

# Constructing a prognostic immune-related lncRNA model for colon cancer

Xinyun Li, MM<sup>a</sup>, Lin Yang, MM<sup>b</sup>, Wen Wang, BA<sup>a</sup>, Xiangshu Rao, MM<sup>b</sup>, Yu Lai, MD<sup>b,\*</sup> 

## Abstract

Colon cancer is a common digestive tract tumor. Although many gene prognostic indicators have been used to predict the prognosis of colon cancer patients, the accuracy of these prognostic indicators is still uncertain. Thus, it is necessary to construct a model for the prognostic analysis of colon cancer. We downloaded the original transcriptome data of colon cancer and performed a differential coexpression analysis of immune-related genes to obtain different immune-related long noncoding RNAs, which were paired as differentially expressed immune-related lncRNA pairs (DEirlncRNAPs). Then, the 1-year overall survival rate receiver operating characteristic curve was calculated, and the Akaike information criterion value was evaluated to determine the maximum inflection point, which was used as the cutoff point to identify groups of colon cancer patients at high and low risk for death. Subsequently, the optimal prediction model was established. Finally, we used the patients' survival times, clinicopathological features, tumor-infiltrating immune cells, chemotherapy responses, and immunosuppressive biomarkers to verify the DEirlncRNAP model. Seventy-one DEirlncRNAPs were obtained to build the risk assessment model. The patients were divided into a high-risk group and a low-risk group according to the cutoff point. Then, the DEirlncRNAP model was verified using patient survival times, clinicopathological features, tumor-infiltrating immune cells, chemotherapy responses, and immunosuppressive biomarkers. A new DEirlncRNAP model for predicting the prognosis of colon cancer patients was established, which could reveal new insights into the relationships of colon cancer with tumor-infiltrating immune cells and antitumor immunotherapy.

**Abbreviations:** AIC = Akaike information criterion, AUCs = area under the curves, COAD = colon adenocarcinoma, DEirlncRNAPs = differentially expressed immune-related lncRNA pairs, DEirlncRNAs = differentially expressed irlncRNAs, FC = fold-change, FDR = false discovery rate, FPKM = per kilobase per million, GTF = gene transfer format, IC<sub>50</sub> = inhibitory concentration 50, ICI = immune checkpoint inhibitor, ir-genes = immune-related genes, irlncRNA = immune-related lncRNA, lncRNAs = Long noncoding ribonucleic acids, OS = overall survival, ROC = receiver operating characteristic, TCGA = The Cancer Genome Atlas.

**Keywords:** colon cancer, immune, immune checkpoint inhibitors, long noncoding RNA, prognostic model

## 1. Introduction

Colon cancer is a common intestinal epithelial malignant tumor with a high incidence and poor prognosis that seriously endangers human life. According to the statistics of the International Agency for Research on Cancer of the World Health Organization, the global incidence of colon cancer was approximately 1.93 million, and the death toll was approximately 940,000 in 2020.<sup>[1]</sup> More than half of the cases of colon cancer are related to smoking, unhealthy eating habits, alcoholism, physical inactivity, and obesity.<sup>[2]</sup> The early clinical symptoms of colon cancer are hidden and lack highly specific markers. Therefore, it is important to construct a reliable prognostic model for colon cancer patients.

Recently, studies have suggested that the immune system plays a key role in the occurrence, development and treatment of colon cancer, especially immune checkpoint inhibitor (ICI) therapy, which is of great significance for prolonging the survival time of patients with colon cancer.<sup>[3–5]</sup> Pembrolizumab, an inhibitor of PD-1, was used to treat colorectal cancer patients with type dMMR-MSI-H cancer in the KEYNOTE(KN)-016 trial. The results indicated that the 2-year overall survival (OS) rate was 72%.<sup>[6]</sup> Navuzumab, another PD-1 inhibitor, was also shown to have a significant effect in patients with colorectal cancer in multicenter, open-label II clinical trials.<sup>[7]</sup> In addition, the combination of navuzumab and the CTLA-4 inhibitor ipilimumab achieved 84% control of colon cancer and an 83% OS

XL and LY contributed equally to this work.

The research was supported by the Xinglin Scholar Research Promotion Project of Chengdu University of TCM (grant no. XSGG2020001) and the National Natural Science Foundation of China (grant no. 82074329).

The authors have no conflicts of interest to disclose.

The public datasets to support the results of this subject are available from TCGA.

All data generated or analyzed during this study are included in this published article [and its supplementary information files]; The datasets generated during and/or analyzed during the current study are publicly available.

Ethical approval was not necessary since the raw data were obtained from the TCGA database, which is a public functional genomics data repository and is freely available for users.

Supplemental Digital Content is available for this article.

<sup>a</sup> School of Traditional Chinese Medicine, Sichuan College of Traditional Chinese Medicine, China, <sup>b</sup> School of Basic Medicine, Chengdu University of Traditional Chinese Medicine, Chengdu, China.

\*Correspondence: Yu Lai, School of Basic Medicine, Chengdu University of Traditional Chinese Medicine, No. 1166 Liutai Avenue, Chengdu 611137, China (e-mail: yulai06@cducm.edu.cn).

Copyright © 2022 the Author(s). Published by Wolters Kluwer Health, Inc. This is an open-access article distributed under the terms of the Creative Commons Attribution-Non Commercial License 4.0 (CCBY-NC), where it is permissible to download, share, remix, transform, and buildup the work provided it is properly cited. The work cannot be used commercially without permission from the journal.

How to cite this article: Li X, Yang L, Wang W, Rao X, Lai Y. Constructing a prognostic immune-related lncRNA model for colon cancer. *Medicine* 2022;101:38(e30447).

Received: 7 April 2022 / Received in final form: 25 July 2022 / Accepted: 29 July 2022

<http://dx.doi.org/10.1097/MD.000000000030447>

rate at 12 months, and the degree of side effects associated with the treatment was acceptable.<sup>[8]</sup>

Long noncoding ribonucleic acids (lncRNAs) are transcripts over 200 nucleotides in length that have no protein coding function. Because lncRNAs are closely related to the occurrence and development of tumors and have obviously abnormal expression levels in tumor tissues and precancerous lesions,<sup>[9,10]</sup> lncRNAs are ideal tumor markers with high specificity and sensitivity that are easy to repeatedly detect.<sup>[11]</sup>

A recent study showed that the immune-related lncRNA (irlncRNA) signature possessed independent prognostic significance in colon cancer.<sup>[12]</sup> According to a previous report by Hong et al,<sup>[13]</sup> we screened differentially expressed immune-related lncRNA pairs (DEirlncRNAPs), established and verified a risk assessment model for colon cancer, and analyzed the relationships between the paired DEirlncRNAPs, tumor infiltrating immune cells and clinical treatment outcomes.

## 2. Methods

### 2.1 Data collection and preprocessing

The main process of this research is demonstrated in Figure 1. Transcriptome profiling (RNAseq) data and fragments per kilobase per million (FPKM) data from 449 patients with colon adenocarcinoma (COAD) and clinical data from 452 COAD patients were downloaded from The Cancer Genome Atlas (TCGA) database (<https://tcga-data.nci.nih.gov/tcga/>). Clinical data with no survival times available and duplicate data were deleted. The gene transfer format (GTF) annotation file was downloaded from Ensembl (<http://asia.ensembl.org>). Perl language was used to distinguish mRNA and lncRNA for subsequent analysis. A total of 2483 identified immune-related genes (ir-genes) were downloaded from ImmPort (<http://www.immport.org>). Then, we performed a coexpression analysis of the ir-genes and the previously obtained lncRNAs to identify irlncRNAs. We used the limma R package to analyze the differential expression of irlncRNAs and set the filter parameters as log-

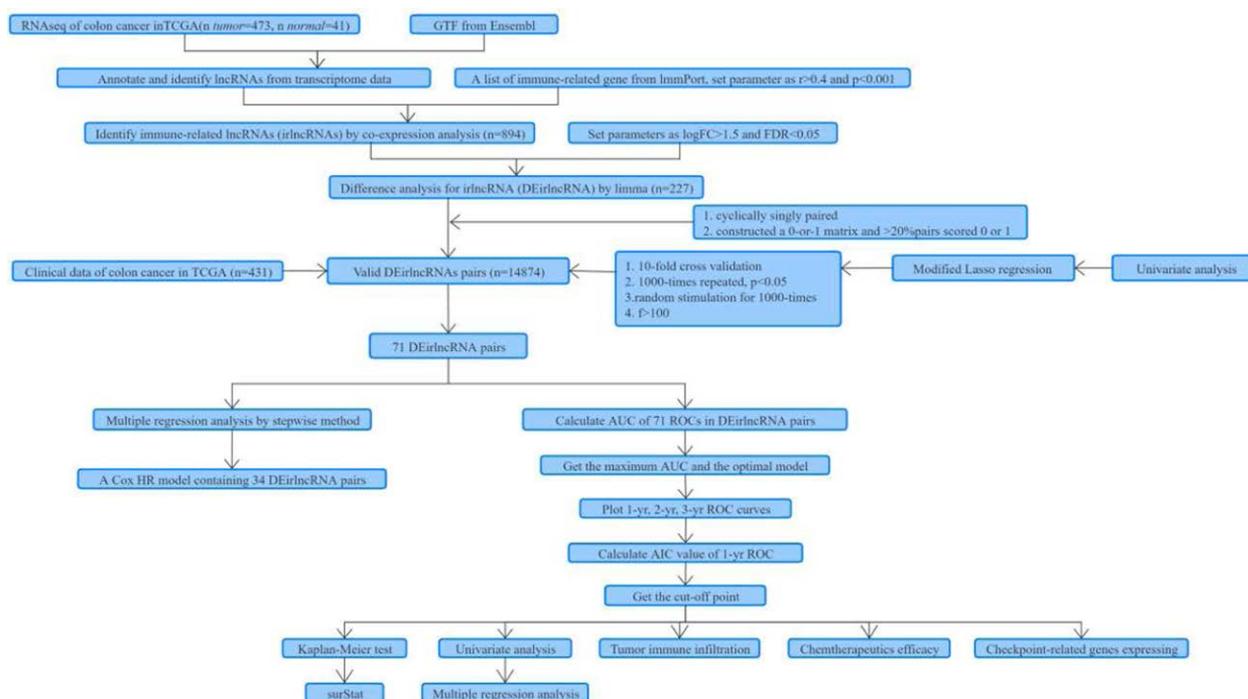
fold-change (FC) > 1.5 and false discovery rate (FDR) < 0.05 to acquire differentially expressed irlncRNAs (DEirlncRNAs).

