Int J Clin Exp Med 2021;14(8):2250-2258
www.ijcem.com /ISSN:1940-5901/IJCEM0129249

Original Article

Integrated bioinformatics revealed 2 novel prognostic biomarkers in colorectal cancer

Haibo Li*, Lang Liu*, Yizhou Jiang, Hua Jiang, Chuangqiang Niu, Jiejun Xia, Zhenyin Liu, Yiqun Guo

*Department of Interventional Radiology and Vascular Anomalies, Guangzhou Women and Children's Medical Center, Guangzhou 510120, China. *Equal contributors.*

Received December 29, 2020; Accepted May 27, 2021; Epub August 15, 2021; Published August 30, 2021

Abstract: Identification of novel molecular biomarker is urgent for the diagnosis and treatment of colorectal cancer (CRC). The aim of our present investigation is to explore potential diagnostic and prognostic factors in CRC. Here, we first downloaded the CRC RNA-seq data (GSE41328 and GSE47076) from the Gene Expression Omnibus (GEO) database. Then, bioinformatics methods were used to screen differentially expressed genes (DEGs). Gene ontology (GO) as well as Kyoto Encyclopedia of Genes and Genomes (KEGG) pathway enrichment analyses were used to investigate their functions. Combined with the Cancer Genome Atlas (TCGA) data of CRC, we performed survival analysis and receiver operating characteristic (ROC) curve analysis. Finally, we found that EIF4E3 and TTYH3 could be independent prognostic biomarkers. However, there still remains the need to further study the function and regulatory mechanism of the two genes in CRC.

Keywords: Colorectal cancer, prognostic biomarkers, EIF4E3, TTYH3

Introduction

Colorectal cancer (CRC) is one of the most prevalent malignancies in the world [1]. Although advances in radiomics, positron emission tomography, immunotherapy and surgery have been widely used, the prognosis of patients with CRC remains unsatisfactory [2, 3]. As a result of lack of effective biomarkers and tumor heterogeneity, good prognosis of CRC faces great challenges [3, 4]. So, it's urgent to find novel and robust biomarkers for CRC diagnosis. Furthermore, deep understanding of the molecular mechanism of CRC can provide novel insights into the pathogenesis of the disease [5].

EIF4E3 is an importnat member of the EIF4E family, which acts as a transcription initiation factor and recruits mRNA to the ribosome [6]. Several studies have reported the function of EIF4E3 in cancers. For example, Osborne et al. found that eIF4E3 relies on methyl-7-guanosine (m (7) G) cap-binding activity to act as a tumor suppressor [7]. Moreover, EIF4E3 overexpression could marginally inhibit eIF4E1-driven translation, which was regulated by MNKs in diffuse large B-cell lymphoma [8]. However, they are still rarely reported in colorectal cancer. Otherwise, TTYH3 encodes a member of the tweety family of proteins, as well as functions as a chloride anion channel and it plays a role in Ca(2+) signal transduction [9]. Some chloride channels have been reported to be associated with the occurrence and development of various tumors [10, 11]. A recent study showed that TTYH3 is highly expressed in gastric cancer (GC) and related to poor prognosis with GC patients [9]. However, the role of TTYH3 in other cancers including CRC remains unclear.

In this study, we first used two GEO datasets GSE41328 and GSE47076 to screen DEGs in CRC. Then, GO and KEGG enrichment analyses were performed to explore the function of these DEGs. Thirdly, we combined the GEO DEGs with TCGA datasets of CRC to identify novel and effective candidate markers. Finally, the GEPIA database was used to validate the expression of TTYH3 and EIF4E3. Kaplan-Meier (KM) survival analysis which is based on TCGA clinical information was conducted to verify the prog-

nostic value of the two genes. Our study identified two novel potential diagnostic and prognostic biomarkers of CRC.

Materials and methods

Raw data download and DEGs screening

The transcriptome expression profiles of mRNA datasets for GSE41328 and GSE47076 were downloaded from the Gene Expression Omnibus (GEO, https://www.ncbi.nlm.nih.gov/geo/). The GSE41328 dataset contains 5 paired colorectal cancer and normal tissues, while GSE47076 contains 8 samples (4 colorectal cancer and paired adjacent tissues). The raw data of these datasets was preprocessed using R limma package. The differentially expressed genes were screened as $P < 0.05$, $|Log_2FC| > 1$.

Functional enrichment analysis of DEGs

Gene ontology and KEGG pathway enrichment analyses of the DEGs were performed using the DAVID Bioinformatics Resources 6.8 (https://david.ncifcrf.gov/). The website provides a comprehensive set of functional annotation tools for investigators to understand biological meaning of the genes, including biological processes (BP), cellular components (CC), molecular functions (MF), functional domains, motifs and signaling pathways [12]. Protein-protein interactions (PPI) analysis was performed by STRING database (https://www.string-db.org/).

