Heliyon 9 (2023) e20172

Contents lists available at ScienceDirect

Heliyon

journal homepage: www.cell.com/heliyon

The predictive significance of a 5-m6A RNA methylation regulator signature in colorectal cancer

Dan Zhang [a], Guotao Wu [b], Lichao Yang [b], Qiang Wu [b], Lianwen Yuan [b,*]

[a] Clinical Nursing Teaching and Research Section, The Second Xiangya Hospital, Central South University, Changsha, 410011, Hunan, PR China
[b] Department of General Surgery, The Second Xiangya Hospital, Central South University, Changsha, Hunan, China

ARTICLE INFO

Keywords:
Colorectal cancer
Least absolute shrinkage and selection operator (LASSO) regression model
RBM15B
KIAA1429
Prognosis

ABSTRACT

Colorectal cancer attacks the colon or rectum, with increasing morbidity and mortality globally. The RNA modification 6-methyladenine (m6A) is related to RNA modifications, playing a critical role in colorectal cancer. We aimed to identify prognostic signatures for colorectal cancer using risk prediction algorithms, and to validate these signatures using independent datasets and clinical samples. In this study, 175 cases in GSE17536 were assigned into two clusters using consistent clustering and PCA analysis. A multivariate Cox risk regression model revealed that among 21 m6A RNA methylation regulators, RBM15B, FTO, IGF2BP2, ZCCHC4, and KIAA1429 were remarkably associated with colorectal cancer patients' overall survival (OS); however, Kaplan–Meier (KM) survival assessment showed no significant association between these five regulators and colorectal cancer patients' prognosis. A 5-m6A RNA methylation regulator signature was established using LASSO algorithm. Risk scores of cases in GSE17536, GSE17537 and GSE75500 were calculated, and lower risk scores were associated with better DSS/OS. receiver operating characteristic (ROC) curve and the nomogram revealed the satisfactory predictive efficiency of the risk score model. The risk score could distinguish cases in Cluster1 and Cluster2 and normal and tumor tissues based on GSE37182. The prognostic variables for colorectal cancer patients were assessed using both univariate and multivariate Cox's proportional hazard regression models, which revealed that the stage and risk score were significant risk factors. In this study, a comprehensive set of integrative bioinformatics analyses was conducted to investigate the prognostic and diagnostic potential of a panel of 5 m6A RNA methylated regulators in colorectal cancer patients. The conducted studies included the use of several statistical methods, such as the LASSO regression model, KM survival evaluation, ROC curve, and univariate and multivariate Cox's proportional hazard regression analyses. The findings from these analyses collectively established the prognostic marker, highlighting its significance in predicting patient outcomes and diagnosing colorectal cancer.

1. Introduction

Colorectal cancer is a malignancy that attacks the colon or rectum, with increasing morbidity and mortality globally [1,2]. The mortality rate of colorectal cancer is mainly due to postoperative recurrence and metastasis [3]. Despite the advances in therapeutic

* Corresponding author. Department of General Surgery, The Second Xiangya Hospital, Central South University, 139 # the Renmin Road, Changsha, 410011, Hunan, China.
E-mail address: yuanlianwen@csu.edu.cn (L. Yuan).

https://doi.org/10.1016/j.heliyon.2023.e20172
Received 11 January 2023; Received in revised form 3 September 2023; Accepted 13 September 2023
Available online 15 September 2023

strategies over the past few years, including improvements in surgical techniques and adjuvant therapies, the prognosis of colorectal cancer remains impaired [4]. Hence, research on the molecular pathways that underlie the onset and progression of colorectal cancer would have significant importance in terms of early detection and tailored treatment approaches.

Genomic epigenetic modifications, including histone tail modification, RNA modification, and DNA methylation, are essential for the initiation and progression of tumors. The RNA modification 6-methyladenine (m6A) is closely related to RNA modifications, such as mRNAs, tRNAs, snRNAs, and long-chain non-coding RNAs [5]; thus, it is the major RNA modifications. The reversible and dynamic regulation of m6A marks on mRNAs is analogous to the regulatory mechanisms seen in DNA methylation and histone modifications. This regulation is achieved via the actions of certain enzymes known as "writers" (methyltransferases), "readers" (binding proteins), and "erasers" (demethylases). The group of methyltransferases that play a significant role in the methylation process are referred to as writers. Notable members of this group include METTL3, METTL14, METTL16, WTAP, KIAA1429, RBM15, RBM15B, and ZC3H13, which catalyze the adenylate mRNA m6A modification [6–10]. Critical binding proteins (readers) consist of YTHDF1/2/3, YTHDC1/2, EIF3A, HNRNPC, and HNRNPA2B1, which decode m6A mark and mediate the downstream effects on post-transcriptional regulation [5,11–13]. Crucial demethylases (erasers) include FTO (fat mass and obesity-associated protein) and ALKBH5, which contribute to removing the m6A methyl group [14,15]. In recent years, increasing evidence showed that m6A modification might be associated with the ability of cancer stem cells to self-renew, tumor cell growth, and chemo- and radio-resistance [16,17] in cervical carcinoma, prostate carcinoma, breast carcinoma, pancreatic carcinoma, hepatic carcinoma, and acute myeloid leukemia [18–20]. The prognostic and diagnostic potential of m6A regulators in colorectal cancer is worth investigating.

The use of gene chips and high-throughput sequencing has led to the discovery that mRNA gene signatures have a significant correlation with the overall survival (OS) of individuals diagnosed with colorectal cancer. The regression approach known as the Least Absolute Shrinkage and Selection Operator (LASSO) was exploited to identify the most significant predictive characteristics within the training dataset [14]. This algorithm is specifically designed for high-dimensional data.

In this study, we obtained four datasets, namely GSE17536, GSE5500, GSE17537, and GSE37182, from the Gene Expression Omnibus (GEO) database (https://www.ncbi.nlm.nih.gov/geo/). A thorough study was undertaken in order to investigate the correlation between 21 m6A RNA regulators and clinicopathological features. Furthermore, we have identified the key regulators whose levels of expression exhibited a strong correlation with the OS of individuals who have been diagnosed with colorectal cancer. The LASSO Cox regression model is used to generate a prognostic signature consisting of five genes, with the aim of investigating the potential utility of m6A changed regulatory variables in the context of colorectal cancer.

2. Materials and methods

2.1. Data acquisition and m6A RNA methylation regulator selection

Four data sets [GSE17536 (n = 175), GSE17537 (n = 54), GSE75500 (n = 114), and GSE37182(n = 172)] were obtained from the GEO database (https://www.ncbi.nlm.nih.gov/geo/), including three sets of chip data with prognostic information and one set of chip data with normal control. In order to mitigate potential statistical bias, our research eliminated colorectal cancer patients who had missing survival data, as well as those with OS values less than 30 days. The training set for this study consisted of GSE17536, which had a total of 175 samples. The remaining three datasets were used as validation sets. The expression matrix was extracted for a total of 21 m6A RNA methylation regulators, which include ZCCHC4, METTL3, METTL14, WTAP, KIAA1429, RBM15, RBM15B, ZC3H13, EIF3A, YTHDC1, YTHDC2, YTHDF1, YTHDF2, YTHDF3, FTO, ALKBH5, HNRNPC, HNRNPA2B1, IGF2BP1, IGF2BP2, and IGF2BP3. Additionally, the associated clinical information for the samples was also extracted.

2.2. Consensus clustering analysis

Exploiting GSE17536, two distinct subgroups, referred to as cluster1 and cluster2, were found using the "Consensus Cluster Plus" package available at http://www.bioconductor.org/. The identification process included using a resampling rate of 80%, doing 50 iterations, and utilizing Pearson correlation as the measure of association. Principal component analysis (PCA) was used to evaluate the gene expression patterns in the two subgroups of colorectal cancer. This analysis was conducted using the R package for R version 3.6.0.

2.3. Integrative bioinformatics analyses

A multivariate Cox risk regression analysis (false discovery rate, FDR<0.05) was performed to analyze the correlation of the 21 regulators with the prognosis in 175 patients in GSE17536. The KM survival assessment was subsequently conducted to analyze the correlation between RBM15B, FTO, IGF2BP2, ZCCHC4, and KIAA1429 and the prognosis in colorectal cancer patients.

The coefficients of RBM15B, FTO, IGF2BP2, ZCCHC4, and KIAA1429 genes were determined by the LASSO algorithm. The LASSO package in R was used to calculate the risk score for each patient's prognostic signature. The formula was as follows: Risk score = $\sum_{i=1}^{n} Expr(Genei) \times Coef\ (Genei)$, where *Coef* (*Genei*) was the coefficient of genes correlated with colorectal cancer survival, and *Expr* (*Genei*) was the expression of genes [21,22]. For the present study, the formula is: Risk score = *Expr*(*RBM15B*) × -36.24446952304233 + *Expr*(*FTO*) × -6.824543600830022 + *Expr(IGF2BP2)* × 7.1102124910427 + *Expr*(*ZCCHC4*) × 10.96525589158451 + *Expr* (*KIAA1429*) × 23.345121748450254.