### 2.2 Screening DEirlncRNAPs

We paired the DEirlncRNAs and called the paired lncRNA pairs DEirlncRNAP. Then, we used DEirlncRNAPs to build a 0 or 1 matrix. According to Hong et al's algorithm, by comparing the expression of 2 lncRNAs in DEirlncRNAP, which can predict the prognosis of patients with COAD, was selected to minimize the error caused by the change in expression.

### 2.3 Establishment of a risk assessment model of colon cancer

First, we performed univariate analysis of the validated DEirlncRNAPs and used the glmnet R package for Lasso regression with 10-fold cross validation, and a P value < .05 was considered statistically significant. The number of iterations was set to 1000, and 1000 random stimuli were set for each cycle. Additionally, the frequency of each DEirlncRNAP in 1000 Lasso regression cycles was recorded, and DEirlncRNAPs with frequencies over 100 were subjected to Cox proportional hazard regression analysis. Then, we calculated the coefficient of DEirlncRNAPs and constructed the risk model. We drew receiver operating characteristic (ROC) curves of the 1-, 2-, and 3-year OS rates of the patients and calculated the area under the curves (AUCs). If the curve met the point pointing to the maximum AUC, the model was regarded as the best candidate, and we ended the calculation process. We used the following formula to determine the risk score of all clinical cases of COAD in the TCGA database: RiskScore =  $\sum_{i=1}^k \beta_i S_i$ . Finally, Akaike information criterion (AIC) values were calculated to evaluate each point of the 1-year ROC curve to determine the maximum inflection point. We divided COAD patients into a high-risk group and a low-risk group according to the cutoff value.



**Figure 1.** The main flow chart of this study. AUCs = area under the curves, DEirlncRNAs = differentially expressed irlncRNAs, GTF = gene transfer format, irlncRNA = immune-related lncRNA, lncRNAs = Long noncoding ribonucleic acids, ROC = receiver operating characteristic, TCGA = The Cancer Genome Atlas.

## 2.4 Verification of the DEirlncRNAP risk assessment model

We used Kaplan–Meier analysis to show the difference in survival between patients in the high-risk and low-risk groups and R language to evaluate the prognostic value of the cutoff point. The R packages used in these operations are survival, glmnet, survivalROC, and survminer.

Then, we analyzed the relationship between the DEirlncRNAP model and clinicopathological features by the chi-square test and drew a band diagram to show the results. The markers were as follows: \* $P < .05$ , \*\* $P < .01$ , \*\*\* $P < .001$ . The Wilcoxon signed rank test was used to calculate the difference in risk scores in different groups with clinicopathological features, and box diagrams were drawn for visualization. Finally, univariate and multivariate Cox proportional hazards regressions were used to compare the DEirlncRNAP model with other clinicopathological features in predicting the prognosis of COAD patients and to evaluate the accuracy of the DEirlncRNAP model. We drew forest maps to show the results. The R packages used in these steps include limma, ComplexHeatmap, survivalROC, and ggpubr.

## 2.5 Study of the relationship between the DEirlncRNAP model and tumor-infiltrating immune cells

We analyzed the relationship between the model and immune cells by the Wilcoxon signed rank test and used 7 common methods to calculate the immune infiltrating cells of COAD samples in the TCGA database, including CIBERSORT,<sup>[14]</sup> CIBERSORT-ABS,<sup>[15]</sup> XCELL,<sup>[16]</sup> QUANTISEQ,<sup>[17]</sup> MCPOUNTER,<sup>[18]</sup> EPIC,<sup>[19]</sup> and TIMER.<sup>[20]</sup> The results are displayed as box plots. Then, a detailed Spearman correlation analysis was performed to study the relationship between immune cells and the risk score. The results are shown as a bubble chart. We set the threshold value as  $P < .05$ . These steps were completed in R language, and the R packages used included limma, scales, ggplot2, and ggtext.

## 2.6 Study of the relationship between the DEirlncRNAP model and clinical treatment

To improve the clinical relevance of the DEirlncRNAP model, we studied the relationship between the model and genes related to immunosuppression checkpoints and drew a violin diagram. This operation used the R packages gamma and ggpubr. In addition to ICI therapy, chemotherapy is also a common treatment for colon cancer. To study the correlation between the risk assessment model and the response to chemotherapy, we calculated the inhibitory concentration 50 ( $IC_{50}$ ) of some commonly used chemotherapy drugs for patients with colon cancer. These drugs include efitinib, thapsigargin, vinorelbine, 5-fluorouracil, and irinotecan. We used the Wilcoxon signed rank test to compare the  $IC_{50}$  values between the high- and low-risk groups, and the results are shown as box charts. In this step, a ridge regression model was constructed by using the R package pRRophetic to predict the  $IC_{50}$  value of chemotherapy drugs.

## 3. Results

### 3.1 Obtaining DEirlncRNAs

In this study, we collected transcriptome data of 514 patients with COAD from the TCGA database, consisting of 41 normal and 473 tumor samples. We used the GTF of Ensemble to annotate the data and analyzed the coexpression of known ir-genes and lncRNAs to obtain irlncRNAs. Then, we analyzed the differential expression of the irlncRNAs. We obtained 894 irlncRNAs (Table S1, Supplemental Digital Content 1, <http://links.lww.com/MD/H261>) and 227 DEirlncRNAs (Fig. 2A), of which the expression of 208 lncRNAs was upregulated and the expression of 19 lncRNAs was downregulated (Fig. 2B).

### 3.2 Obtaining DEirlncRNAs and constructing the risk assessment model

We conducted a pairwise comparison of the DEirlncRNAs and screened the 0-or-1 matrix and obtained 14,874 pairs of valid DEirlncRNAs. Then, we used modified Lasso regression analysis and single factor analysis and extracted 71 DEirlncRNAs (Table 1). We used a stepwise method and obtained 34 DEirlncRNAs to construct a Cox proportional hazards model (Fig. 2C). Next, we used the 71 DEirlncRNAs to draw 1-, 2- and 3-year survival rate ROCs and calculated the AUCs. All AUCs were above 0.95 (Fig. 3A). The maximum AUC of the 1-year ROC curve was 0.955 (Fig. 3B), which was compared with other clinical features of COAD patients (Fig. 3C). Finally, we used AIC values to determine the cutoff point of the 1-year survival rate ROC curve (Fig. 3D). We evaluated the clinical data of 426 COAD patients collected from TCGA. Then, we calculated the risk scores of all patients and divided them into groups. Those with scores higher than the cutoff value were categorized as high-risk; otherwise, they were categorized as low-risk. As a result, we stratified 367 patients into the low-risk group and into the 59 high-risk group (Table S2, Supplemental Digital Content 2, <http://links.lww.com/MD/H262>).

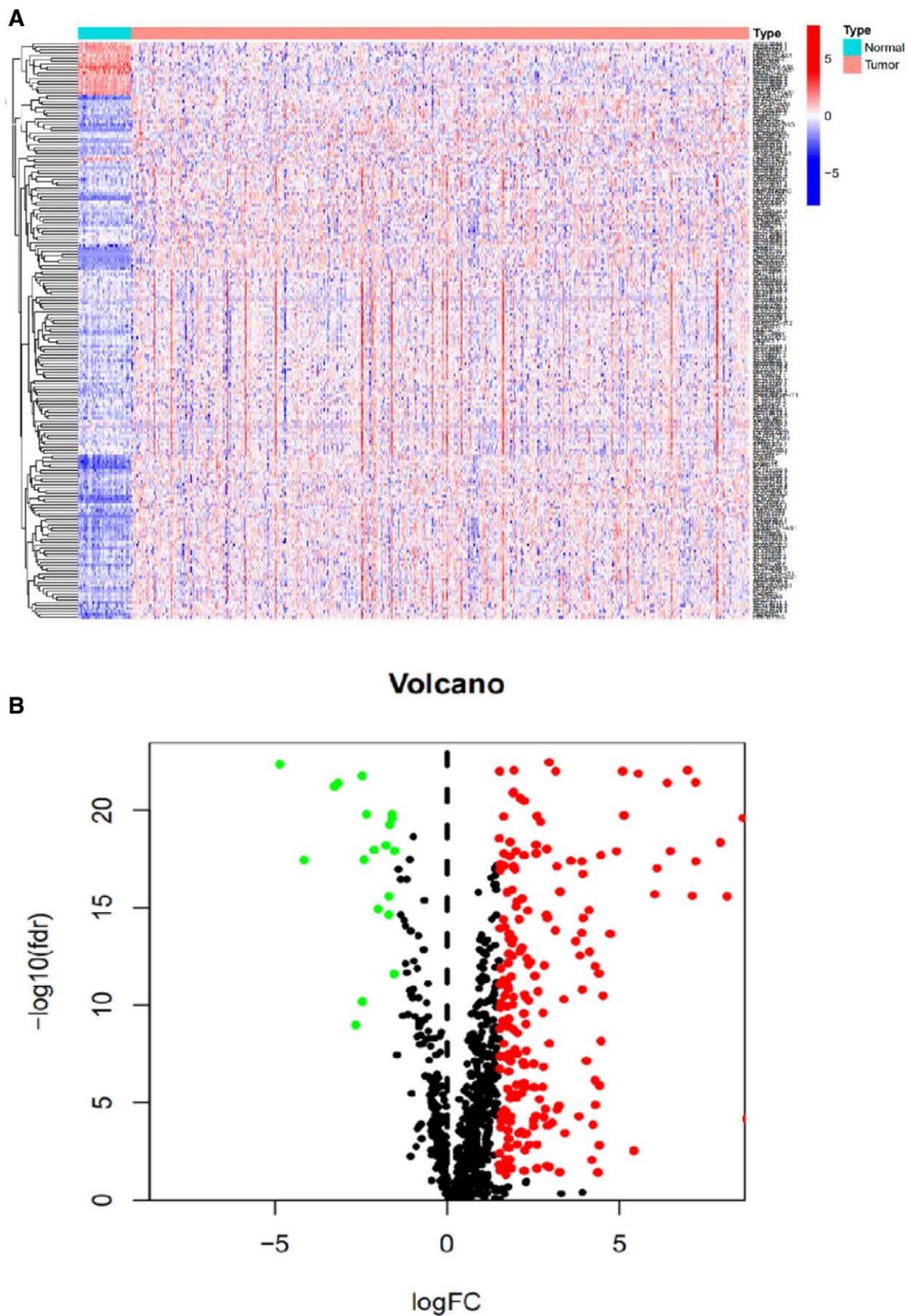
### 3.3 Clinical evaluation based on the DEirlncRNAP risk assessment model