Clinical data acquisition

To verify the RNA-seq results, we used multiple databases to support our findings. The clinical validation data information of the patients was obtained from The Cancer Genome Atlas (TCGA, https://portal.gdc.cancer.gov/). In total we obtained 593 cancer samples and 11 normal samples from TCGA. Moreover, GEPIA (Gene Expression Profiling Interactive Analysis, http://gepia.cancer-pku.cn/), a useful web server for cancer and normal gene expression profiling analyses was used to analyze the differences between tumor samples and normal samples [13]. All the samples had not undergone any treatment such as chemotherapy or radiotherapy.

Survival analysis

To study the survival and prognosis related to TTYH3 and EIF4E3 expression, survival packages were analyzed by R software. Also, KM plotter univariate and multivariate cox analyses were used to filter gene expression data and survival data at a significant standard of $P < 0.05$.

Statistical analysis

The results are displayed with HR and *P*-values from a log-rank test. Unpaired t-test was used to distinguish the differences between CRC and adjacent tissues. $P < 0.05$ was considered as a statistically significant difference with 2-tailed tests. Receiver operating characteristic (ROC) curve, including the area under the curve (AUC) was used to analyze the diagnosis value of the candidates.

Results

Screening of DEGs

With the cut off criteria of an adjusted *P* value < 0.05 and $|log_2FC| > 1$, a total of 1795 DEGs including 780 upregulated and 995 downregulated genes were identified in GSE41328. In addition, there were 1267 DEGs (777 upregulated and 490 downregulated) which were identified in GSE47076. The volcano and heat map plots shows the DEGs in the two datasets (**Figure 1**).

Protein-protein interaction analysis

A total of 195 common DEGs were identified in the two datasets, including 123 up-regulated and 72 down-regulated DEGs (**Figure 2A**). Next, we obtained a protein-protein interaction network using String database. Then, the relationship was visualized by the Cytoscape (https://cytoscape.org/) bioinformatic tool. As shown in **Figure 2B**, the PPI network of the DEGs included 129 nodes and 349 interactions.

Functional enrichment analysis of DEGs

To obtain further insight into the function of DEGs in the network, KEGG and GO enrichment analyses were performed. The BP, CC and MF analyses were shown in **Figure 3A-C**. Top 10 biological processes mainly include regulation of cell proliferation and responses to inflammation (**Figure 3A**). The top 10 results in terms of cellular components and molecular functions are shown in **Figure 3B**, **3C**. Moreover, the top 10 KEGG pathways were shown in **Figure 3D**, mainly including the PI3K-Akt signaling path-

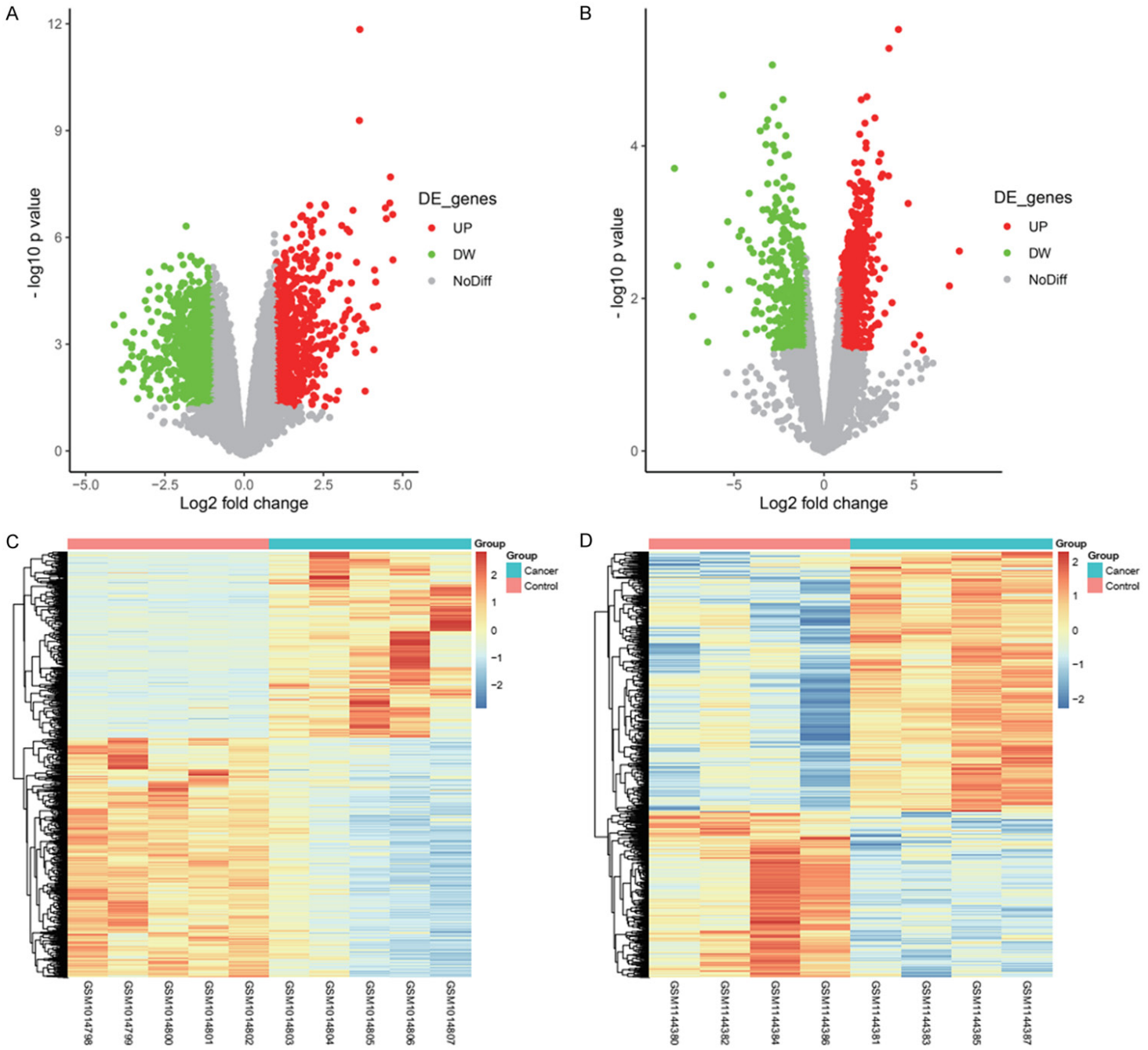
A
- log10 p value
Log2 fold change
DE_genes
UP
DW
NoDiff
B
- log10 p value
Log2 fold change
DE_genes
UP
DW
NoDiff
C
Group
Cancer
Control
GSM1014798
GSM1014799
GSM1014800
GSM1014801
GSM1014802
GSM1014803
GSM1014804
GSM1014805
GSM1014806
GSM1014807
D
Group
Cancer
Control
GSM1144380
GSM1144382
GSM1144384
GSM1144386
GSM1144381
GSM1144383
GSM1144385
GSM1144387

