



Identification of Hub Genes and Construction of a Transcriptional Regulatory Network Associated With Tumor Recurrence in Colorectal Cancer by Weighted Gene Co-expression Network Analysis

Shengwei Liu^{1,2}, Fanping Zeng¹, Guangwen Fan¹ and Qiyong Dong^{1*}

¹ Department of Pharmacy, Yongchuan Hospital of Chongqing Medical University, Chongqing, China, ² Chongqing Key Laboratory of Biochemistry and Molecular Pharmacology, School of Pharmacy, Chongqing Medical University, Chongqing, China

OPEN ACCESS

Edited by:

Michael F. Miles,
Virginia Commonwealth University,
United States

Reviewed by:

Maria Rosaria Sapienza,
European Institute of Oncology (IEO),
Italy
Emil Bulatov,
Kazan Federal University, Russia

*Correspondence:

Qiyong Dong
1345263385@qq.com

Specialty section:

This article was submitted to
Cancer Genetics and Oncogenomics,
a section of the journal
Frontiers in Genetics

Received: 05 January 2021

Accepted: 15 March 2021

Published: 07 April 2021

Citation:

Liu S, Zeng F, Fan G and Dong Q
(2021) Identification of Hub Genes
and Construction of a Transcriptional
Regulatory Network Associated With
Tumor Recurrence in Colorectal
Cancer by Weighted Gene
Co-expression Network Analysis.
Front. Genet. 12:649752.
doi: 10.3389/fgene.2021.649752

Tumor recurrence is one of the most important risk factors that can negatively affect the survival rate of colorectal cancer (CRC) patients. However, the key regulators dictating this process and their exact mechanisms are understudied. This study aimed to construct a gene co-expression network to predict the hub genes affecting CRC recurrence and to inspect the regulatory network of hub genes and transcription factors (TFs). A total of 177 cases from the GSE17536 dataset were analyzed via weighted gene co-expression network analysis to explore the modules related to CRC recurrence. Functional annotation of the key module genes was assessed through Gene Ontology and Kyoto Encyclopedia of Genes and Genomes analyses. The protein and protein interaction network was then built to screen hub genes. Samples from the Cancer Genome Atlas (TCGA) were further used to validate the hub genes. Construction of a TFs-miRNAs-hub genes network was also conducted using StarBase and Cytoscape approaches. After identification and validation, a total of five genes (TIMP1, SPARCL1, MYL9, TPM2, and CNN1) were selected as hub genes. A regulatory network of TFs-miRNAs-targets with 29 TFs, 58 miRNAs, and five hub genes was instituted, including model GATA6-MIR106A-CNN1, SP4-MIR424-TPM2, SP4-MIR326-MYL9, ETS1-MIR22-TIMP1, and ETS1-MIR22-SPARCL1. In conclusion, the identification of these hub genes and the prediction of the Regulatory relationship of TFs-miRNAs-hub genes may provide a novel insight for understanding the underlying mechanism for CRC recurrence.

Keywords: colorectal cancer, tumor recurrence, hub genes, transcription factors, weighted gene co-expression network analysis

INTRODUCTION

Colorectal cancer (CRC) ranks as the third most common type of tumor around the world and accounts for more than 7% of overall cancer-related death in China (Liu et al., 2015; Siegel et al., 2020). Despite the massive progress in understanding its genetic mechanism and improvement of surgical techniques, the overall survival time of CRC patients is still not remarkably improved,

which could be ascribed primarily to the high recurrence rates (Yu et al., 2017; Ha et al., 2019; Mirgayazova et al., 2019). Therefore, identifying the key regulators involved in CRC recurrence and deciphering their underlying mechanisms are critical for CRC prognosis and the development of novel therapeutic drugs.

With the recent developments in bioinformatics, a number of well-designed and effective methods are now available for the comprehensive identification of biomarkers in cancer and the prediction of cancer-related signaling pathways (Carter et al., 2004; Carey et al., 2005; Liu et al., 2007). Among them, the weighted gene co-expression network analysis (WGCNA) approach provides a systematic biology strategy to identify modules of highly correlated genes and construct a co-expression gene network (Langfelder and Horvath, 2008). By utilizing the WGCNA algorithm, genes with similar co-expression patterns are classified into a set of modules, in which the most central genes could be further identified as hub genes. Compared with the other methodologies applied to analyze the high-throughput sequencing data, WGCNA implements methods for both weighted and unweighted correlation networks and provides a more effective mean to explore the potential association between modules and sample traits.

In recent years, WGCNA has been widely used to uncover the potential biomarkers associated with clinical parameters in various cancer types (Liang et al., 2020; Wang et al., 2020). In terms of CRC, PIGU was identified as a key modulator that is closely related to KRAS mutant CRC patients (Zhang M. et al., 2020), while FBXW4 was reported to be associated with chemotherapy resistance and prognosis of CRC (Zhang Y. et al., 2020). With the help of WGCNA, a set of four long non-coding RNAs (lncRNAs) were found to be significantly correlated with the carcinogenesis and progression of colon adenocarcinoma (COAD) (Jiang et al., 2019). In addition, some efforts have also been taken to identify the crucial regulators associated with the tumor recurrence of CRC patients through WGCNA (Qiu et al., 2020; Wu et al., 2020). However, the number of hub genes identified in the existed studies is relatively low, and the important upstream mediators of these hub genes remain to be fully investigated.

In the current study, WGCNA was constructed based on the GSE17536 dataset containing gene expression profiling results from 177 CRC patients. The specific module associated with the recurrence status of CRC was identified. More importantly, hub genes that play essential roles during CRC recurrence were dug out, and their upstream miRNAs and transcription factors (TFs) were also explored.

MATERIALS AND METHODS

Data Collection and Pre-processing

The mRNA expression profiling dataset GSE17536 of human CRC with patient clinical information was downloaded from Gene Expression Omnibus (GEO) online database (Smith et al., 2010). GSE17536 dataset was performed on platform GPL570 Affymetrix Human Genome U133 Plus 2.0 Array

(HG-U133_Plus_2). GSE17536 was used to construct a co-expression network then distinguishing hub genes, which included 177 CRC samples. R packages were used to annotate the original data, generate an expression matrix, and match probes to target gene symbols. Median absolute deviations (MADs) were arranged from large to small, and the expression of the top 25% genes with the greatest differences in samples was selected for WGCNA.

Construction of WGCNA

R package “WGCNA” was conducted (Langfelder and Horvath, 2008). Firstly, a similarity matrix was constructed by calculating the correlation of all gene pairs. Secondly, using the `pickSoftThreshold` function in R language, a suitable soft thresholding power was determined, and the parameters that provide appropriate soft threshold power (β) for network construction were acquired (Botía et al., 2017). After choosing the appropriate $\beta = 4$, subsequently, the adjacency of the gene network was transformed into a topological overlap matrix (TOM) followed by a calculation of corresponding dissimilarity (1-TOM). Subsequently, a hierarchical clustering method is used to classify genes with similar expression profiles into the same modules, and, by default, the minimum number of genes for the genes dendrogram is set to 30. To further analyze the module, the dissimilarity of module eigengenes (ME) was calculated the `dynamicTreeCut` algorithm of WGCNA, and some highly similar modules with the height of ME in the clustering lower than 0.25 were merged. Finally, the characteristic gene network was visualized.

Identification of Modules Associated With Clinical Features

In this study, two methods were used to identify key modules associated with CRC recurrence. ME are the first principal component of a given module and could be considered as a representative of gene expression profile in a module. Module membership (MM) represents the correlation between genes and ME. Gene significance (GS) refers to the correlation between the genes and clinical data, and the average GS for all the genes in a module was defined as module significance (MS), and the module with the highest absolute MS value was regarded as the module with the most significant association with clinical information. Finally, the dissimilarity of the ME was calculated using the `moduleEigengenes` function in the R WGCNA package. GS was calculated by linear regression between gene expression and CRC recurrence, and MS related to clinical CRC recurrence was obtained.

Functional Enrichment Analysis of the Key Module Genes

Gene Ontology enrichment analysis has become a widely used method in functional gene annotation. To further investigate the function of differentially expressed genes (DEGs) in the key module, `clusterProfiler` R package (Yu et al., 2012)¹ with

¹<http://bioconductor.org/packages/release/bioc/html/clusterProfiler.html>

a strict cut-off of $FDR < 0.05$ was used to present GO and Kyoto Encyclopedia of Genes and Genomes (KEGG) pathway analysis. Then, a “ggplot2” R package was used to perform the first 10 enrichment terms of GO analysis and KEGG pathway analysis. The cutoff criterion of $P < 0.05$ was considered to be statistically significant.

Protein–Protein Interaction of the Key Module Genes

Search Tool for the Retrieval of Interacting Genes (Szkarczyk et al., 2019) (STRING, Version: 11.0²) was employed to identify protein–protein interactions (PPIs) with a medium confidence interaction score of 0.4 (the turquoise module). In addition, Cytoscape software (Shannon et al., 2003) was applied to visualize the PPI networks³. The Molecular Complex Detection (MCODE)

plug-in of Cytoscape tool was used to visualize the significant gene modules in CRC, as default, with degree cut-off = 2, node score cut-off = 0.2, k-core = 2, and max.depth = 100. Furthermore, the criteria for selecting the top four significant modules were set as follows: MCODE scores ≥ 8 and number of nodes ≥ 10 . In addition, STRING was used to present the co-expression analysis of hub genes.

