Oncology and Translational Medicine
DOI 10.1007/s10330-020-0472-2
June 2021, Vol. 7, No. 3, P130–P135

ORIGINAL ARTICLE

Construction and validation of an immune-related lncRNA prognostic model for rectal adenocarcinomas*

Danni Jian[1], Yi Cheng[2], Jing Zhang[2], Kai Qin[2] (✉)

[1] Department of Otolaryngology, Union Hospital, Tongji Medical College, Huazhong University of Science and Technology, Wuhan 430022, China
[2] Department of Oncology, Tongji Hospital, Tongji Medical College, Huazhong University of Science and Technology, Wuhan 430030, China

Abstract

Objective This study aimed to construct a prognostic model for rectal adenocarcinomas based on immune-related long noncoding RNAs (lncRNAs) and verify its prediction efficiency.

Methods Transcript data and clinical data of rectal adenocarcinomas were downloaded from The Cancer Genome Atlas (TCGA) database. Perl software (strawberry version) and R language (version 3.6.1) were used to analyze the immune-related genes and immune-related lncRNAs of rectal adenocarcinomas, and the differentially expressed immune-related lncRNAs were screened according to the criteria |log2FC| > 1 and $P < 0.05$. The key immune-related lncRNAs were screened using single-factor Cox regression analysis and lasso regression analysis. Multivariate Cox regression analysis was performed to construct an immune-related lncRNA prognostic model using the risk scores. Next, we evaluated the effectiveness of the model through Kaplan-Meier (K-M) survival analysis, ROC curve analysis, and independent prognostic analysis of clinical features. In addition, prognostic biomarkers of immune-related lncRNAs in the model were analyzed by K-M survival analysis.

Results In this study, we obtained gene expression profile matrices of 89 rectal adenocarcinomas and 2 paracancerous specimens from TCGA database and applied immunologic signatures to these transcripts. Through R and Perl software analysis, we obtained 847 immune-related lncRNAs and 331 protein-encoded immune-related genes in rectal adenocarcinomas. Eight important immune-related lncRNAs related to the prognosis of rectal adenocarcinomas were identified using univariate Cox regression and lasso regression analysis. Furthermore, four immune-related lncRNAs were identified as prognostic markers of rectal adenocarcinomas via multivariate Cox regression analysis. The prognostic risk model was as follows: risk score = (-4.084) * expression LINC01871 + (3.112) * expression AL158152.2 + (7.616) * expression PXN-AS1 + (-0.867) * expression HCP5. The independent prognostic effect of the rectal adenocarcinoma risk score model was revealed through K-M analysis, ROC curve analysis, and univariate, and multivariate Cox regression analysis ($P = 0.035$). LINC01871 ($P = 0.006$), PXN-AS1 ($P = 0.008$), and AL158152.2 ($P = 0.0386$) were closely correlated with the prognosis of rectal adenocarcinomas through the K-M survival analysis.

Conclusion We constructed a prognostic model of rectal adenocarcinomas based on four immune-related lncRNAs by analyzing the data based on TCGA database, with high prediction accuracy. We also identified two biomarkers with poor prognosis (PXN-AS1 and AL158152.2) and one biomarker with good prognosis (LINC01871).

Received: 26 October 2020
Revised: 15 March 2021
Accepted: 4 April 2021

Key words: rectal adenocarcinoma; immune-related lncRNA; prognostic model; The Cancer Genome Atlas (TCGA) database

Rectal cancer is the eighth most common cancer in the world and the tenth leading cause of cancer-related death. In 2018, there were 704,376 new cases and 310,394 deaths [1]. In recent years, the incidence of rectal cancer has increased in China [2]. Although the widespread comprehensive treatment involving total mesorectal excision (TME) surgery and chemoradiotherapy has made progress in patient survival, the long-term survival rate

✉ Correspondence to: Kai Qin. Email: qinkaitj@126.com
* Supported by a grant from the Health Commission of Hubei Province Scientific Research Project (No. WJ2019M118).

is still unsatisfactory, especially for patients with locally advanced and distant metastases, where the overall 5-year survival rate patients with rectal cancer is about 53% [3]. Therefore, there is an urgent need to identify new biomarkers to predict the prognosis of patients and guide precise treatment.

