

Investigation of potential prognostic biomarkers for colorectal cancer

Hui Li¹, Jie Liu¹, WenHui Liu², Liang Zheng¹, JuHui Chen^{1*}

¹Department of Abdominal Radiotherapy, Clinical Oncology School of Fujian Medical University, Fujian Cancer Hospital, China

²Department of Radiation Oncology, Mengchao Hepatobiliary Hospital of Fujian Medical University, China

Submitted: 17 November 2022; **Accepted:** 3 June 2023

Online publication: 1 July 2023

Arch Med Sci 2025; 21 (2): 425–436

DOI: <https://doi.org/10.5114/aoms/167397>

Copyright © 2023 Termedia & Banach

***Corresponding author:**

JuHui Chen
Department of
Abdominal Radiotherapy
Clinical Oncology
School of Fujian
Medical University
Fujian Cancer Hospital
350011, China
E-mail: chen_juhui@outlook.com

Abstract

Introduction: Colorectal cancer (CRC) is the third leading cause of cancer-related death. Since CRC is largely asymptomatic until the alert features develop to an advanced stage, implementation of a screening program is important to reduce cancer morbidity and mortality. Current screening methods have significant limitations.

Material and methods: CRC-related microarray datasets were collected from the GEO database and differentially expressed genes (DEGs) were identified. Next, Venn analysis, functional enrichment analysis, protein interaction network (PPI) analysis, and survival analysis were performed.

Results: A total of 5267 and 4233 DEGs were identified in two datasets (GSE20916, GSE33133). The intersection of up-regulated genes in the two datasets was obtained by Venn Analysis as 1058 DEGs. Among the 1058 genes, 992 genes with survival and clinical information in TCGA were screened. Eleven DEGs were identified as potential prognostic markers. Model results show that the time period with the most obvious prognostic effect is 5 years, and the AUC value is the highest. ROC curve results are consistent with the model results of the survival analysis. The survival curve showed that LRR8A, PCAT6, PLA2G15, SRD5A1, T1GD1 may be oncogenes, and DSN1, ERI1, EIT1, GLMN, MAPKAPK, NOP14 may be tumor suppressor genes.

Conclusions: This study discovers novel prognostic markers through Cox regression and survival analysis, and provides a theoretical basis for the treatment of CRC.

Key words: colorectal cancer, treatment, prognostic biomarker.

Introduction

Colorectal cancer (CRC) is currently the third and second most common cancer in men and women, respectively [1, 2], and one of the leading causes of cancer-related deaths [3, 4]. Colorectal cancer (CRC) is the third most common cancer (9%) and the second most common cause of cancer-related mortality worldwide, according to the 2020 Global Cancer Statistics [5]. According to statistics, in 2008, there were nearly 150,000 newly diagnosed cases and 50,000 deaths [6]. By 2012, the Global Cancer Observatory estimated that 1.3 million new cases were diagnosed with CRC and nearly 700,000 patients died from it [7]. CRC is more common in Western countries, accounting for more than 65% of cases [8]. Within 4 years, both morbidity and mortality had risen at an alarming rate.

According to research, CRC usually occurs in the lining of the colon (large intestine) or rectum [9], and it usually grows on the lining of the colon and rectum as polyps, which are lumps that protrude into the lumen. Not all polyps are neoplastic and develop into cancer, but it is well known that most CRCs develop from adenomatous polyps, the so-called adenoma-to-cancer sequence [10]. Most CRCs occur in patients over the age of 50, with more than 75% of those over the age of 60. Risk factors associated with CRC include older age, male gender, lifestyle, inflammatory bowel disease, and previous history of CRC [4]. The risk of CRC increases with age, is 1.5–2 times higher in men than in women, and the mortality rate in the male population has been steadily increasing over the past three decades [11]. Despite the high mortality rate and high prevalence of CRC in cancer, it is an indolent disease in its early stages and usually develops symptoms when it progresses to a more advanced stage [12]. Therefore, mortality can be reduced through early prevention and detection.

Therefore, the ultimate goal should be to implement and improve screening strategies [13, 14]. Numerous efforts have been made to establish appropriate screening methods. Screening techniques can be divided into non-invasive and invasive with varying sensitivity and specificity, including the fecal occult blood test (FOBT), fecal DNA test, double contrast barium enema (DCBE) and colonoscopy [11]. Among them, colonoscopy is an invasive procedure with a risk of bowel perforation. These methods are either invasive, resulting in low participation rates in healthy populations [13], or have large limitations, limited by insufficient sensitivity or a high percentage of false positive results [15]. Currently, new biologics for CRC screening tests are sought. The development of markers is the subject of intensive research. In recent years, many biomarkers that aid in the treatment and prognosis of colorectal cancer have received more attention. In order to discover these potential biomarkers that can be detected in blood and stool by non-invasive methods, it is important to study the genetics and pathogenesis of CRC [16, 17].

A biomarker, or biomarker, is “a defined characteristic measured as an indicator of normal biological processes, pathogenic processes, or responses to exposures or interventions” [18]. There are several subtypes of biomarkers that can be identified and quantified to aid in the diagnosis, monitoring, assessment and prediction of disease and patient response to therapy, including diagnostic biomarkers, pharmacodynamic biomarkers, predictive biomarkers, prognostic biomarkers, and susceptibility biomarkers [19]. Among them, diagnostic biomarkers refer to detecting the presence of a certain disease or condition, for example, the

hemoglobin A_{1c} (HbA_{1c}) level for the diagnosis of type 2 diabetes. Prognostic biomarkers are used to identify the likelihood of disease recurrence or progression in people who already have the disease, e.g., prostate-specific antigen (PSA) levels, to assess the likelihood of cancer progression. This study aimed to identify potential novel prognostic tests and biomarkers for CRC. Two microarray datasets obtained from the Gene Expression Omnibus (GEO) database were analyzed to further identify differentially expressed genes (DEGs) in CRC by comparing expression levels between CRC and normal samples, and subsequently, differences shared across the datasets were identified using Venn analysis to compare DEGs. Next, enrichment analysis was performed to further analyze the biological processes of CRC-related DEGs and construct a protein-protein interaction (PPI) network.

Survival analysis refers to the method of analyzing and inferring the survival time of organisms or people based on the data obtained from experiments or surveys, and studying the relationship between survival time and outcomes and many influencing factors and their degree, also known as survival rate analysis. Survival rate analysis is often used in the evaluation of common clinical efficacy in tumors, such as exploring how long the remission period lasts for leukemia patients or the observation time of death [20–22]. In survival analysis, what we get is not a single specific value, but a curve, called a survival curve [23], and the function corresponding to the curve is called a survival function. To validate the prognostic role of DEG in CRC, CRC samples from The Cancer Genome Atlas (TCGA) and GEO databases were collected. Survival analysis was performed based on the sample segmentation method and Cox regression model to obtain prognostic genes and their survival curves. Next, using the dataset from TCGA as the training set and the GEO dataset as the validation set, the identified key genes and risk score values were used to predict the survival rate of CRC patients. Finally, the diagnostic value of genetic biomarkers for CRC was assessed using ROC curve analysis [24, 25], with complete identification of prognostic biomarkers for CRC.

Material and methods

Data selection

The Gene Expression Omnibus (GEO) database (<https://www.ncbi.nlm.nih.gov/geo/>) is a publicly available repository of gene expression datasets. Its data sources include microarray and next-generation sequencing [26]. The CRC dataset for this study was derived from GEO’s microarray dataset. In the GEO database, according to the keyword “colorectal cancer (CRC)”, the sample size is sort-

ed from large to small, and the sample size is uniformly distributed, and two GSE series, GSE20916 and GSE33133, are selected from the GPL570 microarray data platform as a research sample for CRC. The microarray data were also obtained from The Cancer Genome Atlas TCGA (<http://cancergenome.nih.gov/>) [27] and the GSE72968 dataset in GEO for survival analysis.