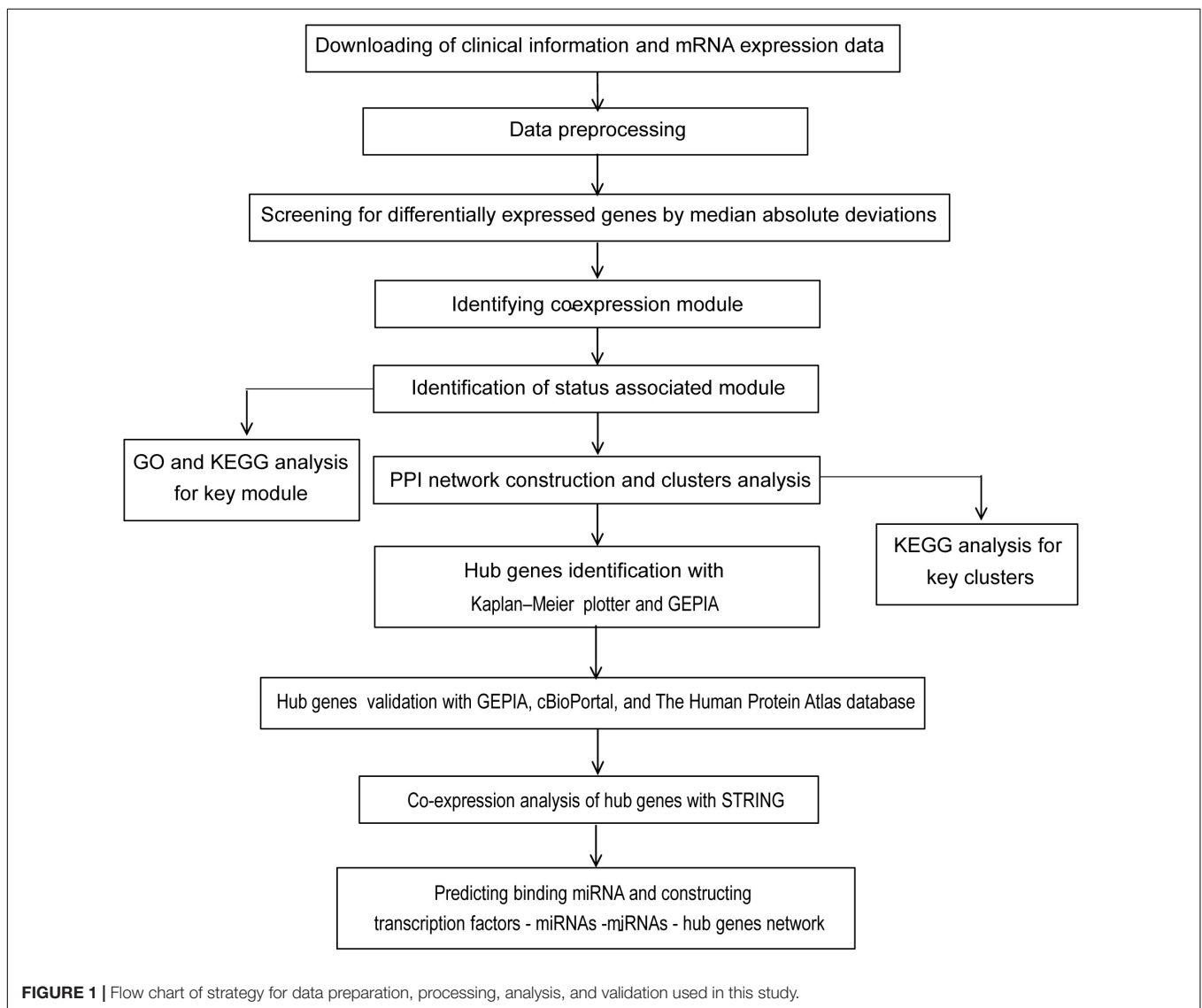
Identification and Validation of Hub Genes

Based on the MCODE analysis, the genes of the top clusters were selected as candidate genes for further analysis. The GEPIA (Tang et al., 2017)⁴ is a webserver for analyzing gene expression profiles of 9736 tumors and 8587 normal samples from the Cancer Genome Atlas (TCGA) and the genotype-tissue expression (GTEx) projects. In this study, the

²<http://string-db.org/>

³<http://cytoscape.org/>

⁴<http://gepia.cancer-pku.cn/>



Kaplan–Meier plotter was used to plot survival analyses of the top module genes, then the GEPIA webserver was used to confirm outcomes of survival analyses. For each gene, cancer patients were divided into two groups according to the median values of mRNA expression. Moreover, TCGA data of CRC were used to validate the mRNA expression of identified hub genes comparing with normal colon tissues. Therefore, we can verify the transcriptional levels of hub genes in CRC tissues. $P < 0.01$ was considered to be statistically significant.

The Human Protein Atlas⁵ was applied to validate the hub genes by immunohistochemistry (IHC). The cBio Cancer

Genomics Portal⁶(version: 2.2.0) is an open-access tool that provides analysis, visualization, and downloading of cancer genomics data sets for multiple tumor types. By using the cBioPortal tool, complex cancer genome profiles can be accessed, and this enables us to compare the genetic variations of the selected hub genes in CRC.

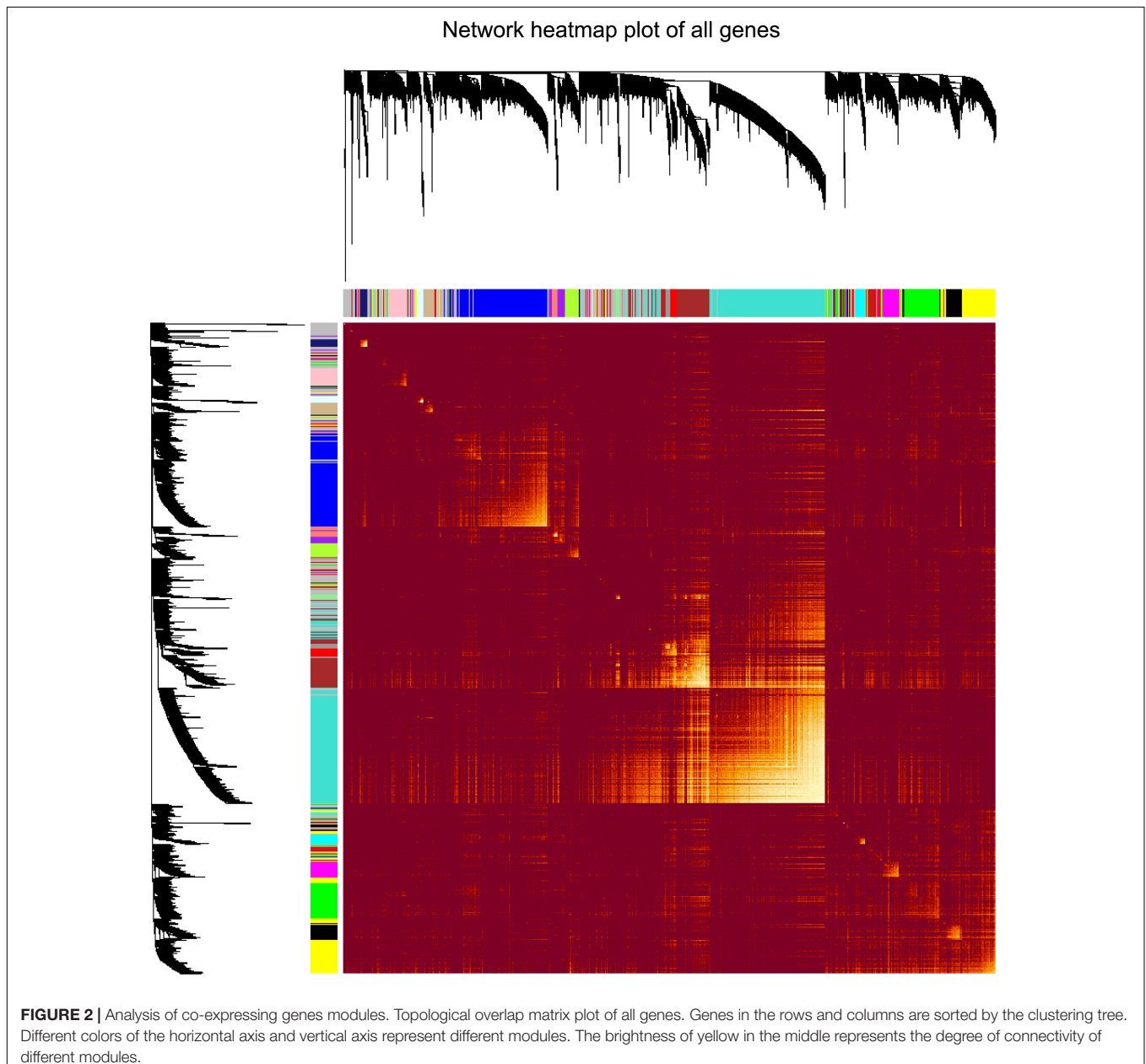
Construction of TFs–miRNAs–Hub Genes Network

StarBase (Li et al., 2014)⁷, a database for exploring microRNA–mRNA interaction maps, was used to predict miRNAs that

⁵<http://www.proteinatlas.org>

⁶<http://www.cbioportal.org/>

⁷<http://starbase.sysu.edu.cn/index.php>



bind to hub genes based on the screening criteria that CLIP Data ≥ 1 and expression was present in at least one tumor sample. Then, miRNAs with most intersections in seven databases (PITA, RNA22, miRmap, microT, miRanda, PicTar, and TargetScan) were selected. A co-expression network on account of correlation analysis of hub genes and cancer-related miRNAs was constructed by Cytoscape software. Then, the plugin iRegulon of Cytoscape is applied to forecast TF regulation networks.

