Construction of immune-related IncRNA signature to predict aggressiveness, immune landscape, and drug resistance of colorectal cancer

Yonggan Xue  
Chinese PLA General Hospital

Bobin Ning  
Chinese PLA General Hospital

Hongyi Liu  
Chinese PLA General Hospital

Baoqing Jia (baoqingjia@126.com)  
Chinese PLA General Hospital

Research Article

Keywords: colorectal cancer (CRC), irlncRNA-based risk models, prognosis

Posted Date: September 13th, 2021

DOI: https://doi.org/10.21203/rs.3.rs-808813/v2

License: This work is licensed under a Creative Commons Attribution 4.0 International License. Read Full License
Abstract

Background:

Colorectal cancer (CRC) remains one of the most common malignancies across the world. Thus far, a biomarker, which can comprehensively predict the survival outcomes, clinical characteristics, and therapeutic sensitivity, is still lacking.

Methods:

We leveraged retrieved transcriptomic data of CRC from the public database, and constructed irlncRNA pairs. After integrating with clinical survival data, we performed differential analysis and constructed 11 irlncRNA pairs signature using Lasso regression analysis. We next drew the 1-, 5-, 10-year curve line of receiver operating characteristics, calculated the areas under the curve, and recognized the optimal cutoff point. Then, we validated the pair-risk model in terms of the survival outcomes of the patients. Moreover, we tested the reliability of the pair-risk model for predicting tumor aggressiveness and therapeutic responsiveness of CRC. Additionally, we reemployed the 11 of irlncRNAs involved in the pair-risk model to construct an expression risk model that was also highly predictive of prognostic outcomes of CRC patients.

Results:

We recognized a total of 377 DEirlcRNAs, including 28 low-expressed and 349 high-expressed irlncRNAs in CRC patients. After performing a univariant Cox analysis, we identified 115 risk irlncRNAs that were significantly correlated with survival outcomes of patients with CRC. By taking the intersection of the DEirlcRNAs and the risk irlncRNAs, we ultimately recognized 55 irlncRNA as core irlncRNAs. Then, we established a Cox HR model (pair-risk model) as well as an expression HR model (exp-risk model). We found that both of the two models significantly outperformed the common-used clinical characteristics, including age, T, N, and M stages, in terms of predicting survival outcomes. Moreover, we validated the pair-risk model can serve as a potential tool for studying the tumor microenvironment of CRC and drug response. Additionally, we noticed that combining the pair-risk model and exp-risk model yielded a more robust approach for predicting the survival outcomes of patients with CRC.

Conclusions:

We suggest that the irlncRNA-based risk models can be utilized as prognostic tools to predict survival outcomes and clinical characteristics and guide treatment regimens of CRC.

Background

Although remarkable progress has been achieved in tackling CRC over the past decades, it remains the second in terms of mortality and the third in terms of incidence [1]. It is reported that CRC takes a toll on almost 700,000 people every year [2]. The global burden of CRC is predicted to increase to 2.2 million new
cases by 2030, with 1.1 million deaths from the disease [3]. It is well established that both genetic predisposition and environmental factors play a role in developing CRC [4]. Nearly a fifth of patients with CRC might have a positive family history of gene-related diseases like familial adenomatous polyposis and hereditary non-polyposis CRC, which are associated with this cancer [5]. Environmental factors, such as smoking, excessive alcohol consumption, and increased red meat intake, are mostly considered contributors to CRC [6]. While the highest rates of incidence are found in developed countries, the incidence of CRC appears to rapidly increase in developing nations due to more people shifting to Western dietary patterns and lifestyles [7].

Surgical resection of the tumor is still considered the cornerstone of curative-intent treatment. Benefiting from the improving surgical performance and equipment, some patients, especially those who are at early stages, can be cured without the intervention like radiotherapy and chemotherapy [8]. By contrast, for those who are diagnosed with a local advanced or distant metastatic tumor, systemic treatment, like adjuvant chemotherapy and immune therapy, should be considered or even prioritized [9]. However, the hard question facing the gastroenterologists and oncologists during the treatment of CRC is how to stratify patients that are suitable to a specific regimen. Therefore, it is urgently required a parameter to robustly predict the clinical characteristics and drug sensitiveness of patients with CRC.

Long non-coding RNA (lncRNA), referred to as transcripts larger than 200 nucleotides, are transcribed by RNA polymerase II but not translated into proteins [10]. Biologically, lncRNA exerts a variety of functions at different levels of gene expression, including transcription, post-transcription, and chromatin modification [11]. When it comes to their role in cancer, lncRNA plays an important role in the process of tumorigenesis and cancer cell evolution via their influence on gene expression, immune response, and drug resistance [12]. Recently, with the improved acknowledge of the tumor microenvironment (TME) and the prevalent implementation of immune therapeutics, such as anti-PD-1 and anti-PD-L1 checkpoint blockade, immune-related lncRNA has increasingly attracted researchers’ attention and is expected to be the next hotspot of tumor research.