Figure 1. The differentially expressed genes in CRC and adjacent paracancerous tissue. A, B: The Volcano plots showed all the genes in GSE41328 and GSE47076 datasets. Red dots indicate significantly upregulated genes, green dots indicate significantly downregulated genes. UP, upregulated genes; DW, downregulated genes; NoDiff, no significantly differential expression. C, D: Heat map of differentially expressed genes in datasets GSE41328 and GSE47076.

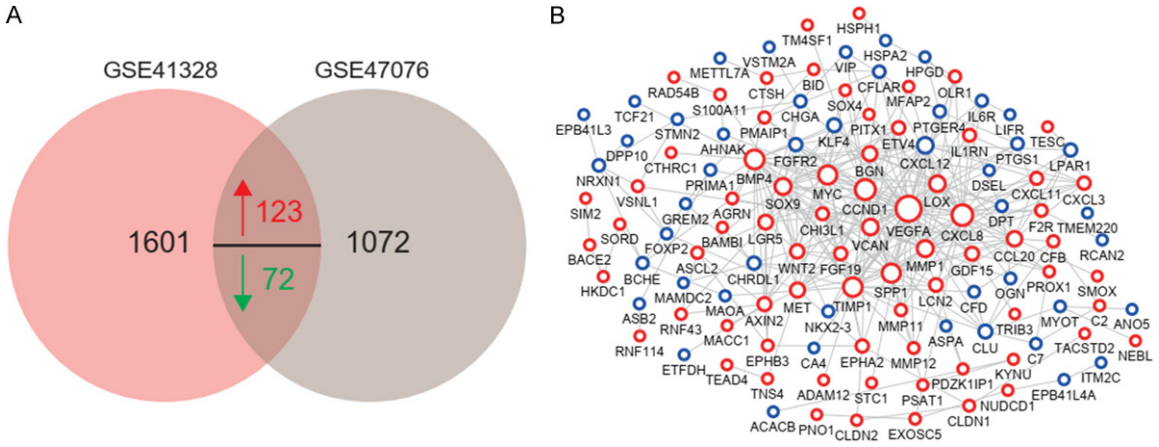

Figure 2. The protein-protein interaction analysis of common DEGs in GSE41328 and GSE47076 datasets. A: The venn map showed the common DEGs in GSE41328 and GSE47076 datasets, including 123 upregulated DEGs and 72 downregulated DEGs. B: The protein-protein interaction network of the DEGs in green module. The red circles represent upregulated genes and blue circles indicate downregulated genes. The greater the degree of the change in the gene, the larger the circle.

way, pathways in cancer and other important metabolism pathways. These results indicated variable pathway changes during the development of colorectal cancer.