2.4. Prognostic potential of the 5-gene risk score model

The risk scores of cases in GSE17536GSE17537, GSE75500 and GSE37182, were calculated following the formula mentioned above. The median risk score was used to classify cases in GSE17536, GSE17537, and GSE75500 into high and low risk categories. The KM survival curves were established for the purpose of evaluating the risk score model's predictive ability.

The prediction effectiveness of the risk score model in GSE17536, GSE17537, and GSE75500 was evaluated using the receiver operating characteristic (ROC) curve [23]. The study included both univariate and multivariate Cox's proportional hazard regression models to examine the prognostic risk variables among patients with colorectal cancer.

Based on the cases with prognostic information and clinical characteristics from GSE17536 and GSE17537, we used the R package rms and regplot (https://cran.r-project.org/package=rms, https://cran.r-project.org/package=regplot) to generate a nomogram and calibration curves. The calculation of the consistency index (C index) was performed using the survcomp program [23] to evaluate the discriminatory capacity of the model, namely its ability to differentiate between patients who survived and those who did not. Subsequently, calibration curves were generated for the nomogram, specifically for the 1-, 3-, and 5-year overall survival (OS) periods. These curves were used to assess the precision of the anticipated survival probabilities in relation to the observed rates.

2.5. Clinical sample collection

Colorectal cancer patients having received surgical resection in the Second Xiangya Hospital were selected. 10 cases of colon cancer and tumor-adjacent (used as normal control) were collected (age 64.8 + 10.3 years, F/M 2/8). Samples were fixed in formalin immediately after sampling until further experimental analysis. The sample collection procedure was approved by the Ethic Committee of Second Xiangya Hospital (approval number: 2021jj40844) and conducted after all the patients signed the informed consent.

2.6. Immunohistochemical staining (IHC staining)

The tumor tissues were subjected to fixation and afterwards fixed in paraffin prior to being sliced into sections with a thickness of 4 μm. These sections were then baked at 60 °C for an hour, dewaxed in xylene, and hydrated in a series of ethanol solutions. The process of antigen retrieval was conducted using citrate buffer in a microwave, while the activity of endogenous peroxidase was inhibited via the use of hydrogen peroxide. The sections were subsequently blocked using goat serum and left to incubate overnight with primary antibodies targeting RBM15B, FTO, IGF2BP2, ZCCHC4, and KIAA1429 (obtained from Proteintech, Wuhan, China and CUSABIO, Wuhan, China) at a temperature of 4 °C. Following the washing step, the sections were subjected to incubation with the secondary antibody labeled with horseradish peroxidase (HRP). Subsequently, another round of washing was performed, and the sections were then treated with a diaminobenzidine (DAB) solution obtained from Boster, located in Wuhan, China. The counterstaining process included the use of hematoxylin. Finally, the sections were dehydrated, transparent in xylene, and sealed with neutral resin before observation and analysis. The ImageJ software (NIH, USA) was used for determining the average optical density (AOD).

2.7. Statistical analyses

SPSS (version 22.0) and the R programming language (version 3.5.1; https://www.r-project.org/) were exploited for the statistics. A p-value of less than 0.05 is generally regarded as statistically significant. Remove the clinical data that is absent from the list and exclude the whole sample from the analysis if any parameter value is missing. The term "overall survival" (OS) refers to the duration of time between the first diagnosis of a medical condition and the occurrence of death. The t-test is used to conduct a mean value comparison of continuous variables. Kaplan-Meier estimates were used to compare high-risk and low-risk groups' survival rates using a two-tailed log-rank test in the programming language R.

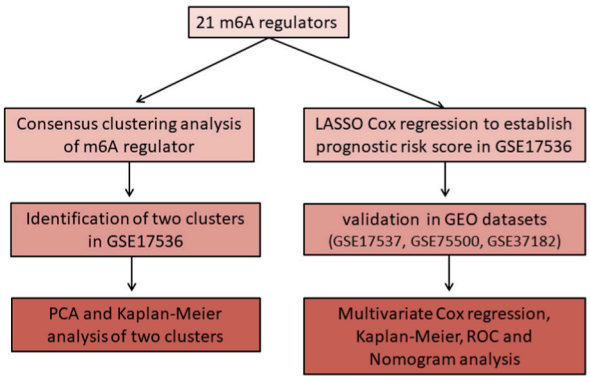

Fig. 1. Schematical diagram illustrating the process of bioinformatics analyses.

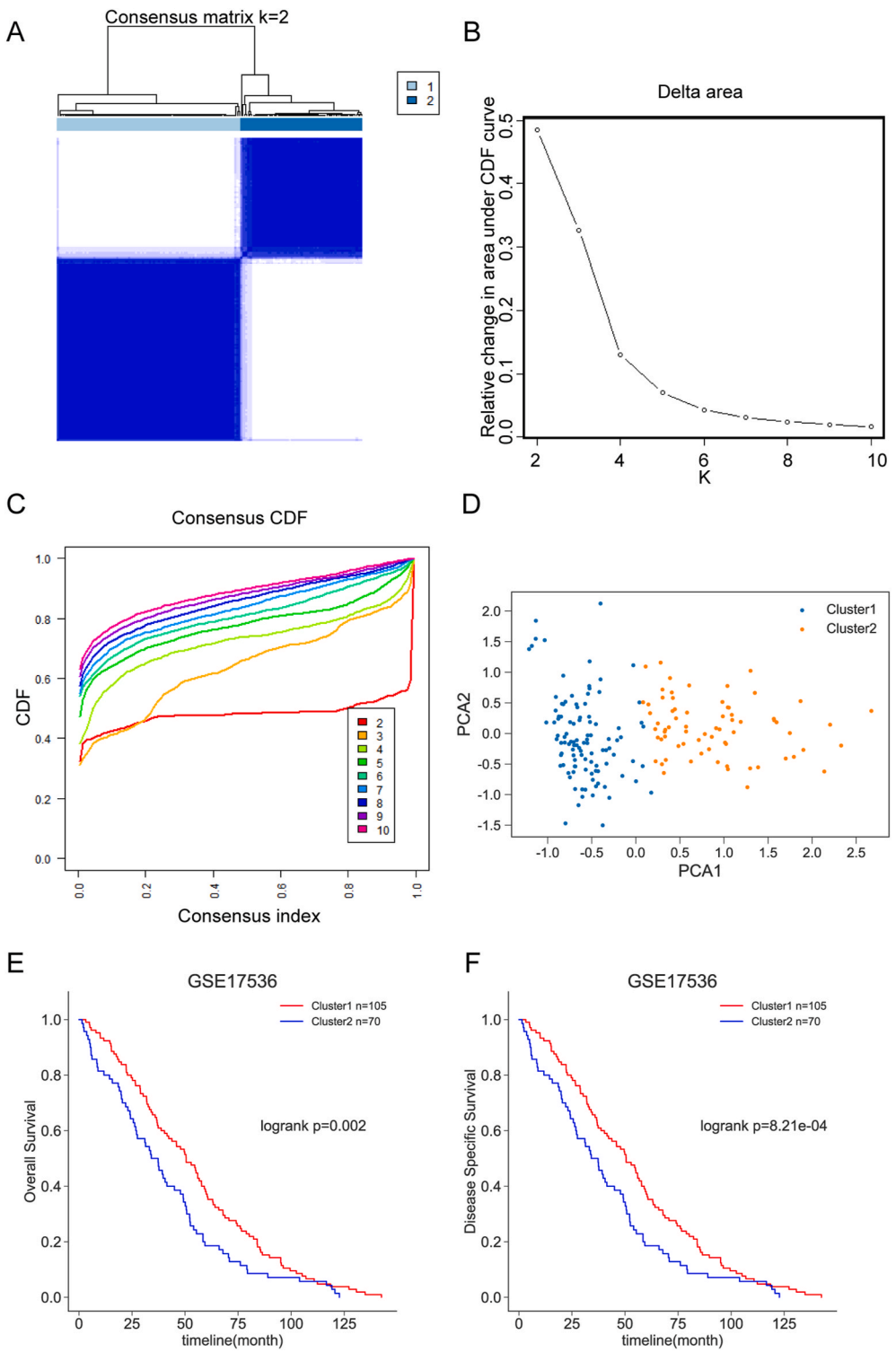

Fig. 2. Consistent cluster analysis of colorectal cancer (A) The correlation between groups when k equals 2. (B) Cumulative distribution function (CDF) Delta area. The relative change in the area under the CDF curve for each category number k compared with k-1. (C) The CDF curve from 2 to 10 of k. (D) The distribution of the sample when k is between 2 and 10. (E–F) The Kaplan–Meier (KM) overall survival curves were established to analyze the overall survival (OS) and disease specific survival (DSS) in GSE17536 cases in cluster1 and cluster2.