Previous studies suggested that lncRNA expression patterns can induce tumorigenesis, regulate immune landscape, and alter survival outcomes in various types of cancer. Shi et al. reported that lncRNA KIMAT1 facilitates lung cancer tumorigenesis and progression via sustaining KRAS signaling by modulating miRNA processing [13]. Overexpression of lncRNA, LINC01503, was identified to drive the growth and metastasis of a variety of squamous cell carcinoma, including esophageal and head and neck, through activating ERK and ATK signaling [14]. The influence of lincRNA expression profiles on CRC is rarely evaluated. In this study, the lincRNA expression signature in CRC was analyzed from the TCGA dataset. After a series of analyses and validation, we eventually established an 11 lincRNA pairs-based risk assessment model that is closely associated with tumor aggressiveness, survival, and drug resistance of CRC. Additionally, based on the same pool of core lincRNAs, we constructed an expression risk model that is indicative of clinical traits and prognostic outcomes. Importantly, by integrating those two risk models, we could more robustly foresee the survival outcomes of patients with CRC.
Methods

Data download and processing

The transcriptomic data including 473 tumor samples and 41 para-cancer normal samples, as well as corresponding clinical materials on COAD, were downloaded from the TCGA (https://portal.gdc.cancer.gov/). For subsequent analysis, the lncRNA and mRNA were distinguished into two separate matrices according to GTF files obtained from Ensemble (http://asia.ensembl.org).

Immune-related genes and irlncRNA

A total of 2483 immune genes were identified from the IMMPORT (https://www.immport.org) and the corresponding expression matrix was extracted by Perl language. Then a co-expression analysis was conducted between immune-related genes and the lncRNAs through the R package “limma”. The lncRNAs with the filtering conditions of Pearson correlation coefficient >0.5 and p value <0.001 were closely related to the immune-related genes and regarded as irlncRNA.

Differential and survival analysis of irlncRNAs and construction of irlncRNA pairs

Differential analysis of the irlncRNAs was conducted by “limma”. The thresholds were configured as logFC>2 and pvalue<0.05. Focusing on the irlncRNAs, a univariate Cox analysis was conducted. The intersection was performed between the DEirlncRNA and RISKirlncRNA. R packages “pheatmap” and “survival” were utilized for the operation. To reduce the complex calculations caused by data correction and its resulting bias, the core irlncRNAs were circularly selected, and a 0-or-1 matrix was constructed. For each cycle, two IncRNAs were selected and their expression levels were compared. Assuming C is defined as 1 if the expression level of irlncRNA A is higher than irlncRNA B, otherwise, C is defined as 0. The irlncRNA pair was regarded as a valid match only when the proportion of irlncRNA-pairs with an expression of 0 or 1 accounted for more than 20% of the total pairs.

Establishment of irlncRNA-pairs signature

After merging the irlncRNA pairs and the associated clinical profiles, an uniCox survival analysis of irlncRNA pairs through the R package “survivor” P value < 0.05 were identified as prognostically relevant. After obtaining the prognosis-related immune IncRNA pairs, we further cross-validated the model and operated a 1000-times-repeated Lasso regression model to filter the ultimate irlncRNA pairs without redundant information and construct the combination of irlncRNA pairs in cross-validation by R package “glmnet”. Then the package “survminer” was performed to construct the uni- and multi- Cox analysis to obtain the optimized risk formula.

Validation of the model and clinical relevance

To verify the stability of the signature, we perform the ROC curves for 1-year, 5-year, and 10-year survival using the R package “survivalROC”, following which the best cutoff point based on the 5-year survival
was identified based on riskScore formula: RiskScore. The COAD patients from the TCGA database were divided into high- and low-risk groups based on the cutoff that maximizes the area under the curve (AUC). To further visualize the risk model, we arranged the patients in terms of their risk values and presented their survival status with different colored dots. Then we generated the 12-year K-M survival curves for two groups of patients by the R package “survivor” and “survminer” to compare the survival differences. The package “survivor” in R language was used to calculate the risk value, survival time, and survival status of each CRC patient, and a p value less than 0.05 was considered statistically significant. We then used the package “survivalROC” to input both the clinical information and the risk model to compare the accuracy of the risk model with other clinical traits. Using the “limma” and “ggpubr” in the R package, we further revealed the correlation between the risk scores and the different clinical traits. Clinical traits are also observed between high- and low-risk groups by virtue of chi-square tests, and the statistically significant results were shown in the clinical correlation heat map by the R package “ComplexHeatmap”.

**Immunocorrelation analysis**

To explore the correlation between TME and the lncRNA pairs model, we calculated the stromal and immune infiltration cells among the CRC patients by seven acknowledged algorithms including XCELL, TIMER, QUANTISEQ, MCPCOUNTER, EPIC, CIBERSORT-ABS, and CIBERSORT. R packages “limma”, “scales”, “ggplot2”, and “ggtext” were used to present these results. Then, a Wilcoxon signed-rank test was performed to analyze the immuno-differences between high- and low-risk groups by “limma” and “ggpubr” packages.

**Correlation analysis of immune genes**

TGFB1 and LAG3 were typical immune checkpoints, which were regarded as having a critical role in immune therapy. The distinct expressions of them between the high- and low-risk groups were analyzed by “limma” and demonstrated by “ggpubr”. We labeled *** if the P value was less than 0.001, ** if the P value was less than 0.01, and * if the P value was less than 0.05.

