Iran J Public Health, Vol. 51, No.2, Feb 2022, pp.364-374

Original Article

Identification of Key Carcinogenic Genes in Colon Adenocarcinoma

Yangbin Dai, Zhenjian Jiang, Yanru Qiu, Yiping Kang, Huangzhen Xu, *Tianwen Xu

Department of Medical Oncology, The Second Affiliated Hospital of Fujian Medical University, Quanzhou 362000, China

***Corresponding Author:** Email: xutianwen53@163.com

(Received 18 Feb 2021; accepted 15 Apr 2021)

Abstract

Background: We aimed to probe carcinogenic genes associated with colon adenocarcinoma (COAD) development.

Methods: The gene expression profile of COAD were downloaded from TCGA. Differentially expressed genes (DEGs) were identified; GO and KEGG pathway enrichment were analyzed. Applying the up-mRNA-and-down-miRNA pairs and the down-mRNA-and-up-miRNA pairs, the miRNA target network was generated. The important genes were further analyzed towards the influence on overall survival and immune infiltration. In addition, essential miRNAs were selected for expression validation using real-time qPCR.

Results: Together, from 2020-2021, in Central Laboratory of the Second Affiliated Hospital of Fujian Medical University, we found 3060 up-regulated transcripts and 2254 down-regulated transcripts in mRNA expression, with 235 up-regulated and 263 down-regulated miRNAs. We discovered 98 enriched GO terms using the up-regulated DEGs and 315 enriched GO terms using downregulated DEGs. There were 14 enriched KEGG pathways based on the down-regulated DEGs and only one pathway based on the up-regulated DEGs. There were 61 up-mRNA-and-down-miRNA pairs, including 7 miRNAs and 41 carcinogenic targets, among which *HOXC13*, *FOXL2NB*, *ALOXE3*, and *ZIC2* were found related to a poorer OS. *ZIC2* located at the subnet with the most targets (the miR-129-5p subnet). *ZIC2* expression was correlated with immune-cell infiltration.

Conclusion: These risk genes, interaction networks, and enrichments may provide a better understanding of the complex molecular mechanisms in COAD development and potential therapeutic targets for clinical treatment of COAD.

Keywords: Colon cancer; Database; Carcinogenic gene; Bioinformatics analysis

Introduction

Colorectal cancer (CRC), including colon adenocarcinoma (COAD) and rectal adenocarcinoma, is one of the most prevalent malignancies worldwide. CRC is especially prevalent among middle-aged and elderly people. Generally, COAD originates from epithelial dysplasia in the colonic mucosa followed by malignant infiltration and growth (1).

Currently, there are extremely few effective sensitive biomarkers for early diagnosis, treatment, and prevention of metastasis for COAD, especially the advanced colon cancer (2). Although

surgery-based com-prehensive treatment is being accumulated, research into specific biomarkers and therapeutic pathways is still of great value for improving patient prognosis (3). In recent years, with the development of gene-sequencing platforms and various network databases (4-6), such as The Cancer Genome Atlas (TCGA) and the Gene Expression Omnibus (GEO), carcinogenic genes (including not only mRNAs but also microRNAs and lncRNAs) that influence the prognosis for survival have been discovered through bioinformatics analysis. An examination of differentially expressed genes (DEGs) between tumor and para-carcinoma tissue may help identify critical biomarkers for COAD.

Currently, a single gene or molecular marker does not provide a good diagnosis or predict the progression and increasing scholars have adopted multiple genes to build predictive models for disease diagnosis. However, due to the insufficient molecular studies, the exact mechanism of COAD occurrence and development is not fully clear, and there have been limited specific markers in the prognosis prediction, miRNA-mRNA interaction network, immune infiltration features, and so on.

We aimed to probe more carcinogenic genes and pathways associated with COAD onset and malignancy progression. In this study, the COAD mRNA and microRNA profiles were downloaded from the TCGA database and multiple bioinformatics methods were conducted. Key mRNAs and microRNAs, as well as mRNA-microRNA interaction pairing, were identified, followed by enrichment analysis, protein-protein interaction (PPI) network establishment, overall survival (OS) analysis. Moreover, some particularly important factors were selected for immune infiltration analysis and expression verification.

Materials and Methods

The mRNA and miRNA data in TCGA COAD datasets

First, we downloaded the COAD mRNA data from the TCGA database, in which there are 496 clinical samples, including normal samples in the adjacent tissues, 78 stage-I samples, 182 stage-II samples, 131 stage-III samples, and 65 stage-IV samples. Next, we downloaded the relevant miRNA data of the TCGA COAD datasets, and 446 samples with the miRNA expression were acquired in comparison between the normal and cancer tissues, including 8 normal tissue samples and 438 cancer tissue samples. Moreover, the 438 samples can be divided into 74 stage-I samples, 172 stage-II samples, 127 stage-III samples, and 65 stage-IV samples.

Differentially expressed gene identification

Differentially expressed gene (DEG) analysis was performed according to the following standards. A DEG was regarded as a gene with a fold-change (FC) >=2 or <= 0.5 and adjusted *P* (*P*-adj) < 0.05. The volcanic plots were produced to present DEGs (DE mRNAs and DE miRNAs) using the ggplot2 R package.