RESULTS

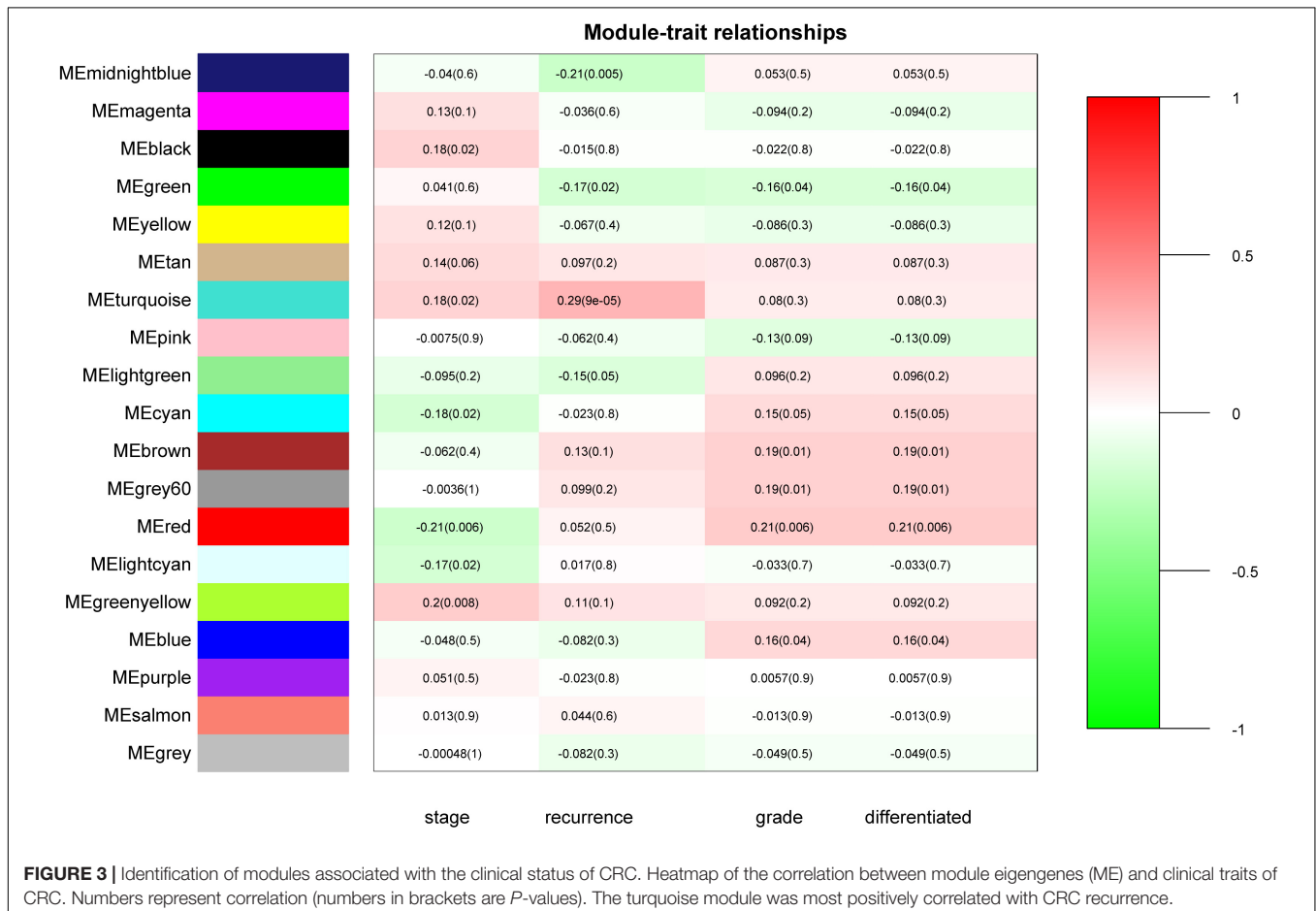
Data Pre-processing and Quality Assessment

The flow diagram of the study design is depicted in **Figure 1**. In general, a total of 177 samples in the GSE17536 dataset were downloaded from the GEO database. After the primary quality control by the WGCNA R package, one outlier sample (**Supplementary Figure 1**) was removed and a total of 176 qualified CRC samples with clinical data (**Supplementary Table 1**) were included (**Supplementary Figure 2A**, the upper panel). Clinical characteristics of tumor pathological stage, histological grade, recurrence, and differentiated

status of CRC patients were denoted (**Supplementary Figure 2A**, the lower panel). After screened by MADs arranged from large to small, the expression of the top 25% genes (5044 genes) with the greatest differences in samples were analyzed by WGCNA.

WGCNA Construction and Identification of Key Modules for CRC Recurrence

“WGCNA” package in R was applied to classify the DEGs with similar expression patterns into modules by average linkage clustering. As a result, a total of 19 modules were identified by merging similar modules when the MedissThres was set at 0.25 (**Supplementary Figure 3**). The network heatmap is presented in **Figure 2**. The relevance between the key module and CRC recurrence was tested using two methods. Our results indicated that the ME of the turquoise module possessed the highest correlation with tumor recurrence [$(P = 9 \times 10^{-5}$, $R^2 = 0.29$), **Figure 3**]. Furthermore, we also showed that the MS of the turquoise module was the highest among all modules (**Supplementary Figure 4**), which was considered to be more associated with tumor recurrence. Therefore, we identified the turquoise module to be a clinically significant module of interest in connection with CRC recurrence in the training set. The correlations between module members and



GS in the turquoise module is demonstrated by scatter plots in **Supplementary Figure 5**.

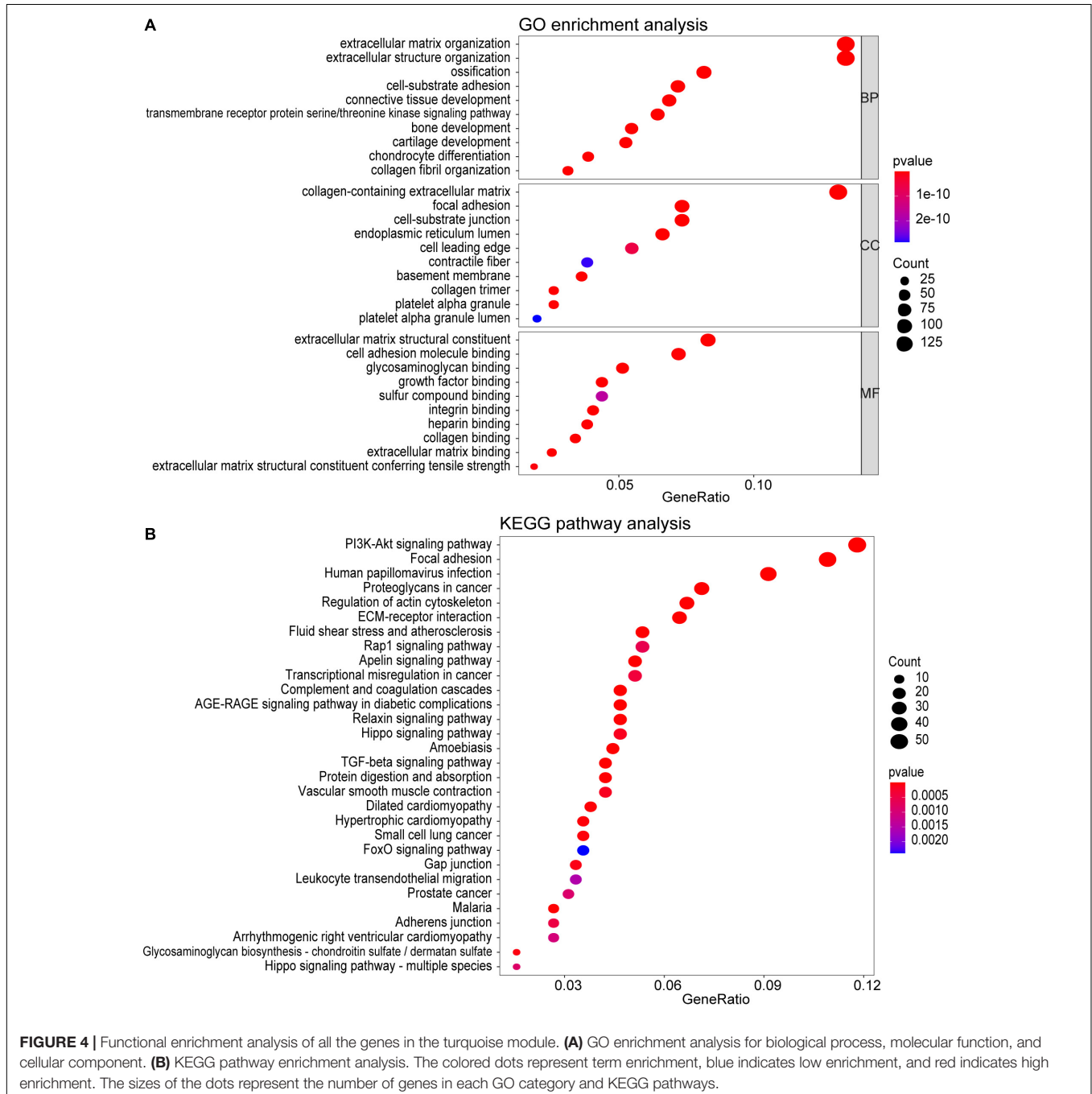
GO and KEGG Analyses of the Turquoise Module

To investigate the potential function of the genes in the turquoise module, GO and KEGG pathway analyses were performed. As shown in **Figure 4A**, genes in the turquoise module were predicted to exert their functions in the fields of extracellular structure organization, collagen-containing extracellular matrix,

and extracellular matrix structural constituent, etc. In addition, these genes were found to be significantly associated with cellular pathways including the PI3K-Akt signaling pathway and regulation of actin cytoskeleton (**Figure 4B**).

