Constructing a prognostic immune-related IncRNA model for colon cancer

Xinyun Li, MM\(^a\), Lin Yang, MM\(^b\), Wen Wang, BA\(^a\), Xiangshu Rao, MM\(^b\), Yu Lai, MD\(^b\)*

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 IncRNA 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 IncRNA pairs, DEirlncRNAs = differentially expressed irncRNAs, FC = fold-change, FDR = false discovery rate, FPKM = per kilobase per million, GTF = gene transfer format, IC\(_{50}\) = inhibitory concentration 50, ICI = immune checkpoint inhibitor, igenes = immune-related genes, irncRNA = immune-related IncRNA, IncRNAs = 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.\(^1\) More than half of the cases of colon cancer are related to smoking, unhealthy eating habits, alcoholism, physical inactivity, and obesity.\(^2\) 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.\(^3–5\) 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%.\(^6\) 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.\(^7\) In addition, the combination of navuzumab and the CTLA-4 inhibitor ipilimumab achieved 84% control of colon cancer and an 83% OS rate.
rate at 12 months, and the degree of side effects associated with the treatment was acceptable.[8]

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,[9,10] lncRNAs are ideal tumor markers with high specificity and sensitivity that are easy to repeatedly detect.[11]

A recent study showed that the immune-related lncRNA (irlncRNA) signature possessed independent prognostic significance in colon cancer.[12] According to a previous report by Hong et al.[13] 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 \beta iSi. 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 DEIrlnRNA 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 DEIrlnRNA 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 DEIrlnRNA model with other clinicopathological features in predicting the prognosis of COAD patients and to evaluate the accuracy of the DEIrlnRNA 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 DEIrlnRNA 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,[14] CIBERSORT-ABS,[15] XCELL,[16] QUANTISEQ,[17] MCPCOUNTER,[18] EPIC,[19] and TIMER.[20] 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 DEIrlnRNA model and clinical treatment

To improve the clinical relevance of the DEIrlnRNA 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 inhibition concentration 50 (IC50) of some commonly used chemotherapy drugs for patients with colon cancer. These drugs include etinbimb, thapsigargin, vinorelbine, 5-fluorouracil, and irinotecan. We used the Wilcoxon signed rank test to compare the IC50 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 IC50 value of chemotherapy drugs.

3. Results

3.1 Obtaining DEIrlnRNAs

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 irlnRNAs. Then, we analyzed the differential expression of the irlnRNAs. We obtained 894 irlnRNAs (Table S1, Supplemental Digital Content 1, http://links.lww.com/MD/H261) and 227 DEIrlnRNAs (Fig. 2A), of which the expression of 208 lncRNAs was upregulated and the expression of 19 IncRNAs was downregulated (Fig. 2B).

3.2 Obtaining DEIrlnRNAs and constructing the risk assessment model

We conducted a pairwise comparison of the DEIrlnRNAs and screened the 0-or-1 matrix and obtained 14,874 pairs of valid DEIrlnRNAs. Then, we used modified Lasso regression analysis and single factor analysis and extracted 71 DEIrlnRNAs (Table 1). We used a stepwise method and obtained 34 DEIrlnRNAs to construct a Cox proportional hazards model (Fig. 2C). Next, we used the 71 DEIrlnRNAs 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 0.935 (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 DEIrlnRNA 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 DEIrlnRNA 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 [1.001–1.001]) 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 DEIrlnRNA model

The DEIrlnRNA 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 DElncRNAs and DElncRNAPs. The DElncRNAs 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 DElncRNAPs obtained by multivariate stepwise regression analysis, which were used to construct a Cox proportional hazards model. COAD = colon adenocarcinoma, DElncRNAPs = differentially expressed immune-related IncRNA pairs, DElncRNAs = differentially expressed lncRNAs, TCGA = The Cancer Genome Atlas.
Figure 2. Continued

| Table 1 | Seventy one DEirlncRNAPs for establishing risk assessment models. |
|---------|---------------------------------------------------------------|
| 1       | AC018628.2 AC026356.1 25 AC092115.2 TSP0AP1-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 AC029783.3 29 LINC01311 AC245100.7 53 AC121761.1 LINC01876 |
| 6       | C0H026-AS1 AL42129.2 30 AL121832.3 AC010536.2 54 CD44-AS1 ARHGEF38-IT1 |
| 7       | AL13869.1 AC121761.1 31 AC145423.2 LINC01748 55 AC009032.1 AC089849.2 |
| 8       | AL13869.1 MIR3142HG 32 AL136115.2 AC004837.2 56 AC009032.1 AL590483.1 |
| 9       | AL13869.1 AL590483.1 33 AL135905.1 LINC01876 57 AC092112.1 AP005271.1 |
| 10      | AL390195.1 AL513-AS1 34 FENDRR NKILA 58 AC092112.1 AC089849.2 |
| 11      | LINC02163 AC015813.1 35 FENDRR AL139083.1 59 AC112496.1 ALG13-AS1 |
| 12      | AC055717.2 AP001469.3 36 MIR17HG AL161729.4 60 LINC03262 AL354932.2 |
| 13      | AP001554.1 AC105460.1 37 MIR17HG AL590483.1 61 AP000786.1 AL354932.2 |
| 14      | AP001554.1 AC092112.1 38 LINC00460 AC007099.1 62 AC007099.1 LINC01748 |
| 15      | MNX1-AS1 AC029783.3 39 AP002336.1 AC029783.3 63 LINC01748 LINC00513 |
| 16      | AC011676.1 AC092115.2 40 MIR181A2HG AL162724.2 64 AL42125.2 AC104964.1 |
| 17      | GK-AS1 AP000866.6 41 LINC02577 AC026368.1 65 TSP0AP1-AS1 LINC00513 |
| 18      | SCARNA9 AL354932.2 42 AC020979.4 AC010973.2 LM07-AS1 LINC00513 |
| 19      | AC007128.1 B4GALT1-AS1 43 GApLINC LINC01876 67 AC031672.1 AC10973.2 |
| 20      | AC007128.1 AC010536.2 44 GApLINC AL080317.1 68 AP001469.3 AL137782.1 |
| 21      | AC245858.8 AL137782.1 45 GApLINC AL080317.1 69 LINC0041 AC108865.2 |
| 22      | B4GALT1-AS1 NKILA 46 AL161729.4 LINC01716.1 70 AL590483.1 AC093585.1 |
| 23      | B4GALT1-AS1 AP001469.3 47 AL163953.1 AC007099.1 71 AC010973.2 LINC00513 |
| 24      | AC029115.2 LINC-PINT 48 AL131971.1 AC104964.1 |