We combined the risk scores of all COAD patients with survival time and visualized the results (Fig. 4A and B) and found that the survival time and number of survivors in the low-risk group were significantly higher than those in the high-risk group. Kaplan–Meier analysis also indicated that the survival time of patients in the low-risk group was better than that of patients in the high-risk group, and the difference was statistically significant ( $P < .001$ ; Fig. 4C).

According to the risk score of the DEirlncRNAP model, we conducted a sequence of chi-square tests to study the relationship between clinicopathological features and risk in COAD patients. We obtained a band diagram (Fig. 5A). We obtained a group of scatter plots using the Wilcoxon signed rank test. The results suggested that risk was related to clinical stage (Fig. 5B), T stage (Fig. 5C), M stage (Fig. 5D), N stage (Fig. 5E) and age (Fig. 5F). Univariate Cox regression analysis (Fig. 5G) showed that the differences in patient age ( $P = .04$ , HR = 1.021, 95% CI [1.001–1.042]), clinical stage ( $P < .001$ , HR = 2.222, 95% CI [1.714–2.881]) and risk score ( $P < .001$ , HR = 1.001, 95% CI [1.000–1.001]) were statistically significant. Multivariate Cox regression analysis (Fig. 5H) showed significant differences in age ( $P = .012$ , HR = 1.027, 95% CI [1.006–1.048]), clinical stage ( $P < .001$ , HR = 2.417, 95% CI [1.851–3.155]) and risk score ( $P < .001$ , HR = 1.001, 95% CI [1.000–1.001]). Univariate and multivariate Cox regression analyses indicated that the risk score could be used as an independent prognostic indicator.

### 3.4 Study of tumor-infiltrating immune cells with the DEirlncRNAP model

The DEirlncRNAP model is related to ir-genes, so we used 7 different online analysis platforms to predict the enrichment of immune cells in different risk groups. The results of the Wilcoxon signed rank test showed that high risk negatively correlated with the infiltration of plasmacytoid dendritic cells and CD8+ T cells (Figure S1, Supplemental Digital Content 3, <http://links.lww.com/MD/H263>). Then, we carried out a Spearman correlation analysis and found that high risk positively correlated with hematopoietic stem cells, macrophages, monocytes, endothelial cells and cancer-associated fibroblast cells and negatively correlated with class switch memory B cells, plasmacytoid dendritic cells, CD8+ T cells and mast cells (Fig. 6) (Table S3, Supplemental Digital Content 4, <http://links.lww.com/MD/H264>).



**Figure 2.** Screening DEirIncRNAs and DEirIncRNAPs. The DEirIncRNAs selected from 473 COAD tumor and 41 normal samples from TCGA were analyzed by heatmap (A) and volcano plots (B). (C) The forest map shows 34 DEirIncRNAPs obtained by multivariate stepwise regression analysis, which were used to construct a Cox proportional hazards model. COAD = colon adenocarcinoma, DEirIncRNAPs = differentially expressed immune-related lncRNA pairs, DEirIncRNAs = differentially expressed irIncRNAs, TCGA = The Cancer Genome Atlas.

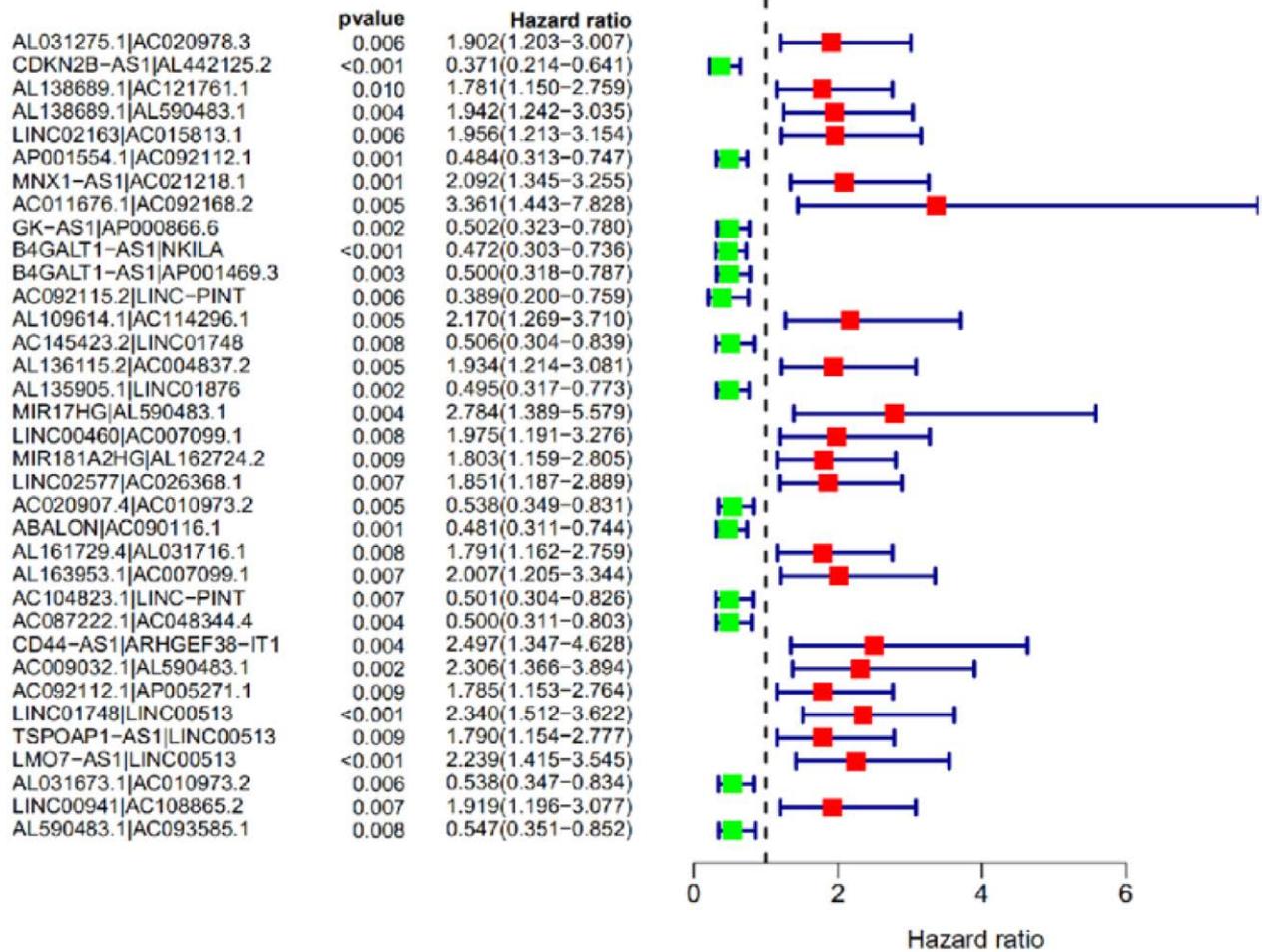
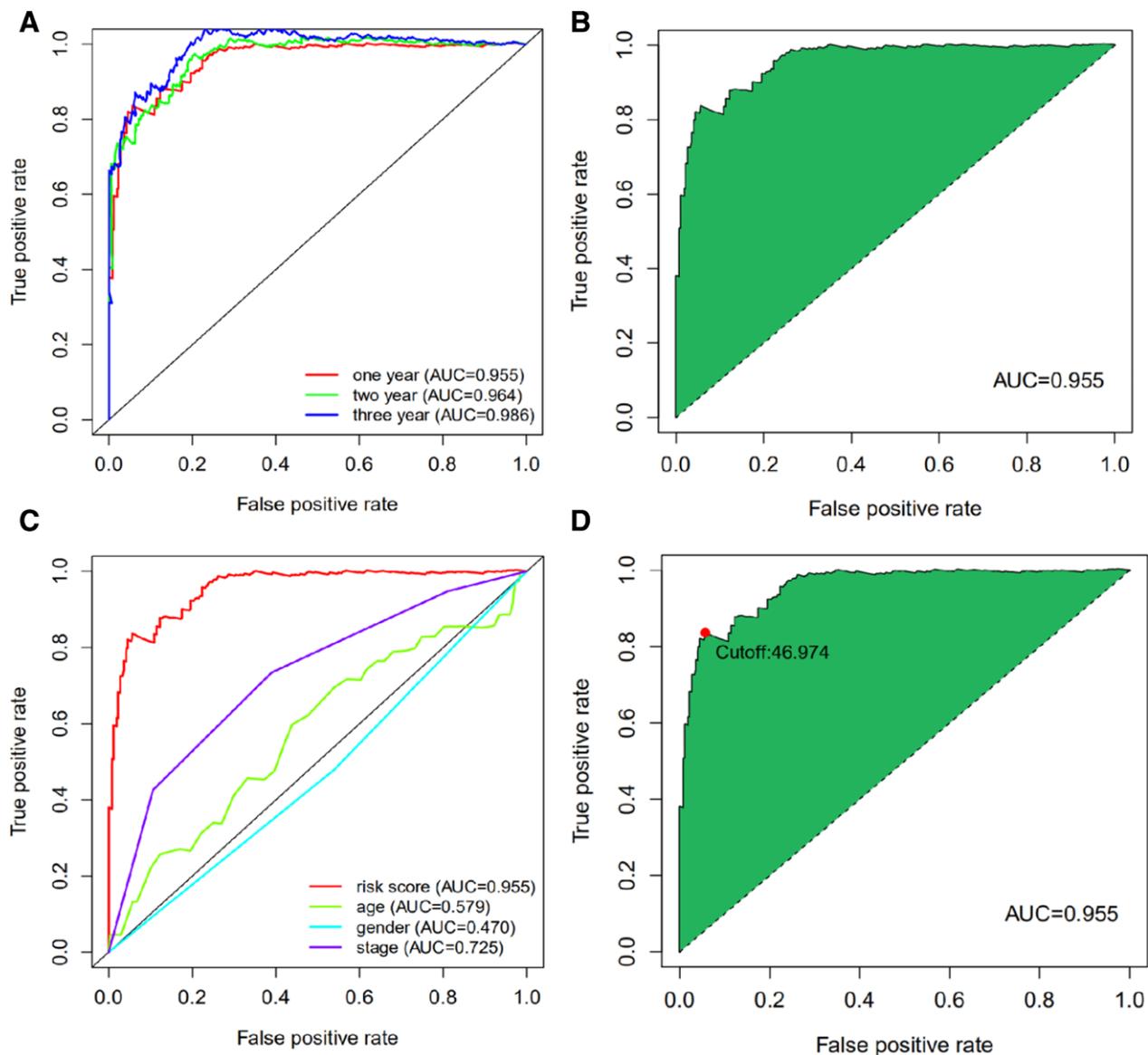