Combined with the depth analysis of TCGA dataset

Subsequently, we compared our DEGs and TCGA cohort (593 cancer and 11 normal) DEGs of CRC. There are 151 common DEGs, with 99 up-regulated and 52 down-regulated genes (**Figure 4A**). Next, the expression model of the DEGs in TCGA was shown by heat map (**Figure 4B**). Then, a forest plot was performed to evaluate the DEGs which can be used as prognostic factors. The univariate analysis showed the *P* value, hazard ratio (HR), 95% CI of 13 significant DEGs in networks (**Figure 4C**). The Hazard ratios of RNF43, MYC, EIF4E3, CXCL11 and CXCL8 were less than 1, which means these genes were not risk factors for carcinogenesis in CRC. While seven genes (FXYD5, BAMB1, TIMP1, VEGFA, TTYH3, CFCL2 and LDLRAD3) maybe risk factors for CRC. Then, multivariate analysis of these 13 genes was performed. We found that TIMP1, RNF43, VEGFA, TTYH3, EIF4E3, CFL2, CSCL8 and ZBTB7C were significantly correlated with survival of CRC patients. These eight genes could be independent prognostic factors in CRC.

Clinical prognostic value analysis of TTYH3 and EIF4E3

Among the 8 potential prognostic factors, TTYH3 and EIF4E3 were novel genes that have not been deeply studied. The two genes were selected as our study subjects. The mRNA expression of the two genes was validated by the GEPIA database. EIF4E3 was significantly downregulated (**Figure 5A**) and TTYH3 (**Figure 5B**) was significantly upregulated in CRC (READ and COAD). To explore the correlation between EIF4E3, TTYH3 expression and prognosis in CRC, we investigated the effects of the expression of the two genes on survival by TCGA data. The overall survival was based on the expression level of the 2 genes in CRC, individually. EIF4E3 (**Figure 5C**) and TTYH3 (**Figure 5D**) had a significant effect on OS of CRC. To assess the potential diagnosis value of EIF4E3 and TTYH3, we further performed ROC curve analysis (**Figure 5E**). We found that the ROC curve of EIF4E3 and TTYH3 showed a significant distinguishing efficiency with AUC values of 0.98 (***P < 0.001) and 0.94 (***P < 0.001), which