GO functional and KEGG pathway enrichments

We used the DAVID tool (https://david.ncifcrf.gov/) to calculated Gene Ontology (GO) functional enrichment and Kyoto Encyclopedia of Genes and Genomes (KEGG) pathway enrichment. The up-regulated DEGs and the down-regulated DEGs were conducted with GO and KEGG pathway enrichment analysis, respectively. Homo sapiens was the background for enrichment analysis of all DEGs. Any term with a *P*-value <0.05 and gene number >= 2 was selected as an enriched term. The *P* values of enrichment acquired from the DAVID tool were used to draw bar graphs of the GO enrichment, and the bubble graphs were generated to present the *P* values and the counts of included DEGs.

The miRNA target network and protein-protein interaction network

The targets of DE-miRNAs were predicted mainly based on the miRDB and TargetScan databases. The Starbase (http://starbase.sysu.edu.cn) was used for con-

firmation. Next, two types of mRNA-miRNA pairs were divided: the up-mRNA-and-down-miRNA pairs and the down-mRNA-and-up-miRNA pairs. The miRNA target network was generated by the Cytoscape software. For the key genes screened out (*ZIC2* in our result), the associated protein-protein interaction (PPI) network was drawn using the Search Tool for the Retrieval of Interacting Genes (STRING) database (http://www.string-db.org). The online tool of multiple proteins (organism: Homo sapiens) was used for batch analysis, and the active interaction source included experiments, databases, co-expression, neighborhood, gene fusion, and co-occurrence. The criterion of node connection was the combined >= 0.4. The PPI data was acquired from the STRING database and the network was drawn using the Cytoscape software.

Survival analysis

For all the DE-mRNAs included in the miRNA target network, the overall survival (OS) regarding the expression of DEGs was analyzed. For each gene, patients were divided into two groups: the high-expression group (> median) and the low-expression group (<= median). The OS time was compared between two groups, and the Kaplan-Meier survival curve was drawn for each significantly risk DEG.

Immune-infiltration analysis

The association between the expression of the carcinogenic gene (*ZIC2* as finally screened for analysis) and immune cell infiltration was analyzed by the Tumor immune to assess resource (TIMER, www.cistrome.shinyapps.io). This database includes gene expression profiles from the TCGA database, and its online tool can estimate the immune cell infiltration and evaluate its clinical impact for different cancer types. The links between *ZIC2* transcription level and immune cell infiltration (including B cells, neutrophils, CD4 + T cells, macrophages, CD8 + T cells, and dendritic cells, as well as the tumor purity) were analyzed using Pearson's correlation method. Moreover, the relationship between the copy number of *ZIC2* and the infiltration level of the associated immune cells were explored using the on tool of TIMER.

Expression validation of key miRNA by real-time qPCR

The expression of key miRNA (miR-129-5p) was assessed by our clinical samples. We collected three normal samples and five-colon cancer samples, and the total RNA was isolated utilizing the TRIzol reagent (Qiagen, Valencia, CA, U.S.A.). CDNA was synthesized with RNA reverse transcription kit (TIANGEN BIOTECH., China). The real-time PCR program was performed with an ABI 7300 Real-Time PCR System (Applied Biosystems Life Technologies, USA). The expression of miR-129-5p was referenced to U6, and all samples were normalized to the average of three normal samples.

Results

Differentially expressed mRNAs and miRNAs

Utilizing the TCGA data, we observed 3060 up-regulated transcripts and 2254 down-regulated transcripts in mRNA expression (Fig. 1A). Meanwhile, there were 235 up-regulated and 263 down-regulated miRNAs (Fig. 1B).

GO and KEGG enrichment

Based on the above up-regulated and down-regulated DEGs, GO functional and KEGG pathway enrichments were performed. When using the upregulated DEGs, we discovered 98 significant GO functional terms (the top terms were shown in Fig. 2A), including keratinization, cornification, keratin filament, intermediate filament, intermediate filament cytoskeleton, pre-mRNA 5'-splice site binding, etc. Based on the downregulated DEGs, there were 315 enriched GO terms (the top terms were shown in Fig. 2B), such as regulation of immunoglobulin complex, passive transmembrane transporter activity, ion channel activity, channel activity, and transmembrane transporter complex. Additionally, the only

enriched KEGG pathway based on up-regulated DEGs was Staphylococcus aureus infection (Fig. 3A), and the 14 enriched pathways based on the down-regulated DEGs included Neuroactive ligand-receptor interaction, Nicotine addiction, Bile secretion, Ascorbate and aldarate metabolism, Taste transduction, cAMP signaling pathway, etc.