Figure 2. Continued

**Table 1**  
Seventy one DEirncRNAs for establishing risk assessment models.

1	AC018628.2	AC026356.1	25	AC092115.2	TSPOAP1-AS1	49	AC067930.4	AL138963.1
2	Z82243.1	AL590483.1	26	AL109614.1	AC114296.1	50	AC104823.1	LINC-PINT
3	AL031275.1	AP001554.1	27	AL109614.1	AC007099.1	51	AC087222.1	AC048344.4
4	AL031275.1	AC092115.2	28	AL117379.1	NKILA	52	AP002336.2	LM07-AS1
5	AL031275.1	AC020978.3	29	LINC01311	AC245100.7	53	AC121761.1	LINC01876
6	CDKN2B-AS1	AL442125.2	30	AL121832.3	AC010536.2	54	CD44-AS1	ARHGEF38-IT1
7	AL138689.1	AC121761.1	31	AC145423.2	LINC01748	55	AC009032.1	AC008649.2
8	AL138689.1	MIR3142HG	32	AL136115.2	AC004837.2	56	AC009032.1	AL590483.1
9	AL138689.1	AL590483.1	33	AL135905.1	LINC01876	57	AC092112.1	AP005271.1
10	AL390195.1	ALG13-AS1	34	FENDRR	NKILA	58	AC092112.1	AC008649.2
11	LINC02163	AC015813.1	35	FENDRR	AL139089.1	59	AC112496.1	ALG13-AS1
12	AC055717.2	AP001469.3	36	MIR17HG	AL161729.4	60	LINC02362	AL354993.2
13	AP001554.1	AC105460.1	37	MIR17HG	AL590483.1	61	AP000786.1	AL354993.2
14	AP001554.1	AC092112.1	38	LINC00460	AC007099.1	62	AC007099.1	LINC01748
15	MNX1-AS1	AC021218.1	39	AP002907.1	AC020978.3	63	LINC01748	LINC00513
16	AC011676.1	AC092168.2	40	MIR181A2HG	AL162724.2	64	AL442125.2	AC104964.1
17	GK-AS1	AP000866.6	41	LINC02577	AC026368.1	65	TSPOAP1-AS1	LINC00513
18	SCARNA9	AL354993.2	42	AC020907.4	AC010973.2	66	LM07-AS1	LINC00513
19	AC007128.1	B4GALT1-AS1	43	GAPLINC	LINC01876	67	AL031673.1	AC010973.2
20	AC007128.1	AC104823.1	44	ABALON	AL080317.1	68	AP001469.3	AL137782.1
21	AC245884.8	AL137782.1	45	ABALON	AC090116.1	69	LINC00941	AL108865.2
22	B4GALT1-AS1	NKILA	46	AL161729.4	AL031716.1	70	AL590483.1	AC093585.1
23	B4GALT1-AS1	AP001469.3	47	AL163953.1	AC007099.1	71	AC010973.2	LINC00513
24	AC092115.2	LINC-PINT	48	AC131971.1	AC104964.1			

DEirncRNAs = differentially expressed immune-related lncRNA pairs.



**Figure 3.** Constructing the risk assessment model with DElncRNAPs. (A) For the 1-, 2-, and 3-year survival rate ROC curves, the AUCs were all >0.95. (B) The AUC value of the 1-year survival rate ROC curve was 0.955. (C) Comparison of the DElncRNAP model with ROC curves of other clinical features shows that the DElncRNAP model has the highest accuracy. (D) AIC was used to determine the maximum inflection point of the 1-year survival rate ROC curve, and the cutoff point was 46.974. AIC = Akaike information criterion, AUCs = area under the curves, DElncRNAPs = differentially expressed immune-related lncRNA pairs, ROC = receiver operating characteristic.

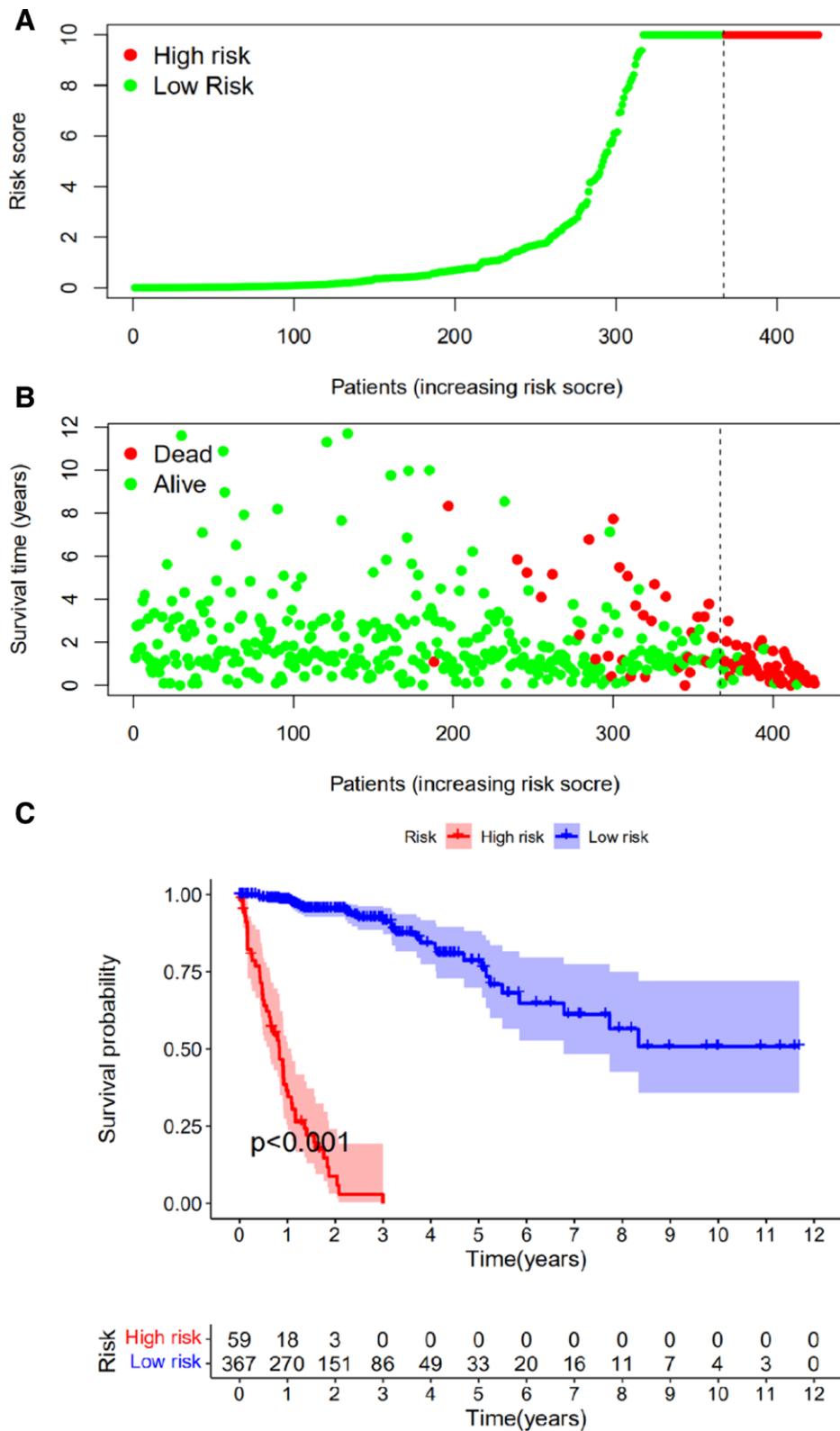
### 3.5 Study of the relationship between clinical treatment and the DElncRNAP model

Because ICIs have been shown to have a beneficial effect on the treatment of colon cancer in clinical trials, we studied whether the DElncRNAP model is related to immunosuppressive point biomarkers. The results showed that there was a positive correlation between high risk and high expression of CTLA-4 ( $P < .05$ ; Fig. 7A), BTLA ( $P > .05$ ; Fig. 7B), HAVCR-2 ( $P > .05$ ; Fig. 7C), and LAG-3 ( $P > .05$ ; Fig. 7D) expression levels were not significantly different. In addition to ICI therapy, chemotherapy is also a common treatment for colon cancer; thus, we studied the sensitivity of patients with high-risk colon cancer to common chemotherapeutic drugs. The results showed that the  $IC_{50}$  of gefitinib ( $P = .019$ ), thapsigargin ( $P = .025$ ) and vinorelbine ( $P = .012$ ) negatively correlated with high risk, but there was no significant difference with 5-fluorouracil ( $P = .42$ ) and irinotecan ( $P = .9$ ). The results showed that the risk score model

can also be used to predict drug sensitivity in COAD patients. (Fig. 7E) (Table S4, Supplemental Digital Content 5, <http://links.lww.com/MD/H265>).