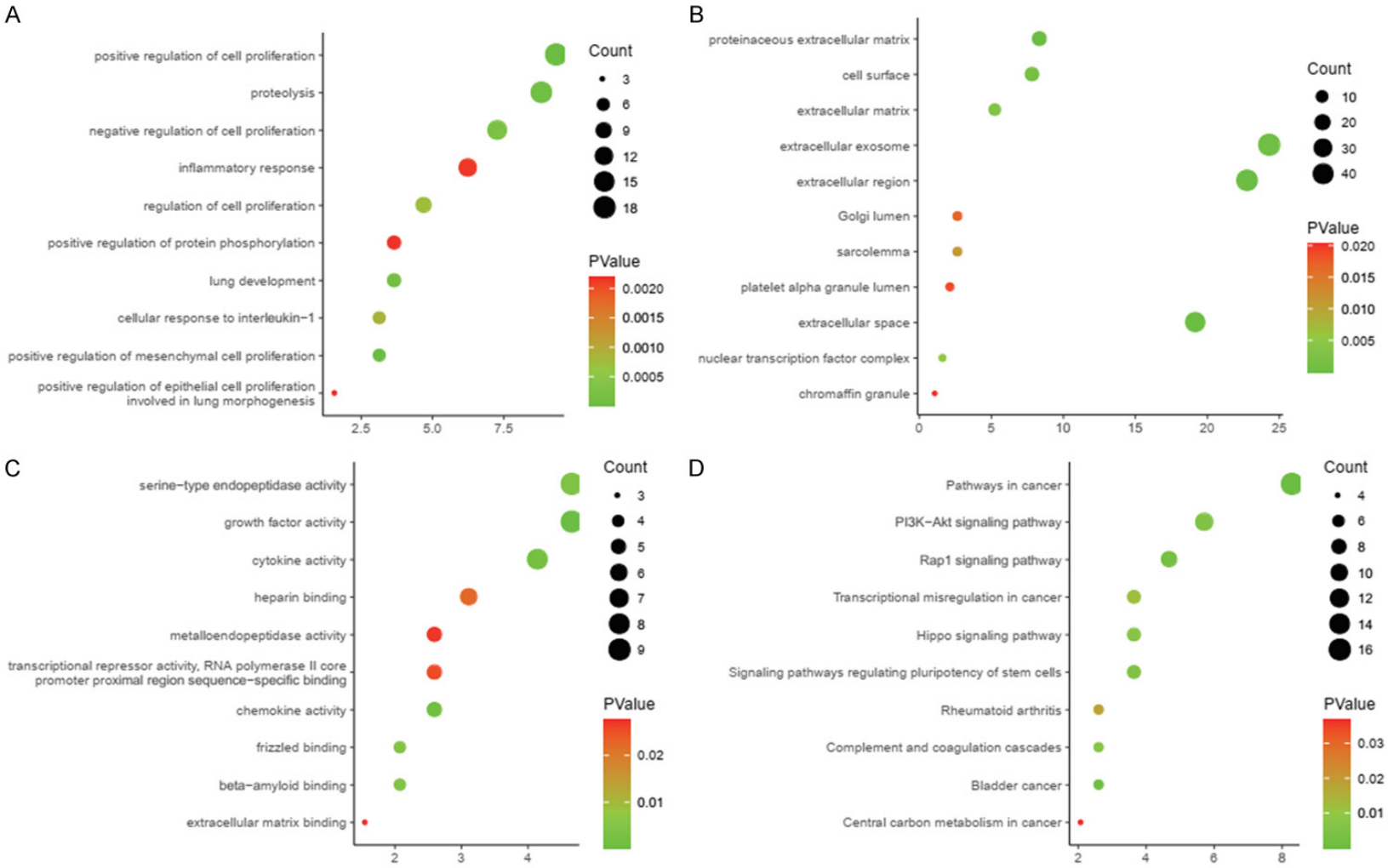

Figure 3. The functional analysis of common DEGs in GSE41328 and GSE47076 datasets. A: The top 10 most significantly enriched biological processes. B: The top 10 most significantly enriched cellular component terms. C: The top 10 most significantly enriched molecular functions. D: The top 10 significant KEGG pathways.

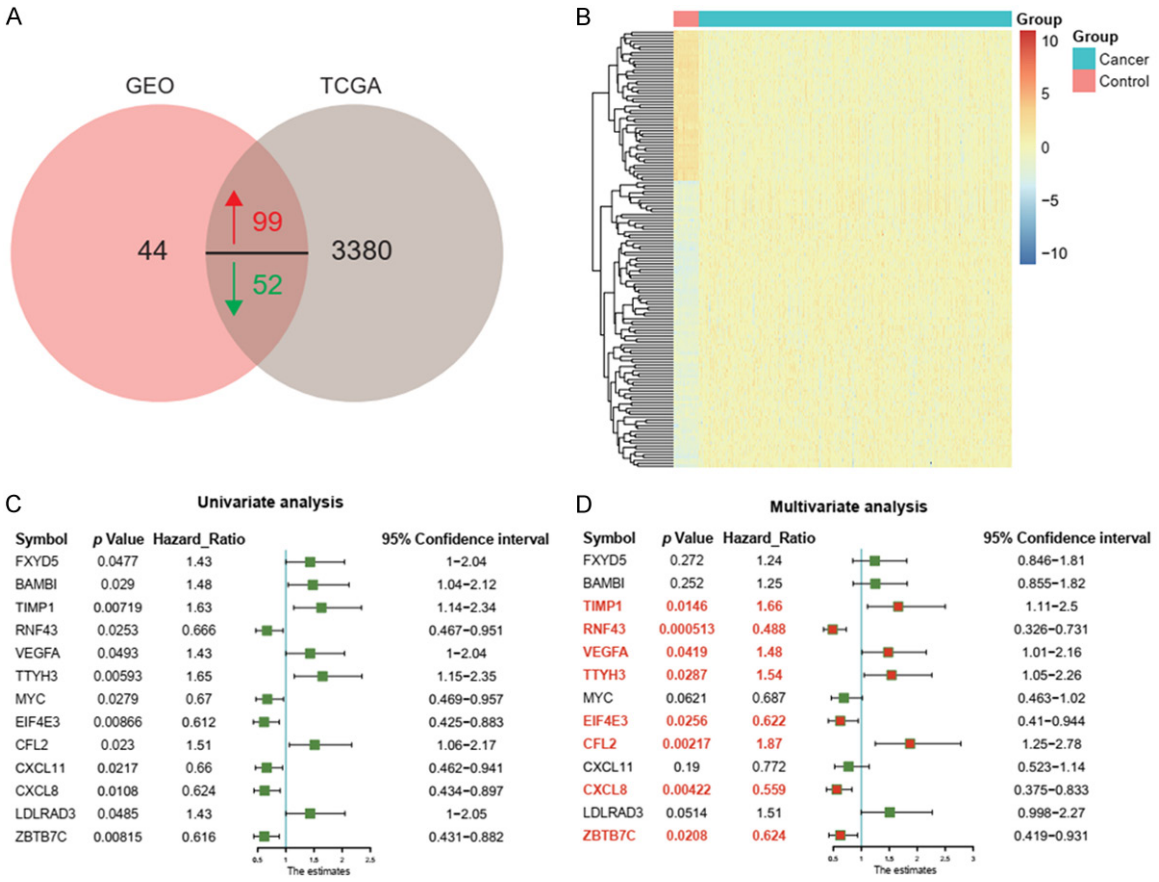

Figure 4. The forest plot of potential prognosis genes. A: The common 151 DEGs of the two datasets and CRC TCGA cohort, containing 99 upregulated genes and 52 down-regulated genes in the venn map. B: Heat map of the 151 differentially expressed genes. C, D: Univariate and multivariate analysis of the 13 significant genes. The forest plot showed the *P* value, hazard ratio (HR), 95% CI of the DEGs.

suggested the two genes could be independent candidate biomarkers for CRC diagnosis, respectively.

Discussion

So far, CRC accounts for approximately 10% of new cancer cases annually, worldwide; remaining the second highest rate of cancer-related deaths [14]. Tumor metastasis is the major cause of a poor outcome of CRC patients. Metastasis dramatically reduces the outcome of CRC patients. So, it is necessary to find new specific diagnostic biomarkers in early stages of CRC [5, 15]. RNA-seq combined bioinformatics have been a robust method to find novel biomarkers in different kinds of diseases, such as breast cancer [16], liver cancer [17], and colon cancer [18, 19]. In the present study, we obtained RNA-seq data from GEO and TCGA datasets, combined with bioinformatics to discover novel potential biomarkers for CRC.

