapping genes between downregulated genes of the drug and upregulated DEGs in CRC by hypergeometric test, and similarly, calculate the  $p$ -value of the overlapping genes between upregulated genes of the drug and downregulated DEGs in CRC. The drugs with any of the two  $p$ -value lower than 0.05 could be taken as candidates that could reverse the DEGs and might treat the CRC.

## Survival analysis

We obtained the OS time of all patients in TCGA database, and estimated the survival probability of CRC patients using Kaplan-

**TABLE 1** General clinical information of CRC patients included in this study.

Characteristics	No	
Type	Tumor	478
	Normal	41
Average age	67.04	
Gender	Female	247
	Male	272
Tumor stage	I	85
	II	209
	III	140
	IV	73
	Unknown	12

Meier method. Kaplan-Meier survival curve was used to estimate the 50th percentile (median) of survival time and compare the survival distribution of two or more groups. Log-rank test was also used to compare the survival differences between groups.  $p$ -value <0.05 was considered to have significant differences between groups. The data were analyzed by R software.

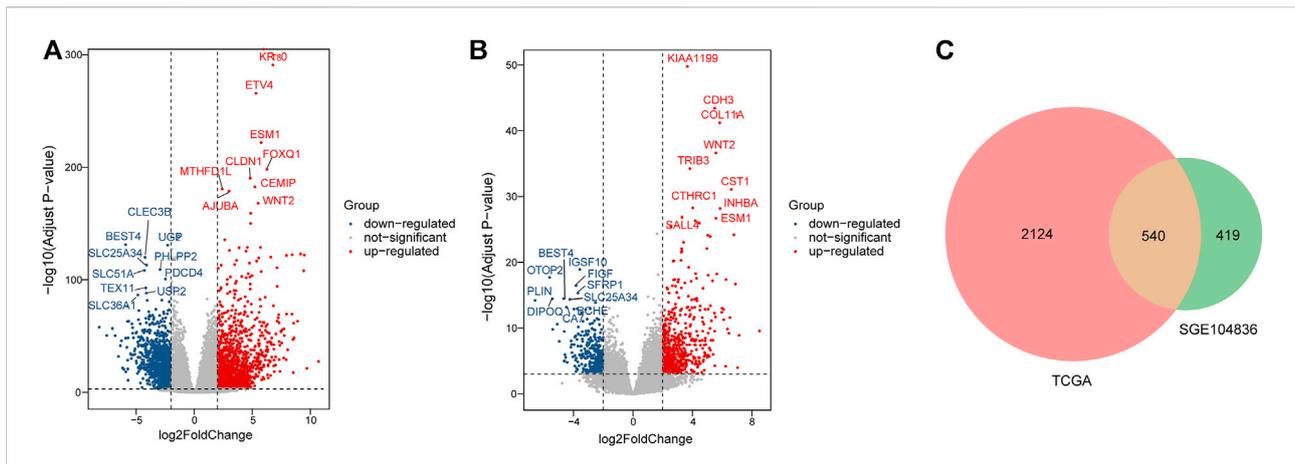
## Results

### A framework of CRC related drugs repurposing

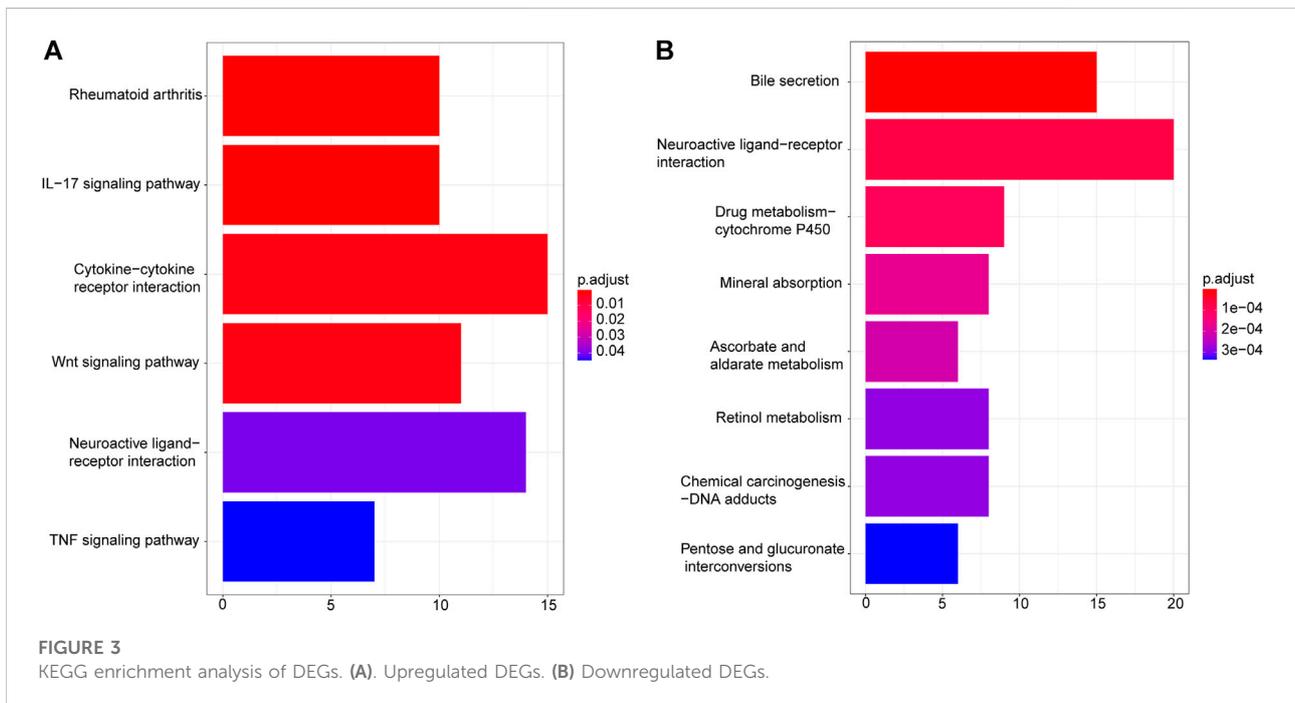
To find drugs that can be used to treat CRC, we proposed a bioinformatics pipeline of drug repurposing based on transcriptome data. The workflow was shown in Figure 1. After downloading the RNA-seq data from TCGA and GEO databases, we performed DEG analysis and pathway enrichment analysis. Then, the hub genes of DEGs were identified and survival analysis was done on the hub genes. According to the DEGs and CREEDS, drugs that could reverse the DEGs were identified, and 10 drugs can reverse the survival-related hub gene were further investigated. Finally, according to some previous studies, the effectiveness of the newly discovered drugs was verified.

### Patient characteristics

The RNA-seq data involved 478 tumor samples and 41 normal samples. There were 247 women and 272 men. 85 cases were at clinical stage I, 209 cases were at stage II, 140 cases were at stage III and 73 cases were at stage IV. Their average age was ~67 years old. The clinical features of patients from the TCGA dataset were shown in Table 1.