PPI Network Construction and Key Clusters Analysis of the Turquoise Module

Using the STRING database and Cytoscape software, a total of 926 DEGs of the turquoise module were mapped into the



PPI network, including 926 nodes and 7045 edges. The top four significant clusters within PPI network were selected using MCODE plug-in in Cytoscape software (Clusters 1, MCODE score = 24.435; Clusters 2, MCODE score = 20; Clusters 3,

MCODE score = 11.562, Clusters 4, MCODE score = 8.958). The functions of each module were further analyzed (Figure 5). Pathway enrichment analysis demonstrated that Module 1 consisted of 47 nodes and 562 edges (Figures 5A,B), which

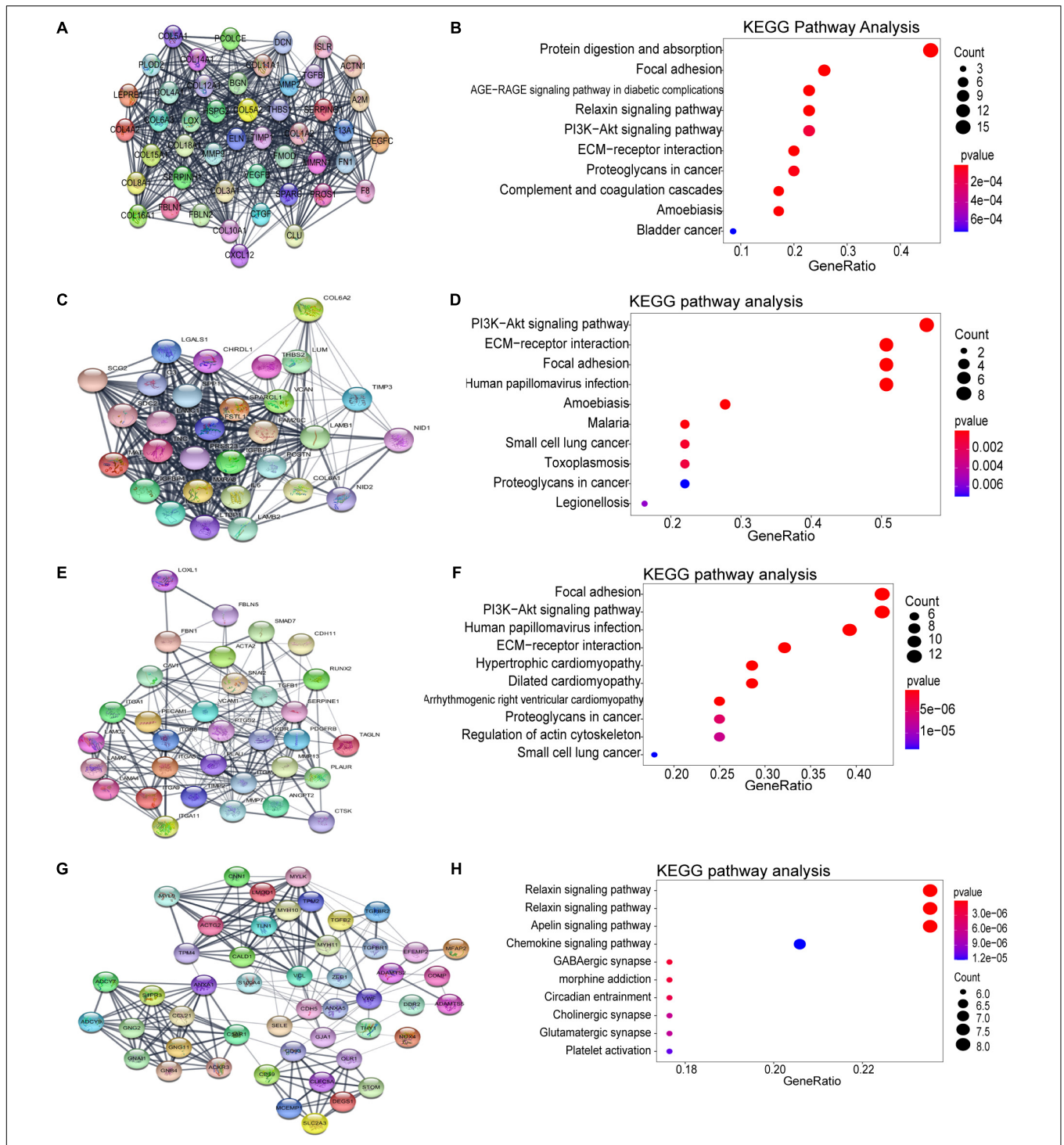
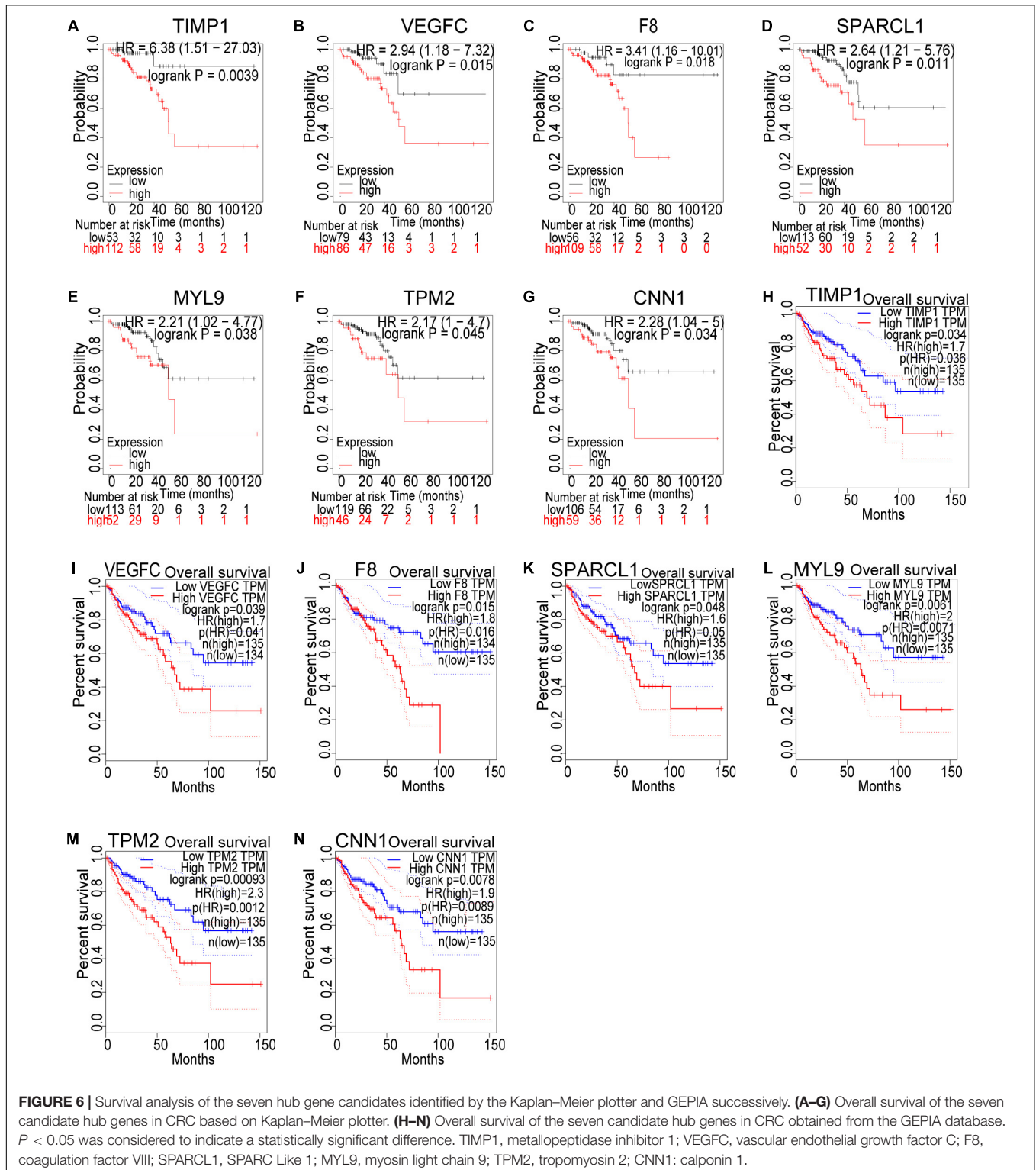


FIGURE 5 | Identification of the top four modules (159 genes in total) and correlated pathways from the PPI network. **(A,B)** Module 1 (47 genes) and its top 10 enriched KEGG pathways. **(C,D)** Module 2 (30 genes) and its top 10 enriched KEGG pathways. **(E,F)** Module 3 (33 genes) and its top 10 enriched KEGG pathways. **(G,H)** Module 4 (49 genes) and its top 10 enriched KEGG pathways.

were mainly associated with protein digestion and absorption, focal adhesion, PI3K-Akt signaling pathway, ECM-receptor interaction, proteoglycans in cancer, and bladder cancer. Module 2 consisted of 30 nodes and 290 edges (Figures 5C,D), which were mainly associated with the PI3K-Akt signaling

pathway, ECM-receptor interaction, focal adhesion, and small cell lung cancer. Module 3 consisted of 33 nodes and 185 edges (Figures 5E,F), which were associated with focal adhesion, PI3K-Akt signaling pathway, ECM-receptor interaction, dilated cardiomyopathy, proteoglycans in cancer, regulation of actin



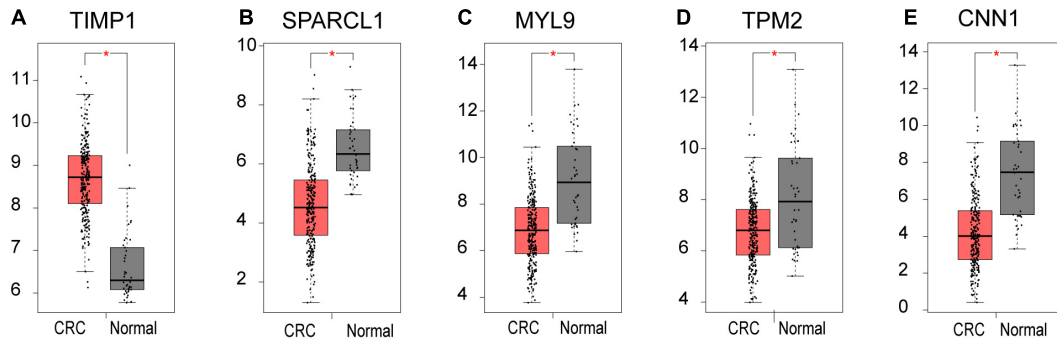


FIGURE 7 | The mRNA level of five hub genes in CRC. Validation of the mRNA expression levels of **(A)** TIMP1, **(B)** SPARCL1, **(C)** MYL9, **(D)** TPM2, and **(E)** CNN1 in CRC tissues compared with normal colon tissues from GEPIA database. These five box plots are based on 275 CRC samples (marked in red) and 41 normal samples (marked in gray). * $P < 0.01$ was considered statistically significant. CRC, colorectal cancer. VEGFC and F8 were filtered out according to the P -value of the mRNA expression levels.