The eIF4E family is comprised of three components: eIF4E1, eIF4E2 and eIF4E3 [20], which are essential for the initiation of translation. Previous studies showed EIF4E family plays an important role in proliferation, survival [21], mRNA export and oncogenic transformation, and these functions are dependent on its cap-binding activity [22]. The family is reported to be a oncogene elevated in an estimated 30% of human cancer [23]. Among the family members, the most commonly studied is eIF4E1 [24]. However, several studies reported that eIF4E3 may be a tumor suppressing factor, rather than a promoter, of both target transcript expression and oncogenic transformation [7, 25]. The high expression of eIF4E3 marginally suppresses eIF4E1-driven translation, which exhibits a new role for eIF4E3 in translation

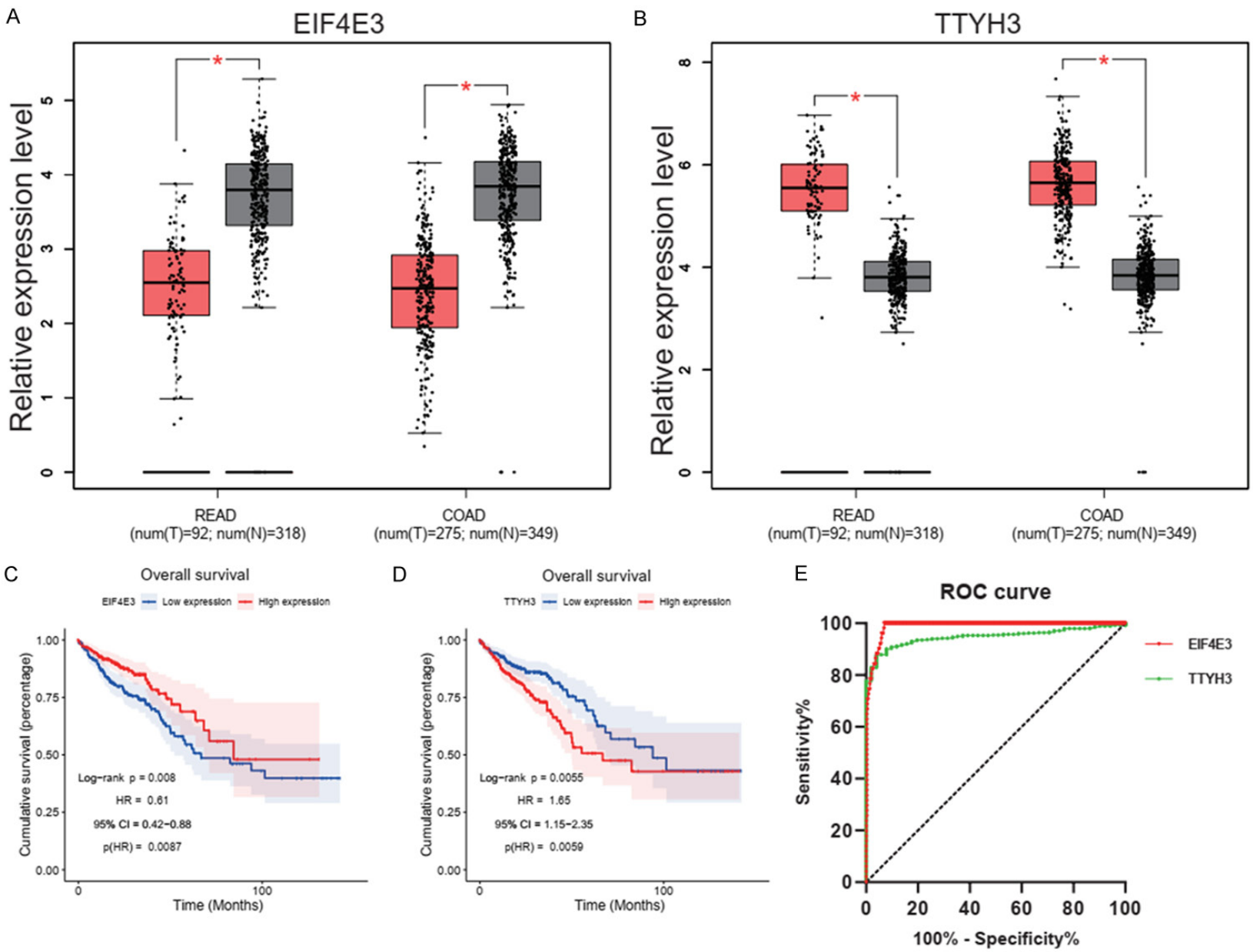
A
EIF4E3
Relative expression level
READ
(num(T)=92; num(N)=318)
COAD
(num(T)=275; num(N)=349)
B
TTYH3
Relative expression level
READ
(num(T)=92; num(N)=318)
COAD
(num(T)=275; num(N)=349)
C
Overall survival
EIF4E3 Low expression High expression
Cumulative survival (percentage)
Log-rank p = 0.008
HR = 0.61
95% CI = 0.42−0.88
p(HR) = 0.0087
Time (Months)
D
Overall survival
TTYH3 Low expression High expression
Cumulative survival (percentage)
Log-rank p = 0.0055
HR = 1.65
95% CI = 1.15−2.35
p(HR) = 0.0059
Time (Months)
E
ROC curve
Sensitivity%
100% - Specificity%
EIF4E3
TTYH3