Immune-related long noncoding RNAs (lncRNAs), which are located near or overlapping the coding gene clusters of immune-related proteins, play an important role in guiding the development, differentiation, and activation of a variety of immune cells [4]. However, to date, only a few immune-related lncRNAs have been implicated in cancer [5]. Therefore, it is of great significance to study the role of immune-related lncRNAs in immune regulation. Although some reports have shown that an lncRNA is recognized as a biomarker to predict the prognosis of rectal adenocarcinomas [6-8], there are few studies on immune-related lncRNAs in rectal adenocarcinomas. In this study, immune-related lncRNAs in rectal adenocarcinomas were obtained by analyzing the transcripts and immune-related gene sets in The Cancer Genome Atlas (TCGA) database. We used univariate/multivariate Cox regression analysis to screen immune-related lncRNAs associated with the prognosis of rectal adenocarcinomas. We constructed a prognostic model composed of four immune-related lncRNAs and identified prognostic biomarkers for rectal adenocarcinoma.

Materials and methods

Patients and datasets of rectal adenocarcinomas

The transcripts and clinical data of rectal adenocarcinomas were downloaded from TCGA database (https://portal.gdc.cancer.gov/) on March 13, 2020. The screening conditions were as follows: (a) primary tumor site: rectal carcinoma; (b) project: TCGA-READ; (c) disease type: adenocarcinoma or adenoma; (d) data classification: transcriptome profiling; (e) data type: quantitative data of gene expression; and (f) workflow type: HTSeq-FPKM. The transcription data of rectal adenocarcinomas were sorted and transformed into a matrix according to the Strawberry Perl software (version 5.30.1.1). The corresponding clinical data of rectal adenocarcinomas were obtained from TCGA program (including patient number, sex, clinical stage, survival time, survival status, and TNM stage).

Immune-related genes and lncRNAs of rectal adenocarcinomas

The mRNA matrix and long noncoding RNA matrix of the coding protein were obtained by sorting the previous matrix (gene and sample names) using Perl software. We searched the immune-related gene set in the MSigDB database (http://software.broadinstitute.org/gsea/msigdb): IMMUNE_RESPONSE (M19817) and IMMUNE_SYSTEM_PROCESS (M13664), which was used to extract the immune-related genes encoding the protein. The R (version 3.6.1) and Bioconductor (https://www.bioconductor.org/) packages were used for data processing and analysis to obtain immune-related lncRNAs.

Establishment and evaluation of the prognostic model and independent prognostic analysis of clinical characteristics

The differential expression of immune-related lncRNAs in rectal adenocarcinomas was analyzed using Software Package EdgeR (http://bioconductor.org/packages/release/bioc/html/edgeR.html), filtered by the criteria |log2 FC (fold change)| > 1 and false discovery rate (FDR) < 0.05. Clinical data from TCGA were analyzed using univariate Cox proportional hazard regression (PHR), and survival-related lncRNAs were screened according to $P < 0.001$. Furthermore, through lasso-Cox analysis, the lncRNAs most related to overall survival were determined and cross-validation was performed to prevent overfitting. Then, multivariate Cox-PHR analysis was used to construct prognostic indicators and calculate risk scores. According to the median risk score, patients with rectal adenocarcinomas were divided into high- and low-risk groups. Kaplan-Meier (K-M) analysis was used to compare the differences in survival rate between the two groups. The risk score of each patient was calculated according to the expression levels of lncRNAs. The risk score model was calculated using the following formula:

$$\text{Risk score} = \sum_{i=1}^{n} \text{coefi X id}$$

To determine if the risk score could be driven by other clinical cofactors, we used a multivariate model (Cox proportional hazards) to account for age, sex, grade, clinical stage, and T stage in TCGA cohort. Receiver operating characteristic (ROC) and area under the curve (AUC) of 5-year overall survival rate and other clinical characteristics (gender, stage, TNM, and risk score) were calculated by R-package "survival ROC." Furthermore, K-M survival analysis was performed to identify lncRNAs associated with prognosis and to explore predictive lncRNAs.