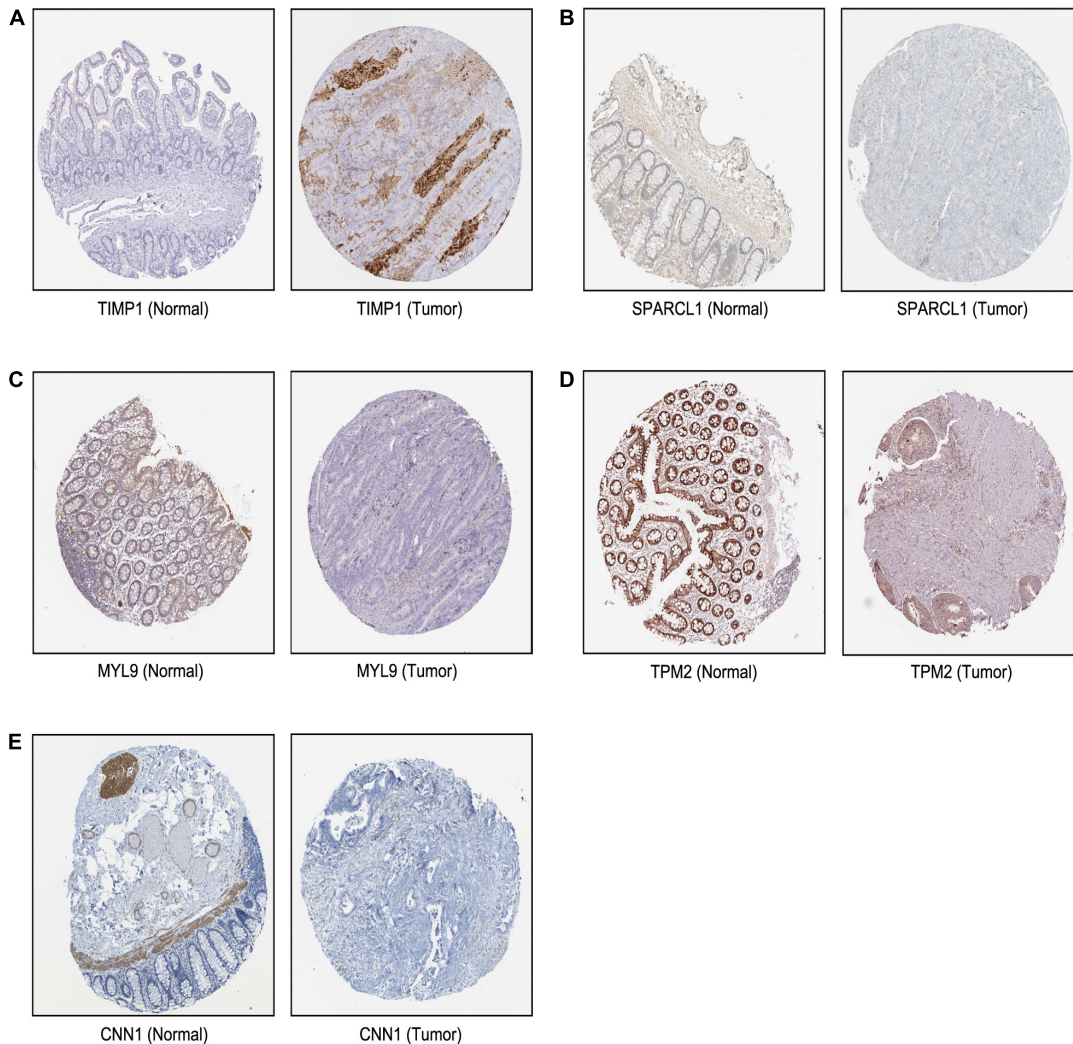


FIGURE 8 | The translational differences of five hub genes in CRC. The expression of **(A)** TIMP1, **(B)** SPARCL1, **(C)** MYL9, **(D)** TPM2, and **(E)** CNN1 on translational levels in CRC tissues compared with normal colon tissues from The Human Protein Atlas database (immunohistochemistry).

cytoskeleton, and small cell lung cancer. Module 4 consisted of 49 nodes and 215 edges (Figures 5G,H), which were associated with the relaxin signaling pathway, vascular smooth muscle contraction, apelin signaling pathway, chemokine signaling pathway, and platelet activation.

Multiple Identification and Validation of Key Genes With Poor Prognosis for CRC Recurrence

Based on our findings above, a total of 159 genes (47 in Cluster 1, 30 in Cluster 2, 33 in Cluster 3, and 49 in Cluster 4, respectively) were identified as potential hub genes. To reduce the scope and further validate these observations, CRC patient survival analysis using Kaplan–Meier plotter was first conducted to unveil the prognostic information of these genes. After the above validation, only genes with significantly deteriorated survival curves were further validated by survival analysis with GEPIA. As shown in

Figure 6, the higher expression of seven candidate hub genes was significantly correlated with poor CRC patient survival. Subsequently, in order to further verify the genes identified by the above methods, the expression level of seven specific genes was evaluated by GEPIA. As a result, elevated expression of TIMP1 was observed in CRC samples as compared to normal colon samples, while reduced expressions of SPARCL1, MYL9, TPM2, and CNN1 were found in CRC samples as compared to normal control (Figure 7). Unexpectedly, VEGFC and F8 were filtered out according to the *P*-value of the mRNA expression levels. To further confirm these results, the protein level of these five hub gene candidates was checked using IHC staining data obtained from The Human Protein Atlas database. Consistent with the GEPIA results, IHC data also showed the dysregulated protein level of these genes in CRC tissues (Figure 8). In summary, five candidate genes (TIMP1, SPARCL1, MYL9, TPM2, and CNN1) were finally characterized as the hub gene associated with tumor recurrence in CRC.

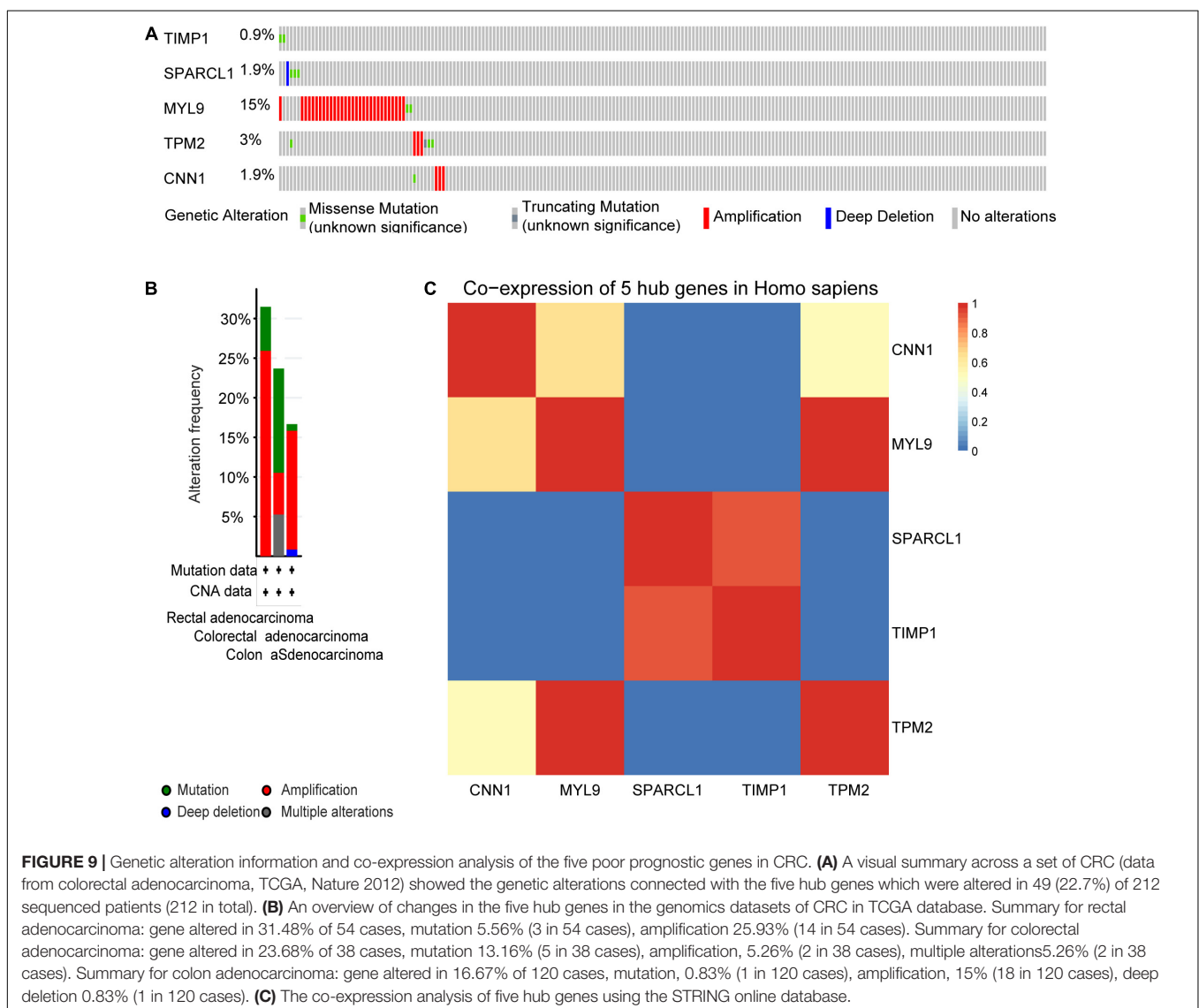


FIGURE 9 | Genetic alteration information and co-expression analysis of the five poor prognostic genes in CRC. **(A)** A visual summary across a set of CRC (data from colorectal adenocarcinoma, TCGA, Nature 2012) showed the genetic alterations connected with the five hub genes which were altered in 49 (22.7%) of 212 sequenced patients (212 in total). **(B)** An overview of changes in the five hub genes in the genomics datasets of CRC in TCGA database. Summary for rectal adenocarcinoma: gene altered in 31.48% of 54 cases, mutation 5.56% (3 in 54 cases), amplification 25.93% (14 in 54 cases). Summary for colorectal adenocarcinoma: gene altered in 23.68% of 38 cases, mutation 13.16% (5 in 38 cases), amplification, 5.26% (2 in 38 cases), multiple alterations 5.26% (2 in 38 cases). Summary for colon adenocarcinoma: gene altered in 16.67% of 120 cases, mutation, 0.83% (1 in 120 cases), amplification, 15% (18 in 120 cases), deep deletion 0.83% (1 in 120 cases). **(C)** The co-expression analysis of five hub genes using the STRING online database.