## 4. Discussion

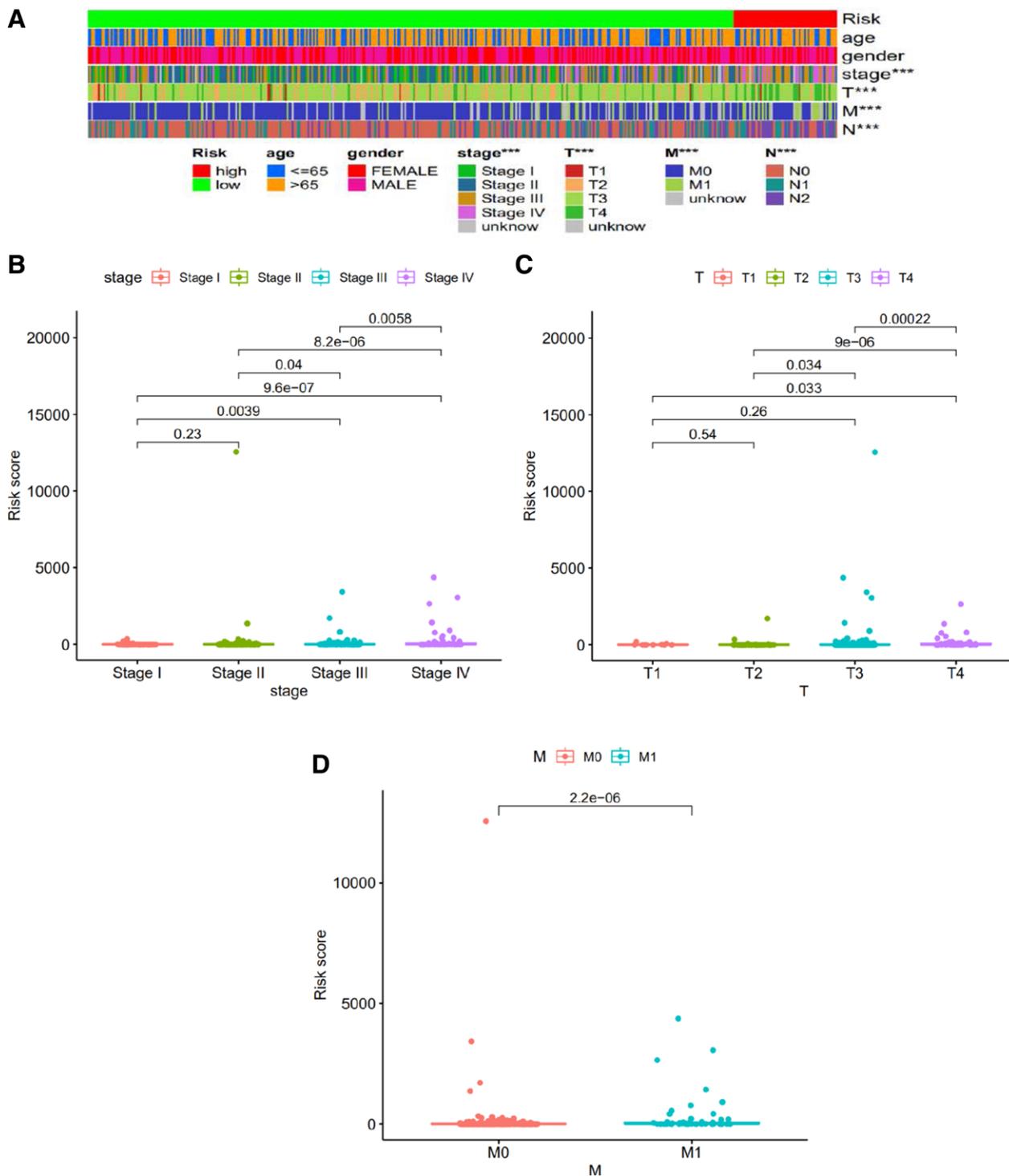
As a common malignant tumor of the digestive tract, colon cancer has the characteristics of a high incidence rate, frequent postoperative recurrence and high mortality. At present, pathophysiological evaluation, treatment decision-making and prognosis evaluation of colon cancer mainly depend on cancer cell-centered evaluation factors, such as the TNM staging system, and serum molecular markers, such as CEA and CA19-9.<sup>[21,22]</sup> Although many gene prognostic indicators have been used to predict the prognosis of patients with colon cancer, their accuracy is still uncertain.<sup>[23,24]</sup> Therefore, it is urgent to explore reliable prognostic biomarkers to predict the survival of COAD patients.



**Figure 4.** Prognostic prediction ability of the DElncRNAP model. (A) The risk score of each COAD patient; those with a risk score >46.974 were categorized as high-risk; otherwise, they were categorized as low-risk. (B) Comparison of survival time between the high- and low-risk groups. (C) Kaplan–Meier analysis indicated that the survival time of patients in the high-risk group was shorter than that of patients in the low-risk group ( $P < .001$ ). COAD = colon adenocarcinoma, DElncRNAPs = differentially expressed immune-related lncRNA pairs.

Interestingly, lncRNAs can regulate the expression of coding genes by affecting adjacent genes or distant genes on other chromosomes.<sup>[25]</sup> Because of the obvious abnormal expression of lncRNAs in the tumors of patients with colon cancer, lncRNAs such as SNHG7,<sup>[26]</sup> SNHG11,<sup>[27]</sup> and POU6F2<sup>[28]</sup> are highly specific and sensitive potential biomarkers. As a specific biomarker of colon cancer, lncRNAs have prominent

effects on diagnosis, curative effect prediction and prognosis. Importantly, it was suggested that the accuracy of a cancer diagnosis model with a combination of multiple biomarkers was significantly better than that of a single biomarker prediction model.<sup>[29]</sup> Due to the biological heterogeneity of tumors and the technical deviation caused by the use of different sequencing platforms, previous risk models of colon



**Figure 5.** Clinical validation of the DElncRNAP risk assessment model. (A) Band diagram indicating that the risk score was significantly correlated with clinical stage, T stage, M stage and N stage. Scatter charts show that the risk score is related not only to clinical stage (B), T stage (C), M stage (D), and N stage (E) but also to patient age (F). (G) The forest map suggested that there were significant differences in age, clinical stage, and risk score in the univariate Cox hazard ratio analysis. (H) The multivariate Cox hazard ratio analysis indicated that there were significant differences in age, clinical stage, and risk score. DElncRNAPs = differentially expressed immune-related lncRNA pairs.