3. Results

3.1. Data clustering

Fig. 1 illustrates the comprehensive methodology used for bioinformatics analysis. Consistent clustering and principal component analysis (PCA) were conducted on the GSE17536 dataset, which comprises expression profile data and clinical information of 175 colorectal cancer patients who had an overall survival of at least 30 days. The dataset also includes information on 21 m6A RNA methylation regulators (Fig. 2A–D). The study demonstrates the relative alteration in the cumulative distribution function (CDF) of the consensus cluster while transitioning from k = 2 to 10. Additionally, it examines the area under the CDF curve for the same range of k values. The results indicate that k = 2 is the optimal option for dividing the colon cancer patient cohort into two clusters, as seen in Fig. 2A–C. The classification of colorectal cancer patients into two distinct groupings, Cluster1 (N = 105) and Cluster2 (N = 70), may be substantially determined by considering 21 specific criteria (Fig. 2D). Kaplan-Meier survival analysis using data from GSE17536 demonstrated a significantly better overall survival (OS; $P < 0.05$; Fig. 2E) and disease-specific survival (DSS; $P < 0.05$; Fig. 2F) in patients in Cluster1. ().

A multivariate Cox risk regression model analyzing the prognostic potential of the 21 m6A RNA methylation regulators.

A multivariate Cox risk regression analysis was performed to analyze the correlation of the 21 regulators with the prognosis in 175 patients in GSE17536; Fig. 3 shows that RBM15B, ZCCHC4, KIAA1429, IGF2BP2 and FTO were remarkably correlated with the OS in colorectal cancer patients. KM survival assessment was subsequently conducted to analyze the correlation between the prognosis and these five factors respectively in colorectal cancer patients. However, Fig. 4A–E shows that these five factors alone could not predict the OS in colorectal cancer patients. Thus, it is necessary to perform integrative bioinformatics analyses to establish a multi-genes prognostic model.

3.2. m6A RNA methylated regulatory-based risk model established using the LASSO algorithm

RBM15B, FTO, IGF2BP2, ZCCHC4, and KIAA1429 coefficients were computed using the LASSO method. The coefficient of each of the 5 genes that form the signature was shown in Fig. 5A and B and the formula for the risk scores is as follows: Risk score = *Expr*(*RBM15B*) × -36.24446952304233 + *Expr*(*FTO*) × -6.824543600830022 + *Expr*(*IGF2BP2*) × 7.1102124910427 + *Expr*(*ZCCHC4*) × 10.96525589158451 + *Expr*(*KIAA1429*) × 23.345121748450254.

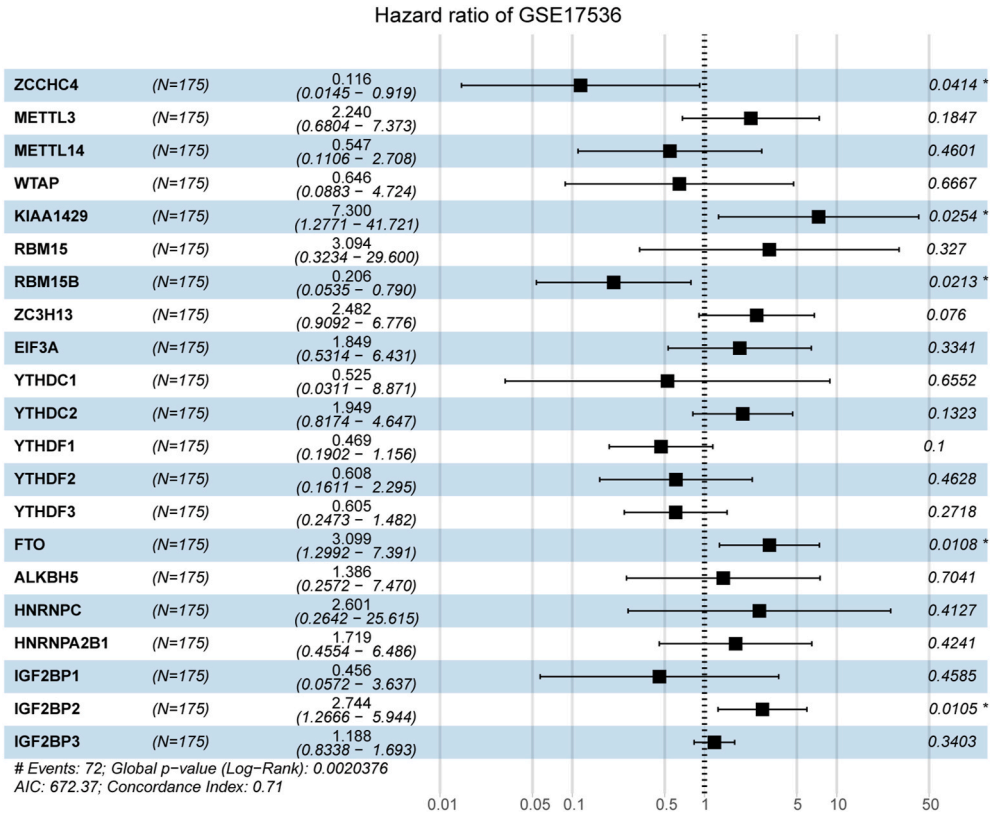

Fig. 3. Multivariate Cox risk regression analysis on the correlation of 21 m6A RNA methylation regulators with the prognosis in 175 colorectal cancer cases according to GSE17536.

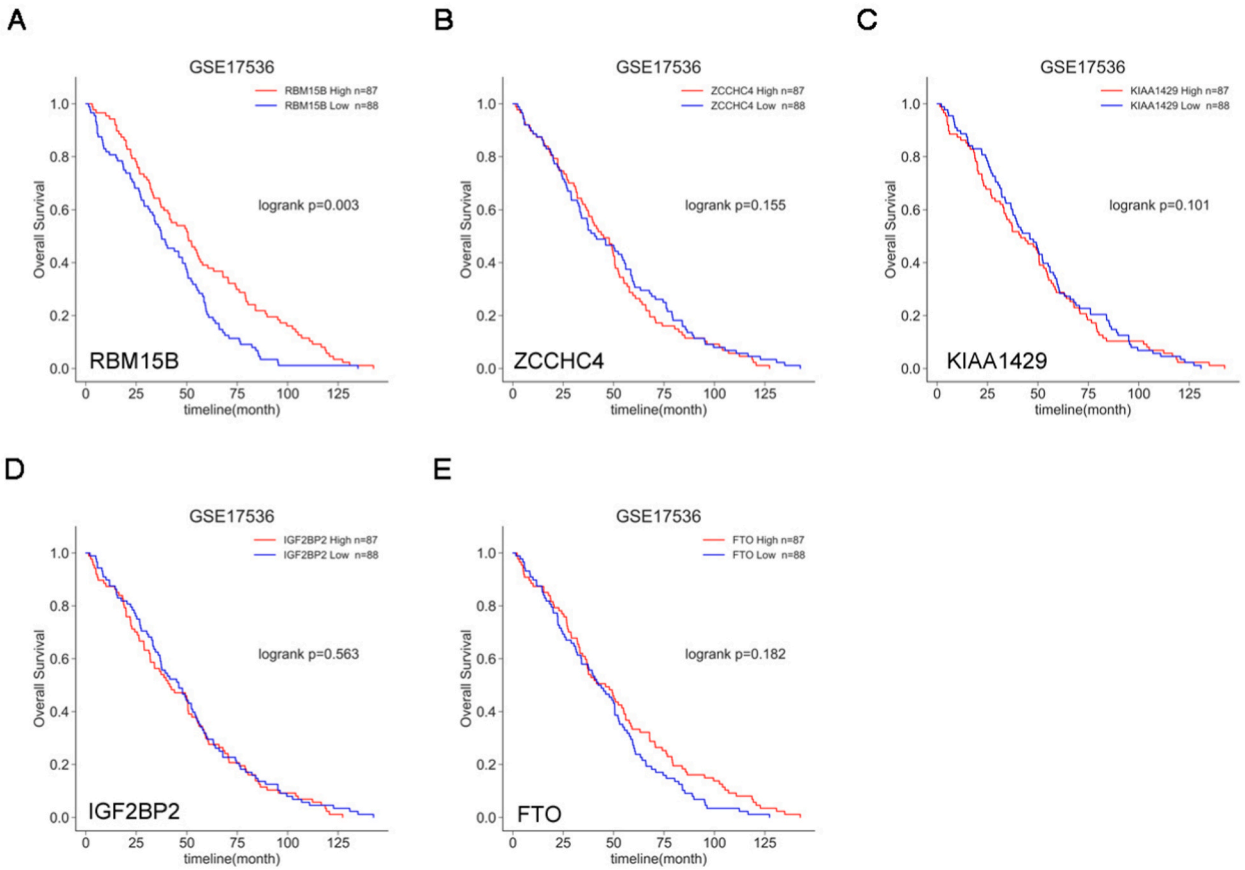

Fig. 4. Survival analysis on colorectal cancer patients based on m6A RNA methylation regulator expression The assignment of cases to high- and low-expression groups was conducted based on the median expression value of RBM15B (A), ZCCHC4 (B), KIAA1429 (C) IGF2BP2 (D), and FTO (E), respectively, as the cut-off. The KM survival curves were established to analyze the prognostic value of these five genes on 175 colorectal cancer cases in GSE17536.