Figure 5. Expression and prognosis of patients with eIF4E3 and TTYH3 expression. A, B: The mRNA expression level of eIF4E3 and TTYH3 in CRC (READ and COAD). C, D: The overall survival analysis between eIF4E3 ($P = 0.008$), TTYH3 ($P = 0.0055$) expression and survival time in TCGA ($^{*}P < 0.05$). E: Receiver operating characteristic curves of eIF4E3 and TTYH3 between AMI patients and healthy controls ($^{***}P < 0.001$).

initiation [8]. These results were matched the results in our CRC investigation. We demonstrated that eIF4E3 is downregulated in CRC and high expression of eIF4E3 showed a better prognosis.

The tweety protein family also comprised of three members (TTYH1, TTYH2 and TTYH3) in human [9]. High TTYH2 expression has been reported to be associated with the development of several cancers, such as renal cell carcinoma [26], colon carcinoma [27] and osteosarcoma [28]. Otherwise, we found high expression of TTYH3 was associated with poor prognosis in GC patients [9]. The finding was consistent with our results in CRC. Combined with ROC analysis, TTYH3 could be an independent candidate biomarker for CRC diagnosis.

In conclusion, our study confirmed that elevated expression of TTYH3 and EIF4E3 could be an independent prognosis factor of CRC patients. However, the elevated expression of the two genes involved in the progression of CRC is still unknown. In the future, we will investigate how the two genes influence the occurrence and development of CRC both *in vivo* and *in vitro*.

Acknowledgements

The work was funded by Guangzhou Women and Children's Medical Center.

Disclosure of conflict of interest

None.

Address correspondence to: Yiqun Guo, Department of Interventional Radiology and Vascular Anomalies, Guangzhou Women and Children's Medical Center, No. 9 Jinsui Road, Zhujiang New Town, Tianhe District, Guangzhou, Guangdong, China. Tel: +86-15013875250; E-mail: 531889323@qq.com

References

[1] Bray F, Ferlay J, Soerjomataram I, Siegel RL, Torre LA and Jemal A. Global cancer statistics 2018: GLOBOCAN estimates of incidence and mortality worldwide for 36 cancers in 185 countries. CA Cancer J Clin 2018; 68: 394-424.

[2] Miller KD, Siegel RL, Lin CC, Mariotto AB, Kramer JL, Rowland JH, Stein KD, Alteri R and Jemal A. Cancer treatment and survivorship statistics, 2016. CA Cancer J Clin 2016; 66: 271-289.

[3] Zhang N, Hu X, Du Y and Du J. The role of miRNAs in colorectal cancer progression and chemoradiotherapy. Biomed Pharmacother 2021; 134: 111099.

[4] Staal FCR, van der Reijd DJ, Taghavi M, Lambregts DMJ, Beets-Tan RGH and Maas M. Radiomics for the prediction of treatment outcome and survival in patients with colorectal cancer: a systematic review. Clin Colorectal Cancer 2021; 20: 52-71.

[5] Niu L, Yang W, Duan L, Wang X, Li Y, Xu C, Liu C, Zhang Y, Zhou W, Liu J, Zhao Q, Hong L and Fan D. Biological implications and clinical potential of metastasis-related miRNA in colorectal cancer. Mol Ther Nucleic Acids 2020; 23: 42-54.

[6] Joshi B, Cameron A and Jagus R. Characterization of mammalian eIF4E-family members. Eur J Biochem 2004; 271: 2189-2203.

[7] Osborne MJ, Volpon L, Kornblatt JA, Culjkovic-Kraljacic B, Baguet A and Borden KL. eIF4E3 acts as a tumor suppressor by utilizing an atypical mode of methyl-7-guanosine cap recognition. Proc Natl Acad Sci U S A 2013; 110: 3877-3882.

[8] Landon AL, Muniandy PA, Shetty AC, Lehrmann E, Volpon L, Houng S, Zhang Y, Dai B, Peroutka R, Mazan-Mamczarz K, Steinhardt J, Mahurkar A, Becker KG, Borden KL and Gartenhaus RB. MNKs act as a regulatory switch for eIF4E1 and eIF4E3 driven mRNA translation in DLBCL. Nat Commun 2014; 5: 5413.

[9] He Y, Hryciw DH, Carroll ML, Myers SA, Whitbread AK, Kumar S, Poronnik P and Hooper JD. The ubiquitin-protein ligase Nedd4-2 differentially interacts with and regulates members of the tweety family of chloride ion channels. J Biol Chem 2008; 283: 24000-24010.