Identification of differentially expressed genes (DEGs)

The number of GSE20916 samples was 145, including 111 cancer tissues and 34 normal controls. The number of GSE33133 samples was 96, including 90 cancer tissues and 6 normal controls. Differential genes were defined with adj. p -value < 0.05 for differential analysis. In the TCGA colorectal cancer data, the number of samples was 624, and 35 samples with expression but no clinical survival information or with clinical survival information but no expression were excluded, and 589 samples were finally included.

Venn analysis

In order to obtain the intersection of up-regulated genes of the two datasets, the Venn diagram network tool (<http://bioinformatics.psb.ugent.be/webtools/Venn>) [28] was used to analyze the gene expression profiles in the two groups, and the identification of genes with CRC prognosis related genes.

Functional and pathway enrichment analysis

GO is a database established by the Gene Ontology Consortium, which aims to keep the functional description of gene products in each database consistent, so that the properties of gene products can be queried at different levels [29], and GO analysis has become a popular tool for large-scale transcriptome data, being a common method for functional studies and genomic data analysis [30]. The Kyoto Encyclopedia of Genes and Genomes (KEGG; <http://www.genome.jp/kegg/pathway>) is a reference database for systematic analysis of gene function, linking genomic information with metabolism, membrane transport, and signal transduction linked to higher-order functional information such as the cell cycle [31]. In this study, GO functional and KEGG pathway enrichment analysis was performed on the identified DEGs using R4.0.3, with $p < 0.005$.

Protein-protein interaction network analysis

The protein-protein interaction network (PPI network) is composed of individual protein interactions and participates in all aspects of life pro-

cesses such as biological signal transmission, gene expression regulation, energy and material metabolism, and cell cycle regulation [32]. The Search Tool for Retrieving Interacting Genes (STRING; <http://string-db.org>) database is an online software tool, and this study used the STRING database to build a PPI network of differential genes related to survival to present interactions [33].

Survival analysis of hub genes

CRC datasets from TCGA and GEO were used for survival analysis based on the sample segmentation method and a Cox regression model. The identified key genes can be used to characterize CRC survival and serve as potential therapeutic targets and prognostic biomarkers. The above studies have screened 1058 up-regulated genes, and due to sequencing factors, 922 genes were sequenced in the clinical data of 589 CRC samples in the TCGA database. In order to study the prognostic impact of the identified hub genes on CRC patients, univariate Cox analysis is usually performed to screen out the associated variables, and then a multivariate model is constructed to further confirm whether the association between variables and survival is independent [34]. However, this approach does not take into account the effect of multicollinearity among variables, leading to the possibility of contradictory results of hazard ratios obtained by univariate and multivariate Cox regression. Traditional variable selection methods such as stepwise regression, forward regression, and backward regression in Cox regression are no longer applicable [35].

Compared with Cox analysis method, LASSO (Least absolute shrinkage and selection operator) carries out variable selection while estimating model parameters, and performs better in solving the multicollinearity problem in regression analysis, and can explain the results well [36]. Therefore, when there is multicollinearity between variables or the number of variables is larger than the sample size, the method of Cox regression combined with LASSO is recommended to initially screen variables, and then further establish a LASSO regression model to analyze its impact on prognosis. In a word, the LASSO-Cox survival analysis mode was used, and the differential genes related to prognosis were analyzed using R4.0.3, the glmnet package and the survival package in this study.

Results

Screening for DEGs

In this study, CRC-related microarray datasets were collected from the GEO database, and a total of 5267 and 4233 DEGs were identified from the GSE20916 and GSE33133 datasets, respectively ($p < 0.05$). Among them, GSE20916 has 2847 up-reg-

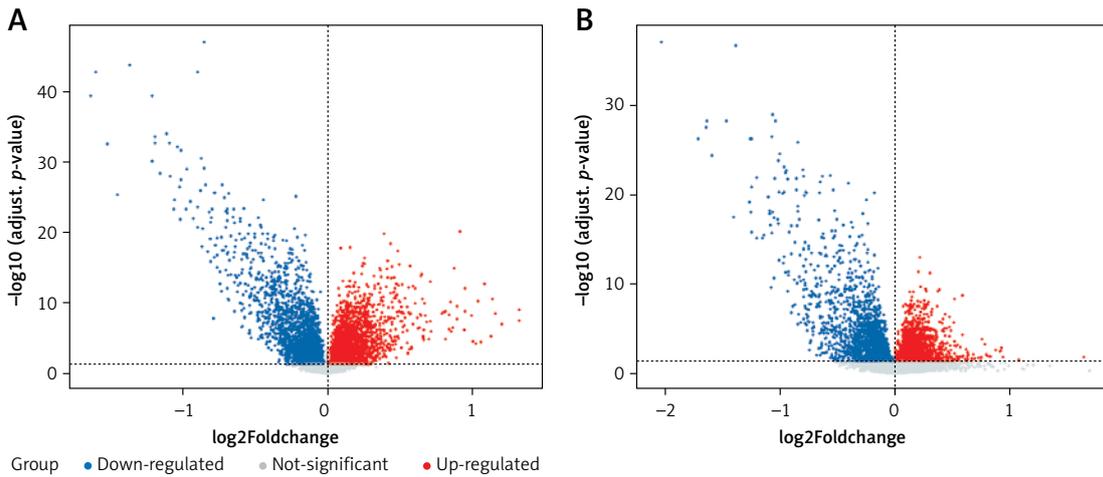


Figure 1. Distribution of differences in gene expression levels. **A** – Volcano plot showing expression data of CRC and normal tissue in GSE20916 microarray profiles. **B** – Volcano plot showing expression data in CRC and normal tissue in GSE33133. The x-axis is the average difference between CRC and normal samples (log2 fold change differential expression fold, according to the positive and negative log2 fold change value to determine whether the expression of these genes increases or decreases); the larger gene difference is distributed on the x-axis at both ends, top; the y-axis represents log-transformed *p*-values converted to $-\log_{10}$

ulated genes and 2420 down-regulated genes; GSE33133 has 2104 up-regulated genes and 2129 down-regulated genes. We created volcano plots of DEG distributions for these two datasets (Figure 1 A). Volcano plots (Figure 1 B) show significant and non-significant DEGs for the dataset, where blue dots represent down-regulated genes, red dots represent up-regulated genes, and gray dots represent genes that were not significantly different.

Venn analysis

Venn analysis of gene expression in tumor and normal tissues, taking the intersection of all up-regulated genes in GSE20916 and GSE33133 expression data, identified 1058 up-regulated genes (Figure 2).

Functional enrichment analysis of DEGs

To fully understand the biological roles of these DEGs in CRC, we continue to analyze the GO functions and pathways in which they are involved.

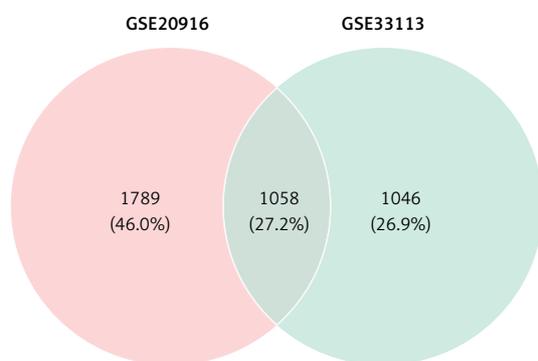


Figure 2. Intersection of up-regulated DEGs. Intersection of all up-regulated DEGs in the expression data of GSE20916 and GSE33133

The results of GO enrichment analysis included biological process (BP), cellular component (CC), and molecular function (MF).