We next sought to compare the genetic alterations of the selected five hub genes in CRC from cBioPortal. As presented in **Figure 9A**, amplification or missense mutation of the MYL9 gene was observed in 15% of CRC patients, while the genetic alteration level of the other hub genes was relatively low. A combined analysis revealed that genetic alteration of five hub genes was found in over 30% of patients with rectal adenocarcinoma, more than 20% in colorectal adenocarcinoma, and over 15% in COAD (**Figure 9B**). The gene co-expression analysis of the five hub genes was then performed using STRING database. The result in **Figure 9C** showed that these genes might be actively interacted with each other, especially for MYL9 and TPM2.

Construction of TFs-miRNA-Hub Genes Network Associated With CRC Recurrence

We next sought to establish the transcriptional regulatory network of hub genes, miRNAs, and TFs by starBase. As revealed in **Figure 10**, a total of 29 TFs, 58 miRNAs, and five hub genes were involved in this network, such as models GATA6-MIR106A-CNN1, SP4-MIR424-TPM2, SP4-MIR326-MYL9, HSF1-MIR424-TPM2, ETS1-MIR22-TIMP1, and ETS1-MIR22-SPARCL1. To further understand the regulatory relationship, central regulatory biomolecules (TFs and miRNAs) were detected using topological parameters (**Table 1**). A large number of TFs

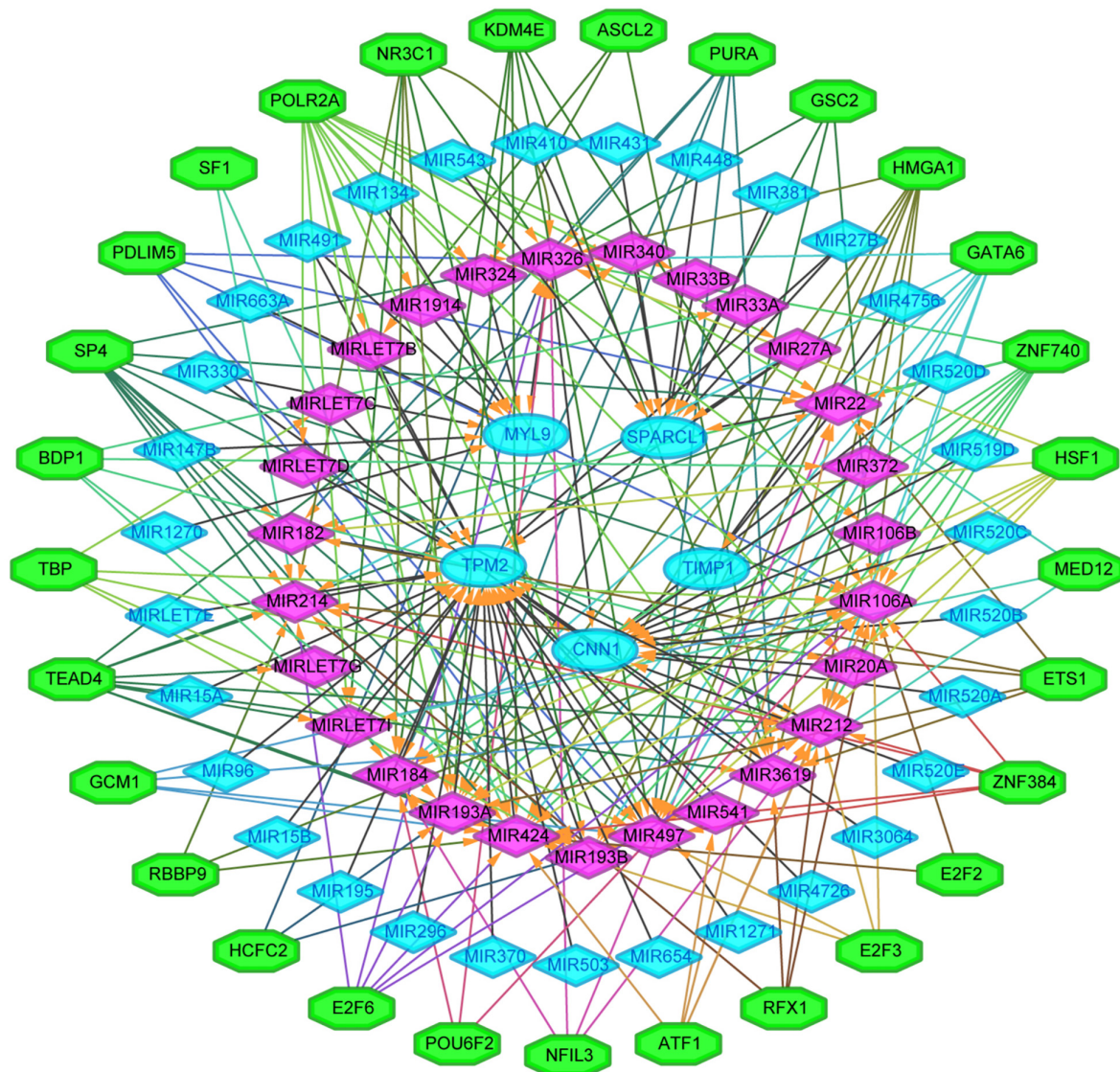


FIGURE 10 | The transcriptional regulatory network of the five hub genes, miRNAs, and TFs. miRNAs, microRNAs; TFs, transcription factors. A green hexagon node represented the TFs, a pink diamond node represented regulated miRNA, a light blue circular node represented the hub gene, their interaction was represented by an arrow. The numbers of arrows in the networks demonstrate the contribution of one TFs to miRNAs or one miRNA to the hub genes, and the higher the degree the more central the nodes were within the network.

and miRNAs that could regulate hub gene expression may reflect the complexity of the mechanisms that lead to CRC recurrence.

DISCUSSION

It is generally accepted that CRC recurrence is possibly the most important factor that influences the survival of CRC patients (Zare-Bandamiri et al., 2017). By performing the effective WGCNA approach, a total of five hub genes were identified to be significantly associated with CRC recurrence in this study. Among them, MYL9 and CNN1 were previously characterized by other groups as CRC recurrence-correlated genes (Qiu et al., 2020), while the others (TIMP1, SPARCL1, and TPM2) have never been reported before.

As a member of the tissue inhibitor of the metalloproteinase (TIMP) family, metalloproteinase inhibitor 1 (TIMP1) was originally discovered as a serum protein that plays a role in collagenase inhibition and erythroid progenitor cell growth (Eckfeld et al., 2019). It has been reported that overexpression of TIMP1 could increase the phosphorylation of the c-Kit and thus promotes proliferation and migration of CRC cells (Nordgaard et al., 2019). Since TIMP1 was expressed mainly in the serum which could be easily detected, its potential role as a CRC diagnostic biomarker has been widely studied (Vocka et al., 2019; Yang et al., 2019). In this study, we further unveiled the association between TIMP1 and CRC recurrence, which may strengthen the clinical relevance of this biomarker in the diagnosis and prognosis of CRC.

Secreted protein acidic and rich in cysteine-like 1 (SPARCL1) belongs to the SPARC-associated family of matricellular proteins and is frequently found to be decreased in a number of cancer types (Gagliardi et al., 2017). Till now, the role of SPARCL1 in CRC is still controversial. Some studies indicated SPARCL1 as a potential CRC suppressor gene that is associated with a good prognosis (Kotti et al., 2014; Han et al., 2018), while others identified SPARCL1 as a potential oncogene in CRC (Zhang et al., 2011). This discrepancy was also observed from our results, in which higher SPARCL1 expression predicts poorer CRC patient survival through Kaplan–Meier analysis, but down-regulated SPARCL1 protein level was observed in CRC tissues using GEPIA and Human Protein Atlas database.

Beta-tropomyosin (β -tropomyosin, TPM2) encodes a thin filament-associated protein which has been proved to play a crucial role in the regulation of muscle contraction (Karpicheva et al., 2016). Single-cell multiomics sequencing revealed TPM2 as one of the fibroblast-specific biomarkers representing a poorer prognosis of CRC (Zhou et al., 2020). Another research suggested that down-regulation of TPM2 was associated with RhoA activation and proliferation of CRC cells (Cui et al., 2016). In the current study, strong interaction was predicted between TPM2 and myosin light chain 9 (MYL9), which deserves to be further investigated to unveil its potential impact on the tumor recurrence of CRC.

In this study, a regulation network of TFs-miRNAs-hub genes was constructed using StarBase and Cytoscape. We first established the regulatory network of TFs-miRNAs-target genes for the recurrence of CRC, involving 29 TFs, 58 miRNAs, and five hub genes, such as models GATA6-MIR106A-CNN1, SP4-MIR424-TPM2, SP4-MIR326-MYL9, HSF1-MIR424-TPM2, ETS1-MIR22-TIMP1, and ETS1-MIR22-SPARCL1. Interestingly, TFs and miRNAs with high connective degrees in regulatory networks have been reported to be closely related to CRC, shown as in **Table 1**.

TABLE 1 | Summary of top 10 regulatory biomolecules (TFs and miRNAs) of the five hub genes in CRC identified from TFs-miRNAs-hub genes interactions.