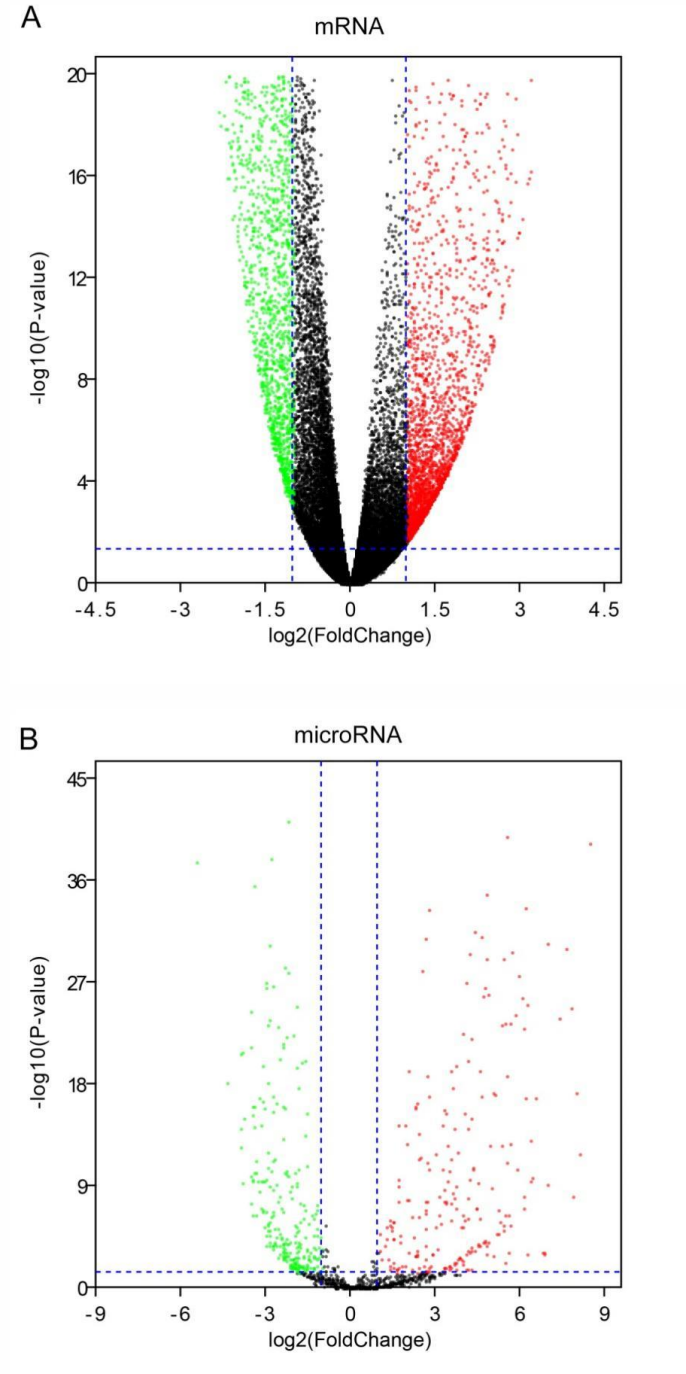

Fig. 1: Differentially expressed genes (DEGs) of COAD versus normal control using the TCGA data. (A) the volcano plot of DE mRNAs. (B) the volcano plot of DE miRNAs

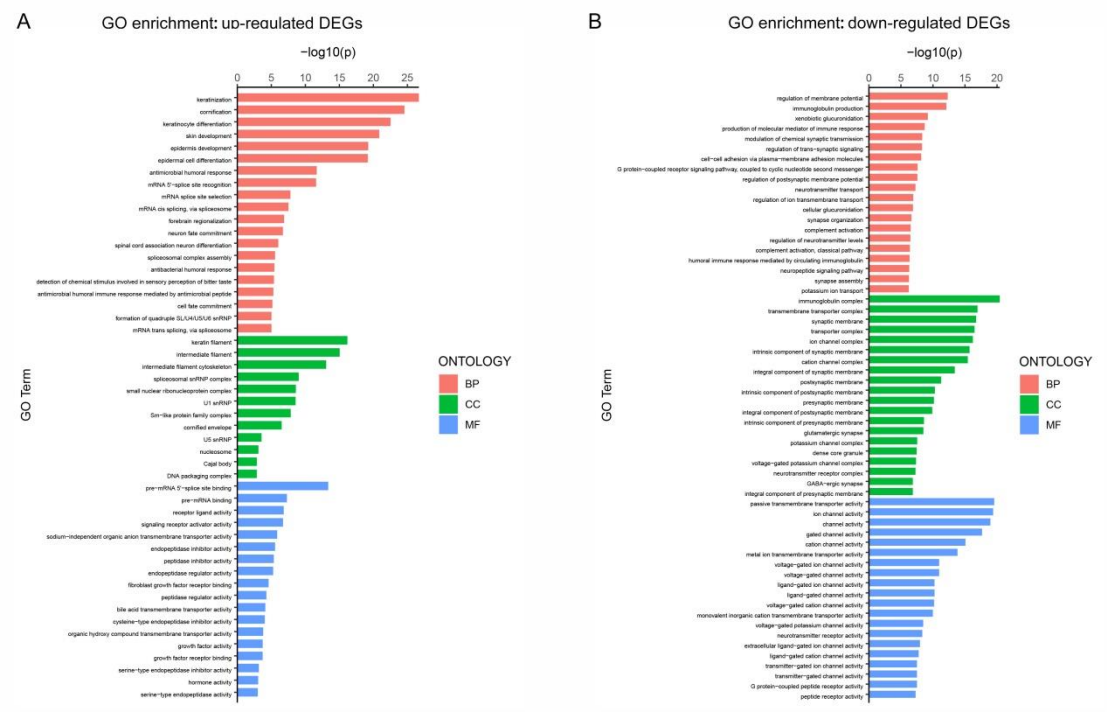

Fig. 2: The enriched GO function terms based on DEGs in COAD. (A) The GO enrichment based on up-regulated DEGs. (B) The GO enrichment based on down-regulated DEGs