GO analysis

Figure 3 (sorted by *p* from small to large) shows the results of the GO enrichment analysis. Up-regulation of DEGs mainly involves chromosomal region, spindle, condensed chromosome, chromosome, centromeric region, spindle pole, kinetochore, pre-ribosome, condensed chromosome, centromeric region, replication fork spindle microtubule cell components, ribonucleoprotein complex biogenesis, ncRNA metabolic process, ncRNA processing, ribosome biogenesis, chromosome segregation, mitotic nuclear division, rRNA processing, rRNA metabolic process, sister chromatid segregation, mitotic sister chromatid segregation and biological process significantly enriched. Molecular functions such as catalytic activity acting on RNA, catalytic activity acting on DNA, helicase activity, single-stranded DNA binding, ribonucleoprotein complex binding, damaged DNA binding, RNA methyltransferase activity, snoRNA binding, DNA replication origin binding, and DNA polymerase binding are active.

KEGG pathway analysis

The 10 key pathways involved in DEG are sorted by the value of *p* adjust from small to large, as shown in Figure 4. As the results show, the most significantly enriched signaling pathways for up-regulated DEGs are: Cell cycle, Ribosome biogenesis in eukaryotes, DNA replication, Nucleocytoplasmic transport, Proteasome, Nucleotide excision repair, Base excision repair, Spliceosome, Mismatch repair and IL-17 signaling pathway.

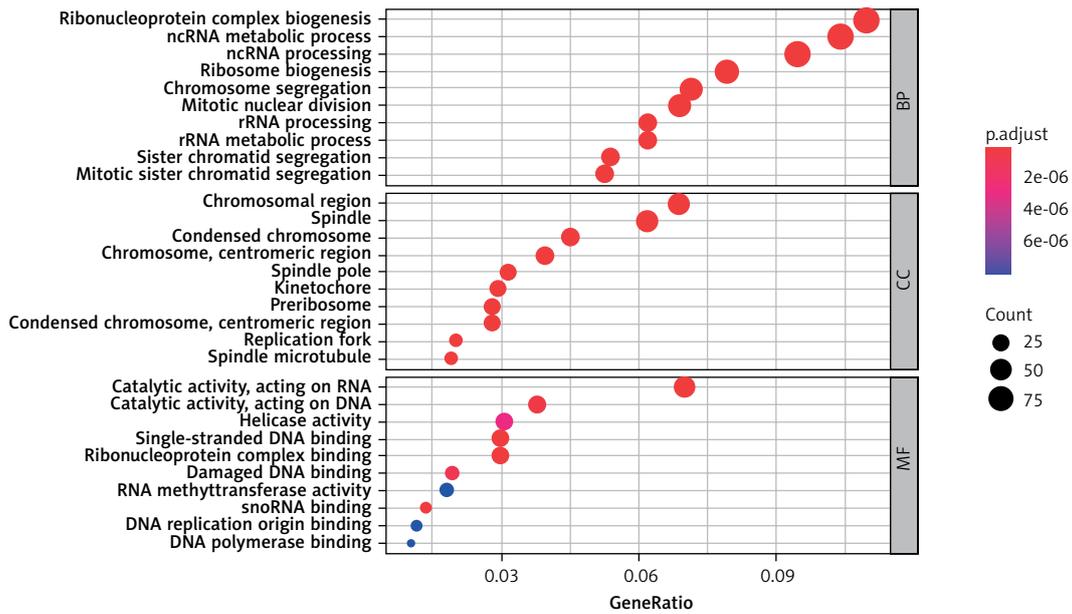


Figure 3. GO cluster analysis of DEGs. Biological processes, cellular component and molecular functions of up-DEGs

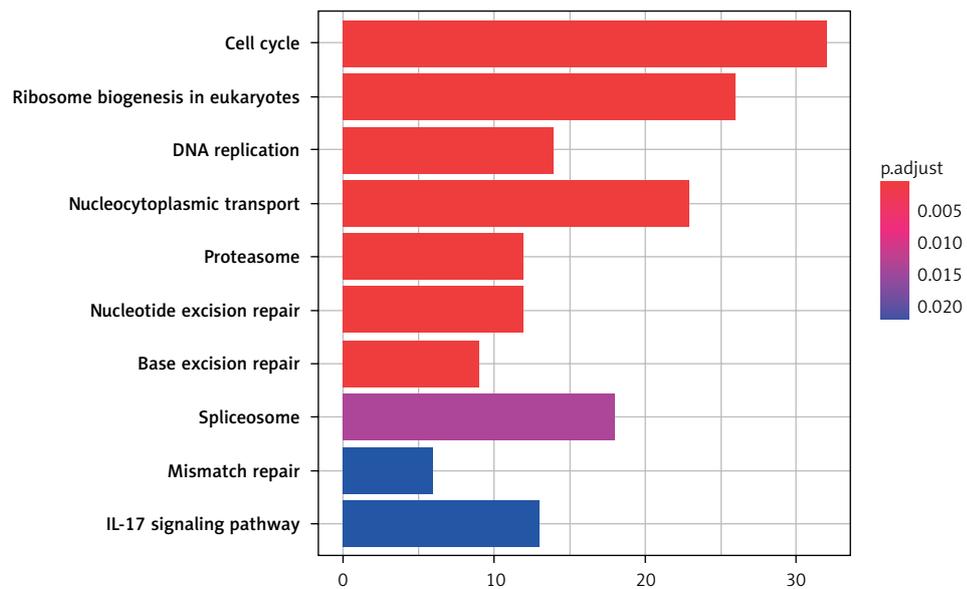


Figure 4. KEGG pathway analysis

Survival analysis of DEGs

Variables were first screened by univariate Cox, followed by Lasso regression analysis to determine final variables and build models to analyze the prognostic impact. Univariate Cox regression analysis was performed on 922 DEGs obtained in this study, and 88 prognostic variables were identified. Differential genes, followed by LASSO regression analysis, yielded 11 DEGs (Figure 5), which can be used as potential prognostic indicators for CRC.

Survival analysis and modeling

Subsequently, survival analysis was performed on the 11 DEGs, and the Kaplan-Meier curves were obtained as shown in Figure 6. Kaplan-Meier curves overall survival (OS) of these genes showed that high LRRC8A, PCAT6, PLA2G15, SRD5A1, T1GD1 expression was significantly associated with low OS in colorectal cancer patients, and high PLA2G15 expression was associated with the lowest survival after 5 years; high DSN1, ERI1, EIT1, GLMN, MAPKAPK, NOP14 expression was significantly associated with high OS in patients with colorectal cancer, and high expression of NOP14

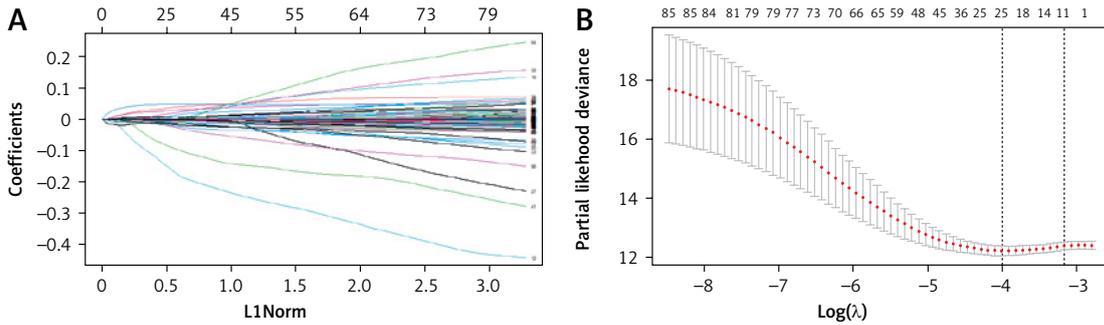


Figure 5. LASSO-Cox regression variable filtration. **A** – Cross-validation. Screening genes with prognostic efficacy from 88 genes, sample size $N = 88$; these genes are highly co-expressed in the sample; that is, there is collinearity between variables, and 11 genes are obtained, see Table I. **B** – Multivariate Cox regression analysis to construct a prognostic model and risk scoring. The abscissa is the $\log \lambda$ value, and the ordinate is the evaluation index corresponding to each λ value. The error bar is used to show the evaluation index of multiple models. Mean + standard error; two vertical dotted lines can be seen in the figure; the dotted line on the left corresponds to the best λ value of the evaluation index, that is, λ_{\min} ; the larger the c-index value, the better, and the smaller the deviance value, the better. The dotted line on the right represents the λ value of the model whose evaluation index is within 1 standard error of the optimal value, namely λ_{1se}

was associated with the highest survival rate after 5 years. The former may be oncogenes, while the latter may be potential tumor suppressor genes.