**FIGURE 2** Identification of DEGs between tumor tissues and normal tissues in CRC patients. (A–B). Differential expression volcanic map of (A)TCGA and (B) GEO dataset. Red dots indicate significant up-regulation, blue dots indicate significant down-regulation, and gray dots indicate genes with no significant changes. (C). Venn plot for DEGs detected in two datasets.



**FIGURE 3** KEGG enrichment analysis of DEGs. (A). Upregulated DEGs. (B). Downregulated DEGs.

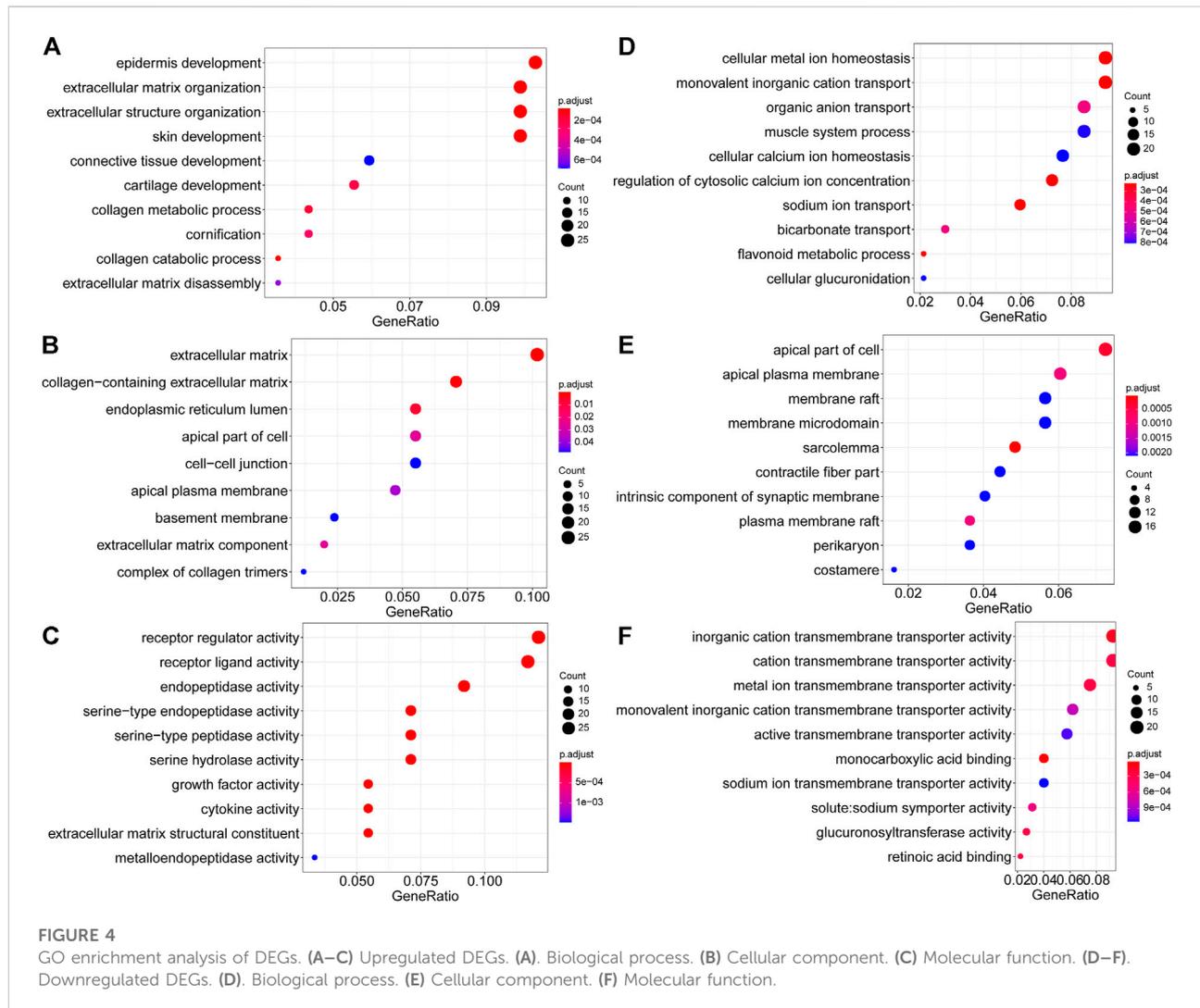
## DEGs identification

In total, 2664 DEGs (1537 upregulated genes and 1127 downregulated genes) and 959 DEGs (563 upregulated genes and 396 downregulated genes) were extracted from TCGA (Figure 2A) and GSE104836 (Figure 2B) datasets respectively using  $p\text{-value} < 0.001$  and  $|\log_2[\text{fold change}]| \geq 2$  as the cut-off criteria. A total of 540 DEGs (276 upregulated

genes and 264 downregulated genes) were identified in both datasets (Figure 2C).

## Enrichment Analysis

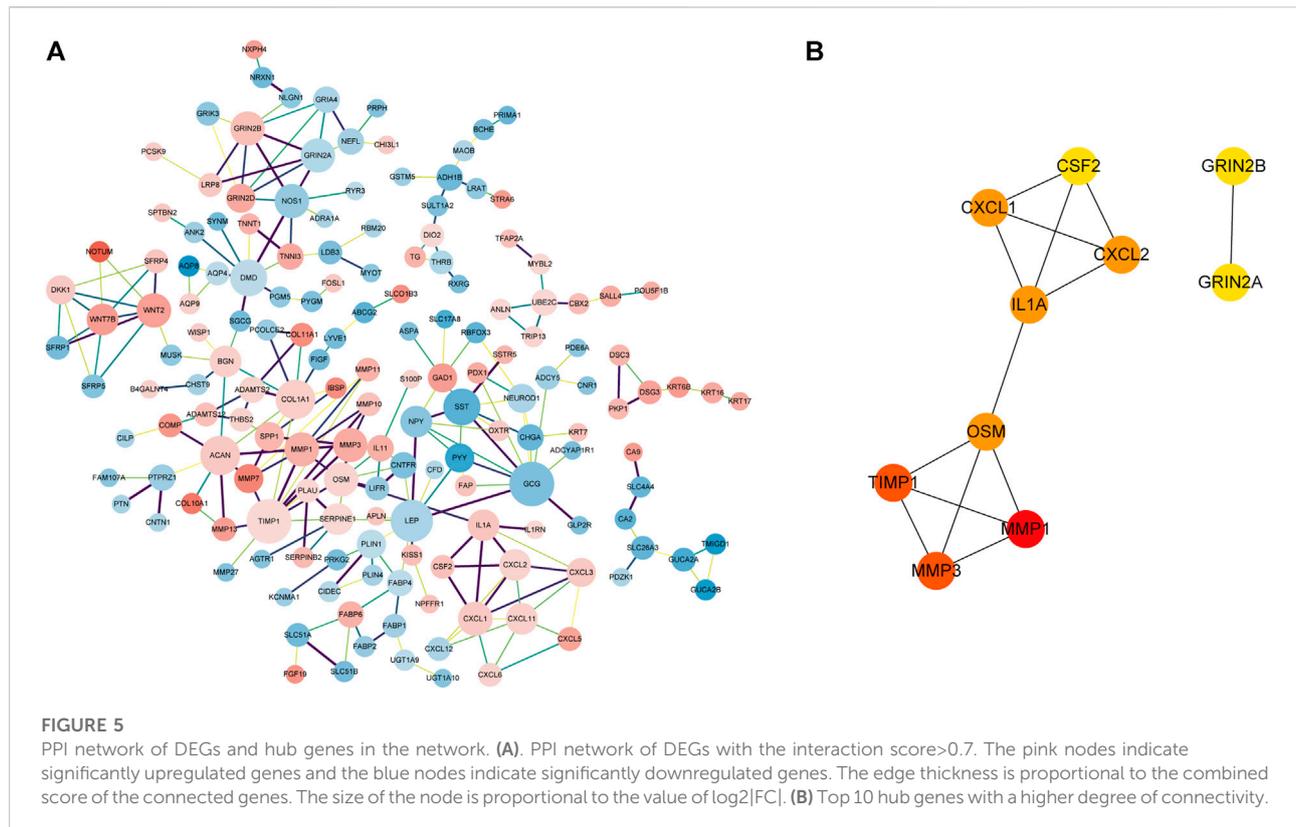
To understand the possible biological mechanisms that cause the identified changes in the transcriptome data, we conducted