**Drug sensitivity analysis**

To identify the relationship between this model and therapeutic efficacy, we performed a chemotherapeutic analysis of the CRC patients from the TCGA database by calculating the IC50 which has a strong relationship with drug sensitivity. With the packages “limma”, “ggpubr”, “pRRophetic”, and “ggplot2”, drug sensitivities between the high- and low-risk groups were calculated by Wilcoxon signed-rank test and the statistically significant results were shown as box drawings.

**Expression risk model construction**

After identifying the irlncRNAs in colon cancer, we construct a exp-risk model to obtain the risk values of each patient and divide them into high-risk and low-risk groups. By using "survival" and "survminer", we can compare whether there is a difference in survival between the two groups. We can also use the ROC curve to predict the accuracy of our model. Next, we can use the "pheatmap" package to plot risk curves.
To make our study more convincing, we combined pair-risk score and expression-risk score, and divided the patients into 4 subgroups according to their different risk values in the two groups, and used the "survminer" and "survivor" packages to statistical analysis and presentation of the survival of patients in each subgroup was performed using the "survminer" and "survival" packages.

**Independent prognostic analysis**

Next, with the "survival" package, we performed an independent prognostic analysis. We compared this model with our patient's clinical data to our prognosis to determine if the lncRNA-based model can be used as an independent prognostic factor in colon cancer.

**GSEA enrichment analysis**

Finally, the GSEA enrichment analysis allowed us to determine in which gene sets were enriched in the high- and low-risk groups.

**Results**

**Identification of differentially expressed lincRNAs (DElncRNAs) and risk lncRNA in CRC**

The overall workflow is illustrated in Figure 1. Briefly, after collecting transcriptional profiles from the publicly available database, we annotated those transcriptomic data based on GTF from Ensembl. Considering it is difficult to directly find out the lncRNA identity, Pearson Correlation Analysis was conducted and a total of 739 lncRNAs were defined with the setting correlation coefficients > 0.5 (Table S1). We next performed differentially expressed analysis to evaluate the differences in those lncRNAs between cancerous and paraneoplastic tissues. We recognized a total of 377 DElncRNAs, including 28 low-expressed and 349 high-expressed lncRNAs in CRC patients, with the filter condition setting to logFC > 2 and P value < 0.05 (Figure 2A). We then performed a univariate Cox analysis and identified 115 risk lncRNAs that were significantly correlated with survival outcomes of patients with CRC. By taking the intersection of the DElncRNAs and the risk lncRNAs, we ultimately recognized 55 lncRNA as core lncRNAs (Figure 2C, Table S1). Consequently, we then constructed a Cox HR model (pair-risk model) containing 11 core lncRNA pairs as well as an expression HR model (exp-risk model) containing 11 core lncRNAs (Table 1, 2). Eventually, we, separately, examined the relationship between the risk scores obtained from the two models and disease-related characteristics, including clinical traits, survival states, immune microenvironment, chemotherapeutic sensitivity, and signaling pathways.

**Construction of core lncRNA pairs and establish of a pair-risk model**

To ensure that our model can be adopted without considering the differences between the profiles, we paired those core lncRNAs and then correlated them with the clinical survival information. With the setting condition of p value < 0.05 in the univariate cox analysis, a total of 75 pairs of core lncRNAs were determined to further construct the risk assessment model (Table S2). After Lasso regression analysis of these 75 core lncRNA pairs, 11 of them were finally selected for constructing the pair-risk
model (Figure 3A, 3B). We then performed univariate and multivariate Cox analyses of these 11 core irlncRNA pairs to confirm their ability to serve as independent clinical prognostic factors (Figure 3C, 3D).

**Evaluation of the clinical predictivity of the pair-risk model**

We drew the 1-year ROC curve and found the area under the curve (AUC) is 0.789, indicating that the model can be recognized with high accuracy to reflect the short-term overall survival of CRC patients (Figure 4A). To further validate the stability and generalizability of the model, we plotted the 1-, 5-, and 10-year ROC curves (Figure 4B). We noticed that the 5-year ROC curve was optimal with the best cutoff value of 0.944 (Figure 4C). Based on this optimal cutoff value, we then categorized the CRC patients from the TCGA database into two high- and low-risk groups. Of these, 237 patients were classified into the high-risk group, while the remaining 209 patients were allocated to the low-risk group. Next, by virtue of multi-metric ROC curves, we plotted the ROC curve of the risk model together with that of clinical characteristics including age, T, N, and M stages in the identical diagram for comparison. The results showed that the AUC value of the risk model dramatically surpassed that of other clinical parameters, indicative of the high performance of this model (Figure 4D).

The risk score of individual patients and their survival statuses were illustrated in Figures 5A and 5B, from which we can recognize apparent differences in survival outcomes between the two subgroups of CRC patients. Concordantly, we conducted the Kaplan-Meier analysis and observed that the life expectancy of patients in the high-risk group was substantially decreased when compared to patients in the low-risk group (P value < 0.001) (Figure 5C). To estimate the relationship between the risk model and the clinic traits, we charted a clinic-related heat map using the chi-square test. Our findings suggested that the T grade and N grade were significantly different between the two groups with p-values less than 0.01 and 0.001, respectively (Figure 5D). We next examined the risk score of patients across different clinic traits including tumor stage, T grade, and N grade, respectively. We observed that the differences across tumor stage, between distinct T grades, as well as different N grades were all of the statistical significance (Figure 5E-G). Finally, we conducted the univariate and multivariate cox analyses and found the risk assessment model can serve as an independent prognostic predictor (Figure 5H and 5I).