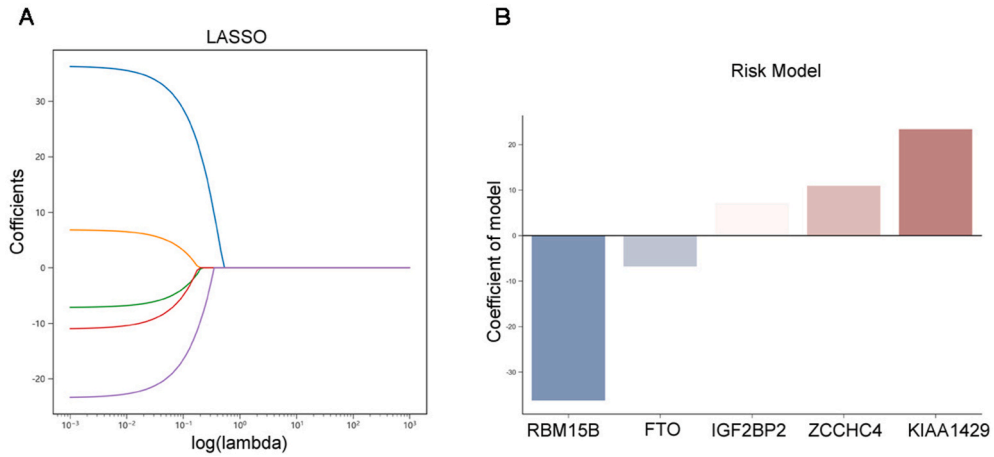

Fig. 5. A risk model based on m6A methylation regulators, which was created using the Least Absolute Shrinkage and Selection Operator (LASSO) method (A–B) The LASSO algorithm was performed to calculate the coefficients of RBM15B, FTO, IGF2BP2, ZCCHC4, and KIAA1429. The coefficient of each of the 5 genes that form the signature was shown.

The prognostic potential of the risk model was subsequently validated. The risk scores of cases in GSE17536 (n = 175, Fig. 6A), GSE17537 (n = 54, Fig. 6B), and GSE75500 (n = 114, Fig. 6C) were calculated following the formula mentioned above. Cases in GSE17536, GSE17537, and GSE75500 were allocated into high- and low-risk score groups taking the median value of the risk score as the cut-off. To evaluate the risk score model's predictive ability, Kaplan-Meier overall survival curves were constructed. As shown in Fig. 6A-C, a lower risk score was remarkably correlated with better OS/DSS.

Secondly, the ROC curve [24] was exploited to test the prediction efficiency of the risk score model. As shown in Fig. 7A-C, the risk model for the training set GSE17536 demonstrates a 10-year OS prediction performance ranging from 0.59 to 0.81. The model demonstrates a 5-year OS prediction performance of 0.57–0.75 for GSE17537 and a 5-year DFS prediction performance of 0.52–0.66 for GSE75500. Additionally, the model achieves a 5-year DFS prediction performance of 0.55–0.76 for GSE75500. Moreover, clinical correlation analysis indicated that cases in Cluster1, which had a better prognosis, obtained significantly lower risk scores ($P < 0.001$, Fig. 8A), whereas colorectal cancer cases in GSE37182 obtained higher risk scores than normal controls ($P < 0.001$, Fig. 8B).

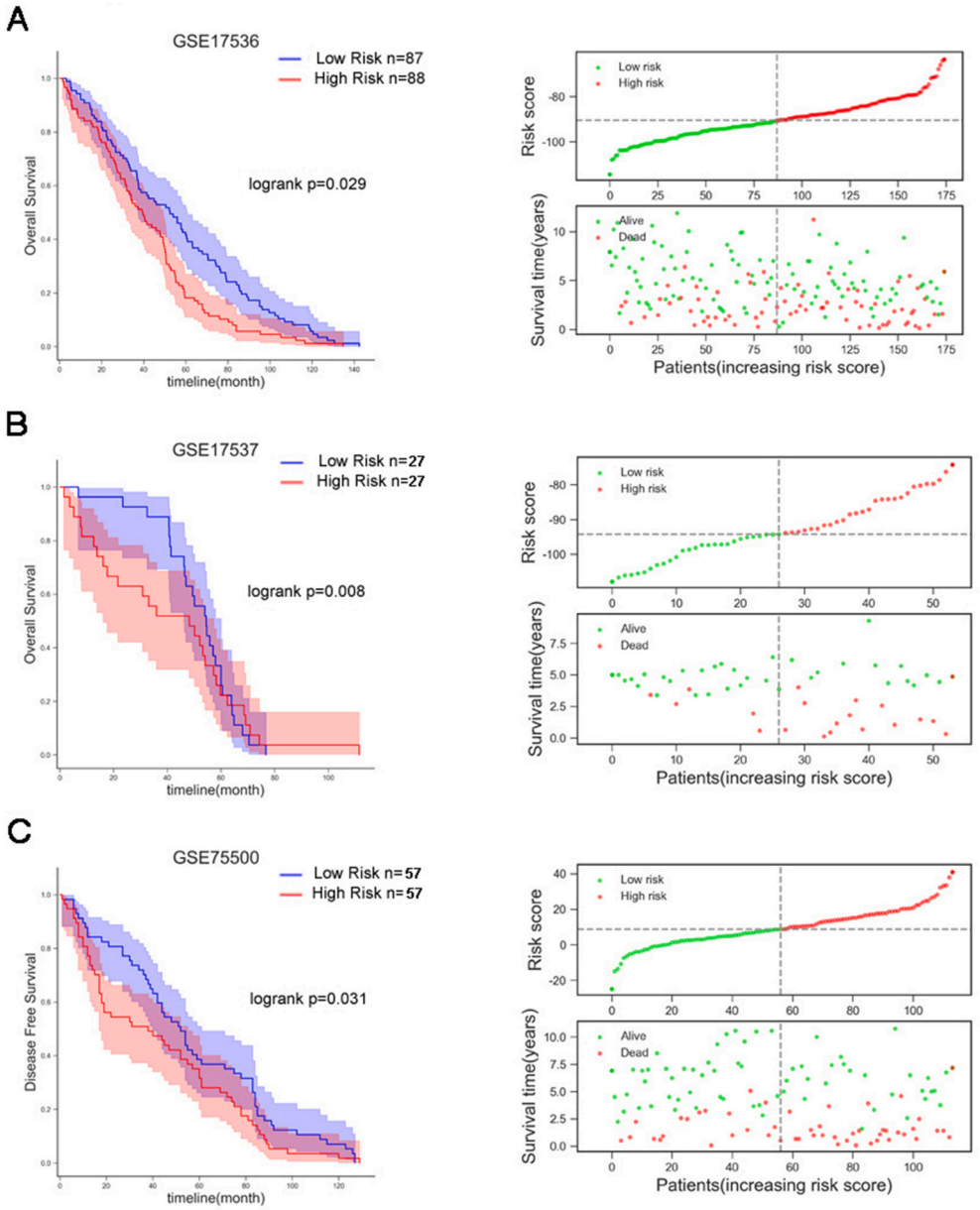

Fig. 6. Correlation of risk scores with the prognosis in patients with colorectal cancer The risk scores of cases in GSE17536 (n = 175) (A), GSE17537 (n = 54) (B), and GSE75500 (n = 114) (C) were calculated following the formula mentioned in the M&M section. Cases in GSE17536, GSE17537, and GSE75500 were assigned into high- and low-risk score groups using the median risk score as the cut-off. The prognostic significance of the risk score model was assessed using the KM survival curves.

Thirdly, the study included both univariate and multivariate Cox's proportional hazard regression models to examine the prognostic risk variables among patients with colorectal cancer. The study examined a series of clinical factors (age, gender, and stage), and risk score. According to the data shown in Table 1 and Fig. 9A-B, both the stage of the disease and the risk score have predictive capabilities for patients' prognosis. Moreover, it is evident that both the stage and risk score may be considered as risk factors.