the enrichment analysis on the overlapped DEGs using KEGG and GO databases. KEGG pathway enrichment results showed that upregulated DEGs were enriched in “Rheumatoid arthritis”, “IL-17 signaling pathway”, “Cytokine–cytokine receptor interaction”, “Wnt signaling pathway”, “Neuroactive ligand–receptor interaction”, and “TNF signaling pathway” (Figure 3A), while downregulated DEGs were enriched in “Bile secretion”, “Neuroactive ligand–receptor interaction”, “Drug metabolism – cytochrome P450”, “Mineral absorption”, “Ascorbate and aldarate metabolism”, “Retinol metabolism”, “Chemical carcinogenesis – DNA adducts”, and “Pentose and glucuronate interconversions” (Figure 3B).

GO terms cover biological process (BP), cellular component (CC), and molecular function (MF). For upregulated DEGs, the enriched BP terms included “epidermis development”, “extracellular matrix organization”, “extracellular structure organization”, “skin development”, “connective tissue development”, “cartilage development”, “collagen metabolic process”, “cornification”, “collagen catabolic process”, and “extracellular matrix disassembly” (Figure 4A). In the CC group, upregulated DEGs were primarily enriched in “extracellular matrix”, “collagen-containing extracellular matrix”, “endoplasmic reticulum lumen”, “apical part of cell”, “cell–cell junction”, “apical plasma membrane”, “basement membrane”, “extracellular matrix component” and “complex of collagen trimers” (Figure 4B). And enriched MF-related terms of upregulated DEGs were “receptor regulator activity”, “receptor ligand activity”, “endopeptidase activity”, “serine-type endopeptidase activity”, “serine-type peptidase activity”, “serine hydrolase activity”, “growth factor activity”, “cytokine activity” and “extracellular matrix structural constituent” (Figure 4C). For downregulated DEGs, the enriched BP terms were “cellular metal ion homeostasis”, “monovalent inorganic cation transport”, “organic anion transport”, “muscle system process”, “cellular calcium ion homeostasis”, “regulation of cytosolic calcium ion concentration”, “sodium ion transport”, “bicarbonate transport”, “flavonoid metabolic process”, and “cellular glucuronidation” (Figure 4D). In the CC group, downregulated DEGs were primarily enriched in “apical part of cell”, “apical plasma membrane”, “membrane raft”, “membrane microdomain”, “sarcolemma”, “contractile fiber part”, “intrinsic component of synaptic membrane”, “plasma membrane raft”, “perikaryon”, and “costamere” (Figure 4E). And enriched MF-related terms of downregulated DEGs were “inorganic cation transmembrane transporter activity”, “cation transmembrane transporter activity”, “metal ion transmembrane transporter activity”, “monovalent inorganic cation transmembrane transporter activity”, “active transmembrane transporter activity”, “monocarboxylic acid binding”, “sodium ion transmembrane transporter activity”, “solute:sodium symporter activity”, “glucuronosyltransferase activity”, and “retinoic acid binding” (Figure 4F).

development”, “collagen metabolic process”, “cornification”, “collagen catabolic process” (Figure 4A). In the CC group, upregulated DEGs were primarily enriched in “extracellular matrix”, “collagen-containing extracellular matrix”, “endoplasmic reticulum lumen”, “apical part of cell”, “cell–cell junction”, “apical plasma membrane”, “basement membrane”, “extracellular matrix component” and “complex of collagen trimers” (Figure 4B). And enriched MF-related terms of upregulated DEGs were “receptor regulator activity”, “receptor ligand activity”, “endopeptidase activity”, “serine-type endopeptidase activity”, “serine-type peptidase activity”, “serine hydrolase activity”, “growth factor activity”, “cytokine activity” and “extracellular matrix structural constituent” (Figure 4C). For downregulated DEGs, the enriched BP terms were “cellular metal ion homeostasis”, “monovalent inorganic cation transport”, “organic anion transport”, “muscle system process”, “cellular calcium ion homeostasis”, “regulation of cytosolic calcium ion concentration”, “sodium ion transport”, “bicarbonate transport”, “flavonoid metabolic process”, and “cellular glucuronidation” (Figure 4D). In the CC group, downregulated DEGs were primarily enriched in “apical part of cell”, “apical plasma membrane”, “membrane raft”, “membrane microdomain”, “sarcolemma”, “contractile fiber part”, “intrinsic component of synaptic membrane”, “plasma membrane raft”, “perikaryon”, and “costamere” (Figure 4E). And enriched MF-related terms of downregulated DEGs were “inorganic cation transmembrane transporter activity”, “cation transmembrane transporter activity”, “metal ion transmembrane transporter activity”, “monovalent inorganic cation transmembrane transporter activity”, “active transmembrane transporter activity”, “monocarboxylic acid binding”, “sodium ion transmembrane transporter activity”, “solute:sodium symporter activity”, “glucuronosyltransferase activity”, and “retinoic acid binding” (Figure 4F).



concentration”, “sodium ion transport”, “bicarbonate transport”, “flavonoid metabolic process” and “cellular glucuronidation” (Figure 4D). In the CC group, the downregulated DEGs were enriched in “apical part of cell”, “apical plasma membrane”, “membrane raft”, “membrane microdomain”, “sarcolemma”, “contractile fiber part”, “intrinsic component of synaptic membrane”, “plasma membrane raft”, “perikaryon” and “costamere” (Figure 4E). The enriched MF-related terms of the downregulated DEGs were “inorganic cation transmembrane transporter activity”, “cation transmembrane transporter activity”, “metal ion transmembrane transporter activity”, “monovalent inorganic cation transmembrane transporter activity”, “active transmembrane transporter activity”, “monocarboxylic acid binding”, “sodium ion transmembrane transporter activity”, “solute:sodium symporter activity”, “glucuronosyltransferase activity” and “retinoic acid binding” (Figure 4F).