Taken together, the irlncRNAs-based pair-risk assessment model not least can be utilized as a robust predictor of the survival outcomes, but also serve as a reliable indicator of tumor aggressiveness of the CRC patients.

**Assessment of the pair-risk model with tumor immune microenvironment**

To probe whether this model can reflect the tumor-infiltrating immune cells in the context of CRC, we first performed the correlation analysis of immune cells via 7 different algorithms, which is illustrated with distinct colors (Figure 6A). We observed that granulocytes, M1 macrophage, dendric cell, neutrophil, Th1 cell, CD8+ T cell were positively correlated with the risk value of the model. On the contrary, B cells, hematopoietic stem cell, neutrophil, NK cell resting, T cell CD4+ memory resting showed a negative
correlation (Figure 6B-X). This means that the risk model can be applied to predict the landscape of immune microenvironment in patients with CRC.

**Estimation of the chemotherapeutic responsiveness of CRC by the pair-risk model**

To investigate whether this risk model can be employed to predict the chemotherapeutic sensitivity for CRC patients, we examined the capacity of our risk model for forecasting drug sensitivity by analyzing the immune-related gene expression as well as the differences of IC50 between the high- and low-risk groups. Our observations indicated that some immune-correlative genes, such as TGFB2 and LAG3, were shown a significantly positive correlation with risk score value (Figure 7A and B). Furthermore, we identified 5 chemotherapeutic drugs, including CCT007093, CCT018159, CGP.60474, CGP.082996, JNK.9L, that appeared to have an inverse correlation between risk score and IC50 value (Figure 7C-G), implicating that our risk model can serve as a tool to assist the chemotherapeutic medication.

**Construction and validation of an exp-risk model based on the selected core lirncRNAs**

We next constructed an lirncRNA exp-risk model based on the 11 core lirncRNAs. Again, we divided the patients into a high-risk group and a low-risk group based on the median risk value of the expression risk model (Figure 8A). Notably, patients in the high-risk group tended to have an apparent shorter life expectancy when comparing to those in the low-risk group (Figure 8B). Moreover, we illustrated the expression levels of the core IncRNAs involved in those risk groups (Figure 8C). Additionally, we plotted the Kaplan-Meier curves comparing the survival outcomes of the groups and observed consistent results (Figure 8D).

**Independent prognostic analysis of the exp-risk model**

To evaluate whether this exp-risk model can be applied to predict prognostic outcomes of patients with CRC, we performed univariate independent prognostic analysis. Our observations suggested that the risk score, alongside tumor stage, T grade, N grade, and M grade, was significant associated with survival outcomes of CRC patients (Figure 9A). Moreover, we performed the multifactorial independent prognostic analysis and recognized that only the risk score could serve as an independent high-risk factor for predicting the prognosis of colorectal cancer (Figure 9B). Then, we plotted the curve of the exp-risk model together with the curves of other clinic traits. By comparing the AUC of the exp-risk model to those of other clinic traits, we found the AUC value of this risk model is 0.751, while the AUC values obtained for other clinical traits are less than 0.75, indicating that this model outperforms other clinical traits for predicting the survival outcomes of CRC patients (Figure 9C).

Additionally, to explore whether combining the pairs-risk model with the exp-risk model could produce a more robust prognostic predictability, we categorized the patients involved into four groups: high pair score + high expression score; high pair score + low expression score; low pair score + high expression score; low pair score + low expression score. As expected, we found patients in the high pair score + high
expression score group had the dramatically worst life expectancy, with a 5-year survival rate of only around 25% (Figure 9D).

**GSEA enrichment analysis**

Finally, we performed a GSEA enrichment analysis and found that immune-related signaling and cancer-associated signaling pathways were significantly overactivated in the high-risk group. These results further confirmed the accuracy and reliability of the exp-risk model (Figure 10A-L).

**Discussion**

The advent of high-throughput technologies and the evolution of conventional approaches have left a burst of non-coding RNA recognition and exploration. Accumulating studies indicate that lncRNA participates in the development of malignant tumors via regulating the construction of the tumor immune microenvironment [15]. Furthermore, a wealth of scientists has engaged in building up a model to predict the prognosis of cancer patients based on quantifying the expression levels of lncRNA. Hong and colleagues constructed a risk assessment model based on the expression pattern of irlncRNA to predict the immune landscape and survival outcomes of hepatocellular carcinoma [16]. Zhong et al. investigated RNA sequencing profiles and identified a prognostic irlncRNA-scoring system and three molecular subtypes in clear cell renal cell carcinoma [17]. Similarly, Zhou et al. discovered an irlncRNA biomarker that can be utilized not least to predict prognostic outcomes of patients with diffuse large B cell lymphoma but can distinguish germinal center B-cell-like and activated B-cell-like subtypes [18]. However, due to the discrepancy of samples among different databases, the lncRNA-based risk assessment models mentioned above can hardly be transplanted to prevalent clinical practice. In the present study, we explored setting up a risk assessment model based on the irlncRNA pairing approach that circumvents the external interference among individual patients and, simultaneously, has the potential to foresee the aggressiveness, immune signature, and drug sensitivity of CRC.