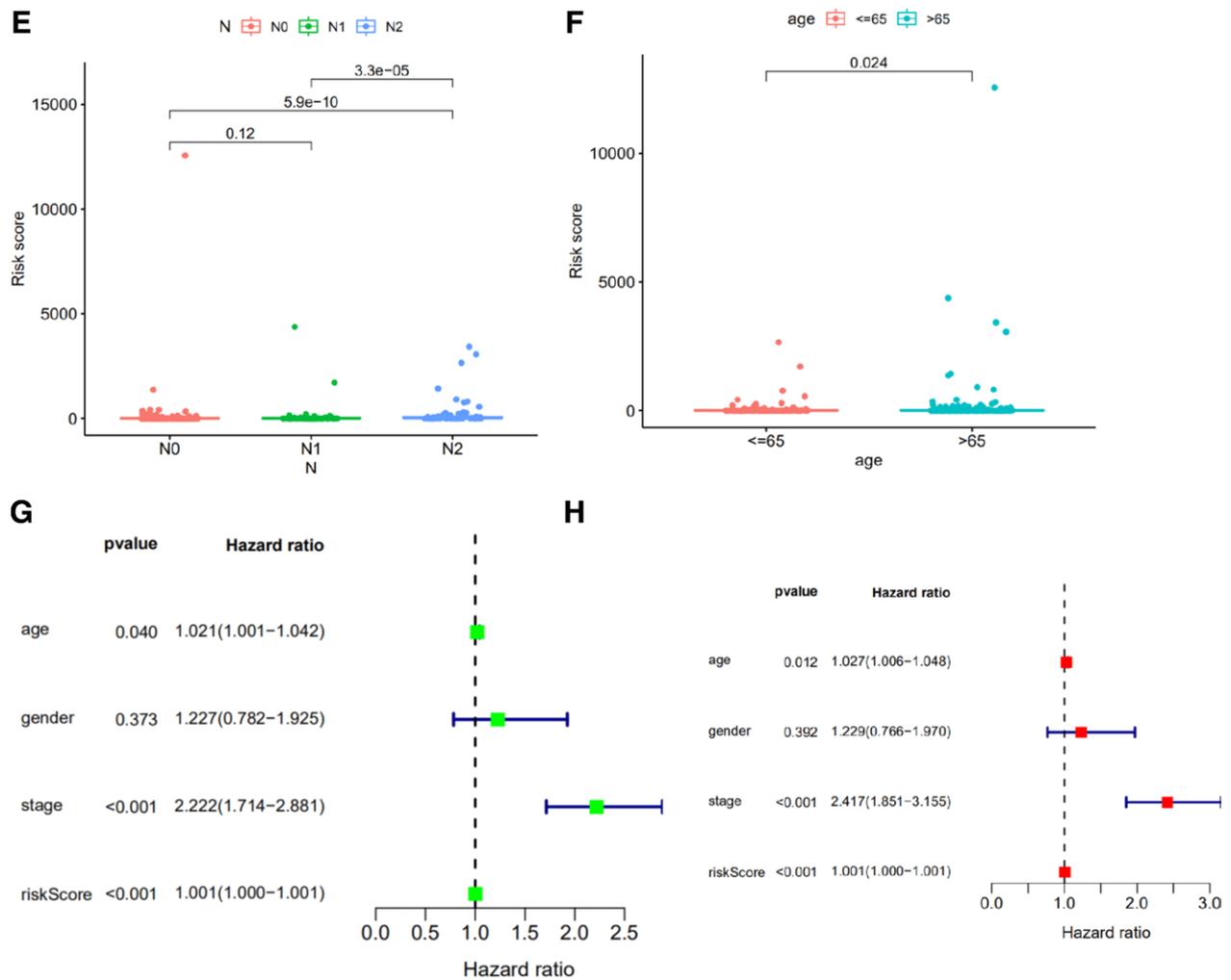


Figure 5. Continued

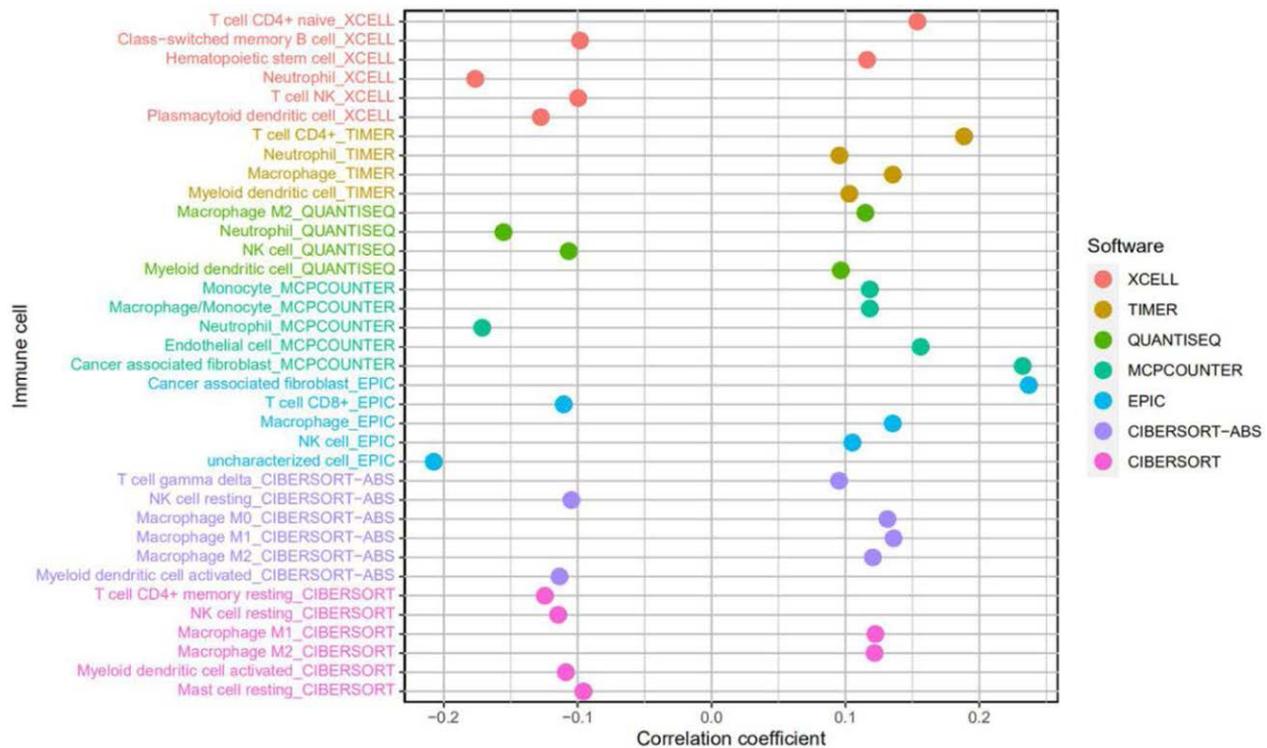
cancer need to be standardized to gene expression profiles. To achieve robustness of prediction, we used a new method without the need to consider the data deviation arising from the use of different platforms. Moreover, the new analysis method does not require the specific expression level of lncRNAs or the scaling and normalization of the expression levels of lncRNAs but rather uses relative ranking and paired comparisons of lncRNA expression levels. This method obtained reliable results in a previous study of human hepatocellular carcinoma.<sup>[13]</sup>

In this study, we identified a prognostic DEirlncRNAP model to predict the prognosis of COAD patients. The prognostic model consisted of 71 DEirlncRNAPs, including 90 independent DEirlncRNAs. It has been shown that lncRNAs are key regulators of the immune response, participate in gene activation and regulate the immunophenotype.<sup>[30,31]</sup> With the remarkable achievements of immunotherapy for cancer treatment, irlncRNAs have gradually become a new hot spot and have been found to be prognostic factors for breast cancer,<sup>[32]</sup> renal cancer<sup>[33]</sup> and pancreatic cancer.<sup>[34]</sup>

A model for predicting the prognosis of colon cancer with 10 lncRNAs was established, which suggested that lncRNAs with high expression levels have important biological functions.<sup>[35]</sup> Wu et al<sup>[36]</sup> created a prognostic model composed of 12 irlncRNAs, and the internal verification C index of the nomogram was 0.807. We obtained DEirlncRNAPs by differential coexpression analysis and established a DEirlncRNAP model

by modified Lasso penalized regression. This algorithm only requires analysis of the expression level of paired lncRNAs, and it is not necessary to calculate the specific expression level of each lncRNA, so it has higher clinical practical value than other models. Some DEirlncRNAs involved in our study have been found to play an important role in the occurrence, progression and metastasis of various cancers, such as CDKN2B-AS1 in lung cancer and cervical cancer,<sup>[37,38]</sup> MNX1-AS1 in esophageal squamous cell carcinoma and gastric cancer,<sup>[39,40]</sup> B4GALT1-AS1 in colon cancer,<sup>[41]</sup> NKILA in hepatocellular carcinoma and breast cancer<sup>[42,43]</sup> and MIR17HG in glioma and osteosarcoma.<sup>[44,45]</sup> Among them, the expression of NKILA in colorectal cancer was lower than that in normal tissues and adenomas.<sup>[46]</sup> MNX1-AS1 regulates the miR-218-5p/SEC61A1 axis to promote the development of colon cancer.<sup>[47]</sup> Both B4GALT1-AS1 and MIR17HG promote the migration of colon cancer cells.<sup>[41,48]</sup> The abnormal expression of these lncRNAs facilitates the development of colon cancer. We are the first to identify the other lncRNAs used to build this model that may be associated with cancer and could be novel biomarkers for further colon cancer research.

We used 7 common methods to evaluate the difference in infiltrating immune cells between the high- and low-risk groups, including XCELL, TIMER, QUANTISEQ, MCPOUNTER, EPIC, CIBERSORT-ABS, and CIBERSORT. Because the algorithms of each platform are complex and different, we integrated the results. The results suggested that high risk positively correlated with hematopoietic stem cells, macrophages,



**Figure 6.** Analysis of tumor infiltrating cells in the DEIRlncRNAP model. The results of Spearman correlation analysis suggested that high risk positively correlated with hematopoietic stem cells, macrophages, monocytes, endothelial cells, and cancer-associated fibroblast cells and negatively correlated with class switch memory B cells, plasmacytoid dendritic cells, CD8 + T cells and mast cells. DEIRlncRNAPs = differentially expressed immune-related lncRNA pairs.

monocytes, endothelial cells and cancer-associated fibroblast cells and negatively correlated with class-switch memory B cells, plasmacytoid dendritic cells, CD8+ T cells and mast cells. It has been reported that endothelial cells can form new tumor blood vessels and provide nutrition for tumors.<sup>[49]</sup> Macrophages induce the proliferation and migration of colon cancer cells.<sup>[50]</sup> Monocytes have also been found to be closely associated with liver metastasis of colon cancer.<sup>[51]</sup> Plasmacytoid dendritic cells produce type I interferon, which has a strong immune response to tumors and contributes to the establishment of an immunosuppressive tumor microenvironment.<sup>[52]</sup> Notably, tumor-infiltrating CD8+T cells often indicate a better immunotherapy response and prognosis.<sup>[53]</sup> It was also found that the prognosis of high-risk patients with tumor-infiltrating immune cells was poor in the DEIRlncRNAP model. Galon et al<sup>[54]</sup> proposed that an immune score based on tumor immune cell density analysis can predict the prognosis of patients and that this method is more accurate than TNM analysis.