[10] Peng JM, Lin SH, Yu MC and Hsieh SY. CLIC1 recruits PIP5K1A/C to induce cell-matrix adhesions for tumor metastasis. J Clin Invest 2020; 131: e133525.

[11] Wasson CW, Ross RL, Morton R, Mankouri J and Del Galdo F. The intracellular chloride channel 4 (CLIC4) activates systemic sclerosis fibroblasts. Rheumatology (Oxford) 2020; keaa797.

[12] Huang da W, Sherman BT and Lempicki RA. Systematic and integrative analysis of large gene lists using DAVID bioinformatics resources. Nat Protoc 2009; 4: 44-57.
[13] Tang Z, Li C, Kang B, Gao G and Zhang Z. GEPIA: a web server for cancer and normal gene expression profiling and interactive analyses. Nucleic Acids Res 2017; 45: W98-W102.
[14] Yarla NS, Madka V, Pathuri G and Rao CV. Molecular targets in precision chemoprevention of colorectal cancer: an update from pre-clinical to clinical trials. Int J Mol Sci 2020; 21: 9609.
[15] Luo XJ, Zhao Q, Liu J, Zheng JB, Qiu MZ, Ju HQ and Xu RH. Novel genetic and epigenetic biomarkers of prognostic and predictive significance in stage II/III colorectal cancer. Mol Ther 2021; 29: 587-596.
[16] Zhang F, Deng CK, Wang M, Deng B, Barber R and Huang G. Identification of novel alternative splicing biomarkers for breast cancer with LC/MS/MS and RNA-Seq. BMC Bioinformatics 2020; 21: 541.
[17] Li W, Kong X, Huang T, Shen L, Wu P and Chen QF. Bioinformatic analysis and in vitro validation of a five-microRNA signature as a prognostic biomarker of hepatocellular carcinoma. Ann Transl Med 2020; 8: 1422.
[18] Ren Y, Lv Y, Li T and Jiang Q. High expression of PLAC1 in colon cancer as a predictor of poor prognosis: a study based on TCGA data. Gene 2020; 763: 145072.
[19] Silva-Fisher JM, Dang HX, White NM, Strand MS, Krasnick BA, Rozycki EB, Jeffers GGL, Grossman JG, Highkin MK, Tang C, Cabanski CR, Eteleeb A, Mudd J, Goedegebuure SP, Luo J, Mardis ER, Wilson RK, Ley TJ, Lockhart AC, Fields RC and Maher CA. Long non-coding RNA RAMS11 promotes metastatic colorectal cancer progression. Nat Commun 2020; 11: 2156.
[20] Joshi B, Lee K, Maeder DL and Jagus R. Phylogenetic analysis of eIF4E-family members. BMC Evol Biol 2005; 5: 48.
[21] Culjkovic B, Topisirovic I and Borden KL. Controlling gene expression through RNA regulons: the role of the eukaryotic translation initiation factor eIF4E. Cell Cycle 2007; 6: 65-69.
[22] Cohen N, Sharma M, Kentsis A, Perez JM, Strudwick S and Borden KL. PML RING suppresses oncogenic transformation by reducing the affinity of eIF4E for mRNA. EMBO J 2001; 20: 4547-4559.
[23] Graff JR and Zimmer SG. Translational control and metastatic progression: enhanced activity of the mRNA cap-binding protein eIF-4E selectively enhances translation of metastasis-related mRNAs. Clin Exp Metastasis 2003; 20: 265-273.
[24] Yi T, Papadopoulos E, Hagner PR and Wagner G. Hypoxia-inducible factor-1alpha (HIF-1alpha) promotes cap-dependent translation of selective mRNAs through up-regulating initiation factor eIF4E1 in breast cancer cells under hypoxia conditions. J Biol Chem 2013; 288: 18732-18742.
[25] Volpon L, Osborne MJ, Culjkovic-Kraljacic B and Borden KL. eIF4E3, a new actor in mRNA metabolism and tumor suppression. Cell Cycle 2013; 12: 1159-1160.
[26] Rae FK, Hooper JD, Eyre HJ, Sutherland GR, Nicol DL and Clements JA. TTYH2, a human homologue of the drosophila melanogaster gene tweety, is located on 17q24 and upregulated in renal cell carcinoma. Genomics 2001; 77: 200-207.
[27] Toiyama Y, Mizoguchi A, Kimura K, Hiro J, Inoue Y, Tutumi T, Miki C and Kusunoki M. TTYH2, a human homologue of the drosophila melanogaster gene tweety, is up-regulated in colon carcinoma and involved in cell proliferation and cell aggregation. World J Gastroenterol 2007; 13: 2717-2721.
[28] Moon DK, Bae YJ, Jeong GR, Cho CH and Hwang SC. Upregulated TTYH2 expression is critical for the invasion and migration of U2OS human osteosarcoma cell lines. Biochem Biophys Res Commun 2019; 516: 521-525.