Growing evidence suggests that lncRNAs participate in a variety of immune-related biological activities. For example, Carpenter et al. reported that LincRNA-Cox2 acts as an essential mediator regulating both activation and repression of a series of immune gene expression in macrophages [19]. Huang et al. found that lncRNA MorbId subtly controls the life expectancy of neutrophils, eosinophils, and monocytes through regulating the expression of Bim in response to extracellular stimulation [20]. In the context of the tumor, Li et al. performed a comprehensive examination of the landscape of lncRNAs across 33 cancer types and found that the expression of lncRNAs tended to show a cancer-type specific pattern. Moreover, the perturbed transcription of lncRNA was significantly associated with the signature of immune cell infiltration [21].

In this study, we initially sorted out the differentially expressed immune-related gene and coordinated those genes with lncRNA transcription profiles to identify irlncRNAs. After correlating with survival data, we determined the differentially expressed irlncRNAs and established the most significant irlncRNA pairs. Instead of evaluating the specific expression values of individual lncRNAs, we only needed to consider
the relative value of an lncRNA pair. Thus, this novel strategy bears a superiority of feasibility and generalizability to clinical practice among different patients. After stratifying the patients into high- and low-risk groups in the light of the cutoff value, we reevaluated the survival outcomes, clinical characteristics, tumor microenvironment components, immune-related gene expression, and chemotherapeutical sensitivity of CRC between the two groups. Our findings suggested that the pair-risk model was significantly associated with the survival rates and clinical traits, such as tumor aggressiveness, of CRC, implying the reliable performance of this modeling algorithm.

It is well documented that the tumor microenvironment (TME) plays a crucial role in the development and progression of malignant carcinoma [22]. Immune infiltrates make up an essential part of the tumor microenvironment, exerting various functions that impact a series of bioactivities of cancer cells, such as promoting immune escape, accelerating tumor progression, and boosting drug resistance [23]. The composition of the tumor microenvironment is closely associated with the dynamics of tumorigenesis, malignant progression, and chemotherapeutic response [22, 24]. Intertumoral immune cells and fibroblasts constitute the major part of the TME and play crucial roles in determining the biological activities of cancer cells [25–27]. Therefore, improved acknowledge of the architecture of tumor components, such as immune cells and fibroblasts, can help oncologists to choose a suitable therapeutic regimen to treat cancer patients [28]. To determine the relationship between risk scores and tumor microenvironment components, we used seven distinct computational methods. Our observations indicated that the pair-risk score was robustly indicative of landscape of immune infiltration of CRC.

Finally, we scrutinized the performance of this model to predict antitumor therapeutic resistance and found statistically significant differences in immune-related gene expression and drug efficacy between high- and low-risk groups. For example, activation of JNK signaling has been confirmed to render chemotherapy in a variety of malignancies, including hepatocellular carcinoma [29], gastric cancer [30], and pancreatic cancer [31]. In the setting of colon cancer, activating JNK signaling pathway contributes to 5-fluorouracil resistance in p53-defective or mutant colon cancer cells by inducing pro-survival autophagy [32]. Our findings suggest that a high score of lncRNA pairs is positively associated with JNN inhibitor resistance. Compensatory to previous investigations, this lncRNA-based risk assessment model can serve as a chemotherapeutics-selecting tool to assist clinicians to formulate suitable regimens.

Additionally, we reconstructed an expression risk model by using the same pool of core lncRNAs and found that the exp-risk model was also predictive of survival outcomes of patients with CRC with high performance. Importantly, the combination of pair-risk model and exp-risk model can yield a more robust strategy to forecast the prognostic outcomes of CRC patients. Those results mean that the lncRNAs involved might play a vital role in the progression of CRC, which further highlights the potential of studying those lncRNAs to obtain deeper insight into the evolution of the disease.

However, the present study has some shortcomings and limitations. For example, the risk assessment models were constructed entirely based on the raw dataset from TCGA, but not validated by the clinical samples, implying the lack of clinical evidence. Additionally, this IncRNA pairs-based risk model can
merely be acted as a biomarker to predict survival outcomes, tumor aggressiveness, and therapeutic resistance, whereas it failed to spot the specific biological function of individual lncRNAs involved in this model.

**Conclusion**

In conclusion, this study suggested a pair-risk assessment model, which was built based on 11 pairs of irlncRNAs. This risk model is not merely associated with the survival outcomes of the patients with CRC, but also linked to the tumor grade, tumor microenvironment, and chemotherapeutic resistance as well. We reused the core irlncRNAs involved to establish an exp-risk model, which is also indicative of survival outcomes of CRC patients. The identification of those models provides alternative measures to predict the biological characteristics of CRC and guide the treatment scheme.