Number	Symbol	Degree	Feature	References
TFs				
1	POLR2A	14	Afflicted with cancer	Xu et al., 2019
2	SP4	10	Afflicted with CRC	Meng et al., 2020
3	TEAD4	10	Afflicted with CRC	Kim et al., 2020
4	HMGA1	8	Afflicted with CRC	Yang et al., 2020
5	GATA6	8	Afflicted with CRC	Lai et al., 2020
6	HSF1	7	Afflicted with CRC	Song et al., 2020
7	NR3C1	7	Afflicted with CRC	Sun et al., 2020
8	ATF1	5	Afflicted with CRC	Zhao et al., 2020
9	PDLIM5	5	Afflicted with CRC	Shi et al., 2020
10	ETS1	5	Afflicted with CRC	Xu et al., 2020
11	TBP	5	Afflicted with CRC	Wei et al., 2020
miRNAs				
1	MIR106A	18	Afflicted with CRC	Liu et al., 2020
2	MIR424	17	Afflicted with CRC	Di et al., 2020
3	MIR497	16	Afflicted with CRC	Bai et al., 2020
4	MIR326	16	Afflicted with CRC	Xian et al., 2020
5	MIR212	15	Afflicted with CRC	Mou et al., 2019
6	MIR193A	14	Afflicted with CRC	Hejazi et al., 2020
7	MIR214	11	Afflicted with CRC	Liu et al., 2019
8	MIR22	11	Afflicted with CRC	Cong et al., 2020
9	MIR193B	10	Afflicted with cancer	Tian et al., 2020
10	MIR182	6	Afflicted with CRC	Lin et al., 2020

CONCLUSION

Our current study identified a number of five hub genes including TIMP1, SPARCL1, MYL9, TPM2, and CNN1, which may play vital roles during CRC recurrence. Their PPI network and upstream TF and miRNA regulators were also investigated to unveil the underlying mechanism by which these hub genes modulate the progression of CRC recurrence. It is anticipated that targeting these hub genes solely or combined therapy with currently available anti-cancer drugs may be served as an alternative method to benefit patients diagnosed with CRC. In general, this study may pave a novel way for the diagnosis, prognosis, and treatment strategies of the devastating disease.

DATA AVAILABILITY STATEMENT

The original contributions presented in the study are included in the article/**Supplementary Material**, further inquiries can be directed to the corresponding author/s.

AUTHOR CONTRIBUTIONS

QD conceived the idea. SL finished the bioinformatics analysis and drafted the manuscript. FZ and GF involved in the data analysis and interpretation. All authors contributed to the article and approved the submitted version.

FUNDING

This work was supported by the Chongqing Basic Research and Frontier Exploration Special Project (Chongqing Natural Science Foundation) (cstc2018jcyjAX0378), the Technology Innovation and Application Development Project of Chongqing Kewei United Chinese Medicine (2021ZY3591), and the Research Project of Yongchuan Hospital of Chongqing Medical University (YJJC202009).

ACKNOWLEDGMENTS

We thank the GEO, STRING, GEPIA2, cBioPortal, The Human Protein Atlas, Kaplan–Meier plotter, TCGA, and StarBase databases for providing their platforms and contributors for their valuable data sets.

REFERENCES

- Bai, J., Xu, J., Zhao, J., and Zhang, R. (2020). lncRNA SNHG1 cooperated with miR-497/miR-195-5p to modify epithelial-mesenchymal transition underlying colorectal cancer exacerbation. *J. Cell Physiol.* 235, 1453–1468. doi: 10.1002/jcp.29065
- Botía, J. A., Vandrovčova, J., Forabosco, P., Guelfi, S., D'Sa, K., Hardy, J., et al. (2017). An additional k-means clustering step improves the biological features of WGCNA gene co-expression networks. *BMC Syst. Biol.* 11:47. doi: 10.1186/s12918-017-0420-6
- Carey, V. J., Gentry, J., Whalen, E., and Gentleman, R. (2005). Network structures and algorithms in Bioconductor. *Bioinformatics* 21, 135–136. doi: 10.1093/bioinformatics/bth458
- Carter, S. L., Brechbuhler, C. M., Griffin, M., and Bond, A. T. (2004). Gene co-expression network topology provides a framework for molecular characterization of cellular state. *Bioinformatics* 20, 2242–2250. doi: 10.1093/bioinformatics/bth234
- Cong, J., Gong, J., Yang, C., Xia, Z., and Zhang, H. (2020). miR-22 suppresses tumor invasion and metastasis in colorectal cancer by targeting NLRP3. *Cancer Manag. Res.* 12, 5419–5429. doi: 10.2147/cmar.S255125
- Cui, J., Cai, Y., Hu, Y., Huang, Z., Luo, Y., Kaz, A. M., et al. (2016). Epigenetic silencing of TPM2 contributes to colorectal cancer progression upon RhoA activation. *Tumour. Biol.* 37, 12477–12483. doi: 10.1007/s13277-016-5103-1
- Di, Z., Di, M., Fu, W., Tang, Q., Liu, Y., Lei, P., et al. (2020). Integrated analysis identifies a Nine-microRNA signature biomarker for diagnosis and prognosis in colorectal cancer. *Front. Genet.* 11:192. doi: 10.3389/fgene.2020.00192
- Eckfeld, C., Haussler, D., Schoeps, B., Hermann, C. D., and Kruger, A. (2019). Functional disparities within the TIMP family in cancer: hints from molecular divergence. *Cancer Metastasis Rev.* 38, 469–481. doi: 10.1007/s10555-019-09812-6
- Gagliardi, F., Narayanan, A., and Mortini, P. (2017). SPARCL1 a novel player in cancer biology. *Crit. Rev. Oncol. Hematol.* 109, 63–68. doi: 10.1016/j.critrevonc.2016.11.013
- Ha, Y. J., Tak, K. H., Kim, S. K., Kim, C. W., Lee, J. L., Roh, S. A., et al. (2019). Biological characteristics and clinical significance of ITGB1 and RHOC in patients with recurrent colorectal cancer. *Anticancer Res.* 39, 4853–4864. doi: 10.21873/anticancer.13671

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: <https://www.frontiersin.org/articles/10.3389/fgene.2021.649752/full#supplementary-material>

Supplementary Figure 1 | Clustering of samples to detect outliers.

Supplementary Figure 2 | Sample clustering tree and soft threshold power analysis in the WGCNA. **(A)** The clustering was based on the expression data in the GSE17536 dataset ($n = 176$). The color intensity was proportional to tumor pathological stage, histological grade, recurrence, and differentiated status. In recurrence status, red means cancer recurrence, white means no recurrence, and gray means unknown. **(B)** Scale independence analysis of the scale-free fit index for various soft-thresholding powers (β). The power of $\beta = 4$ was selected. **(C)** Analysis of the mean connectivity for various soft-thresholding powers. determination of soft-thresholding power by analyzing network topology for thresholding powers from 1 to 20. **(D)** Checking the scale free topology when $\beta = 4$. K, connectivity; p(k), possibility of the connectivity.

Supplementary Figure 3 | Dendrogram of all differentially expressed genes clustered based on a dissimilarity measure (1-TOM). Each branch in the figure represented one gene, and the colors below represented co-expression modules.

Supplementary Figure 4 | Module clusters for CRC recurrence. There are total 19 Module memberships vs. gene significance clusters for CRC recurrence.

Supplementary Figure 5 | Correlation of the key turquoise module with CRC recurrence. Scatter plot of the correlation between METurquoise membership and gene significance.

- Han, W., Cao, F., Ding, W., Gao, X. J., Chen, F., Hu, Y. W., et al. (2018). Prognostic value of SPARCL1 in patients with colorectal cancer. *Oncol. Lett.* 15, 1429–1434. doi: 10.3892/ol.2017.7511
- Hejazi, M., Baghbani, E., Amini, M., Rezaei, T., Aghanejad, A., Mosafar, J., et al. (2020). MicroRNA-193a and taxol combination: a new strategy for treatment of colorectal cancer. *J. Cell Biochem.* 121, 1388–1399. doi: 10.1002/jcb.29374
- Jiang, S., Tan, B., and Zhang, X. (2019). Identification of key lncRNAs in the carcinogenesis and progression of colon adenocarcinoma by co-expression network analysis. *J. Cell Biochem.* 120, 6490–6501. doi: 10.1002/jcb.27940
- Karpicheva, O. E., Simonyan, A. O., Kuleva, N. V., Redwood, C. S., and Borovikov, Y. S. (2016). Myopathy-causing Q147P TPM2 mutation shifts tropomyosin strands further towards the open position and increases the proportion of strong-binding cross-bridges during the ATPase cycle. *Biochim. Biophys. Acta* 1864, 260–267. doi: 10.1016/j.bbapap.2015.12.004
- Kim, J. Y., Kim, E. K., Lee, W. M., Hong, Y. O., and Lee, H. (2020). VGLL4 with low YAP expression is associated with favorable prognosis in colorectal cancer. *Apmis* 128, 543–551. doi: 10.1111/apm.13070
- Kotti, A., Holmqvist, A., Albertsson, M., and Sun, X. F. (2014). SPARCL1 expression increases with preoperative radiation therapy and predicts better survival in rectal cancer patients. *Int. J. Radiat. Oncol. Biol. Phys.* 88, 1196–1202. doi: 10.1016/j.ijrobp.2013.12.041
- Lai, H. T., Chiang, C. T., Tseng, W. K., Chao, T. C., and Su, Y. (2020). GATA6 enhances the stemness of human colon cancer cells by creating a metabolic symbiosis through upregulating LRH-1 expression. *Mol. Oncol.* 14, 1327–1347. doi: 10.1002/1878-0261.12647
- Langfelder, P., and Horvath, S. (2008). WGCNA: an R package for weighted correlation network analysis. *BMC Bioinformatics* 9:559. doi: 10.1186/1471-2105-9-559
- Li, J. H., Liu, S., Zhou, H., Qu, L. H., and Yang, J. H. (2014). starBase v2.0: decoding miRNA-ceRNA, miRNA-ncRNA and protein-RNA interaction networks from large-scale CLIP-Seq data. *Nucleic Acids Res.* 42, D92–D97. doi: 10.1093/nar/gkt1248
- Liang, Q., Guan, G., Li, X., Wei, C., Wu, J., Cheng, P., et al. (2020). Profiling pro-neural to mesenchymal transition identifies a lncRNA signature in glioma. *J. Transl. Med.* 18:378. doi: 10.1186/s12967-020-02552-0
- Lin, M., Li, Y., Xian, J., Chen, J., Feng, Y., Mao, C., et al. (2020). Long non-coding RNA AGER-1 inhibits colorectal cancer progression through