## Hub genes in the PPI network of DEGs

Based on the STRING online database (<http://string-db.org>) and Cytoscape software, a PPI network of 164 DEGs and 241 edges was constructed. The minimum required

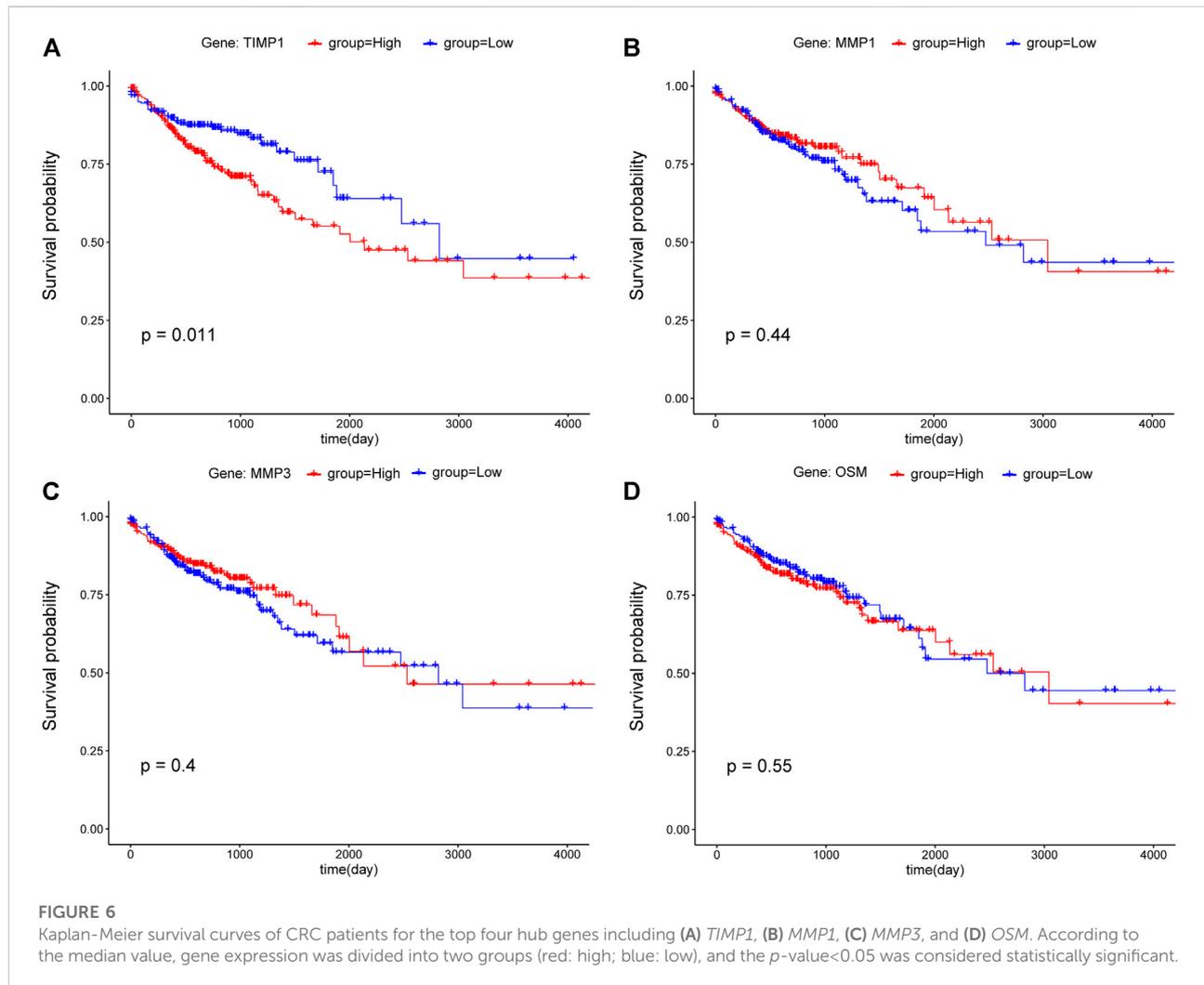
interaction score of each edge were bigger than 0.7 (Figure 5A) which excludes 376 DEGs. The top 10 hub genes according to the node degree were *MMP1*, *MMP3*, *TIMP1*, *OSM*, *IL1A*, *CXCL1*, *CXCL2*, *CSF2*, *GRIN2A*, and *GRIN2B* (Figure 5B).

## Correlation between hub genes expression and overall survival

To examine the potential relationship between DEGs and overall survival (OS), a weighted Kaplan Meier survival curves were generated from TCGA data. The survival curves of the top four hub genes were shown in Figure 6, which shown that only *TIMP1* is associated with OS ( $p$ -value < 0.05), and its high expression led to poor prognosis (Figure 6A). Other hub genes are not significantly associated with OS (Figures 6B–D) and Supplementary Figure S1.

## Identification of potential drugs

249 potential drugs were predicted according to the DEGs. For example, we plotted five drugs for upregulated DEGs and



five drugs for downregulated DEGs in Figure 7. Figure 7 indicated that formaldehyde, glucocorticoid[dexamethasone, paclitaxel|eribulin, messenger RNA|inhibitor, and eribulin|paclitaxel could reverse upregulated DEGs. fluoxetine|sucrose|antidepressant|imipramine, nevirapine, sucrose|antidepressant|imipramine|L-proline residue|, imipramine|sucrose|antidepressant|, and histone|N-methyl-D-aspartic acid could reverse the downregulated DEGs.

Since *TIMP1* is significantly related to the OS of CRC patients, and the high expression of *TIMP1* is correlated to a poor prognosis, we next looked for drugs/small molecules that can reverse the expression of *TIMP1*, which might improve the prognosis of CRC patients. We provided details of the top 10 drugs that can reverse the hub gene *TIMP1* in Table 2, including formaldehyde, paclitaxel|eribulin, erlotinib|dimethyl sulfoxide, glucocorticoid[dexamethasone, antagonist, trichostatin A, rosiglitazone, inhibitor, retinoic acid, and cisplatin. Among them, eight drugs/small molecules were

confirmed to be related to *TIMP1* or CRC. It is reported that exposure to formaldehyde can reduce *TIMP1* expression (Kang et al., 2022).