**Abbreviations**

CRC: Colorectal cancer; irlncRNA: immune-related lncRNA; TCGA: The Cancer Genome Atlas; ROC: Receiver operating characteristic; AUC: Areas under the curve; IncRNA: Long non-coding RNA; TME: The tumor microenvironment; DEirIncRNA: Different expressed immune-related lncRNA.

**Declarations**

**Ethics approval and consent to participate**

Not applicable

**Consent for publication**

Not applicable

**Availability of data and materials**

The datasets analyzed during the current study are available in TCGA (https://portal.gdc.cancer.gov/).

**Conflict of Interest:** The authors declare no conflict of interest.

**Funding:** This research did not receive any specific grant from funding agencies in the public, commercial, or not-for-profit sectors.

**Autor contributions:** Yonggan Xue: Validation, Writing - Original Draft, Writing - Review & Editing, Bobin Ning: Methodology, Software, Formal analysis, Hongyi Liu: Resources, Data Curation, Visualization, Baoqing Jia: Conceptualization, Supervision, Project administration.

**Acknowledgment:** None
References

1. Sung H, Ferlay J, Siegel RL, Laversanne M, Soerjomataram I, Jemal A, Bray F: Global cancer statistics 2020: GLOBOCAN estimates of incidence and mortality worldwide for 36 cancers in 185 countries. CA Cancer J Clin 2021.

2. Brody H: Colorectal cancer. Nature 2015, 521(7551):S1.

3. Arnold M, Sierra MS, Laversanne M, Soerjomataram I, Jemal A, Bray F: Global patterns and trends in colorectal cancer incidence and mortality. Gut 2017, 66(4):683–691.

4. Brenner H, Kloor M, Pox CP: Colorectal cancer. Lancet 2014, 383(9927):1490–1502.

5. Henrikson NB, Webber EM, Goddard KA, Scrol A, Piper M, Williams MS, Zallen DT, Calonge N, Ganiats TG, Janssens AC et al: Family history and the natural history of colorectal cancer: systematic review. Genet Med 2015, 17(9):702–712.

6. Dekker E, Tanis PJ, Vleugels JLA, Kasi PM, Wallace MB: Colorectal cancer. The Lancet 2019, 394(10207):1467–1480.

7. Ferlay J, Colombet M, Soerjomataram I, Dyba T, Randi G, Bettio M, Gavin A, Visser O, Bray F: Cancer incidence and mortality patterns in Europe: Estimates for 40 countries and 25 major cancers in 2018. Eur J Cancer 2018, 103:356–387.

8. Buccafusca G, Proserpio I, Tralongo AC, Rametta Giuliano S, Tralongo P: Early colorectal cancer: diagnosis, treatment and survivorship care. Crit Rev Oncol Hematol 2019, 136:20–30.

9. Modest DP, Pant S, Sartore-Bianchi A: Treatment sequencing in metastatic colorectal cancer. Eur J Cancer 2019, 109:70–83.

10. Huarte M: The emerging role of IncRNAs in cancer. Nat Med 2015, 21(11):1253–1261.

11. Kopp F, Mendell JT: Functional Classification and Experimental Dissection of Long Noncoding RNAs. Cell 2018, 172(3):393–407.

12. Bhan A, Soleimani M, Mandal SS: Long Noncoding RNA and Cancer: A New Paradigm. Cancer Res 2017, 77(15):3965–3981.

13. Shi L, Magee P, Fassan M, Sahoo S, Leong HS, Lee D, Sellers R, Brullé-Soumaré L, Cairo S, Monteverde T et al: A KRAS-responsive long non-coding RNA controls microRNA processing. Nat Commun 2021, 12(1):2038.

14. Xie JJ, Jiang YY, Jiang Y, Li CQ, Lim MC, An O, Mayakonda A, Ding LW, Long L, Sun C et al: SuperEnhancer-Driven Long Non-Coding RNA LINC01503, Regulated by TP63, Is Over-Expressed and Oncogenic in Squamous Cell Carcinoma. Gastroenterology 2018, 154(8):2137–2151.e2131.
15. Denaro N, Merlano MC, Lo Nigro C: Long noncoding RNAs as regulators of cancer immunity. Mol Oncol 2019, 13(1):61–73.

16. Hong W, Liang L, Gu Y, Qi Z, Qiu H, Yang X, Zeng W, Ma L, Xie J: Immune-Related lncRNA to Construct Novel Signature and Predict the Immune Landscape of Human Hepatocellular Carcinoma. Mol Ther Nucleic Acids 2020, 22:937–947.

17. Zhong W, Chen B, Zhong H, Huang C, Lin J, Zhu M, Chen M, Lin Y, Lin Y, Huang J: Identification of 12 immune-related lncRNAs and molecular subtypes for the clear cell renal cell carcinoma based on RNA sequencing data. Sci Rep 2020, 10(1):14412.

18. Zhou M, Zhao H, Xu W, Bao S, Cheng L, Sun J: Discovery and validation of immune-associated long non-coding RNA biomarkers associated with clinically molecular subtype and prognosis in diffuse large B cell lymphoma. Mol Cancer 2017, 16(1):16.