PPI analysis results of DEGs

The PPI network of 88 survival-related DEGs was constructed using the String online tool (Figure 7), and Figure 8 was obtained after removing unrelated genes.

Survival analysis and modeling

Then, based on the 11 DEGs and coefficients that can be used as prognostic indicators obtained by LASSO-Cox analysis (Table I), these potential prognostic indicators were further studied. A prognostic risk model was constructed: Risk score = $DSN1^* (-0.0006) + ERI1^* (-0.0035) + ETF1^* (-0.0005) + GLMN^* (-0.0124) + LRR-C8A^* 0.0015 + MAPKAPK3^* (-0.0029) + NOP14^*$

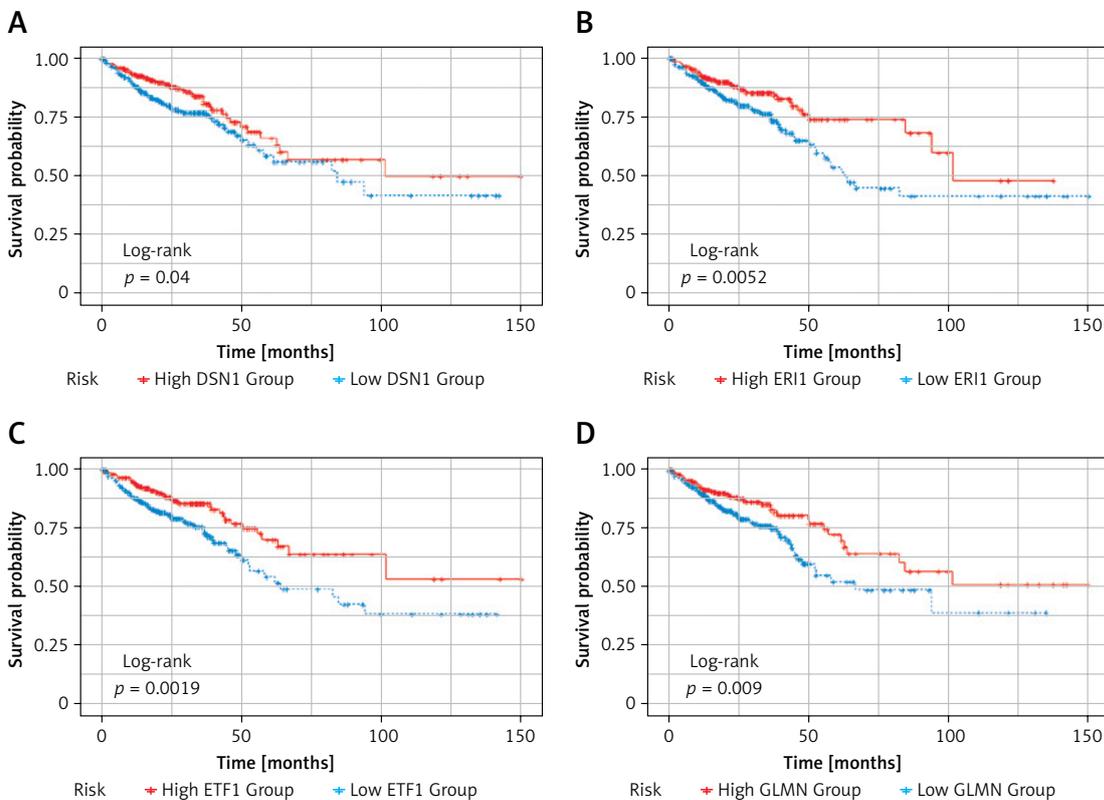


Figure 6. Kaplan-Meier OS curves. Kaplan-Meier OS curves of 11 DEGs. **A** – Kaplan-Meier OS curves of DSN1. **B** – Kaplan-Meier OS curves of ERI1. **C** – Kaplan-Meier OS curves of ETF1. **D** – Kaplan-Meier OS curves of GLMN