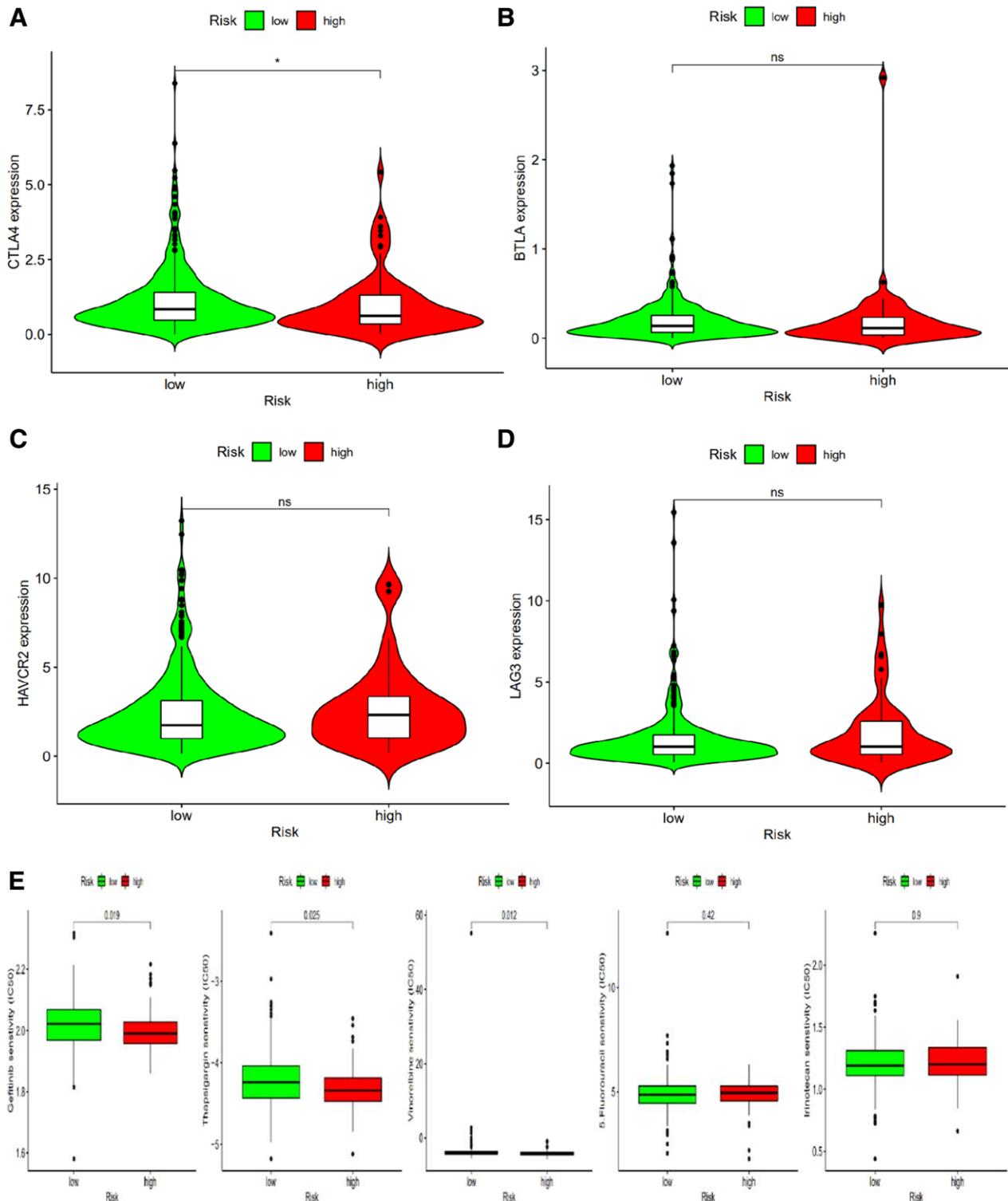
This immune score also contributes to individualized immunotherapy and chemotherapy in the clinic<sup>[55,56]</sup>; thus, we analyzed the IC<sub>50</sub> of chemotherapy drugs and immune checkpoint inhibitors in the high-risk group. The results showed that the high-risk group was more sensitive to gefitinib, thapsigargin and vinorelbine than to 5-fluorouracil and irinotecan. Liu et al<sup>[57]</sup> found that vinorelbine can inhibit the metastasis of colon cancer cells by reducing metastasis, invasion and epithelial-mesenchymal transition. This model may provide a new direction for the further study of chemotherapeutic drugs for colon cancer. However, we believe that immunotherapy can eliminate cancer cells and produce new antigens, which may have an advantage over chemotherapy in inhibiting tumor development. Although the results of our model and immune checkpoint inhibitor analysis only showed that the difference in the expression of CTLA4 was statistically significant, the tumor-infiltrating immune cells and immunophenotype of the patients were different, and the expression of the verified biomarkers should also be different.

Due to the heterogeneity of tumors, we need to effectively and accurately identify the tumor type, tumor invasion matrix, and molecular and functional characteristics of COAD, identify more biomarkers, develop different immunotherapy regimens, and carry out personalized precision treatment.

However, there are still some deficiencies in our research. The TCGA data we adopted have some defects, mainly because the data are only from a small number of countries, the sample size is limited, the data are not updated in a timely manner, and the sequencing technologies and quality control methods are different across the samples, such that all of which have a certain impact on the accuracy of the data. Moreover, we only searched the RNA-seq dataset, which does not fully represent the status of the tumor genes, and we were not able to collect the lncRNA expression level data of the colon cancer patients. The same dataset was used to build and verify the model. The prognostic efficacy and potential mechanism of this model still need to be externally verified through the analysis of real clinical data and basic research. Therefore, we used a 0-or-1 matrix to screen the qualified DEIRlncRNAPs without calculating the expression level of lncRNAs to minimize sample error. In addition, the algorithm was verified through several parameters, such as the survival time of the patients, clinicopathological features, tumor infiltrating immune cells and clinical therapeutics. Nevertheless, external validation with clinical data is still needed to confirm the reliability of the model. Therefore, more functional studies and in vitro and in vivo experiments should be carried out to test the accuracy of the DEIRlncRNAP risk assessment model to improve its clinical applicability.

## 5. Conclusion

In summary, the findings of this study provide an immune-related lncRNA model that does not require the prediction of lncRNA expression levels to predict prognosis for patients with



**Figure 7.** Clinical treatment analysis of the DEIRncRNAP model. (A) Violin plot showing that high risk negatively correlated with CTLA-4. (B-D) There was no significant difference in the expression levels of BTLA (B), HAVCR-2 (C), and LAG-3 (D) between the high- and low-risk groups. (E) Box plots showed that a high-risk score was associated with a lower IC<sub>50</sub> of chemotherapy drugs such as gefitinib, thapsigargin and vinorelbine but not 5-fluorouracil or irinotecan. DEIRncRNAPs = differentially expressed immune-related lncRNA pairs.

colon cancer and increase the understanding of the relationships of colon cancer with tumor infiltrating immune cells and antitumor immunotherapy.