## Discussion

In recent decades, CRC, including colon and rectal cancer, has become one of the main causes of cancer-related death around the world (Fuccio et al., 2018; Røed Skårderud et al., 2018; He et al., 2020a; He et al., 2020b). Therefore, it is urgent to find more effective prevention and treatment to reverse this problem (Teer et al., 2017). With the recent progress in the field of medicine and biotechnology, many preclinical and clinical studies have been carried out to reveal the potential mechanism of CRC liver metastasis. Identifying cancer-related marker genes through gene-targeted therapy is a new and effective potentially powerful treatment for CRC (Okugawa

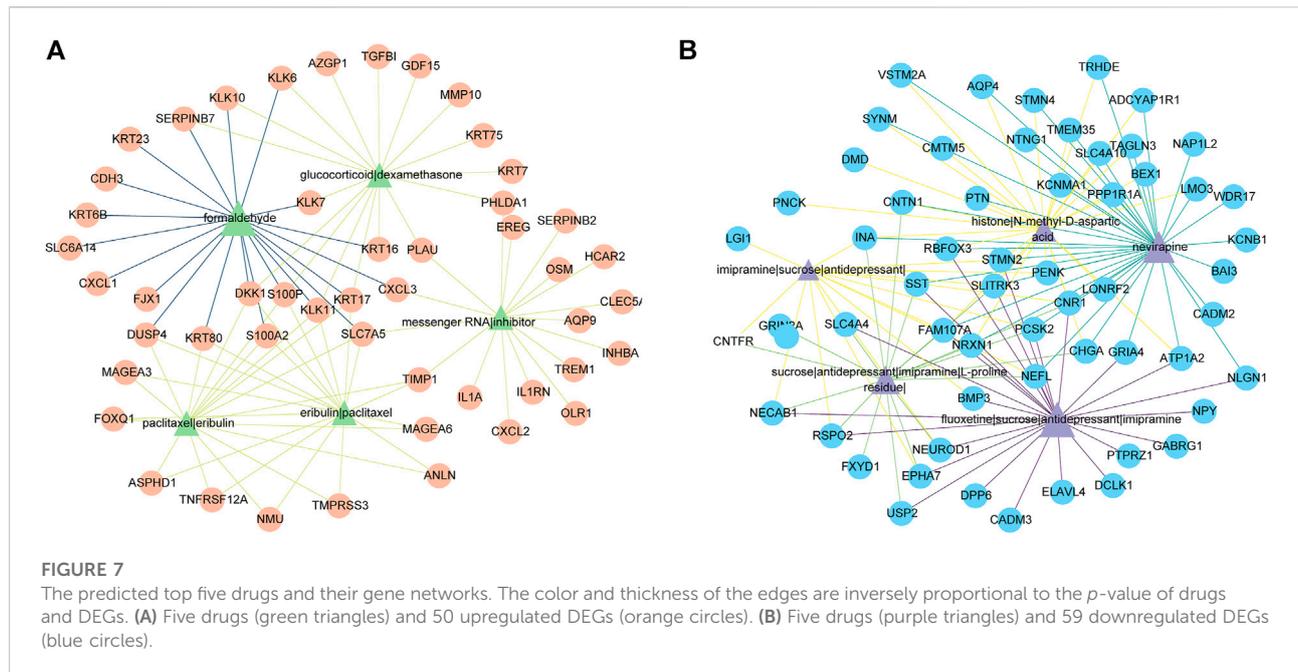


TABLE 2 Top 10 drugs for *TIMP1* that were significantly associated with survival rate of CRC patients.

Gene name	Drug/Small molecule	<i>p</i> -value	Possible effect	Evidence
<i>TIMP1</i>	formaldehyde	1.04403630163397E-06	Formaldehyde is a colorless, irritant, highly active and toxic environmental pollutant, which is used in various industries and products. Inhaled formaldehyde is a human and animal carcinogen that can cause genotoxicity, such as the formation of reactive oxygen species and DNA damage	PMID:35379891
	paclitaxel eribulin	8.64715E-06	A well-known anticancer agent with a unique mechanism of action. It is considered to be one of the most successful natural anticancer drugs	Unconfirmed
	erlotinib dimethyl sulfoxide	1.74273E-05	It can interfere with a variety of cellular processes, such as cell proliferation, differentiation, apoptosis and cycle	PMID: 32911099
	glucocorticoid dexamethasone	3.5055E-05	It has pharmacological effects of anti-inflammatory, anti-endotoxin, inhibiting immunity, anti-shock and enhancing stress response	PMID: 21789017
	antagonist	3.66683E-05	It can bind to receptors and has strong affinity without intrinsic activity ( $\alpha = 0$ ) drugs	Unconfirmed
	trichostatin A	0.0000703775443677834	trichostatin A (TSA), a histone deacetylase (HDAC) inhibitor	PMID: 21520296
	rosiglitazone	0.000141022683961323	Rosiglitazone is a thiazolidinedione insulin sensitizer. Its mechanism of action is similar to that of specific peroxisome proliferator activator $\gamma$ Type a receptor	PMID: 29743857
	Inhibitor	0.000282045	Inhibitors of proteinases or antibodies against certain proteolytic enzymes can prevent tumor invasion and metastasis in experimental conditions	PMID: 23202950
	retinoic acid	0.000564894	Retinoic acid (RA) signal transduction is an important and conservative way to regulate cell proliferation and differentiation. In addition, disturbed RA signaling is associated with the occurrence and progression of cancer	PMID: 34877501
	cisplatin	0.000758387	Cisplatin is an inorganic platinum complex, which can be inhibited by the formation of DNA adducts in tumor cells	PMID: 32329836; PMID: 20607860

et al., 2015; Guo et al., 2017). High throughput sequencing technology provides a new perspective on the genome, transcriptome, and epigenome characteristics of cancer. In this study, we aim to reveal the hub gene of CRC through

bioinformatics methods and identify potential drugs or small molecules, to improve the predictive power of CRC and provide a valuable theoretical basis for the clinical treatment of CRC patients.

First, RNA-seq data and clinical information of 478 CRC tumor samples and 41 healthy control samples were downloaded from TCGA. In addition, RNA-seq data of 10 tumor samples and 10 normal samples were obtained from the GSE104836 dataset. Using DESeq2 to detect the DEGs from TCGA and GEO respectively, 2664 DEGs were identified from TCGA, 959 DEGs were identified from the GSE104836 data set, and 540 DEGs appeared in both datasets, including 276 upregulated genes and 264 downregulated genes. KEGG pathway enrichment results showed that upregulated DEGs are enriched in “Rheumatoid arthritis”, “IL-17 signaling pathway”, “Cytokine–cytokine receptor interaction”, “Wnt signaling pathway”, “Neuroactive ligand–receptor interaction”, and “TNF signaling pathway” (Figure 3A). It has been reported that IL-17 is able to regulate colorectal tumor cells and inhibits their production of cxcl9/10 chemokines, thus prevents the infiltration of CD8 + CTLs and Tregs into CRC tumor, thereby promoting the development of CRC (Chen et al., 2019). Wnt signaling pathway is the key medium of tissue homeostasis and repair. Almost all CRC tumors show overactivation of Wnt pathway (Schatoff et al., 2017; Bian et al., 2020). GO enrichment analysis shows that epidermis development, extracellular matrix, and receptor regulator activity are the most significantly abundant upregulated DEGs in biological processes, cellular components, and molecular function categories. Downregulated DEGs are enriched in “Bile secretion”, “Neuroactive ligand–receptor interaction”, “Drug metabolism–cytochrome P450”, “Mineral absorption”, “Ascorbate and aldarate metabolism”, “Retinol metabolism”, “Chemical carcinogenesis–DNA adducts”, and “Pentose and glucuronate interconversions” (Figure 3B). Previous studies have shown that a high-fat diet promotes the secretion of bile acids, thereby inducing the formation of precancerous lesions and/or aggravating the occurrence of colon tumors (Ocvirk and O’Keefe, 2021). Neuroactive ligand–receptor interactions were associated with other gastrointestinal cancers (Yu et al., 2021). The lack and deficiency of minerals may be related to cancer and increase the risk of cancer; For example, effective absorption of vitamin D can prevent colorectal cancer (Takada and Makishima, 2017).