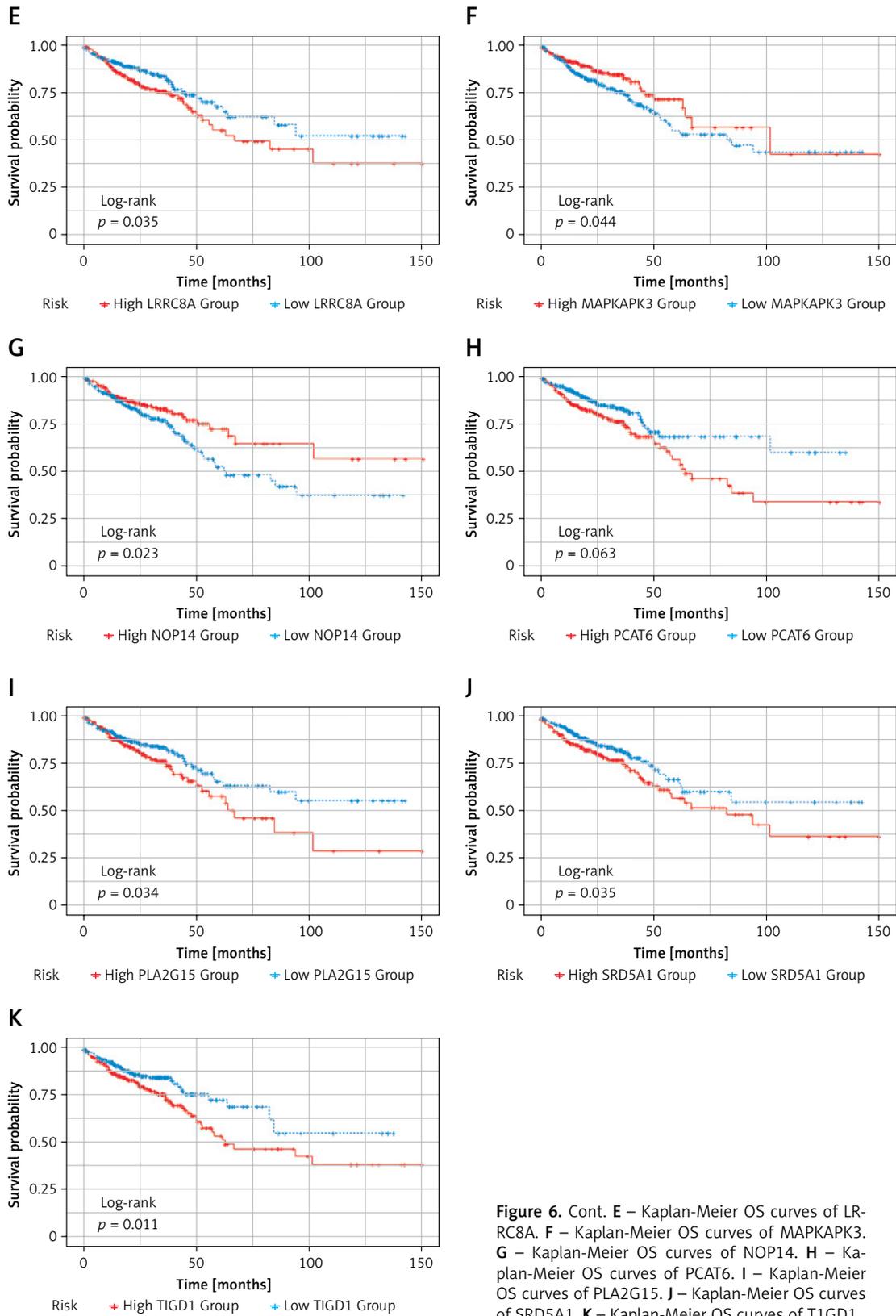


Figure 6. Cont. **E** – Kaplan-Meier OS curves of LRRC8A. **F** – Kaplan-Meier OS curves of MAPKAPK3. **G** – Kaplan-Meier OS curves of NOP14. **H** – Kaplan-Meier OS curves of PCAT6. **I** – Kaplan-Meier OS curves of PLA2G15. **J** – Kaplan-Meier OS curves of SRD5A1. **K** – Kaplan-Meier OS curves of TIGD1

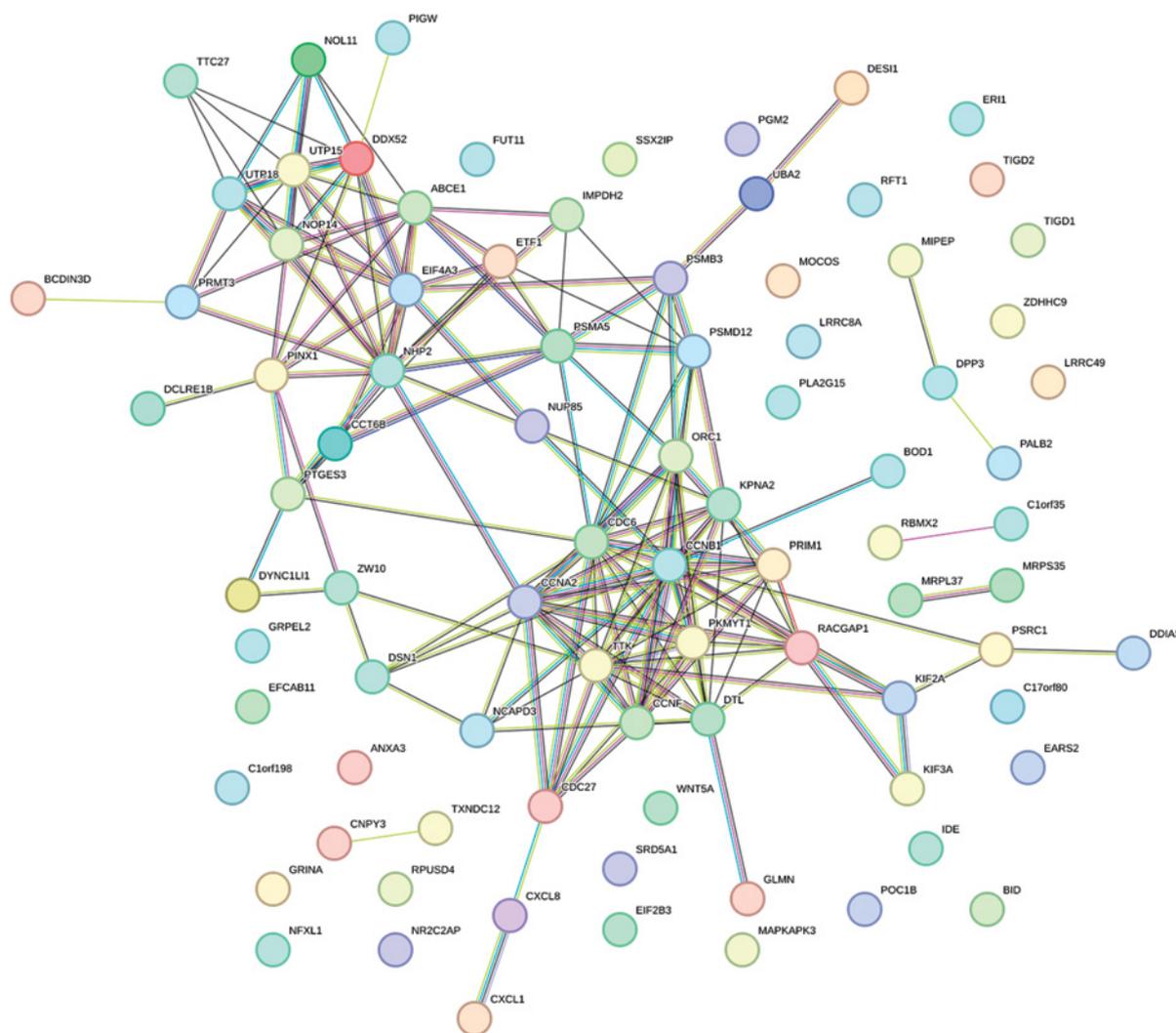


Figure 7. PPI plot of DEG. Differential protein interaction network diagram of up-regulated genes

(-0.0003) + PCAT6*0.0020 + PLA2G15*0.0049 + SRD5A1*0.0015 + TIGD1*0.0231. We multiplied the gene expression by the coefficient to get the risk score, and then predicted the patient's survival according to the score, and obtained the AUC value of the model result, that is, the value of each sample.

Receiver operating characteristic curve analysis

A higher AUC value indicates a better classification performance of the CRC prognostic model for the gene. The model prognostic ability was assessed using the TCGA dataset, and the model was validated using GSE72968. The performance of the model on the training set (TCGA): The AUCs for predicting 1-year, 3-year and 5-year survival are: 0.577, 0.569 and 0.633, respectively. Model performance on the validation set (GSE72968): The AUCs for predicting 1-year, 3-year and 5-year

survival were: 0.509, 0.603 and 0.631, respectively. The results show that the most obvious prognostic effect is in 5 years. Comparing the AUC values of survival in each stage of the validation set and the training set, the difference between the two AUC values in 5 years is the smallest, and the result is more accurate. Finally, the model was validated. Figure 9 shows the ROC curves of the training set TCGA and the validation set GSE72968 sample data. The ROC plot is a graph reflecting the relationship between sensitivity and specificity (Figure 9).