19. Carpenter S, Aiello D, Atianand MK, Ricci EP, Gandhi P, Hall LL, Byron M, Monks B, Henry-Bezy M, Lawrence JB et al: A long noncoding RNA mediates both activation and repression of immune response genes. Science 2013, 341(6147):789–792.

20. Huang W, Thomas B, Flynn RA, Gavzy SJ, Wu L, Kim SV, Hall JA, Miraldi ER, Ng CP, Rigo F et al: DDX5 and its associated IncRNA Rmrp modulate TH17 cell effector functions. Nature 2015, 528(7583):517–522.

21. Li Y, Jiang T, Zhou W, Li J, Li X, Wang Q, Jin X, Yin J, Chen L, Zhang Y et al: Pan-cancer characterization of immune-related lncRNAs identifies potential oncogenic biomarkers. Nat Commun 2020, 11(1):1000.

22. Hinshaw DC, Shevde LA: The Tumor Microenvironment Innately Modulates Cancer Progression. Cancer Res 2019, 79(18):4557–4566.

23. Lei X, Lei Y, Li JK, Du WX, Li RG, Yang J, Li J, Li F, Tan HB: Immune cells within the tumor microenvironment: Biological functions and roles in cancer immunotherapy. Cancer Lett 2020, 470:126–133.

24. Denton AE, Roberts EW, Fearon DT: Stromal Cells in the Tumor Microenvironment. Adv Exp Med Biol 2018, 1060:99–114.

25. Gajewski TF, Schreiber H, Fu YX: Innate and adaptive immune cells in the tumor microenvironment. Nat Immunol 2013, 14(10):1014–1022.

26. Ishii G, Ochiai A, Neri S: Phenotypic and functional heterogeneity of cancer-associated fibroblast within the tumor microenvironment. Adv Drug Deliv Rev 2016, 99(Pt B):186–196.

27. Locy H, de Mey S, de Mey W, De Ridder M, Thielemans K, Maenhout SK: Immunomodulation of the Tumor Microenvironment: Tum Foe Into Friend. Front Immunol 2018, 9:2909.

28. Binnewies M, Roberts EW, Kersten K, Chan V, Fearon DF, Merad M, Coussens LM, Gabrilovich DI, Ostrand-Rosenberg S, Hedrick CC et al: Understanding the tumor immune microenvironment (TIME) for effective therapy. Nat Med 2018, 24(5):541–550.

29. Haga Y, Kanda T, Nakamura M, Nakamoto S, Sasaki R, Takahashi K, Wu S, Yokosuka O: Overexpression of c-Jun contributes to sorafenib resistance in human hepatoma cell lines. PLoS One
30. Ye XL, Zhao YR, Weng GB, Chen YC, Wei XN, Shao JP, Ji H: **IL-33-induced JNK pathway activation confers gastric cancer chemotherapy resistance.** *Oncol Rep* 2015, 33(6):2746–2752.

31. Suzuki S, Okada M, Shibuya K, Seino M, Sato A, Takeda H, Seino S, Yoshioka T, Kitanaka C: **JNK suppression of chemotherapeutic agents-induced ROS confers chemoresistance on pancreatic cancer stem cells.** *Oncotarget* 2015, 6(1):458–470.

32. Sui X, Kong N, Wang X, Fang Y, Hu X, Xu Y, Chen W, Wang K, Li D, Jin W et al: **JNK confers 5-fluorouracil resistance in p53-deficient and mutant p53-expressing colon cancer cells by inducing survival autophagy.** *Sci Rep* 2014, 4:4694.

### Tables

#### Table 1. The risk model of immune lncRNA pairs.

| ID                | coef     | HR        | HR.95L     | HR.95H     | pvalue     |
|-------------------|----------|-----------|------------|------------|------------|
| AC105460.1|AL590483.1 | 0.434869442 | 1.544761365 | 0.972868122 | 2.45283777 | 0.065275621 |
| AP001469.3|AL137782.1 | 1.173749307 | 3.234095554 | 1.349471659 | 7.50717832 | 0.00847151 |
| LINC00941|AL590483.1 | 0.412896035 | 1.51187907 | 0.968800939 | 2.357232325 | 0.068723513 |
| AP001453.2|AC027796.4 | -0.773706228 | 0.461300212 | 0.284593551 | 0.74725607 | 0.001691186 |
| AC007128.1|AL590483.1 | 0.704647609 | 2.023313626 | 1.27659379 | 3.206580434 | 0.002711144 |
| AC124067.4|SNHG7    | -0.467395261 | 0.626632358 | 0.39248827 | 1.000458211 | 0.050228869 |
| LINC00513|AC010973.2 | -0.801575511 | 0.448821599 | 0.279579346 | 0.719874614 | 0.000893026 |
| LINC00513|LENG8-AS1 | -0.518086862 | 0.595690337 | 0.324938192 | 1.09129779 | 0.093828869 |
| FENDRR|NKILA    | -0.310222378 | 0.733283872 | 0.517964308 | 1.03813571 | 0.002956828 |
| AL137782.1|AL354993.2 | -0.544271131 | 1.723351826 | 1.046348925 | 2.838385406 | 0.032522231 |
| AL137782.1|AL354993.2 | -0.554021701 | 0.57463415 | 0.367725261 | 0.898209241 | 0.015056368 |