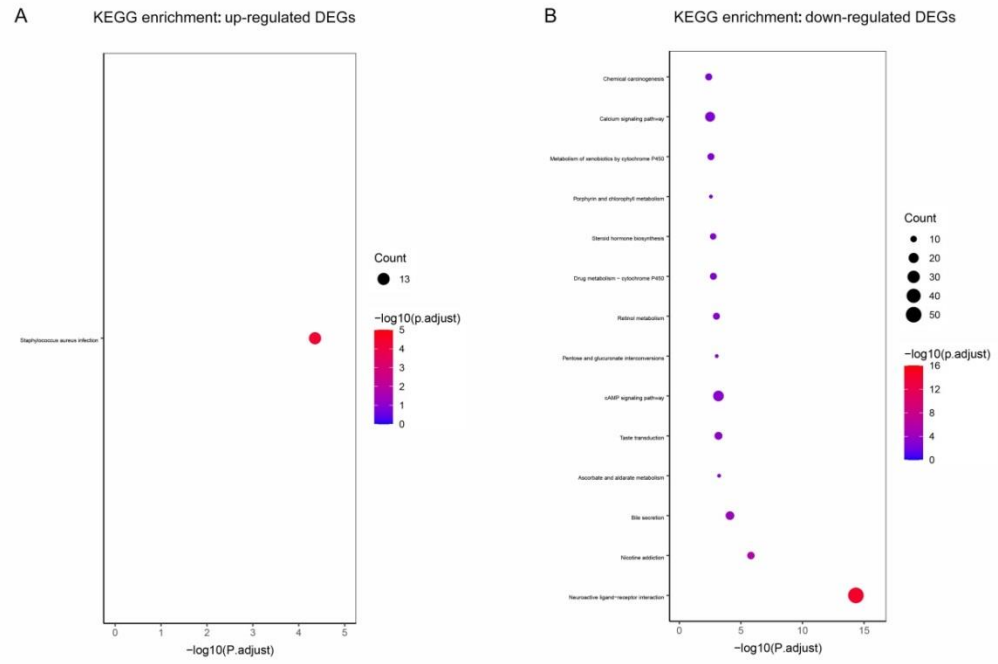

Fig. 3: The KEGG pathway enrichments based on DEGs in COAD. (A) The KEGG pathway enrichment based on up-regulated DEGs. (B) The KEGG enrichment based on down-regulated DEGs

The miRNA target network

Applying the up-mRNA-and-down-miRNA pairs and the down-mRNA-and-up-miRNA pairs, the miRNA target network was generated (Fig. 4 A and B). There were 61 up-mRNA-and-down-miRNA pairs (including 7 miRNAs and 41 mRNA targets) and 236 down-mRNA-and-up-miRNA pairs (including 8 miRNA and 101 mRNA targets). Here, we paid more attention to the carcinogenic genes. Therefore, those mRNAs with a positive relationship with COAD and miRNAs with a negative relationship with COAD were selected for further analysis. As Fig. 4A shown, miR-129-5p was highly important in DEG regulation during COAD development. It may be a powerful tumor suppressor, which targets 21 upregulated mRNAs.