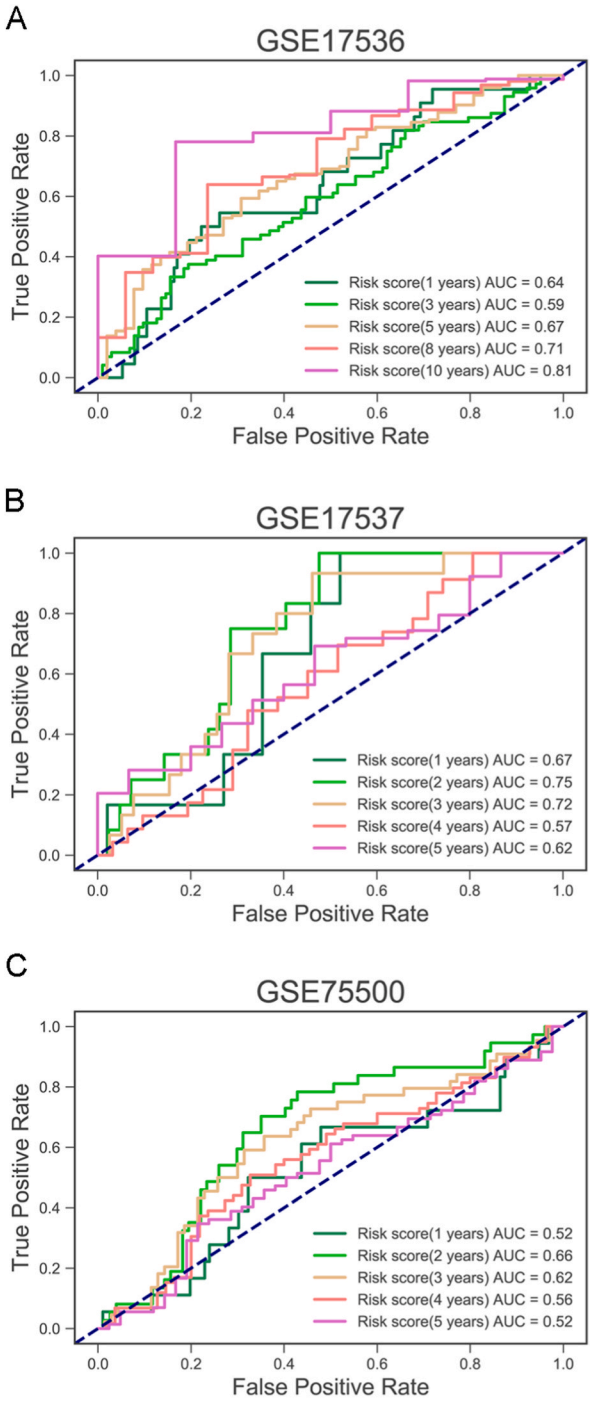

Fig. 7. Model classification performance verification The predicted accuracy of the 1-, 2-, 3-, 4-, and 5-year ROC curve was assessed by using the survival ROC package in the R programming language based on data from GSE17536 (n = 175) (A), GSE17537 (n = 54) (B), and GSE75500 (n = 114) (C).

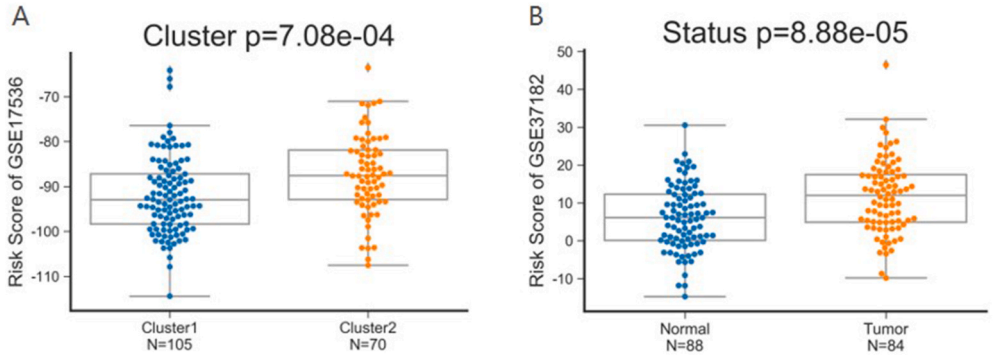

Fig. 8. The diagnostic potential of the risk score (A) The risk score of cases in cluster1 and cluster2 of GSE17536 (n = 175). (B) The risk score of cases in normal colon tissues and tumor tissues in GSE37182 (n = 172).

3.3. Validation of the nomogram in the prediction of patients' prognosis based on GEO datasets

Based on the cases with prognostic information from GSE17536, we established a prognostic nomogram that predicted the 1-, 3-, and 5-year survival chance (Fig. 10A–D). Age, gender, stage, and risk score were parameters included in the nomogram. The C-index of the risk score was 0.76, indicating that the prognostic outcomes of the model are satisfactory. As a further confirmation, similar analyses were performed on GSE17537. The C-index of the risk score was 0.82 for GSE17537 (Fig. 11A–D), indicating that the Nomogram using the model risk score is reliable and accurate.

3.4. Expression validation of the factors using dataset and clinical samples

Lastly, the expression levels of RBM15B, FTO, IGF2BP2, ZCCHC4, and KIAA1429 were validated in normal colon tissue and cancer samples according to GSE37182. Fig. 12A demonstrates a substantial upregulation of the expression of all five m6A regulators in tumor tissues. Furthermore, a total of 10 samples of cancerous tissue and their corresponding surrounding tissues were harvested and the expression levels of RBM15B, FTO, IGF2BP2, ZCCHC4, and KIAA1429 were examined in tissue samples using IHC staining. Fig. 12B shows that the protein levels of RBM15B, FTO, IGF2BP2, ZCCHC4, and KIAA1429 were found to be considerably elevated in cancer samples as compared to the corresponding tumor-adjacent tissues.

4. Discussion

In this study, 175 cases in GSE17536 were assigned into two clusters using consistent clustering and PCA analysis. A multivariate Cox regression analysis was conducted to examine the association between 21 m6A RNA methylation regulators and OS in patients with colorectal cancer. The results revealed that RBM15B, FTO, IGF2BP2, ZCCHC4, and KIAA1429 exhibited significant associations with OS. However, KM survival assessment showed no significant association between these five regulators and colorectal cancer patients' prognosis. A 5-m6A RNA methylation regulator signature was established using LASSO algorithm. Risk scores of cases in GSE17536, GSE17537, and GSE75500 were calculated, and lower risk scores were associated with better DSS/OS. The predictive potential of the risk score model was proved to be favorable by both the ROC curve and the nomogram. Both univariate and multivariate Cox's proportional hazard regression models were exploited to ascertain the significance of the stage and risk score variables as risk factors influencing the prognosis of patients diagnosed with colorectal cancer. Furthermore, the expression levels of RBM15B, FTO, IGF2BP2, ZCCHC4, and KIAA1429 exhibited a significant increase in cancer samples when compared to the normal control samples, as shown by both dataset analysis and clinical sample research. In the context of colorectal cancer, the independent prognostic importance of the risk score obtained from the evaluation of 5-m6A regulators was seen.

Table 1
The Univariate and multivariate Cox risk regression analyses on the correlation between clinical parameters and colorectal cancer patients' prognosis.

Dataset	Factor	Univariate		Multivariate	
		HR(95%CI)	p.value	HR(95%CI)	p.value
GSE17536	Age	1(0.99–1)	0.45	1.02(1.00–1.04)	0.0736
	Gender	1.1(0.71–1.8)	0.58	1.15(0.70–1.89)	0.5736
	Stage	0.24(0.14–0.43)	***9.80E-07***	0.23(0.13–0.41)	***7.61E-07***
	Risk_score	1(1–1.1)	***0.0058***	1.03(1.00–1.05)	***0.0286***
GSE17537	Age	1(0.97–1)	0.84	1.02(0.98–1.06)	0.2945
	Gender	0.59(0.23–1.5)	0.27	0.87(0.32–2.37)	0.7794
	Stage	0.3(0.09–1)	***0.053***	0.25(0.07–0.92)	***0.037***
	Risk_score	1(0.99–1.1)	***0.08***	1.06(1.00–1.12)	***0.0477***

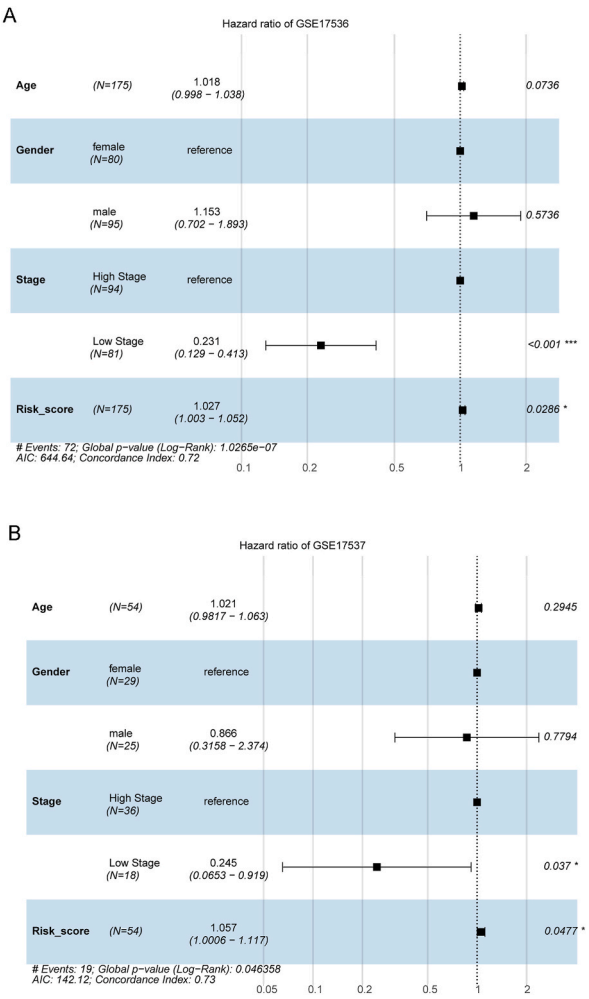

Fig. 9. The present study used univariate and multivariate Cox regression models to examine the association of age, gender, stage, and risk score with prognosis in patients diagnosed with colorectal cancer according to GSE17536 (n = 175) and GSE17537 (n = 54).