#### Table 2. The risk model of immune lncRNA expression.

| ID                | coef     | HR        | HR.95L     | HR.95H     | pvalue     |
|-------------------|----------|-----------|------------|------------|------------|
| AP001469.3        | 0.574505872 | 1.776252615 | 0.998051245 | 3.161233822 | 0.050780302 |
| AP001453.2        | -0.363495071 | 0.695242154 | 0.432289518 | 1.118143358 | 0.133783345 |
| AC007128.1        | 0.427241742 | 1.533023213 | 0.924508885 | 2.542063371 | 0.097768947 |
| AC124067.4        | -0.269272158 | 0.763935316 | 0.611645819 | 0.954142331 | 0.017606693 |
| LINC00513         | -0.419186774 | 0.657581365 | 0.470812319 | 0.918440818 | 0.013930483 |
| FENDRR            | -0.310222378 | 0.733283872 | 0.517964308 | 1.03813571 | 0.002956828 |
| AL137782.1        | -0.474462172 | 0.474847219 | 0.231960611 | 0.972061078 | 0.041601358 |
| NKILA             | 0.514359624 | 1.672567087 | 1.217886075 | 2.296997001 | 0.001484107 |
| AL590483.1        | -0.773741406 | 0.461283985 | 0.268857903 | 0.79143262 | 0.004966124 |
| AC010973.2        | 1.34334251 | 3.831830054 | 2.112475134 | 6.950577227 | 9.80E-06 |
| AL354993.2        | 0.422350358 | 1.525542918 | 0.972653578 | 2.392713343 | 0.065882942 |

### Figures
Figure 1

The process flow of the present study.
Figure 2

Identification of differentially expressed irlncRNAs. A. The heatmap represents a total of irlncRNAs in colorectal cancer using TCGA datasets and annotated by Ensembl. B. The forest map shows the significantly differentially expressed irlncRNAs in colorectal cancer. C. The core irlncRNAs were determined by intersecting DEirlncRNA and RISKirlncRNA.
Figure 3

Construction of differentially expressed irlncRNA pairs. A. Lasso coefficient profiles of the irlncRNAs were determined for establishing a pair-risk model. B. The selection of the tuning parameter was performed via 10 times cross-validation in the Lasso model. C, D. Forest maps show the clinical link of the 11 individual irlncRNA pairs identified by Cox proportional hazard regression to prognosis.
Figure 4

Establishment of a pair-risk assessment model by differentially expressed lncRNA pairs. A. Plot shows the maximum AUC value of the model based on the core lncRNA pairs. B. The 1-, 5-, and 10-year ROC curves of the model suggest that all AUC values are over 0.70. C. The maximum inflection point is recognized by the AIC. D. Comparing risk score ROC curves with other common clinical characteristics indicate the superiority of this pair-risk model.
Figure 5

Survival predictability and clinical values of the risk model. A. The risk plot shows the risk scores of individual patients in the high- and low-risk groups. B. The scatter plot represents the survival status distribution of individual patients in the high- and low-risk groups. C. Kaplan-Meier curves indicate patients in the low-risk group experiencing a longer survival time. D-G. The strip chart (D) together with the box plots revealed that tumor grade (E), T stage (F), and N stage (G) were statistically significantly
related to the risk score. H. The univariate Cox regression analysis demonstrated that the T stage, N stage, M stage, and risk score were statistically different. I. The multivariate Cox regression analyses indicated that the T stage, N stage, M stage, and risk score can serve as the independent prognostic predictor.

**Figure 6**

Investigation of tumor-related immune infiltrates by the risk model. A-X. The scatter chart and box plots revealed that the relationship between the risk model and immune infiltrates of CRC.
Figure 7

Evaluation of the relationship of the pair-risk model with certain immune-related genes and therapeutics
A, B. The violin plots represent the relationship between the pair-risk model and some immune-related
genes. C-G. The box plots reflect the association of the pair-risk model with certain therapeutics.
Figure 8

Illustration of the prognostic predictability of the exp-risk model. A. The risk score distribution of individual CRC patients according to the exp-risk model. B. The survival status of CRC patients involved based on the exp-risks core. C. Expression heatmap of the eleven core lncRNAs among the low-risk and high-risk groups. D. Kaplan–Meier curve of the overall survival between the low-risk and high-risk groups.
Figure 9

Evaluation of the relationship between the exp-risk model and clinical traits of CRC patients. A. The univariate cox analysis revealed that age, stage, T, N, M, and exp-risk score can be regarded as risk factors. B. The multiple cox analysis demonstrated that the exp-risk score remained as a risk factor when considered the whole characters. C. The multiple ROC curves exhibited that the exp-risk score was the most optimal parameter to predict the prognosis of the CRC patients. D. The evaluation of the exp-risk score and the pair-risk score shows robust predictability for the survival status of CRC patients.
Figure 10

GSEA enrichment analysis according to the exp-risk model. A-L GSEA enrichment analysis of signal pathway based on the exp-risk model.

Supplementary Files

This is a list of supplementary files associated with this preprint. Click to download.

- TableS1.docx
- TableS2.docx