Results

Differentially expressed lncRNAs in rectal adenocarcinomas

The clinical data of 90 rectal adenocarcinomas were downloaded from TCGA database (Table 1). We obtained the gene expression matrix of 89 cases of rectal

Table 1 Clinical characteristics of rectal adenocarcinomas

Clinical characteristics	No. of patients (%)
Gender	
Male	51 (56.7)
Female	39 (43.3)
Stage	
I	18 (20)
II	30 (33.3)
III	24 (26.7)
IV	14 (15.6)
Unknow	4 (4.4)
Tumor stage	
T1	4 (4.4)
T2	18 (20)
T3	62 (68.9)
T4	6 (6.7)
Fustat	
Alive	80 (88.9)
Dead	10 (11.1)
Lymph node	
N0	50 (55.6)
N1	26 (28.9)
N2	13 (14.4)
Unknow	1 (1.1)
Metastasis	
M0	70 (77.8)
M1	13 (14.4)
Unknow	7 (7.8)

adenocarcinomas; there were 2 cases of paracancerous specimens and 56754 genes were expressed. A total of 847 immune-related lncRNAs and 331 protein-encoded immune-related genes were processed and analyzed using the R language and the corresponding data packet. Using the edgeR package, 47 differentially expressed immune-related lncRNAs were screened with a threshold of $|\log2FC| > 1$ and FDR < 0.05, including 11 upregulated and 36 downregulated lncRNAs. Eight key lncRNAs related to prognosis were identified using lasso regression analysis and univariate Cox regression (Table 2).

Establishment and verification of the prognostic model of rectal adenocarcinomas

Table 2 Immune-related lncRNAs in rectal adenocarcinomas identified by univariate Cox regression analysis

Immune-related LncRNA	HR (95% CI)	*P* value
LINC01871	0.169 (0.047–0.617)	0.007
LINC02298	6.511 (1.840–23.035)	0.004
AL158152.2	4.66 (1.454–14.939)	0.01
SNHG6	4.362 (1.551–12.272)	0.005
ZFAS1	3.322 (1.383–7.984)	0.007
PXN-AS1	13.001 (1.886–89.641)	0.009
HCP5	0.262 (0.114–0.603)	0.002
Unknow	6.284 (1.890–20.892)	0.003

Table 3 Immune-related lncRNAs in rectal adenocarcinomas identified by multivariate Cox

Immune-related lncRNA	coef	R (95% CI)	*P* value
LINC01871	–4.084	0.0168 (0.001–0.433)	0.014
AL158152.2	3.112	22.457 (2.698–186.904)	0.004
PXN-AS1	7.616	20.261 (15.931–25.432)	0.002
HCP5	–0.867	0.420 (0.175–1.008)	0.052

Eight key immune-related lncRNAs, obtained by univariate Cox regression analysis, were used in the multivariate Cox-PHR regression analysis to calculate the prognosis risk score of each patient, and we constructed a risk score model consisting of four lncRNAs (Table 3): risk score = (–4.084) * expression LINC01871 + (3.112) * expression AL158152.2 + (7.616) * expression PXN-AS1 + (–0.867) * expression HCP5. K-M analysis comparing the survival difference between the high- and low-risk groups showed that the total survival time of patients in the low-risk group was significantly longer than that in the high-risk group (P= 1.93e–03; Fig. 1a). The area under the ROC curve was 0.957 (Fig. 2). Univariate/multivariate independent prognostic analysis of clinical traits showed that the prognostic risk score had an independent prognostic risk effect on rectal adenocarcinomas (P = 0.035; Table 4). The heat map, risk score, and scatter plots of survival time with respect to immune-related lncRNAs in rectal adenocarcinomas showed that the higher the risk score, the shorter the survival time and the more the death (Fig. 3).

Table 4 Univariate and multivariate Cox regression analyses of clinical characters in rectal adenocarcinomas

Clinical characteristics	Univariate Cox regression		Multivariate Cox regression	
	HR (95% CI)	*P* value	HR (95% CI)	*P* value
Age	1.12 (1.028–1.221)	0.01	1.04 (0.925–1.168)	0.512
Gender	0.559 (0.149–2.092)	0.388	0.982 (0.152–6.336)	0.985
Stage	2.355 (1.115–4.973)	0.025	9.757 (0.297–32.08)	0.201
T	2.152 (0.711–6.511)	0.175	1.562 (0.334–7.316)	0.571
M	4.206 (1.110–15.931)	0.035	0.179 (0.002–14.893)	0.446
N	2.074 (0.923–4.664)	0.078	0.587 (0.11–3.12)	0.532
Risk score	1.003 (1.001–1.005)	0.001	1.002 (1–1.003)	0.035