DEirlncRNAPs = differentially expressed immune-related lncRNA pairs.
3.5 Study of the relationship between clinical treatment and the DEirlncRNAP model

Because ICIs have been shown to have a beneficial effect on the treatment of colon cancer in clinical trials, we studied whether the DEirlncRNAP 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.[21,22] Although many gene prognostic indicators have been used to predict the prognosis of patients with colon cancer, their accuracy is still uncertain.[23,24] Therefore, it is urgent to explore reliable prognostic biomarkers to predict the survival of COAD patients.
Figure 4. Prognostic prediction ability of the DEIncRNAP 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, DEIncRNAPs = differentially expressed immune-related IncRNA pairs.
Interestingly, lncRNAs can regulate the expression of coding genes by affecting adjacent genes or distant genes on other chromosomes. Because of the obvious abnormal expression of lncRNAs in the tumors of patients with colon cancer, lncRNAs such as SNHG7, SNHG11, and POU6F2 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. Due to the biological heterogeneity of tumors and the technical deviation caused by the use of different sequencing platforms, previous risk models of colon cancer.
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.[13]

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.[30,31] With the remarkable achievements of immunotherapy for cancer treatment, lirlncRNAs have gradually become a new hot spot and have been found to be prognostic factors for breast cancer,[32] renal cancer,[33] and pancreatic cancer.[34]

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.[35] Wu et al.[36] created a prognostic model composed of 12 lirlncRNAs, 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,[37,38] MNX1-AS1 in esophageal squamous cell carcinoma and gastric cancer,[39,40] B4GALT1-AS1 in colon cancer,[41] NKILA in hepatocellular carcinoma and breast cancer,[42,43] and MIR17HG in glioma and osteosarcoma.[44,45] Among them, the expression of NKILA in colorectal cancer was lower than that in normal tissues and adenomas.[46] MNX1-AS1 regulates the miR-218-5p/SEC61A1 axis to promote the development of colon cancer.[47] Both B4GALT1-AS1 and MIR17HG promote the migration of colon cancer cells.[41,48] 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, MCPCOUNTER, 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,
monocytes, endothelial cells and cancer-associated fibroblast cells and negatively correlated with class-switch memory B cells, plasmacytid 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.[49] Macrophages induce the proliferation and migration of colon cancer cells.[50] Monocytes have also been found to be closely associated with liver metastasis of colon cancer.[51] 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.[52] Notably, tumor-infiltrating CD8+ T cells often indicate a better immunotherapy response and prognosis.[53] 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.[54] 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,[55,56]; thus, we analyzed the IC_{50} 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.[57] 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 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, plasmacytid dendritic cells, CD8+ T cells and mast cells. DEirlncRNAPs = differentially expressed immune-related lncRNA pairs.
Figure 7. Clinical treatment analysis of the DEirIncRNAP 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_{50} of chemotherapy drugs such as gefitinib, thapsigargin and vinorelbine but not 5-fluorouracil or irinotecan. DEirIncRNAPs = differentially expressed immune-related IncRNA 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 S, 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, Hartman LA, 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 K, Stewart GL, et al. Long non-coding RNAs are expressed in oral mucosa and altered in oral premalignant lesions. Oral Oncol. 2011;47:653–61.

[10] Tsang WP, Enders KON, Simon MN, et al. Oncofetal H19-derived miR-675 regulates tumor suppressor BRD 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 transcripted 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 lncRNAs 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;17:243–59.

[15] Tammenga M, Hiltermann T, Ji, Schuuring E, et al. Immune microenvironment composition in non-small cell lung cancer and its association with survival. Clin Transl Immunol. 2020;9:e1142.

[16] Divir 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 immune cell enrichment analysis of bulk transcriptomes using quanTIseq. Methods in Enzymology. Vol 636. Academic Press; 2020:261–85.

[18] Pettypiece 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] Racl J, Jonge K, Baumgaertner P. Simultaneous enumeration of cancer and immune cell types from bulk tumor gene expression data. eLife. 2017;6:e24476.
[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.