Discussion

The morbidity and mortality of CRC remain high despite advances in surgical and medical treatment of CRC [37]. Successful screening techniques and reducing the risk of CRC are critical to helping reduce CRC mortality. Understanding the etiology and mechanisms of colorectal cancer pro-

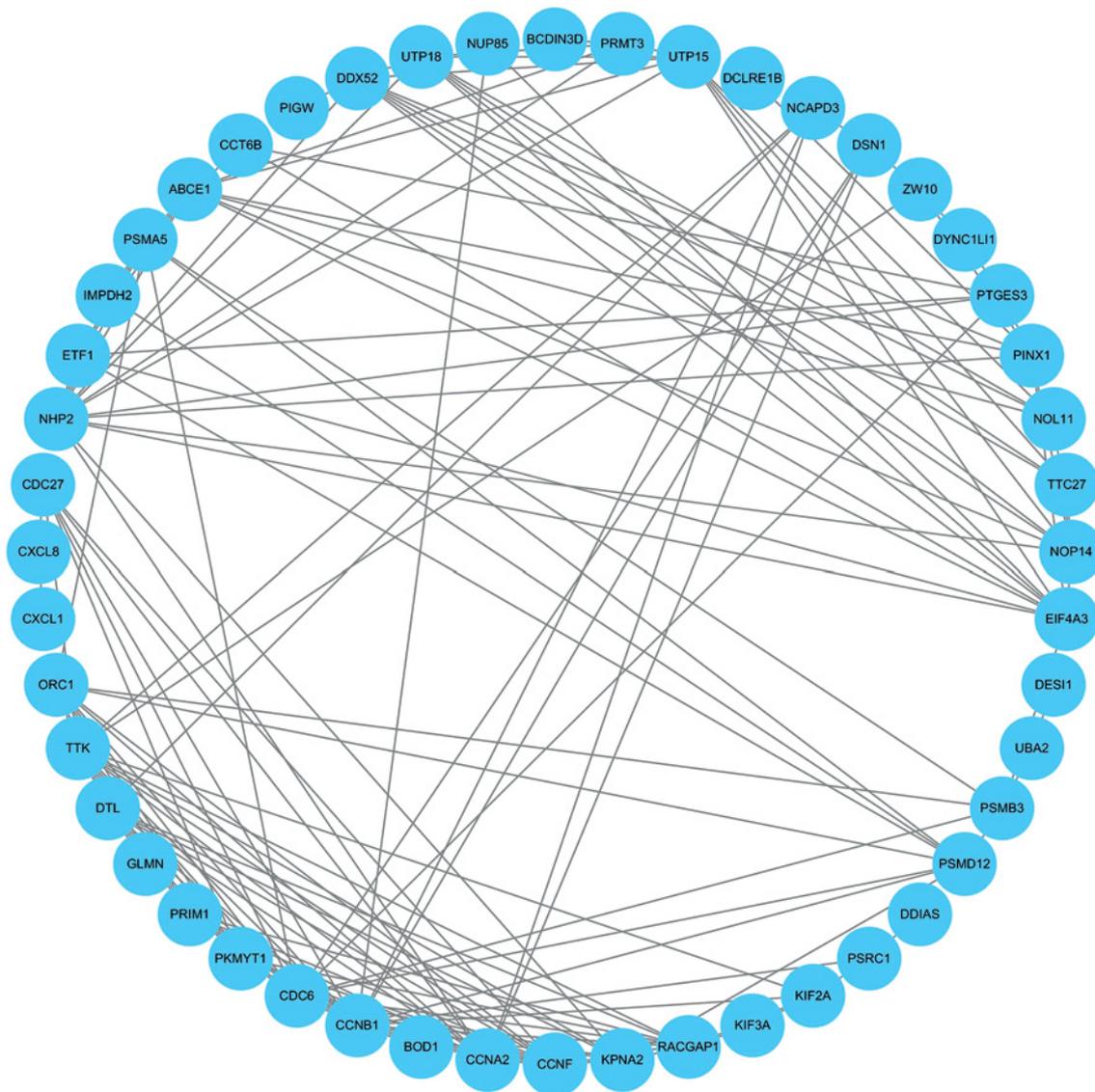


Figure 8. PPI plot of DEG after selection. The differential protein network diagram of up-regulated genes that can form an interaction relationship; the single differential protein that cannot form an interaction relationship no longer appears

gression is extremely important for improving the survival rate of colorectal cancer patients. Recently, the rapidly developing microarray technology has been widely used to compare the expression levels of genes and has been used to predict disease progression to make an accurate assessment of prognosis [38].

In this study, two microarray datasets (GSE20916, GSE33133) were analyzed to identify DEGs in CRC tissue samples. A total of 5267 and 4233 DEGs were identified in the two datasets. GSE20916 included 2847 up-regulated genes and 2420 down-regulated genes. GSE33133 identified a total of 4233 differential genes, including 2104 up-regulated genes. The intersection of up-regulated genes in the two datasets was obtained by Venn analysis as 1058 DEGs. Subsequently, GO en-

Table I. Index and 11 DEGs as prognostic indicators

Gene	Index
DSN1	-0.000625538
ERI1	-0.003511305
ETF1	-0.000545896
GLMN	-0.012356453
LRRC8A	0.001503675
MAPKAPK3	-0.002917185
NOP14	-0.000267735
PCAT6	0.001982782
PLA2G15	0.004877215
SRD5A1	0.001475701
TIGD1	0.023072475

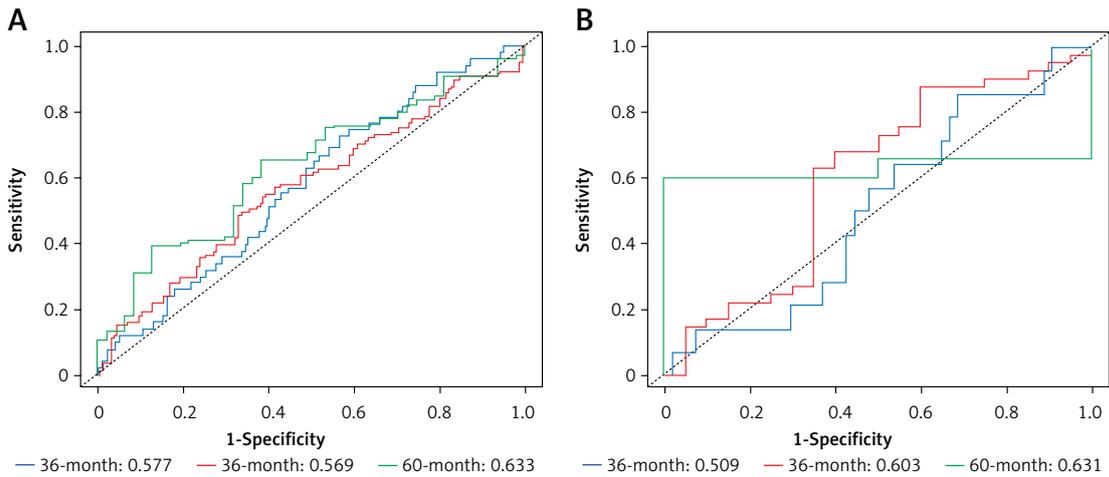


Figure 9. Receiver operating characteristic curve (ROC curve). **A** – ROC curve of ROC sample data with training set TCGA. **B** – ROC curve of the validation set GSE72968 sample data. The X-axis of the abscissa is the specificity, also known as the false positive rate (false positive rate: 1-Specificity); the closer the X-axis is to zero, the higher the accuracy; the Y-axis of the ordinate is called the sensitivity, also known as the true positive rate (Sensitivity: Sensitivity); the larger the Y-axis, the better the accuracy. According to the position of the curve, the whole graph is divided into two parts. The area under the curve (AUC) is used to indicate the prediction accuracy. The higher the AUC value, the larger the area under the curve, indicating higher prediction accuracy. The closer the curve is to the upper left corner (the smaller the X, the larger the Y), the higher the prediction accuracy. The area under the ROC curve, AUC, is normally between 0.5 and 1. In the case of $AUC > 0.5$, the closer the AUC is to 1, the better the diagnostic effect. Note: From Figure 9, it can be concluded that both the training set and the validation set are 60 months, that is, the area under the ROC curve in the 5-year period is the largest, and the AUC value is the highest, which is consistent with the results of the survival analysis and conforms to the above conjecture

richment analysis and pathway analysis were performed. Among the 1058 genes, 992 genes with survival and clinical information in TCGA were screened. Then, survival analysis was performed based on Cox regression and LASSO regression analysis models from 992 genes, and a prognostic risk model was obtained. Eleven DEGs were identified as potential prognostic markers. The results of the model show that the time period with the most obvious prognostic effect is 5 years, and the AUC value is the highest. The verification results of the ROC curve are consistent with the model results of the survival analysis. The survival curve showed that LRR8A, PCAT6, PLA2G15, SRD5A1, T1GD1 may be oncogenes, and DSN1, ERI1, EIT1, GLMN, MAPKAPK, and NOP14 may be tumor suppressor genes. Among them, PLA2G15 and NOP14 have the most significant effect. It can be speculated that DSN1, ERI1, EIT1, GLMN, MAPKAPK, and NOP14 are effective in inhibiting the further development and lethality of CRC, and the most obvious period is 5 years.