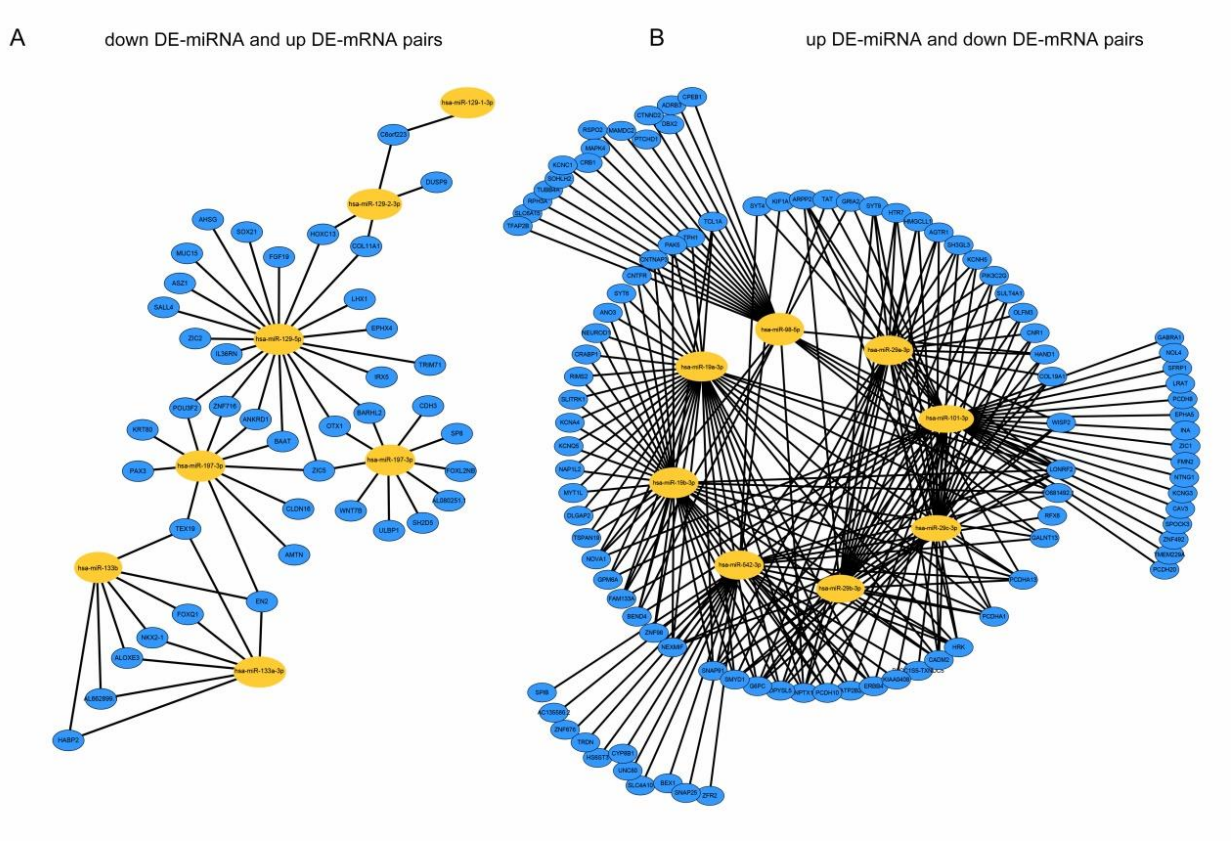

Fig. 4: The miRNA target network established using the up-mRNA-and-down-miRNA pairs and the down-mRNA-and-up-miRNA pairs

Important DEGs associated with COAD survival

Next, for the 41 up-regulated and 101 down-regulated mRNAs in the above miRNA target network, three DEGs were found assessed with the OS of COAD patients: *HOXC13*, *FOXL2NB*, and *ALOXE3*. All these genes are carcinogenic ones that they were in the up-mRNA-and-down-miRNA network and their high expression indicate a poor OS (Fig. 5 A-C).

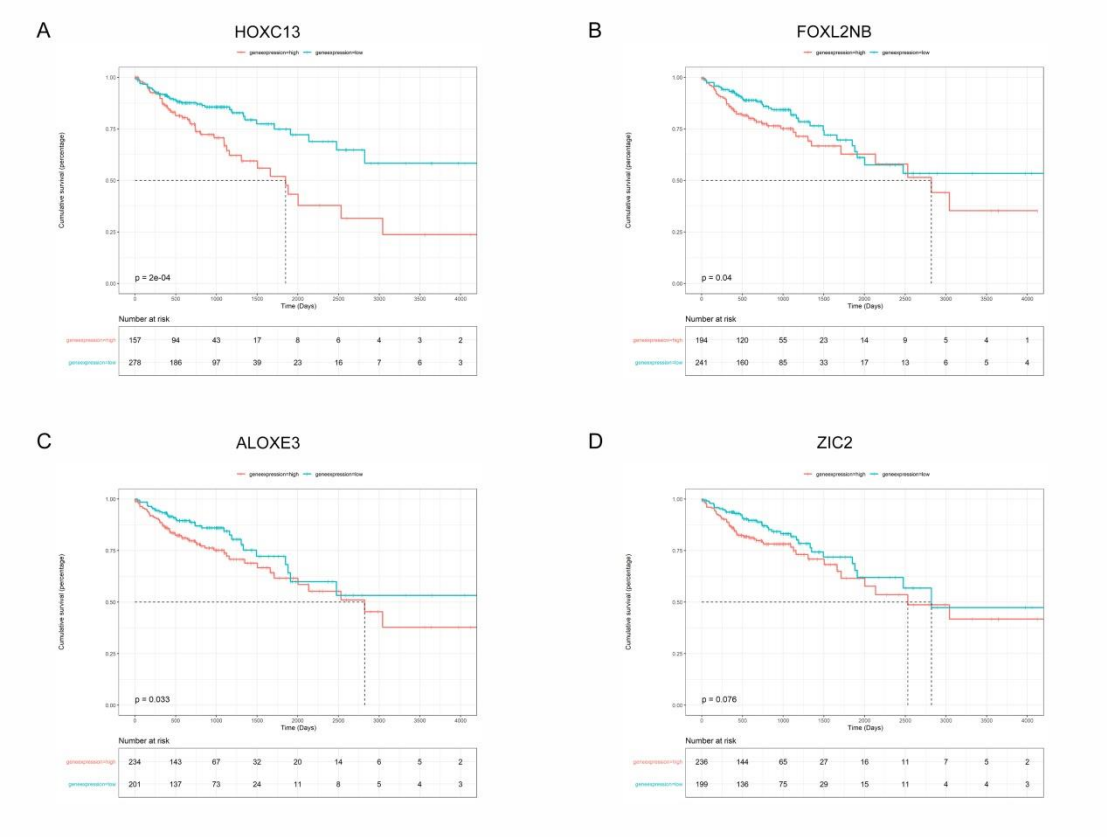

Fig. 5: The differences in overall survival (OS) in COAD patients among different risk genes. (A) *HOXC13*, (B) *FOXL2NB*, (C) *ALOXE3*, (D) *ZIC2*