The association between m6A RNA alteration and several aspects of tumor development, including proliferation, differentiation, incidence, invasion, and metastasis, has been shown. The m6A RNA alteration has a dual role as both oncogenes and antioncogenes in the context of malignancies. Liu et al. [25] analyzed colonic adenocarcinoma cases within TCGA datasets; accordingly, regulatory factors related to m6A RNA methylation in colonic adenocarcinoma were dramatically altered within tumor tissue samples, indicating the critical effect of m6A modification upon colorectal adenocarcinoma. The dysregulation of m6A modification in mRNAs and noncoding RNAs (ncRNAs) seen in colorectal cancer tissues plays a critical role in the initiation, advancement, invasion, and distant metastasis of cancer. Moreover, the identification of m6A regulators and m6A-related RNAs has great potential as biomarkers, prognostic predictors, and therapeutic targets [26]. Herein, based on GEO dataset GSE17536, a Cox proportional hazard regression model was exploited to analyze the association between the m6A RNA methylation-related regulators and colorectal cancer patients' OS and DFS, and we found that RBM15B, FTO, IGF2BP2, ZCCHC4, and KIAA1429 were significantly associated with the overall survival in colorectal cancer patients; however, KM survival assessment showed no significant association between colorectal cancer patients' prognosis and the five regulators respectively.

The protein known as RNA binding motif protein 15B (RBM15B/OTT3) was first discovered as a binding partner of the EpsteinBarr virus mRNA export factor EB2 [27]. Additionally, it has been more recently recognized as a co-factor of the nuclear export receptor NXF1 [28]. Previously, RBM15B was found to be up-regulated in ovarian cancer [29]. FTO, the first m6A demethylase, might be involved in the transcription of adjacent genes [30,31]. Moreover, FTO, as an m6A demethylase, exerts a critical carcinogenic effect on acute myeloid leukemia [31], cervical cancer [32], and breast cancer [33]. IGF2BP2 could be stabilized by lncRNA LINRIS, promoting aerobic glycolysis in colorectal cancer [34]. IGF2BP2 is also involved in the process of METTL3 facilitating tumor progression in colorectal carcinoma [35]. KIAA1429 up-regulation was observed in colonic adenocarcinoma samples [36]. Based on the collective evidence from past and current research, it is observed that although no individual component exhibited a substantial association with patients' prognosis, the m6A RNA methylation-related regulators might potentially serve as prognostic and/or diagnostic indicators in

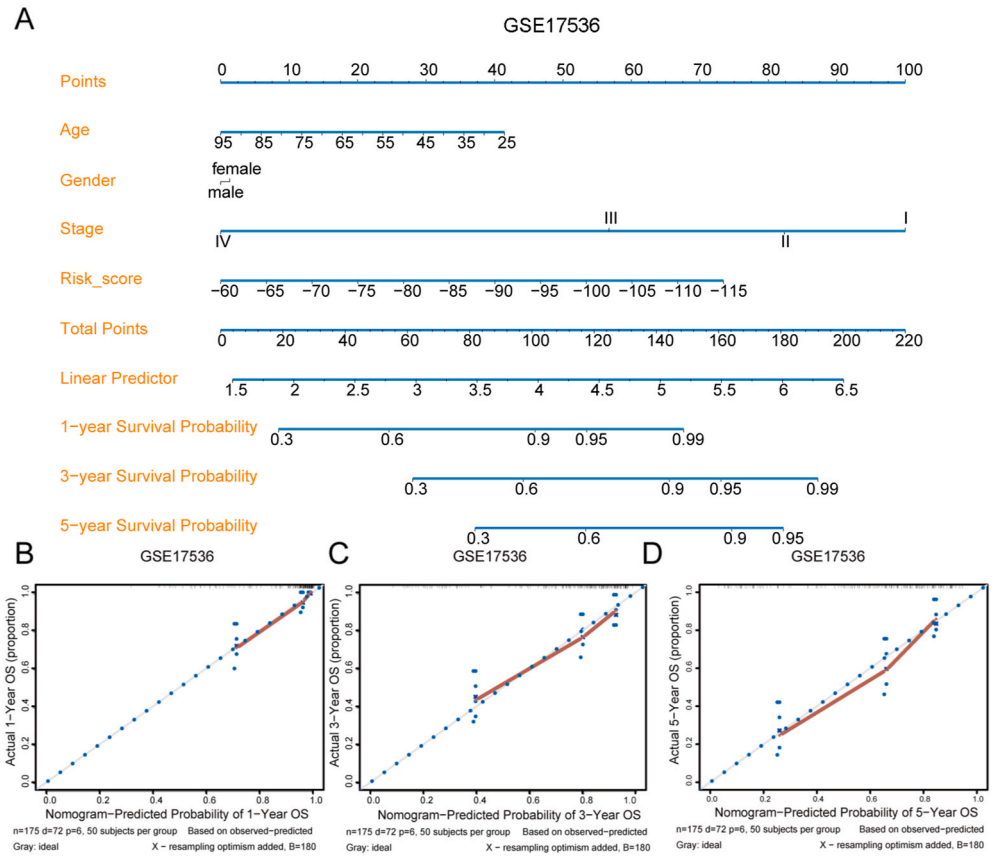

Fig. 10. Nomogram analysis based on GSE17536 (A) The construction of the nomogram involves the use of age, stage, and risk score factors in order to provide an estimation of the probability of survival at certain time intervals, namely 1, 3, and 5 years. (B) The calibration plot was used to assess the performance of the nomogram in predicting the 1-year, 3-years, and 5-years OS.

the context of colorectal cancer.

The LASSO Cox regression method is suitable for data processing and constructing models when dealing with a restricted sample size and a large number of independent variables [37]. Previously, studies used Lasso Cox regression to construct an m6A regulators-based signature for colon cancer prognosis prediction [38,39]. To further validate the role of m6A RNA methylated regulators in the clinicopathological features and prognosis of colorectal cancer, LASSO Cox regression was performed and a 5-gene signature consisting of RBM15B, FTO, IGF2BP2, ZCCHC4, and KIAA1429 was established. Both the training set and the validation set were exploited to determine each patient's risk score. The KM survival curve demonstrated a remarkable association of the risk score with the patients' OS and DFS, respectively. Based on the ROC curve analysis, the curve generated by the risk score exhibited a good level of predictive accuracy. The independent risk factor was determined based on the multivariate Cox's proportional hazard regression analysis. Finally, an association was found between the risk score and DFS in colorectal cancer patients using predictive nomograms created using the risk score. Importantly, by grouping cases in GSE17536, and GSE37182 taking the median value of the risk score as a cut-off, high-risk and low-risk score groups were shown to be overlapped with colorectal cancer and normal non-cancerous groups; in other words, colorectal cancer samples could be distinguished from normal non-cancerous samples using the risk score system as a classifier, suggesting that the risk score system could serve as a classifier distinguishing colorectal cancer samples from healthy controls. This shows that the risk model based on multiple genes can well solve the limitations of single-gene analysis. Several research also develop an m6A RNA methylation regulators-related prognostic signature using different data resource or different algorithms. Zhang et al. establish a 2 gene (YTHDC2 and IGF2BP3) signature with prognostic ability in colorectal cancer [40]. Yu et al. established an 18 prognostic m6A genes signature which could serve as potential prognostic predictor for colorectal cancer survival [38]. Through their signatures are total different with our the 5- five RNA-methylated regulators, also showed great prognostic ability. These evidences confirmed that the dysregulated m6A-related proteins are indeed associated with clinical progression.