NOP14 encodes nucleolar protein, an 875 amino group protein that is highly conserved in eukaryotes and is involved in many biological processes such as DNA repair and replication and cell cycle control. Recent studies support the involvement of NOP14 in cancer initiation and progression. NOP14 expression has been reported in various cancers such as melanoma and breast cancer, bladder cancer and pancreatic cancer [39–41], for

example, NOP14 inhibits the Wnt/ β -catenin signaling pathway, thereby reducing the proliferation and metastasis of melanoma and breast cancer, thus having a tumor suppressor role. Likewise, in bladder cancer, NOP14 attenuated miR-502-5p-mediated inhibition of tumor cell migration and proliferation. Nonetheless, the role of NOP14 depends on the specific type of cancer. In pancreatic cancer, NOP14 is highly expressed and has been shown to promote tumor cell growth and invasion by stabilizing mutant P53, which inhibits P21 expression by inducing miR-17-5p. Apart from that, as a recent analysis reports, increased expression of NOP14 is associated with improved prognosis in CRC [42]. This suggests that it inhibits tumor development in CRC, consistent with the findings of this study. Another study found that the mRNA and protein levels of NOP14 were significantly upregulated in CRC tissues, which suggested that NOP14 could serve as a potential biomarker for CRC. Moreover, its high expression improves the prognosis of CRC patients by inhibiting signaling pathways involved in tumor growth and promoting the immune response [43].

DSN1 is a protein-coding gene that is part of the MIS12 kinetochore complex and is required for proper kinetochore assembly and cell cycle progression [44]. It is critical for malignant transformation associated with chromosome 20q amplification in CRA (colorectal adenomas). Elevated DSN1 expression was positively correlated with chr20 amplification, a high CIN index, and SKA3

overexpression while DSN1 knockdown resulted in slightly increased G2/M phase arrest but did not affect cell growth rate or apoptosis in both CRC cell lines [45]. These results suggest that overexpression of DSN1 contributes to the development of colorectal cancer. In addition, a recent study reported that overexpression of DSN1 is associated with poor survival in patients with colorectal cancer, gastric cancer, liver cancer, and breast cancer, and can be used as a prognostic biomarker for various tumor diseases [46–48].

Lysosomal phospholipase A2 (PLA2G15) is a transacylase that recognizes short-chain ceramides as fatty acyl receptors; it is 50% identical to lecithin cholesterol acyltransferase; these enzymes share close structural homology [49]. Besides, PLA2G15 has broad substrate recognition capacity for a variety of glycerophospholipids, including oxidized phospholipids, it catabolizes pulmonary surfactant and processes lipid antigens for CD1 protein presentation [50]. Moreover, PLA2G15 is a ubiquitous enzyme that is uniquely characterized by its subcellular localization to lysosomes and late endosomes and it is the first and only lysosomal phospholipase A identified to date [51]. Although initially characterized as a member of the PLA2 family, recent work has supported its unique properties as transacylases, clearly demonstrating PLA1 and PLA2 activities. Additionally, according to a recent study, knockdown of NFATC3-PLA2G15 reduces the expression of markers of epithelial-mesenchymal transition (EMT) such as vimentin, twist, and fibronectin in colo-320 cells when studying CRC samples and adjacent normal tissue, mRNA expression, and increased mesenchymal-epithelial transition markers of E-cadherin, claudin-1, and FOXC2. Similarly, the NFATC3-PLA2G15 FT-containing line, NFATC3-PLA2G15 knockdown also inhibited invasion, colony formation capacity, and cell proliferation [52], and NFATC3-PLA2G15 fusion transcript levels were elevated in 19 pairs of CRC tumor and adjacent normal tissue samples [53]. This evidence suggests that NFATC3-PLA2G15 may promote CRC development by enhancing EMT invasion and proliferation, consistent with the results of this study indicating that PLA2G15 might be a considerable prognostic biomarker of CRC.

The results of this study may have clinical implications for the prognosis of CRC. However, the current study has certain limitations. First of all, the AUC value obtained by the survival analysis is not particularly large, between 0.5 and 0.7, and the effect is not particularly significant. Second, among the 11 DEGs screened, the potential roles of some genes are still unknown, and further experimental studies are needed.

In conclusion, 11 DEGs were identified as significantly associated with OS in CRC patients

by performing the above-mentioned analysis of GEO and TCGA. These 11 genes can be used as new independent prognostic biomarkers, among which the most significant ones are NOP14 and PLA2G15, which can be used to predict the clinical outcome of CRC patients. However, further analysis of the role and underlying mechanisms of gene expression in CRC is required to establish its specific role in CRC prognosis.

Funding

No external funding.

Ethical approval

Not applicable.

Conflict of interest

The authors declare no conflict of interest.

References

1. Pourhoseingholi MA, Zali MR. Colorectal cancer screening: time for action in Iran. *World J Gastrointest Oncol* 2012; 4: 82-3.
2. Marley AR, Nan H. Epidemiology of colorectal cancer. *Int J Mol Epidemiol Genet* 2016; 7: 105-14.
3. Voorneveld PW, Jacobs RJ, Kodach LL, et al. A meta-analysis of SMAD4 immunohistochemistry as a prognostic marker in colorectal cancer. *Transl Oncol* 2015; 8: 18-24.
4. Zarkavelis G, Boussios S, Papadaki A, et al. Current and future biomarkers in colorectal cancer. *Ann Gastroenterol* 2017; 30: 613-21.
5. Siegel RL, Miller KD, Jemal A. Cancer statistics, 2019. *CA Cancer J Clin* 2019; 69: 7-34.
6. Lad J, Serra S, Queresy F, et al. Polarimetric biomarkers of peri-tumoral stroma can correlate with 5-year survival in patients with left-sided colorectal cancer. *Sci Rep* 2022; 12: 12652.
7. Huang Y, Zhou J, Zhong H, et al. Identification of a novel lipid metabolism-related gene signature for predicting colorectal cancer survival. *Front Genet* 2022; 13: 989327.
8. Haggard FA, Boushey RP. Colorectal cancer epidemiology: incidence, mortality, survival, and risk factors. *Clin Colon Rectal Surg* 2009; 22: 191-7.
9. Liu H, Li Y, Lv Y, et al. LncRNA AK077216 affects the survival of colorectal adenocarcinoma patients via miR-34a. *Arab J Gastroenterol* 2022; 23: 65-9.
10. Shussman N, Wexner SD. Colorectal polyps and polyposis syndromes. *Gastroenterol Rep* 2014; 2: 1-15.
11. Hibner G, Kimsa-Furdzik M, Francuz T. Relevance of microRNAs as potential diagnostic and prognostic markers in colorectal cancer. *Int J Mol Sci* 2018; 19: 2944.
12. Duan L, Yang W, Wang X, et al. Advances in prognostic markers for colorectal cancer. *Exp Rev Mol Diagnostics* 2019; 19: 313-24.
13. Carlomagno N, Schonauer F, Tammaro V, et al. A multidisciplinary approach to an unusual medical case of locally advanced gastric cancer: a case report. *J Med Case Rep* 2015; 9: 13.
14. Santangelo M, Esposito A, Tammaro V, et al. What indication, morbidity and mortality for central pancreatec-