The significance of ZIC2 and miR-129-5p

Further, a pair of important factors were used to dig the deeper association with COAD progression. In our previous clinical practice, we noticed an increased trend of Zinc-finger of the cerebellum 2 (*ZIC2*) in COAD samples. Interestingly, *ZIC2* located at the subnet with the most targets (the miR-129-5p subnet, Fig. 5A). Therefore, we further evaluated the significance of *ZIC2* in COAD progression. Although it does not significantly influence the OS in the overall time window (only a slight effect was noticed, $P = 0.076$, Fig. 5D), when we focused on the 10-yr survival, the TIMER online tool showed that *ZIC2* is also a significant risk gene for COAD (Fig. 6A). The PPI network (Fig. 6B) of *ZIC2* included SUMO1, POU5F1, PRKDC, and FBXW11. In addition, *ZIC2* expression was negatively correlated with the tumor purity but overall, positively correlated with B cell, CD8+ cell, neutrophil, and DC cell infiltration (Fig. 6C). However, the cases with ultra-high infiltration level may have decreased *ZIC2* level. In particular, only those with deep deletion of *ZIC2* had a high immune-cell infiltration (Fig. 6D); while those with high amplification of *ZIC2* had lowest immune-cell infiltration. Next, given the significance of miR-129-5p in the up-mRNA-and-down-miRNA network, we further explored the role of miR-129-5p in COAD. From stage I to IV, the expression of miR-129-5p consistently showed a dramatically decrease versus normal (Fig. 6E). Thereby, we performed the validation of miR-129-5p by real-time qPCR. A total of three normal control tissues and five-colon cancer tissues were examined, and as expected, miR-129-5p was significantly decreased in all the tumor samples (Fig. 6F).