To identify the key regulating genes in CRC development, a PPI network was constructed based on overlapping DEGs. In this network, edges with association scores <0.7 were filtered out. The PPI network obtained based STRING online database has 164 nodes, and the top 10 hub genes, including *MMP1*, *MMP3*, *TIMP1*, *OSM*, *IL1A*, *CXCL1*, *CXCL2*, *CSF2*, *GRIN2A*, and *GRIN2B*, were identified using Cytoscape. Among them, *TIMP1* is a soluble protein that can be released from endometrial cells, fibroblasts, and cancer cells, which are correlated with the prognosis of various cancers (Peng et al., 2011; Wang et al., 2013). The Kaplan–Meier survival analysis of Zheng et al. showed that *TIMP1* expression was upregulated in CRC tissues and was also connected with poor prognosis in GEPIA datasets ( $p$ -value = 0.02) (Zheng et al., 2020). Song et al. (2016) reported that *TIMP1* depletion can inhibit the proliferation, migration, and invasion of

colon cancer cells, and inhibit the tumorigenesis and metastasis of CRC. Consistent with these studies, our results show that *TIMP1* was up-regulated in CRC samples compared with matched normal tissue samples, and its high expression was associated with poor OS in CRC patients.

Based on DEGs and CREEDS, we made drug predictions for all DEGs (Yang et al., 2020). Previous studies have shown the anti-migration and anti-invasion effects of imipramine, an FDA-approved antidepressant oral drug, on CRC cells (Liu et al., 2016; Albuquerque-González et al., 2020). Fluoxetine has been shown to induce antitumor activity. It was found that fluoxetine could selectively induce concentration-dependent apoptosis in human CRC cells by changing mitochondrial membrane potential and inducing phosphatidylserine translocation to the outer membrane (Marcinkute et al., 2019). In addition, 10 potential drugs were identified to reverse the expression of *TIMP1*. It has been shown that after glucocorticoid treatment, the expression level of *TIMP1* in patients with idiopathic pulmonary fibrosis (IPF) were significantly lower than those before glucocorticoid treatment ( $p < 0.05$ ) (Zhang et al., 2015). Dexamethasone is a synthetic steroid with anti-inflammatory, anti-allergic, and immunosuppressive properties (Sinner, 2019). Trichostatin A is a histone deacetylase (HDAC) inhibitor, which inhibits the growth of CRC cells and induces G1 cell cycle arrest and apoptosis by regulating the downstream target of the JAK2/STAT3 signal (Xiong et al., 2012). A study on the effect of cisplatin on the invasion of ovarian cancer cells showed that the use of cisplatin could reduce the expression of *TIMP1* by 5.0 times ( $p < 0.05$ ) (Karam et al., 2010). It is worth noting that there is no relevant evidence that paclitaxel/eribulin, and Antiagonist are related to the expression of *TIMP1* or the outcome of CRC. Further experiments are needed to verify their effectiveness of action, which may provide a basis for guiding the treatment of CRC patients.

Overall, this study revealed the altered gene expressions and enriched pathways in CRC based on bioinformatics analyses and provides insights for further screening of effective biomolecules for CRC treatment intervention, which is of clinical significance. However, the current research has some limitations. First, because the candidate prognosis-related central DEGs were detected using the data from two independent databases, more datasets were needed to confirm our discoveries. Secondly, experimental methods such as PCR were also needed to verify the DEGs. Third, clinical trials were needed to identify effects of the predicted drugs.

## Conclusion

Our study effectively identified several candidate drug targets through differentially gene expression analysis, hub gene analysis and survival analysis for CRC treatment. We revealed compounds that have the potential to reverse the expressions

of the identified DEGs. These findings provide new directions for the diagnosis and treatment of CRC.

## 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.

## Author contributions

SZ conceived and designed the study. XL, JY, and TL performed the experiments. CZ and ZC analyzed the data. XL wrote the manuscript. CL and JS contributed to literature review

## References

- Albuquerque-González, B., Bernabe-García, M., Montoro-García, S., Bernabe-García, A., Rodrigues, P. C., Ruiz Sanz, J., et al. (2020). New role of the antidepressant imipramine as a Fascin1 inhibitor in colorectal cancer cells. *Exp. Mol. Med.* 52 (2), 281–292. doi:10.1038/s12276-020-0389-x
- Alves Martins, B. A., de Bulhões, G. F., Cavalcanti, I. N., Martins, M. M., de Oliveira, P. G., and Martins, A. M. A. (2019). Biomarkers in colorectal cancer: The role of translational proteomics research. *Front. Oncol.* 9, 1284. doi:10.3389/fonc.2019.01284
- Arnold, M., Sierra, M. S., Laversanne, M., Soerjomataram, I., Jemal, A., and Bray, F. (2017). Global patterns and trends in colorectal cancer incidence and mortality. *Gut* 66 (4), 683–691. doi:10.1136/gutjnl-2015-310912
- Bian, J., Dannappel, M., Wan, C., and Firestein, R. (2020). Transcriptional regulation of wnt/ $\beta$ -catenin pathway in colorectal cancer. *Cells* 9 (9), E2125. doi:10.3390/cells9092125
- Chen, J., Ye, X., Pitmon, E., Lu, M., Wan, J., Jellison, E. R., et al. (2019). IL-17 inhibits CXCL9/10-mediated recruitment of CD8(+) cytotoxic T cells and regulatory T cells to colorectal tumors. *J. Immunother. Cancer* 7 (1), 324. doi:10.1186/s40425-019-0757-z
- Cheng, L., Qi, C., Yang, H., Lu, M., Cai, Y., Fu, T., et al. (2021). gutMGene: a comprehensive database for target genes of gut microbes and microbial metabolites. *Nucleic Acids Res.* 50, D795–D800. doi:10.1093/nar/gkab786
- Cheng, L., Qi, C., Zhuang, H., Fu, T., and Zhang, X. (2020). gutMDisorder: a comprehensive database for dysbiosis of the gut microbiota in disorders and interventions. *Nucleic Acids Res.* 48 (D1), D554–D560. doi:10.1093/nar/gkz843
- De Sousa, E. M. F., Wang, X., Jansen, M., Fessler, E., Trinh, A., de Rooij, L. P. M. H., et al. (2013). Poor-prognosis colon cancer is defined by a molecularly distinct subtype and develops from serrated precursor lesions. *Nat. Med.* 19 (5), 614–618. doi:10.1038/nm.3174
- Deshiere, A., Berthet, N., Lecouturier, F., Gaudaire, D., and Hans, A. (2019). Molecular characterization of Equine Infectious Anemia Viruses using targeted sequence enrichment and next generation sequencing. *Virology* 537, 121–129. doi:10.1016/j.virol.2019.08.016

and language editing. All authors have read and approved this manuscript.