Regarding the limitations of this study, one is the lack of detailed treatment information for the patients included in the datasets. This could potentially influence the prognosis and the risk score, and future studies should aim to incorporate this information to better understand its impact on the risk score. Another limitation is the absence of follow-up data for the clinical samples we recently collected. While these samples were useful for the initial validation of our findings, the lack of longitudinal data prevents us from assessing the long-term reliability and clinical utility of our risk score model. Future work should include long-term follow-up to fully

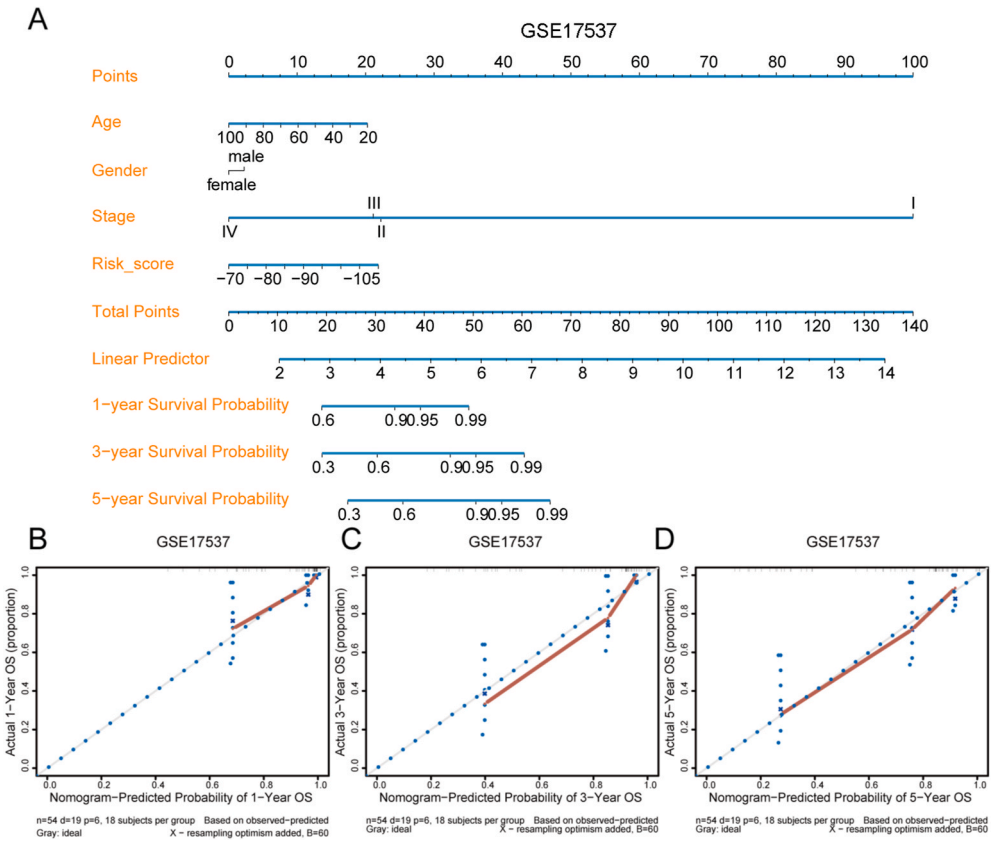

Fig. 11. Nomogram analysis based on GSE17537 (A) The construction of the nomogram involves the use of age, stage, and risk score factors in order to provide an estimation of the probability of survival at certain time intervals, namely 1, 3, and 5 years. (B) The calibration plot was used to assess the performance of the nomogram in predicting the 1-year, 3-years, and 5-years OS.

evaluate the prognostic value of our model. Moreover, while we have successfully discovered five RNA-methylated regulators that exhibit a substantial association with the overall survival of individuals diagnosed with colorectal cancer, the precise functional functions of these regulators in the context of colorectal cancer have yet to be elucidated via experimental investigations. Additional *in vivo* and *in vitro* experimental investigations are necessary to address this issue.

5. Conclusions

Altogether, a prognostic signature comprising of five m6A RNA methylation regulators, which has prognostic and diagnostic values for colorectal cancer patients, was established by the use of the LASSO Cox regression model, Kaplan-Meier survival evaluation, ROC curve analysis, and univariate and multivariate Cox's proportional hazard regression models.

Funding

This study was supported by the Natural Science Foundation of China (Grant Nos. 81970493 and 82270590) and the Natural Science Foundation of Hunan Province (2022JJ70144).

Ethical statement

Our study has been approved by the Ethics Committee of the Second Xiangya Hospital.

Author contribution statement

Dan Zhang: Conceived and designed the experiments; Analyzed and interpreted the data; Wrote the paper.
Lianwen Yuan: Conceived and designed the experiments, Contributed reagents, materials, analysis tools or data.
Lichao Yang; Qiang Wu: Performed the experiments; Analyzed and interpreted the data.
Guotao Wu: Analyzed and interpreted the data.

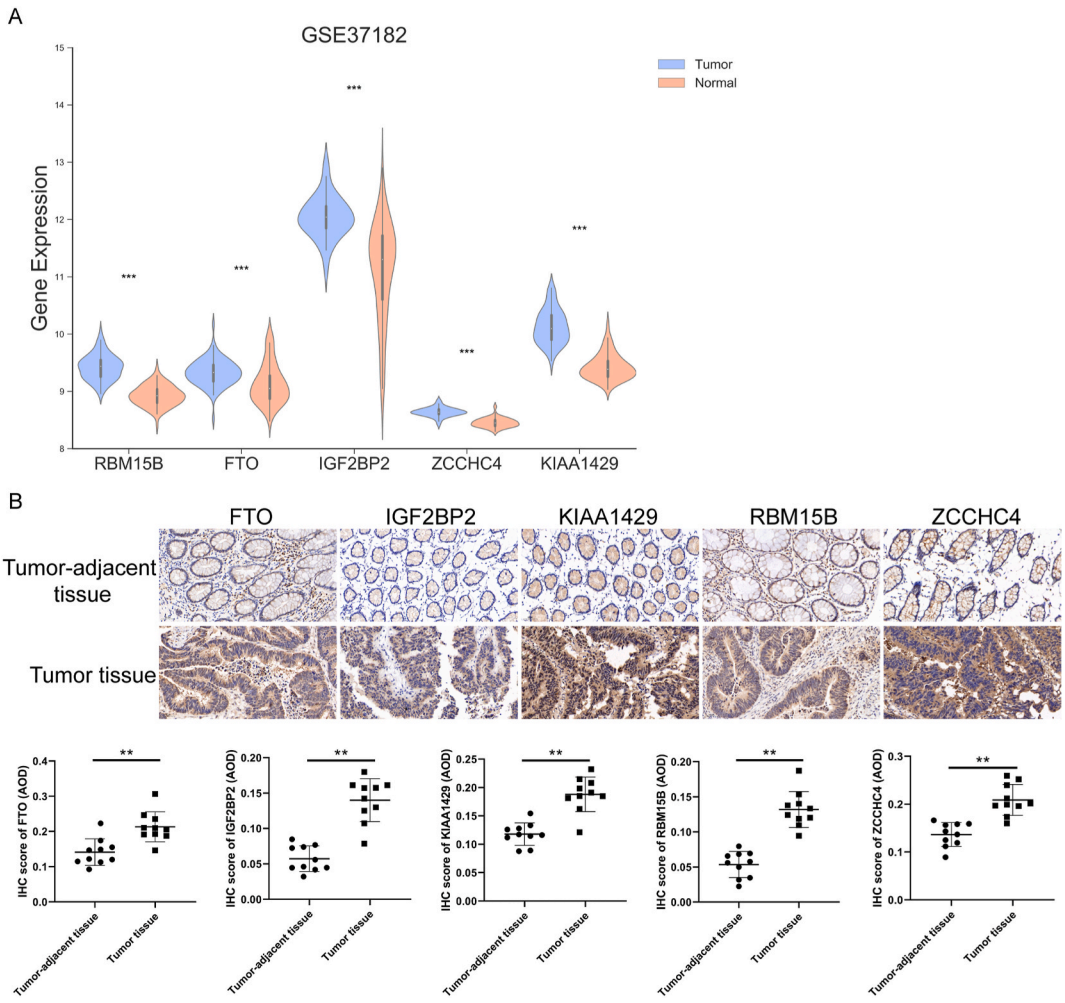

Fig. 12. Expression validation of the factors using dataset and clinical samples (A) The expression levels of RBM15B, FTO, IGF2BP2, ZCCHC4, and KIAA1429 in normal colon tissue and cancer samples according to GSE37182. (B) The expression levels of RBM15B, FTO, IGF2BP2, ZCCHC4, and KIAA1429 were examined in collected colorectal cancer and paired adjacent tissues using Immunohistochemical staining (n = 10). **p < 0.01.

Data availability statement

The data that support the findings of this study are available in GEO database. These data were derived from the following resources available in the public domain: GSE17536, GSE17537, GSE75500, and GSE37182.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Appendix A. Supplementary data

Supplementary data to this article can be found online at https://doi.org/10.1016/j.heliyon.2023.e20172.

References

[1] F. Bray, J. Ferlay, I. Soerjomataram, R.L. Siegel, L.A. Torre, A. Jemal, Global cancer statistics 2018: GLOBOCAN estimates of incidence and mortality worldwide for 36 cancers in 185 countries, Ca - Cancer J. Clin. 68 (8) (2018) 394–424.

[2] J. Ferlay, M. Colombet, I. Soerjomataram, C. Mathers, D.M. Parkin, M. Pieros, et al., Estimating the global cancer incidence and mortality in 2018: GLOBOCAN sources and methods, Int. J. Cancer 144 (8) (2019) 1941–1953.

[3] GBoDC. Collaboration, C. Fitzmaurice, C. Allen, e al, Global, regional, and national cancer incidence, mortality, years of life lost, years lived with disability, and disability-adjusted life-years for 32 cancer groups, 1990 to 2015: a systematic analysis for the global burden of disease study, JAMA Oncol. 3 (4) (2017) 524–548.