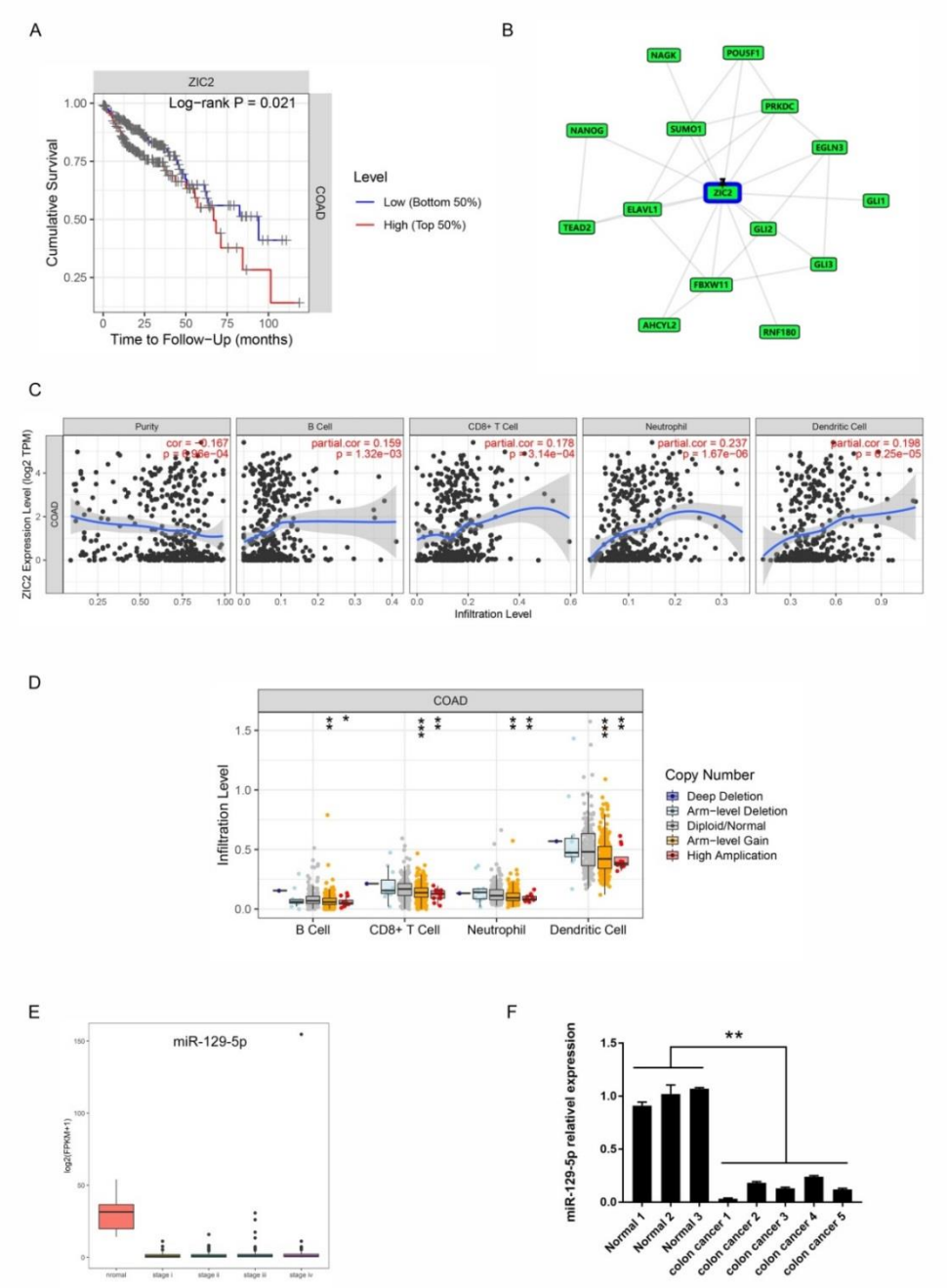

Fig. 6: The significance of *ZIC2* and miR-129-5p. (A) Focusing on the 10-year survival, *ZIC2* is also a significant risk gene. (B) The PPI network of *ZIC2*. (C) *ZIC2* expression was correlated with the tumor purity, B cell, CD8+ cell, neutrophil, and DC cell infiltration. (D) In TCGA database, those with deep deletion of *ZIC2* have a high immune-cell infiltration; while those with high amplification of *ZIC2* have lowest immune-cell infiltration. (E) From stage I to IV, the expression of miR-129-5p consistently show a dramatical decrease versus normal. (F) The validation of miR-129-5p expression by real-time qPCR. A total of three normal control tissues and five colon cancer tissues were examined, and miR-129-5p is significantly decreased in all the tumor samples

Discussion

We used the TCGA data to identify key carcinogenic genes in COAD and discovered 61 up-mRNA-and-down-miRNA pairs (including 7 miRNAs and 41 mRNA targets) and 236 down-mRNA-and-up-miRNA pairs (including 8 miRNA and 101 mRNA targets). The onset of colon cancer might involve important GO functional enrichments such keratinization, cornification, keratin filament, intermediate filament, intermediate filament cytoskeleton immunoglobulin complex, passive transmembrane transporter activity, and KEGG pathway enrichments such as Staphylococcus aureus infection, Neuroactive ligand-receptor interaction, Nicotine addiction, and Bile secretion. Among the key carcinogenic genes of COAD, *HOXC13*, *FOXL2NB*, and *ALOXE3* might influence the overall survival, and the miR-129-5p/*ZIC2* pair may decide the progression and survival through different mechanisms, e.g., affecting the immune infiltration.

Actually, the role of *ZIC2* in colon has been mentioned in several studies. *ZIC2* is implicated in different cancers, but the role of *ZIC2* in tumorigenesis is bilateral. *ZIC2* could render colon cancer cells more resistant to low glucose-induced apoptosis. Recently, it was revealed that *ZIC2* can activate the Wnt signaling by interacting with β-catenin, and *ZIC2* correlates with colon cancer stem cell properties (7). Accordingly, we noticed that *ZIC2* may influence the survival of colon cancer patients, and the similar result was found in the study from Euna Jeong et al, that the disease-free survival is poor in the patients with high expression of *ZIC2* (8). The mechanism of its cancer-promoting effect is not clear so far. Given the deterioration, process of COAD is closely related to the tumor microenvironment, the influence infiltration was explored in combination of the *ZIC2* expression. For the first time, we reported that a possible mechanism is the influence on immune cell infiltration. However, it is still early to drawn a definite conclusion about the causal relationship *ZIC2* expression and immune infiltration changes.

Among all DE miRNAs, we noticed that miR-129-5p might be a tumor suppressor, which targets the most upregulated DE mRNAs. Down-regulation of miR-129-5p was observed in some independent colon cancer studies, such as colon-cancer-stage 2 vs normal (9), advanced colorectal cancer vs normal (10), human colon cancer cell lines vs normal human intestinal epithelial HIEC cells (11), and other studies using the TCGA data (12). According to limited studies, miR-129-5p may be sponged by high motility group box protein 1 (HMGB1) and inhibit the HMGB1 signals (11). Besides, the miR-129-5p/NFAT5 axis and miR-129-5p/SEMA6A axis may represent a novel regulatory mechanism concerning the progression of colon cancer (12, 13).

Besides *ZIC2*, *HOXC13*, *FOXL2NB*, and *ALOXE3* were found to be risk genes for OS. *HOXC13* belongs to the homeobox family of genes. The homeobox genes encode a highly conserved family of transcription factors that play an important role in morphogenesis in all multicellular organisms. GO annotations related to this gene include sequence-specific DNA binding and proximal promoter DNA-binding transcription activator activity, RNA polymerase II-specific. *HOXC13* plays an important role in the progression of breast cancer (14), but its activity in colon cancer has not been reported. Furthermore, the relationship between *FOXL2NB* and colon cancer is not yet investigated. Finally, *ALOXE3* a member of the ALOX lipoxygenase family, which are catabolized by arachidonic acid-derived compounds. To date, there has been one study conducted by Ruan et al showing that *ALOXE3* combined ALOX12 might serve as potential prognosis biomarkers for COAD (15). However, the deep mechanism is still obscure. We believe above three carcinogenic genes can be selected as preferred targets for exploring the mechanism of COAD occurrence and development.