## Conflicts 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.

## Publisher's note

All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

## Supplementary material

The Supplementary Material for this article can be found online at: <https://www.frontiersin.org/articles/10.3389/fgene.2022.1017539/full#supplementary-material>

- Fuccio, L., Repici, A., Hassan, C., Ponchon, T., Bhandari, P., Jover, R., et al. (2018). Why attempt *en bloc* resection of non-pedunculated colorectal adenomas? A systematic review of the prevalence of superficial submucosal invasive cancer after endoscopic submucosal dissection. *Gut* 67 (8), 1464–1474. doi:10.1136/gutjnl-2017-315103
- Grassetto, G., Capirci, C., Marzola, M. C., Rampin, L., Chondrogiannis, S., Musto, A., et al. (2012). Colorectal cancer: Prognostic role of 18F-FDG-PET/CT. *Abdom. Imaging* 37 (4), 575–579. doi:10.1007/s00261-011-9789-7
- Guo, Y., Bao, Y., Ma, M., and Yang, W. (2017). Identification of key candidate genes and pathways in colorectal cancer by integrated bioinformatical analysis. *Int. J. Mol. Sci.* 18 (4), E722. doi:10.3390/ijms18040722
- He, B., Dai, C., Lang, J., Bing, P., Tian, G., Wang, B., et al. (2020). A machine learning framework to trace tumor tissue-of-origin of 13 types of cancer based on DNA somatic mutation. *Biochim. Biophys. Acta. Mol. Basis Dis.* 1866 (11), 165916. doi:10.1016/j.bbdis.2020.165916
- He, B., Lang, J., Wang, B., Liu, X., Lu, Q., He, J., et al. (2020). TOOme: A novel computational framework to infer cancer tissue-of-origin by integrating both gene mutation and expression. *Front. Bioeng. Biotechnol.* 8, 394. doi:10.3389/fbioe.2020.00394
- Hong, J., Lin, X., Hu, X., Wu, X., and Fang, W. (2021). A five-gene signature for predicting the prognosis of colorectal cancer. *Curr. Gene Ther.* 21 (4), 280–289. doi:10.2174/1566523220666201012151803
- Kang, D. S., Lee, N., Shin, D. Y., Jang, Y. J., Lee, S. H., Lim, K. M., et al. (2022). Network-based integrated analysis for toxic effects of high-concentration formaldehyde inhalation exposure through the toxicogenomic approach. *Sci. Rep.* 12 (1), 5645. doi:10.1038/s41598-022-09673-0
- Karam, A. K., Santiskulvong, C., Fekete, M., Zabih, S., Eng, C., and Dorigo, O. (2010). Cisplatin and PI3kinase inhibition decrease invasion and migration of human ovarian carcinoma cells and regulate matrix-metalloproteinase expression. *Cytoskelet. Hob.* 67 (8), 535–544. doi:10.1002/cm.20465
- Kong, F., Zou, H., Liu, X., He, J., Zheng, Y., Xiong, L., et al. (2020). miR-7112-3p targets PERK to regulate the endoplasmic reticulum stress pathway and apoptosis