[4] R.D. Kennedy, M. Bylesjo, P. Kerr, T. Davison, D.P. Harkin, Development and independent validation of a prognostic assay for stage II colon cancer using formalin-fixed paraffin-embedded tissue, J. Clin. Oncol. 29 (35) (2011) 4620–4626.

[5] D. Dan, M.-M. Sharon, S. Schraga, S.-D. Mali, U. Lior, O. Sivan, et al., Topology of the human and mouse m6A RNA methylomes revealed by m6A-seq, Nature 485 (7397) (2012) 201–206.

[6] K.E. Pendleton, B. Chen, K. Liu, O.V. Hunter, Y. Xie, B.P. Tu, et al., The U6 snRNA m(6)A methyltransferase METTL16 regulates SAM synthetase intron retention, Cell 169 (5) (2017) 824–835.

[7] X.L. Ping, B.F. Sun, L. Wang, W. Xiao, X. Yang, W.J. Wang, et al., Mammalian WTAP is a regulatory subunit of the RNA N6-methyladenosine methyltransferase, Cell Res. 24 (2) (2014) 177–189.

[8] S. Schwartz, M. Mumbach, M. Jovanovic, T. Wang, K. Maciag, G.G. Bushkin, et al., Perturbation of m6A writers reveals two distinct classes of mRNA methylation at internal and 5′ sites, Cell Rep. 8 (1) (2014) 284–296.

[9] P. Patil, Deepak, Constanza Jackson, Chen, et al., m(6)A RNA methylation promotes XIST-mediated transcriptional repression, Nature 537 (7620) (2016) 369–373.

[10] T. Wang, S. Kong, M. Tao, S. Ju, The potential role of RNA N6-methyladenosine in Cancer progression, Mol. Cancer 19 (1) (2020) 88.

[11] D. Zhen, Y. Wu, Y. Zhang, K. Chen, J. Meng, m6A reader: epitranscriptome target prediction and functional characterization of N6-methyladenosine (m6A) readers, Front. Cell Dev. Biol. 8 (2020) 741.

[12] Z. Bi, Y. Liu, Y. Zhao, Y. Yao, R. Wu, Q. Liu, et al., A dynamic reversible RNA N 6 -methyladenosine modification: current status and perspectives, J. Cell. Physiol. 234 (6) (2019) 7948–7956.

[13] X.Y. Chen, J. Zhang, J.S. Zhu, The role of m 6 A RNA methylation in human cancer, Mol. Cancer 18 (1) (2019) 103.

[14] G. Jia, Y. Fu, X. Zhao, Q. Dai, G. Zheng, Y. Yang, et al., N6-methyladenosine in nuclear RNA is a major substrate of the obesity-associated FTO, Nat. Chem. Biol. 7 (12) (2011) 885–887.

[15] G. Zheng, J.A. Dahl, Y. Niu, e al, ALKBH5 is a mammalian RNA demethylase that impacts RNA metabolism and mouse fertility, RNA Biol. 49 (1) (2013) 18–29.

[16] D. Dai, H. Wang, L. Zhu, H. Jin, X. Wang, N6-methyladenosine links RNA metabolism to cancer progression, Cell Death Dis. 9 (2) (2018) 124.

[17] Q. Cui, H. Shi, P. Ye, e al, m 6 A RNA methylation regulates the self-renewal and tumorigenesis of glioblastoma stem cells, Cell Rep. 18 (11) (2017) 2622–2634.

[18] S. Zhou, Z.L. Bai, D. Xia, Z.J. Zhao, R. Zhao, Y.Y. Wang, et al., FTO regulates the chemo-radiotherapy resistance of cervical squamous cell carcinoma (CSCC) by targeting β-catenin through mRNA demethylation, Mol. Carcinog. 57 (5) (2018) 590–597.

[19] X. Cheng, M. Li, X. Rao, W. Zhang, X. Li, L. Wang, et al., KIAA1429 regulates the migration and invasion of hepatocellular carcinoma by altering m6A modification of ID2 mRNA, OncoTargets Ther. 12 (2019) 3421–3428.

[20] K. Taketo, M. Konno, A. Asai, J. Koseki, K. Ogawa, The epitranscriptome m6A writer METTL3 promotes chemo- and radioresistance in pancreatic cancer cells, Int. J. Oncol. 52 (2) (2018) 621–629.

[21] L. Li, L. Jiajing, H. Hongying, Identification of potential crucial genes associated with the pathogenesis and prognosis of endometrial cancer, Front. Genet. 10 (2019) 373.

[22] W. Xinyi, F. Xiaorui, Z. Junjia, X. Chengfeng, Z. Shuyong, L. Yunxia, Identification and validation of m6A RNA methylation regulators with clinical prognostic value in Papillary thyroid cancer, Cancer Cell Int. 20 (2020) 203.

[23] M.S. Schröder, A.C. Culhane, J. Quackenbush, B. Haibe-Kains, survcomp: an R/Bioconductor package for performance assessment and comparison of survival models, Bioinformatics 27 (22) (2011) 3206–3208.

[24] A HJ, J MB, The meaning and use of the area under a receiver operating characteristic (ROC) curve, Radiology 143 (1) (1982) 29–36.

[25] T. Liu, C. Li, L. Jin, C. Li, L. Wang, The prognostic value of m6A RNA methylation regulators in colon adenocarcinoma, Medical Science Monitor International Medical Journal of Experimental & Clinical Research 25 (2019) 9435–9445.

[26] J. Li, L. Liang, Y. Yang, X. Li, Y. Ma, N(6)-methyladenosine as a biological and clinical determinant in colorectal cancer: progression and future direction, Theranostics 11 (6) (2021) 2581–2593.

[27] E. Hiriart, Interaction of the Epstein-Barr virus mRNA export factor EB2 with human Spen proteins SHARP, OTT1, and a novel member of the family, OTT3, links Spen proteins with splicing regulation and mRNA export, J. Biol. Chem. 280 (44) (2005) 36935–36945.

[28] H. Uranishi, A.S. Zolotukhin, S. Lindtner, S. Warming, G.M. Zhang, J. Bear, et al., The RNA-binding motif protein 15B (RBM15B/OTT3) acts as cofactor of the nuclear export receptor NXF1, J. Biol. Chem. 284 (38) (2009) 26106–26116.

[29] Q. Wang, Q. Zhang, Q. Li, J. Zhang, J. Zhang, Clinicopathological and immunological characterization of RNA m6A methylation regulators in ovarian cancer, Molecular Genetics & Genomic Medicine (2020) e1547.

[30] R.J.F. Loos, C. Bouchard, FTO: the first gene contributing to common forms of human obesity, Obes. Rev. 9 (3) (2010) 246–250.

[31] Z. Li, H. Weng, R. Su, X. Weng, Z. Zuo, C. Li, et al., FTO plays an oncogenic role in acute myeloid leukemia as a N6-methyladenosine RNA demethylase, Cancer Cell 31 (1) (2016) 127–141.

[32] D. Zou, L. Dong, C. Li, Z. Yin, Q. Zhou, The m6A eraser FTO facilitates proliferation and migration of human cervical cancer cells, Cancer Cell Int. 19 (1) (2019) 321.

[33] Y. Niu, Z. Lin, A. Wan, H. Chen, H. Liang, L. Sun, et al., RNA N6-methyladenosine demethylase FTO promotes breast tumor progression through inhibiting BNIP3, Mol. Cancer 18 (1) (2019) 46.

[34] Y. Wang, J.H. Lu, Q.N. Wu, Y. Jin, D.S. Wang, Y.X. Chen, et al., LncRNA LINRIS stabilizes IGF2BP2 and promotes the aerobic glycolysis in colorectal cancer, Mol. Cancer 18 (1) (2019) 174.

[35] T. Li, P.S. Hu, Z. Zuo, J.F. Lin, X. Li, Q.N. Wu, et al., METTL3 facilitates tumor progression via an m(6)A-IGF2BP2-dependent mechanism in colorectal carcinoma, Mol. Cancer 18 (1) (2019) 112.

[36] D. Xu, J. Shao, H. Song, J. Wang, The YTH domain family of N6-methyladenosine "readers" in the diagnosis and prognosis of colonic adenocarcinoma, BioMed Res. Int. 2020 (2020), 9502560.

[37] S. Lee, M.H. Seo, Y. Shin, The Lasso for high-dimensional regression with a possible change-point, J. Roy. Stat. Soc. 78 (1) (2016) 193–210.

[38] L. Yu, L. Wang, J. Sun, X. Zhou, Y. Hu, L. Hu, et al., N6-methyladenosine related gene expression signatures for predicting the overall survival and immune responses of patients with colorectal cancer, Front. Genet. 14 (2023), 885930.

[39] L. Ji, S. Chen, L. Gu, X. Zhang, Exploration of potential roles of m6A regulators in colorectal cancer prognosis, Front. Oncol. 10 (2020) 768.

[40] J. Zhang, X. Cheng, J. Wang, Y. Huang, J. Yuan, D. Guo, Gene signature and prognostic merit of M6a regulators in colorectal cancer, Exp. Biol. Med. 245 (15) (2020) 1344–1354.