### Author contributions

**Conceptualization:** Lin Yang.

**Data curation:** Xinyun Li.

**Formal analysis:** Xinyun Li, Lin Yang.

**Methodology:** Xiangshu Rao.

**Validation:** Wen Wang.

**Visualization:** Xinyun Li, Lin Yang, Yu Lai.

**Writing – original draft:** Xinyun Li, Lin Yang.

**Writing – review & editing:** Yu Lai.

### References

- [1] Sung H, Ferlay J, Siegel RL, et al. Global cancer statistics 2020: GLOBOCAN estimates of incidence and mortality worldwide for 36 cancers in 185 countries. *CA Cancer J Clin.* 2021;71:209–49.
- [2] Siegel RL, Miller KD, Goding SA, et al. Colorectal cancer statistics, 2020. *CA Cancer J Clin.* 2020;70:145–64.
- [3] Angell H, Galon J. From the immune contexture to the immunoscore: the role of prognostic and predictive immune markers in cancer. *Curr Opin Immunol.* 2013;25:261–7.
- [4] Gentles AJ, Newman AM, Liu CL, et al. The prognostic landscape of genes and infiltrating immune cells across human cancers. *Nat Med.* 2015;21:938–45.
- [5] Salama P, Phillips M, Grieu F, et al. Tumor-infiltrating FOXP3+ T regulatory cells show strong prognostic significance in colorectal cancer. *J Clin Oncol.* 2009;27:186–92.
- [6] Le DT, Durham JN, Smith KN, et al. Mismatch repair deficiency predicts response of solid tumors to PD-1 blockade. *Science.* 2017;357:409–13.
- [7] Overman MJ, McDermott R, Leach JL, et al. Nivolumab in patients with metastatic DNA mismatch repair-deficient or microsatellite instability-high colorectal cancer (CheckMate 142): an open-label, multicentre, phase 2 study. *Lancet Oncol.* 2017;18:1182–91.
- [8] Morse MA, Overman MJ, Hartmana L, et al. Safety of nivolumab plus low-dose ipilimumab in previously treated microsatellite instability-high/mismatch repair-deficient metastatic colorectal cancer. *Oncologist.* 2019;24:1453–61.
- [9] Gibb EA, Enfield KSS, Stewart GL, et al. Long non-coding RNAs are expressed in oral mucosa and altered in oral premalignant lesions. *Oral Oncol.* 2011;47:1055–61.
- [10] Tsang WP, Enders KON, Simon SMN, et al. Oncofetal H19-derived miR-675 regulates tumor suppressor RB in human colorectal cancer. *Carcinogenesis.* 2010;31:195–208.
- [11] Brunner AL, Beck AH, Edris B, et al. Transcriptional profiling of long non-coding RNAs and novel transcribed regions across a diverse panel of archived human cancers. *Genome Biol.* 2012;13:R75.
- [12] Lin Y, Pan X, Chen Z, et al. Identification of an immune-related nine-lncRNA signature predictive of overall survival in colon cancer. *Front Genet.* 2020;11:318.
- [13] Hong W, Liang L, Gu Y, et al. Immune-related lncRNA to construct novel signature and predict the immune landscape of human hepatocellular carcinoma. *Mol Ther Nucleic Acids.* 2020;22:937–47.
- [14] Chen B, Khodadoust MS, Long Liu C, et al. Profiling tumor infiltrating immune cells with CIBERSORT. *Cancer Syst Biol.* 2018;1711:243–59.
- [15] Tamminga M, Hiltermann TJN, Schuurin E, et al. Immune microenvironment composition in non-small cell lung cancer and its association with survival. *Clin Transl Immunol.* 2020;9:e1142.
- [16] Dvir A. Cell-type enrichment analysis of bulk transcriptomes using xCell. *Methods Mol Biol.* 2020;2120:263–76.
- [17] Plattner C, Finotello F, Riede D. Chapter Ten – deconvoluting tumor-infiltrating immune cells from RNA-seq data using quanTIseq. In: Galluzzi L, Rudqvist N-P, eds. *Methods in Enzymology.* Vol 636. Academic Press; 2020:261–85.
- [18] Petitprez F, Vano YA, Becht E, et al. Transcriptomic analysis of the tumor microenvironment to guide prognosis and immunotherapies. *Cancer Immunol Immunother.* 2018;67:981–8.
- [19] Racle J, Jonge K, Baumgaertner P. Simultaneous enumeration of cancer and immune cell types from bulk tumor gene expression data. *eLife.* 2017;6:e26476.
- [20] Li T, Fu J, Zeng Z, et al. TIMER2.0 for analysis of tumor-infiltrating immune cells. *Nucleic Acids Res.* 2020;48:W509–14.
- [21] Hu H, Krasinskas A, Willis J. Perspectives on current Tumor-Node-Metastasis (TNM) staging of cancers of the colon and rectum. *Semin Oncol.* 2011;38:500–10.
- [22] Liu S-X, Tan Q-Y, Song Y-Y, et al. Anti-p53 autoantibody in blood as a diagnostic biomarker for colorectal cancer: a meta-analysis. *Scand J Immunol.* 2020;91:e12829.
- [23] Barrier A, Boelle P-Y, Roser F, et al. Stage II colon cancer prognosis prediction by tumor gene expression profiling. *J Clin Oncol.* 2006;24:4685–91.
- [24] Wang Y-X, Zhang Y, Mutch MG, et al. Gene expression profiles and molecular markers to predict recurrence of Dukes' B colon cancer. *J Clin Oncol.* 2004;22:1564–71.
- [25] Krackowska W, Jagodziński PP. The long non-coding RNA landscape of atherosclerotic plaques. *Mol Diagn Ther.* 2019;23:735–49.
- [26] Liu KL, Wu J, Li WK, et al. LncRNA SNHG7 is an oncogenic biomarker interacting with MicroRNA-193b in colon carcinogenesis. *Clin Lab.* 2019;65:2199–204.
- [27] Xu W, Zhou G, Wang H, et al. Circulating lncRNA SNHG11 as a novel biomarker for early diagnosis and prognosis of colorectal cancer. *Int J Cancer.* 2020;146:2901–12.
- [28] Xu G, Zhu H, Xu J, et al. Long non-coding RNA POU6F2-AS2 promotes cell proliferation and drug resistance in colon cancer by regulating miR-377/BRD4. *J Cell Mol Med.* 2020;24:4136–49.
- [29] Lv Y, Lin SY, Hu FF, et al. Landscape of cancer diagnostic biomarkers from specifically expressed genes. *Brief Bioinform.* 2020;21:2175–84.
- [30] Imamura K, Imachi N, Akizuki G, et al. Long noncoding RNA NEAT1-dependent SFPQ relocation from promoter region to paraspeckle mediates IL8 expression upon immune stimuli. *Mol Cell.* 2014;54:1055–1055.
- [31] Rapicavoli NA, Qu K, Zhang J, et al. A mammalian pseudogene lncRNA at the interface of inflammation and anti-inflammatory therapeutics. *Elife.* 2013;2:e00762.
- [32] Liu Z, Mi M, Li X, et al. lncRNA OSTN-AS1 may represent a novel immune-related prognostic marker for triple-negative breast cancer based on integrated analysis of a ceRNA network. *Front Genet.* 2019;10:850.
- [33] Khadirnaikar S, Kumar P, Pandi SN, et al. Immune associated lncRNAs identify novel prognostic subtypes of renal clear cell carcinoma. *Mol Carcinog.* 2019;58:544–53.
- [34] Wei C, Liang Q, Li X, et al. Bioinformatics profiling utilized a nine immune-related long noncoding RNA signature as a prognostic target for pancreatic cancer. *J Cell Biochem.* 2019;120:14916–27.
- [35] Zhou W, Zhang S, Li HB, et al. Development of prognostic indicator based on autophagy-related lncRNA analysis in colon adenocarcinoma. *Biomed Res Int.* 2020;2020:9807918.
- [36] Meng Y, Li S, Gu D, et al. Genetic variants in m6A modification genes are associated with colorectal cancer risk. *Carcinogenesis.* 2020;41:8–17.
- [37] Du Y, Hao X, Liu X. Low expression of long noncoding RNA CDKN2B-AS1 in patients with idiopathic pulmonary fibrosis predicts lung cancer by regulating the p53-signaling pathway. *Oncol Lett.* 2018;15:4912–8.
- [38] Zhu L, Zhang Q, Li S, et al. Interference of the long noncoding RNA CDKN2B-AS1 upregulates miR-181a-5p/TGFβ1 axis to restrain the metastasis and promote apoptosis and senescence of cervical cancer cells. *Cancer Med.* 2019;8:1721–30.
- [39] Chu J, Li H, Xing Y, et al. LncRNA MNX1-AS1 promotes progression of esophageal squamous cell carcinoma by regulating miR-34a/SIRT1 axis. *Biomed Pharmacother.* 2019;116:109029.
- [40] Zhang W, Huang L, Lu X, et al. Upregulated expression of MNX1-AS1 long noncoding RNA predicts poor prognosis in gastric cancer. *Bosn J Basic Med Sci.* 2019;19:164–71.
- [41] Zhang Y, Fang Z, Guo X, et al. lncRNA B4GALT1-AS1 promotes colon cancer cell stemness and migration by recruiting YAP to the nucleus and enhancing YAP transcriptional activity. *J Cell Physiol.* 2019;234:18524–34.
- [42] Chen R, Cheng Q, Owusu-Ansah KG, et al. NKILA, a prognostic indicator, inhibits tumor metastasis by suppressing NF-κB/Slug mediated epithelial-mesenchymal transition in hepatocellular carcinoma. *Int J Biol Sci.* 2020;16:495–503.
- [43] Luo L-H, Rao L, Luo L-F, et al. Long non-coding RNA NKILA inhibited angiogenesis of breast cancer through NF-κB/IL-6 signaling pathway. *Microvasc Res.* 2020;129:103968.
- [44] Cao S, Zheng J, Liu X, et al. FXR1 promotes the malignant biological behavior of glioma cells via stabilizing MIR17HG. *BioMed Central.* 2019;38:37.

- [45] Meng Y, Hao D, Huang Y, et al. Positive feedback loop SP1/MIR17HG/miR-130a-3p promotes osteosarcoma proliferation and cisplatin resistance. *Biochem Biophys Res Commun.* 2020;521:739–45.
- [46] Peng J, Xiaoting H, Yingnan Z, et al. Long non-coding RNA NKILA serves as a biomarker in the early diagnosis and prognosis of patients with colorectal cancer. *Oncol Lett.* 2019;18:2109–17.
- [47] Ye Y, Gu B, Wang Y, et al. E2F1-mediated MNX1-AS1-miR-218-5p-SEC61A1 feedback loop contributes to the progression of colon adenocarcinoma. *J Cell Biochem.* 2019;120:6145–53.
- [48] Xu J, Meng Q, Li X, et al. Long noncoding RNA MIR17HG promotes colorectal cancer progression via miR-17-5p. *Cancer Res.* 2019;79:4882–95.
- [49] Carmeliet P, Jain RK. Angiogenesis in cancer and other diseases. *Nat Med.* 2000;407:249–57.
- [50] Jedinak A, Dudhgaonkar S, Sliva D. Activated macrophages induce metastatic behavior of colon cancer cells. *Immunobiology.* 2010;215:242–9.
- [51] Hu S, Zou Z, Li H, et al. The preoperative peripheral blood monocyte count is associated with liver metastasis and overall survival in colorectal cancer patients. *PLoS One.* 2016;11:e01574861–11.
- [52] Mitchell D, Chintala S, Dey M. Plasmacytoid dendritic cell in immunity and cancer. *J Neuroimmunol.* 2018;322:63–73.
- [53] Garon EB, Hellmann MD, Rizvi NA, et al. Five-year overall survival for patients with advanced non-small-cell lung cancer treated with pembrolizumab: results from the phase I KEYNOTE-001 study. *J Clin Oncol.* 2019;37:2518–27.
- [54] Galon J, Costes A, Sanchez-Cabo F, et al. Type, density, and location of immune cells within human colorectal tumors predict clinical outcome. *Science.* 2006;313:1960–4.
- [55] Frederick DT, Piris A, Cogdill AP, et al. BRAF inhibition is associated with enhanced melanoma antigen expression and a more favorable tumor microenvironment in patients with metastatic melanoma. *Clin Cancer Res.* 2013;19:1225–31.
- [56] Ma Y, Adjemian S, Mattarollo SR, et al. Anticancer chemotherapy-induced intratumoral recruitment and differentiation of antigen-presenting cells. *Immunity.* 2013;38:729–41.
- [57] Liu H, Fu Q, Lu Y, et al. Anti-tubulin agent vinorelbine inhibits metastasis of cancer cells by regulating epithelial-mesenchymal transition. *Eur J Med Chem.* 2020;200:112332.