- tomy in oncological surgery? A systematic review. *Int J Surg* 2016; 28: S172-6.
15. Issa IA, Nouredine M. Colorectal cancer screening: an updated review of the available options. *World J Gastroenterol* 2017; 23: 5086.
 16. Das V, Kalita J, Pal M. Predictive and prognostic biomarkers in colorectal cancer: a systematic review of recent advances and challenges. *Biomed Pharmacother* 2017; 87: 8-19.
 17. Santangelo M, Romano G, Vescio G, et al. Functional results of colorectal and coloanal anastomosis with and without pouch. *Ann Ital Chirur* 2001; 72: 443-8.
 18. FDA-NIH Biomarker Working Group. BEST (Biomarkers, endpoints, and other tools) resource [Internet] 2016.
 19. Califf RM. Biomarker definitions and their applications. *Exp Biol Med* 2018; 243: 213-21.
 20. Hossain MA, Islam SMS, Quinn JMW, et al. Machine learning and bioinformatics models to identify gene expression patterns of ovarian cancer associated with disease progression and mortality. *J Biomed Inform* 2019; 100: 103313.
 21. Horenblas S, Van Tinteren H. Squamous cell carcinoma of the penis. IV. Prognostic factors of survival: analysis of tumor, nodes and metastasis classification system. *J Urol* 1994; 151: 1239-43.
 22. Chua DT, Sham JS, Kwong DL, et al. Volumetric analysis of tumor extent in nasopharyngeal carcinoma and correlation with treatment outcome. *Int J Radiation Oncol Biol Phys* 1997; 39: 711-9.
 23. Rich JT, Neely JG, Paniello RC, et al. A practical guide to understanding Kaplan-Meier curves. *Otolaryngol Head Neck Surg* 2010; 143: 331-6.
 24. Chen L, Lu D, Sun K, et al. Identification of biomarkers associated with diagnosis and prognosis of colorectal cancer patients based on integrated bioinformatics analysis. *Gene* 2019; 692: 119-25.
 25. Zheng Y, Cai T, Feng Z. Application of the time-dependent ROC curves for prognostic accuracy with multiple biomarkers. *Biometrics* 2006; 62: 279-87.
 26. Barrett T, Wilhite S E, Ledoux P, et al. NCB GEO: archive for functional genomics data sets – update. *Nucleic Acids Res* 2012; 41: D991-5.
 27. Cancer Genome Atlas Research Network. Comprehensive genomic characterization defines human glioblastoma genes and core pathways. *Nature* 2008; 455: 1061-8.
 28. Heberle H, Meirelles GV, da Silva FR, et al. InteractiVenn: a web-based tool for the analysis of sets through Venn diagrams. *BMC Bioinformatics* 2015; 16: 169.
 29. Young MD, Wakefield MJ, Smyth GK, et al. Gene ontology analysis for RNA-seq: accounting for selection bias. *Genome Biol* 2010; 11: R14.
 30. Ashburner M, Ball CA, Blake JA, et al. Gene ontology: tool for the unification of biology. *Nat Genet* 2000; 25: 25-9.
 31. Kanehisa M, Goto S. KEGG: kyoto encyclopedia of genes and genomes. *Nucleic Acids Res* 2000; 28: 27-30.
 32. Safari-Alighiarloo N, Taghizadeh M, Rezaei-Tavirani M, et al. Protein-protein interaction networks (PPI) and complex diseases. *Gastroenterol Hepatol Bed Bench* 2014; 7: 17-31.
 33. Mering C, Huynen M, Jaeggi D, et al. STRING: a database of predicted functional associations between proteins. *Nucleic Acids Res* 2003; 31: 258-61.
 34. Royston P, Lambert P C. Flexible parametric survival analysis using Stata: beyond the Cox model. College Station, TX: Stata press 2011.
 35. Shahraki H R, Salehi A, Zare N. Survival prognostic factors of male breast cancer in Southern Iran: a LASO-Cox regression approach. *Asian Pacific J Cancer Prev* 2015; 16: 6773-7.
 36. Meier L, Van De Geer S, Bühlmann P. The group lasso for logistic regression. *J R Statist Soc B* 2008; 70: 53-71.
 37. Bray F, Ferlay J, Soerjomataram I, et al. 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.
 38. Dai GP, Wang LP, Wen YU, et al. Identification of key genes for predicting colorectal cancer prognosis by integrated bioinformatics analysis. *Oncol Lett* 2020; 19: 388-98.
 39. Li J, Fang R, Wang J, Deng L. NOP14 inhibits melanoma proliferation and metastasis by regulating Wnt/beta-catenin signaling pathway. *Braz J Med Biol Res* 2018; 52: e7952.
 40. Lei JJ, Peng RJ, Kuang BH, et al. NOP14 suppresses breast cancer progression by inhibiting NRIP1/Wnt/beta-catenin pathway. *Oncotarget* 2015; 6: 25701-14.
 41. Du Y, Liu Z, You L, et al. Pancreatic cancer progression relies upon mutant p53-induced oncogenic signaling mediated by NOP14. *Cancer Res* 2017; 77: 2661-73.
 42. Lu C, Liao W, Huang Y, et al. Increased expression of NOP14 is associated with improved prognosis due to immune regulation in colorectal cancer. *BMC Gastroenterol* 2022; 22: 207.
 43. Zhu X, Jia W, Yan Y, et al. NOP14 regulates the growth, migration, and invasion of colorectal cancer cells by modulating the NRIP1/GSK-3 β /beta-catenin signaling pathway. *Eur J Histochem* 2021; 65: 3246.
 44. Navarro-Mendoza MI, Pérez-Arques C, Panchal S, et al. Early diverging fungus *Mucor circinelloides* lacks centromeric histone CENP-A and displays a mosaic of point and regional centromeres. *Curr Biol* 2019; 29: 3791-802.
 45. Chuang TP, Wang JY, Jao SW, et al. Over-expression of AURKA, SKA3 and DSN1 contributes to colorectal adenoma to carcinoma progression. *Oncotarget* 2016; 7: 45803.
 46. Li FN, Zhang QY, Li O, et al. ESRR α promotes gastric cancer development by regulating the CDC25C/CDK1/CyclinB1 pathway via DSN1. *Int J Biol Sci* 2021; 17: 1909.
 47. Sun C, Huang S, Ju W, et al. Elevated DSN1 expression is associated with poor survival in patients with hepatocellular carcinoma. *Human Pathol* 2018; 81: 113-20.
 48. Peng Q, Wen T, Liu D, et al. DSN1 is a prognostic biomarker and correlated with clinical characterize in breast cancer. *Int Immunopharmacol* 2021; 101: 107605.
 49. Murakami M, Sato H, Taketomi Y. Updating phospholipase A2 biology. *Biomolecules* 2020; 10: 1457.
 50. Shayman JA, Tesmer JJ. Lysosomal phospholipase A2. *Biochim Biophys Acta Mol Cell Biol Lipids* 2019; 1864: 932-40.
 51. Shayman JA, Tesmer J J G. Lysosomal phospholipase A2. *Biochim Biophys Acta Mol Cell Biol Lipids* 2019; 1864: 932-40.
 52. Jang JE, Kim HP, Han SW, et al. NFATC3-PLA2G15 fusion transcript identified by RNA sequencing promotes tumor invasion and proliferation in colorectal cancer cell lines. *Cancer Res Treat* 2019; 51: 391-401.
 53. Dalal N, Jalandra R, Sharma M, et al. Omics technologies for improved diagnosis and treatment of colorectal cancer: Technical advancement and major perspectives. *Biomed Pharmacother* 2020; 131: 110648.