- sponging miR-182. *Int. J. Biol. Markers* 35, 10–18. doi: 10.1177/1724600819897079
- Liu, B., Liu, Q., Pan, S., Huang, Y., Qi, Y., Li, S., et al. (2019). The HOTAIR/miR-214/ST6GAL1 crosstalk modulates colorectal cancer progression through mediating sialylated c-Met via JAK2/STAT3 cascade. *J. Exp. Clin. Cancer Res.* 38:455. doi: 10.1186/s13046-019-1468-5
- Liu, M., Liberzon, A., Kong, S. W., Lai, W. R., Park, P. J., Kohane, I. S., et al. (2007). Network-based analysis of affected biological processes in type 2 diabetes models. *PLoS Genet.* 3:e96. doi: 10.1371/journal.pgen.0030096
- Liu, S., Zheng, R., Zhang, M., Zhang, S., Sun, X., and Chen, W. (2015). Incidence and mortality of colorectal cancer in China, 2011. *Chin. J. Cancer Res.* 27, 22–28. doi: 10.3978/j.issn.1000-9604.2015.02.01
- Liu, Z., Qin, Y., Dong, S., Chen, X., Huo, Z., and Zhen, Z. (2020). Overexpression of miR-106a enhances oxaliplatin sensitivity of colorectal cancer through regulation of FOXQ1. *Oncol. Lett.* 19, 663–670. doi: 10.3892/ol.2019.11151
- Meng, N., Chen, M., Chen, D., Chen, X. H., Wang, J. Z., Zhu, S., et al. (2020). Small protein hidden in lncRNA LOC90024 Promotes “Cancerous” RNA splicing and tumorigenesis. *Adv. Sci.* 7:1903233. doi: 10.1002/adv.201903233
- Mirgayazova, R., Khadiullina, R., Mingaleeva, R., Chasov, V., Gomzikova, M., Garanina, E., et al. (2019). Novel Isatin-based activator of p53 transcriptional functions in tumor cells. *Mol. Biol. Res. Commun.* 8, 119–128. doi: 10.22099/mbr.2019.34179.1419
- Mou, T. Y., Zhang, R. R., and Wang, Y. N. (2019). MiRNA-212 acts as a tumor-suppressor in colorectal carcinoma through targeting SOX4. *Eur. Rev. Med. Pharmacol. Sci.* 23, 10751–10760. doi: 10.26355/eurrev_201912_19777
- Nordgaard, C., Doll, S., Matos, A., Hoeberg, M., Kazi, J. U., Friis, S., et al. (2019). Metalloproteinase inhibitor 1 (TIMP-1) promotes receptor tyrosine kinase c-Kit signaling in colorectal cancer. *Mol. Oncol.* 13, 2646–2662. doi: 10.1002/1878-0261.12575
- Qiu, X., Cheng, S. H., Xu, F., Yin, J. W., Wang, L. Y., and Zhang, X. Y. (2020). Weighted gene co-expression network analysis identified MYL9 and CNN1 are associated with recurrence in colorectal cancer. *J. Cancer* 11, 2348–2359. doi: 10.7150/jca.39723
- 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
- Shi, Y., Wang, X., Xu, Z., He, Y., Guo, C., He, L., et al. (2020). PDLIM5 inhibits STUB1-mediated degradation of SMAD3 and promotes the migration and invasion of lung cancer cells. *J. Biol. Chem.* 295, 13798–13811. doi: 10.1074/jbc.RA120.014976
- Siegel, R. L., Miller, K. D., Goding Sauer, A., Fedewa, S. A., Butterly, L. F., Anderson, J. C., et al. (2020). Colorectal cancer statistics, 2020. *CA Cancer J. Clin.* 70, 145–164. doi: 10.3322/caac.21601
- Smith, J. J., Deane, N. G., Wu, F., Merchant, N. B., Zhang, B., Jiang, A., et al. (2010). Experimentally derived metastasis gene expression profile predicts recurrence and death in patients with colon cancer. *Gastroenterology* 138, 958–968. doi: 10.1053/j.gastro.2009.11.005
- Song, P., Feng, L., Li, J., Dai, D., Zhu, L., Wang, C., et al. (2020). β -catenin represses miR455-3p to stimulate m6A modification of HSF1 mRNA and promote its translation in colorectal cancer. *Mol. Cancer* 19:129. doi: 10.1186/s12943-020-01244-z
- Sun, Y. L., Zhang, Y., Guo, Y. C., Yang, Z. H., and Xu, Y. C. (2020). A prognostic model based on six metabolism-related genes in colorectal cancer. *Biomed. Res. Int.* 2020:5974350. doi: 10.1155/2020/5974350
- 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
- Tang, Z., Li, C., Kang, B., Gao, G., 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
- Tian, R., Jiang, H., Shao, L., Yu, Y., Guo, Q., Cao, B., et al. (2020). mir193b promotes apoptosis of gastric cancer cells via directly mediating the akt pathway. *Biomed. Res. Int.* 2020:2863236. doi: 10.1155/2020/2863236
- Vocka, M., Langer, D., Fryba, V., Petrtyl, J., Hanus, T., Kalousova, M., et al. (2019). Serum levels of TIMP-1 and MMP-7 as potential biomarkers in patients with metastatic colorectal cancer. *Int. J. Biol. Markers* 34, 292–301. doi: 10.1177/1724600819866202
- Wang, Y., Cui, K., Zhu, M., and Gu, Y. (2020). Coexpression module construction by weighted gene coexpression network analysis and identify potential prognostic markers of breast cancer. *Cancer Biother. Radiopharm.* doi: 10.1089/cbr.2020.3821 [Epub ahead of print].
- Wei, Y., Gu, X., Sun, Y., Meng, F., Storm, G., and Zhong, Z. (2020). Transferrin-binding peptide functionalized polymersomes mediate targeted doxorubicin delivery to colorectal cancer in vivo. *J. Control Release* 319, 407–415. doi: 10.1016/j.jconrel.2020.01.012
- Wu, Y., Zhang, S., and Yan, J. (2020). IRF1 association with tumor immune microenvironment and use as a diagnostic biomarker for colorectal cancer recurrence. *Oncol. Lett.* 19, 1759–1770. doi: 10.3892/ol.2020.11289
- Xian, Z. Y., Hu, B., Wang, T., Cai, J. L., Zeng, J. Y., Zou, Q., et al. (2020). CircABC10 silencing inhibits the cell ferroptosis and apoptosis by regulating the miR-326/CCL5 axis in rectal cancer. *Neoplasma* 67, 1063–1073. doi: 10.4149/neo_2020_191024N1084
- Xu, J., Liu, Y., Li, Y., Wang, H., Stewart, S., Van der Jeught, K., et al. (2019). Precise targeting of POLR2A as a therapeutic strategy for human triple negative breast cancer. *Nanotechnol.* 14, 388–397. doi: 10.1038/s41565-019-0381-6
- Xu, J., Tang, Y., Sheng, X., Tian, Y., Deng, M., Du, S., et al. (2020). Secreted stromal protein ISLR promotes intestinal regeneration by suppressing epithelial Hippo signaling. *Embo J.* 39:e103255. doi: 10.15252/embj.2019103255
- Yang, L., Jiang, Q., Li, D. Z., Zhou, X., Yu, D. S., and Zhong, J. (2019). TIMP1 mRNA in tumor-educated platelets is diagnostic biomarker for colorectal cancer. *Aging* 11, 8998–9012. doi: 10.18632/aging.102366
- Yang, M., Guo, Y., Liu, X., and Liu, N. (2020). HMG1A1 promotes hepatic metastasis of colorectal cancer by inducing expression of glucose transporter 3 (GLUT3). *Med. Sci. Monit.* 26:e924975. doi: 10.12659/msm.924975
- Yu, G., Wang, L. G., Han, Y., and He, Q. Y. (2012). clusterProfiler: an R package for comparing biological themes among gene clusters. *OMICS* 16, 284–287. doi: 10.1089/omi.2011.0118
- Yu, T., Guo, F., Yu, Y., Sun, T., Ma, D., Han, J., et al. (2017). *Fusobacterium nucleatum* promotes chemoresistance to colorectal cancer by modulating autophagy. *Cell* 170, 548.e16–563.e16. doi: 10.1016/j.cell.2017.07.008
- Zare-Bandamiri, M., Fararouei, M., Zohourinia, S., Daneshi, N., and Dianatinasab, M. (2017). Risk factors predicting colorectal cancer recurrence following initial treatment: a 5-year cohort study. *Asian Pac. J. Cancer Prev.* 18, 2465–2470. doi: 10.22034/APJCP.2017.18.9.2465
- Zhang, H., Widegren, E., Wang, D. W., and Sun, X. F. (2011). SPARCL1: a potential molecule associated with tumor diagnosis, progression and prognosis of colorectal cancer. *Tumour. Biol.* 32, 1225–1231. doi: 10.1007/s13277-011-0226-x
- Zhang, M., Wang, H. Z., Li, H. O., Zhou, Y. J., Peng, R. Y., Liu, J., et al. (2020). Identification of PIGU as the hub gene associated with KRAS mutation in colorectal cancer by coexpression analysis. *DNA Cell Biol.* 39, 1639–1648. doi: 10.1089/dna.2020.5574
- Zhang, Y., Sun, L., Wang, X., Sun, Y., Chen, Y., Xu, M., et al. (2020). FBXW4 acts as a protector of FOLFOX-Based chemotherapy in metastatic colorectal cancer identified by co-expression network analysis. *Front. Genet.* 11:113. doi: 10.3389/fgene.2020.00113
- Zhao, X., Weng, W., Long, Y., Pan, W., Li, Z., and Sun, F. (2020). LINC00665/miR-9-5p/ATF1 is a novel axis involved in the progression of colorectal cancer. *Hum. Cell* 33, 1142–1154. doi: 10.1007/s13577-020-00393-z
- Zhou, Y., Bian, S., Zhou, X., Cui, Y., Wang, W., Wen, L., et al. (2020). Single-cell multiomics sequencing reveals prevalent genomic alterations in tumor stromal cells of human colorectal cancer. *Cancer Cell* 38, 818.e5–828.e5. doi: 10.1016/j.ccell.2020.09.015

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 Liu, Zeng, Fan and Dong. 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.