Still, this study has some limitations. Apart from miR-129-5p, we have not yet performed qPCR validation of the expression of *ZIC2*, *HOXC13*, *FOXL2NB* and *ALOXE3*. Moreover, the influence of above key genes on progression and

treatment efficacy are to be surveyed. In addition, our further work will utilize COAD cell lines to verify the matching relationship between miR-129-5p and the important carcinogenic genes.

Conclusion

This study provided an integrative analysis of gene expression profiles and interaction networks regarding COAD development. We identified several key carcinogenic genes and an important miRNA, miR-129-5p as a tumor suppressor. These genes correlate with the clinical prognosis of patients with COAD. These risk genes, interaction networks, and enrichments may provide a better understanding of the complex molecular mechanisms in COAD development and potential therapeutic targets for clinical treatment of COAD.

Ethical considerations

Ethical issues (Including plagiarism, informed consent, misconduct, data fabrication and/or falsification, double publication and/or submission, redundancy, etc.) have been completely observed by the authors.

Acknowledgements

This work was supported by Quanzhou Science and Technology Bureau High-level Talent Scientific Research Project (No.2018C051R); Talent project of Fujian Provincial Health and Family Planning Commission (No.2017-CXB-10); Quanzhou Science and Technology Plan Project (No.2018Z099).

Conflict of interest

The authors declare that there is no conflict of interest.

References

1. Termuhlen PM, Sweeney-Gotsch BM, Berman RS, et al (2002). Increased apoptosis in metastatic human colonic adenocarcinomas. *Cancer Biol Ther*, 1(1):58-63.
2. Yu D, Gao P, Song Y, et al (2018). The differences on efficacy of oxaliplatin in locally advanced colon cancer between mucinous and nonmucinous adenocarcinoma. *Cancer Med*, 7(3):600-615.
3. Chang Z, Huang R, Fu W, et al (2020). The Construction and Analysis of ceRNA Network and Patterns of Immune Infiltration in Colon Adenocarcinoma Metastasis. *Front Cell Dev Biol*, 8:688.
4. Chen C, Huang H, Wu CH (2017). Protein Bioinformatics Databases and Resources. *Methods Mol Biol*, 1558:3-39.
5. Bhattacharya S, Dunn P, Thomas CG, et al (2018). ImmPort, toward repurposing of open access immunological assay data for translational and clinical research. *Sci Data*, 5:180015.
6. Li B, Severson E, Pignon JC, et al (2016). Comprehensive analyses of tumor immunity: implications for cancer immunotherapy. *Genome Biol*, 17(1):174.
7. Xu Z, Zheng J, Chen Z, et al (2021). Multilevel regulation of Wnt signaling by *Zic2* in colon cancer due to mutation of β-catenin. *Cell Death Dis*, 12(6):584.
8. Jeong E, Lee Y, Kim Y, Lee J, Yoon S (2020). Analysis of Cross-Association between mRNA Expression and RNAi Efficacy for Predictive Target Discovery in Colon Cancers. *Cancers (Basel)*, 12(11):3091
9. Zhang JX, Song W, Chen ZH, et al (2013). Prognostic and predictive value of a microRNA signature in stage II colon cancer: a microRNA expression analysis. *Lancet Oncol*, 14(13):1295-1306.
10. Lu JH, Zuo ZX, Wang W, et al (2018). A two-microRNA-based signature predicts first-line chemotherapy outcomes in advanced colorectal cancer patients. *Cell Death Discov*, 4:116.
11. Wu Q, Meng WY, Jie Y, Zhao H (2018). LncRNA MALAT1 induces colon cancer development by regulating miR-129-

5p/HMGB1 axis. *J Cell Physiol*, 233(9):6750-6757.

12. Hu D, Zhang B, Yu M, Shi W, Zhang L (2020). Identification of prognostic biomarkers and drug target prediction for colon cancer according to a competitive endogenous RNA network. *Mol Med Rep*, 22(2):620-632.
13. Xi G, Ziyu X, Yiting L, Zonghang L, Lifeng Z (2021). Construction of competing endogenous RNA network and identification of novel molecular biomarkers in colon cancer: A bioinformatic analysis. *Medicine (Baltimore)*, 100(14):e25369.
14. Li C, Cui J, Zou L, Zhu L, Wei W (2020). Bioinformatics analysis of the expression of *HOXC13* and its role in the prognosis of breast cancer. *Oncol Lett*, 19(1):899-907.
15. Ruan GT, Gong YZ, Zhu LC, et al (2020). The Perspective of Diagnostic and Prognostic Values of Lipoxygenases mRNA Expression in Colon Adenocarcinoma. *Onco Targets Ther*, 13:9389-9405.