- induced by photodynamic therapy in colorectal cancer CX-1 cells. *Photodiagnosis Photodyn. Ther.* 29, 101663. doi:10.1016/j.pdpdt.2020.101663
- Liu, C., Wei, D., Xiang, J., Ren, F., Huang, L., Lang, J., et al. (2020). An improved anticancer drug-response prediction based on an ensemble method integrating matrix completion and ridge regression. *Mol. Ther. Nucleic Acids* 21, 676–686. doi:10.1016/j.omtn.2020.07.003
- Liu, H., Qiu, C., Wang, B., Bing, P., Tian, G., Zhang, X., et al. (2021). Evaluating DNA methylation, gene expression, somatic mutation, and their combinations in inferring tumor tissue-of-origin. *Front. Cell Dev. Biol.* 9, 619330. doi:10.3389/fcell.2021.619330
- Liu, J., Lan, Y., Tian, G., and Yang, J. (2022). A systematic framework for identifying prognostic genes in the tumor microenvironment of colon cancer. *Front. Oncol.* 12, 899156. doi:10.3389/fonc.2022.899156
- Liu, X., Yang, J., Zhang, Y., Fang, Y., Wang, F., Wang, J., et al. (2016). A systematic study on drug-response associated genes using baseline gene expressions of the Cancer Cell Line Encyclopedia. *Sci. Rep.* 6, 22811. doi:10.1038/srep22811
- Love, M. I., Huber, W., and Anders, S. (2014). Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. *Genome Biol.* 15 (12), 550. doi:10.1186/s13059-014-0550-8
- Marcinkute, M., Afshinjavid, S., Fatokun, A. A., and Javid, F. A. (2019). Fluoxetine selectively induces p53-independent apoptosis in human colorectal cancer cells. *Eur. J. Pharmacol.* 857, 172441. doi:10.1016/j.ejphar.2019.172441
- McQuade, R. M., Stojanovska, V., Bornstein, J. C., and Nurgali, K. (2017). Colorectal cancer chemotherapy: The evolution of treatment and new approaches. *Curr. Med. Chem.* 24 (15), 1537–1557. doi:10.2174/092986732466617011152436
- Meng, Y., Lu, C., Jin, M., Xu, J., Zeng, X., and Yang, J. (2022). A weighted bilinear neural collaborative filtering approach for drug repositioning. *Brief. Bioinform.* 23 (2), bbab581. doi:10.1093/bib/bbab581
- Nguyen, H. T., and Duong, H. Q. (2018). The molecular characteristics of colorectal cancer: Implications for diagnosis and therapy. *Oncol. Lett.* 16 (1), 9–18. doi:10.3892/ol.2018.8679
- O'Connell, J. B., Maggard, M. A., and Ko, C. Y. (2004). Colon cancer survival rates with the new American Joint Committee on Cancer sixth edition staging. *J. Natl. Cancer Inst.* 96 (19), 1420–1425. doi:10.1093/jnci/djh275
- Ocvirk, S., and O'Keefe, S. J. D. (2021). Dietary fat, bile acid metabolism and colorectal cancer. *Semin. Cancer Biol.* 73, 347–355. doi:10.1016/j.semcancer.2020.10.003
- Okugawa, Y., Grady, W. M., and Goel, A. (2015). Epigenetic alterations in colorectal cancer: Emerging biomarkers. *Gastroenterology* 149 (5), 1204–1225. doi:10.1053/j.gastro.2015.07.011.e12
- Peng, L., Yanjiao, M., Ai-guo, W., Pengtao, G., Jianhua, L., Ju, Y., et al. (2011). A fine balance between CCN1 and TIMP1 contributes to the development of breast cancer cells. *Biochem. Biophys. Res. Commun.* 409 (2), 344–349. doi:10.1016/j.bbrc.2011.05.021
- Røed Skårderud, M., Polk, A., Kjeldgaard Vistisen, K., Larsen, F. O., and Nielsen, D. L. (2018). Efficacy and safety of regorafenib in the treatment of metastatic colorectal cancer: A systematic review. *Cancer Treat. Rev.* 62, 61–73. doi:10.1016/j.ctrv.2017.10.011
- Rogler, G. (2014). Chronic ulcerative colitis and colorectal cancer. *Cancer Lett.* 345 (2), 235–241. doi:10.1016/j.canlet.2013.07.032
- Sadanandam, A., Lyssiottis, C. A., Homicsko, K., Collisson, E. A., Gibb, W. J., Wullschlegel, S., et al. (2013). A colorectal cancer classification system that associates cellular phenotype and responses to therapy. *Nat. Med.* 19 (5), 619–625. doi:10.1038/nm.3175
- Saito, M., Momma, T., and Kono, K. (2018). Targeted therapy according to next generation sequencing-based panel sequencing. *Fukushima J. Med. Sci.* 64 (1), 9–14. doi:10.5387/fms.2018-02
- Schatoff, E. M., Leach, B. I., and Dow, L. E. (2017). Wnt signaling and colorectal cancer. *Curr. Colorectal Cancer Rep.* 13 (2), 101–110. doi:10.1007/s11888-017-0354-9
- Siegel, R., DeSantis, C., Virgo, K., Stein, K., Mariotto, A., Smith, T., et al. (2012). Cancer treatment and survivorship statistics, 2012. *Ca. Cancer J. Clin.* 62 (4), 220–241. doi:10.3322/caac.21149
- Sinner, B. (2019). [Perioperative dexamethasone]. *Anaesthesist* 68 (10), 676–682. doi:10.1007/s00101-019-00672-x
- Song, G., Xu, S., Zhang, H., Wang, Y., Xiao, C., Jiang, T., et al. (2016). TIMP1 is a prognostic marker for the progression and metastasis of colon cancer through FAK-PI3K/AKT and MAPK pathway. *J. Exp. Clin. Cancer Res.* 35 (1), 148. doi:10.1186/s13046-016-0427-7
- Sung, H., Ferlay, J., Siegel, R. L., Laversanne, M., Soerjomataram, I., Jemal, A., et al. (2021). Global cancer statistics 2020: GLOBOCAN estimates of incidence and mortality worldwide for 36 cancers in 185 countries. *Ca. Cancer J. Clin.* 71 (3), 209–249. doi:10.3322/caac.21660
- Swiderska, M., Choromanska, B., Dabrowska, E., Konarzewska-Duchnowska, E., Choromanska, K., Szczurko, G., et al. (2014). The diagnostics of colorectal cancer. *Contemp. Oncol.* 18 (1), 1–6. doi:10.5114/wo.2013.39995
- Takada, I., and Makishima, M. (2017). Control of inflammatory bowel disease and colorectal cancer by synthetic vitamin D receptor ligands. *Curr. Med. Chem.* 24 (9), 868–875. doi:10.2174/0929867323666161202145509
- Teer, J. K., Zhang, Y., Chen, L., Welsh, E. A., Cress, W. D., Eschrich, S. A., et al. (2017). Evaluating somatic tumor mutation detection without matched normal samples. *Hum. Genomics* 11 (1), 22. doi:10.1186/s40246-017-0118-2
- The Gene Ontology, C. (2019). The gene Ontology resource: 20 years and still GOing strong. *Nucleic Acids Res.* 47 (D1), D330–D338–D338. doi:10.1093/nar/gky1055
- Wang, Y. Y., Li, L., Zhao, Z. S., and Wang, H. J. (2013). Clinical utility of measuring expression levels of KAP1, TIMP1 and STC2 in peripheral blood of patients with gastric cancer. *World J. Surg. Oncol.* 11, 81. doi:10.1186/1477-7819-11-81
- Xiong, H., Du, W., Zhang, Y. J., Hong, J., Su, W. Y., Tang, J. T., et al. (2012). Trichostatin A, a histone deacetylase inhibitor, suppresses JAK2/STAT3 signaling via inducing the promoter-associated histone acetylation of SOCS1 and SOCS3 in human colorectal cancer cells. *Mol. Carcinog.* 51 (2), 174–184. doi:10.1002/mc.20777
- Yang, J., Peng, S., Zhang, B., Houten, S., Schadt, E., Zhu, J., et al. (2020). Human geroprotector discovery by targeting the converging subnetworks of aging and age-related diseases. *Geroscience* 42 (1), 353–372. doi:10.1007/s11357-019-00106-x
- Yang, M., Yang, H., Ji, L., Hu, X., Tian, G., Wang, B., et al. (2022). A multi-omics machine learning framework in predicting the survival of colorectal cancer patients. *Comput. Biol. Med.* 146, 105516. doi:10.1016/j.combiomed.2022.105516
- Yu, J., Zhang, Q., Wang, M., Liang, S., Huang, H., Xie, L., et al. (2021). Comprehensive analysis of tumor mutation burden and immune microenvironment in gastric cancer. *Biosci. Rep.* 41 (2), BSR20203336. doi:10.1042/BSR20203336
- Zhang, H. T., Fang, S. C., Wang, C. Y., Wang, W., Wu, J., Wang, C., et al. (2015). MMP-9 1562C>T gene polymorphism and efficacy of glucocorticoid therapy in idiopathic pulmonary fibrosis patients. *Genet. Test. Mol. Biomarkers* 19 (11), 591–597. doi:10.1089/gtmb.2015.0057
- Zhang, Y., Xiang, J., Tang, L., Li, J., Lu, Q., Tian, G., et al. (2021). Identifying breast cancer-related genes based on a novel computational framework involving KEGG pathways and PPI network modularity. *Front. Genet.* 12, 596794. doi:10.3389/fgene.2021.596794
- Zhao, B., Baloch, Z., Ma, Y., Wan, Z., Huo, Y., Li, F., et al. (2019). Identification of potential key genes and pathways in early-onset colorectal cancer through bioinformatics analysis. *Cancer Control.* 26 (1), 1073274819831260. doi:10.1177/1073274819831260
- Zhao, T., Hu, Y., Zang, T., and Cheng, L. (2020). MRTFB regulates the expression of NOMO1 in colon. *Proc. Natl. Acad. Sci. U. S. A.* 117 (14), 7568–7569. doi:10.1073/pnas.2000499117
- Zheng, Z., Xie, J., Xiong, L., Gao, M., Qin, L., Dai, C., et al. (2020). Identification of candidate biomarkers and therapeutic drugs of colorectal cancer by integrated bioinformatics analysis. *Med. Oncol.* 37 (11), 104. doi:10.1007/s12032-020-01425-2