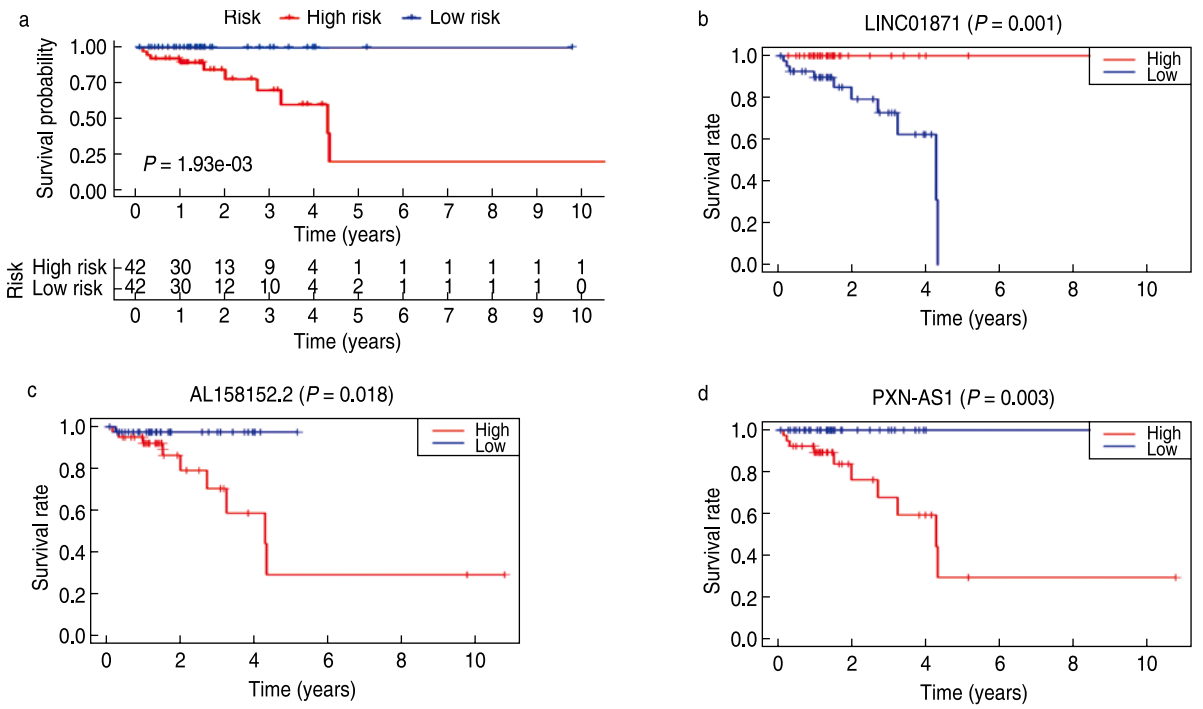

Fig. 1 Three prognostic immune-related lncRNAs identified by the multivariate Cox regression. (a) LINC01871; (b) AL158152.2; (c) PXN-AS1 and K-M survival curves of the rectal adenocarcinoma prognostic model (d)

Correlation between clinical features and key immune-related lncRNAs in the prognosis of rectal adenocarcinomas

Correlation with lymph node staging

We assessed the expression of the four immune-related lncRNAs in different lymph node stages of rectal cancer (AJCC, 8th edition) and found that the expression of AL158152.2 was positively correlated with the lymph node stage and the difference was statistically significant ($P < 0.05$). The expression of HCP5, LINC01871, and PXN–AS1 was not significantly correlated with lymph node staging ($P > 0.05$; Fig. 4a).

Correlation with the T stage and clinical stage of rectal adenocarcinomas

The results showed that there was no significant correlation among immune-related lncRNA, clinical stage, and T stage ($P > 0.05$; Figs. 4b and 4c).

Prognostic biomarkers of rectal adenocarcinomas

LINC01871 ($P = 0.001$), PXN-AS1 ($P = 0.003$), and AL158152.2 ($P = 0.018$) were associated with the prognosis in the prognostic model, as per the K-M survival analysis. LINC01871, PXN-AS1, and AL158152.2 may be independent prognostic factors, while LINC01871 may be a protective prognostic factor for rectal adenocarcinomas (Figs. 1b–1d).

Discussion

Immune-related lncRNAs are important regulators of gene expression in the immune system and play an important role in the occurrence and development of tumors [4–5, 9]. Li *et al* found that immune-related lncRNAs, which have high tissue specificity, are highly expressed in B cells and T cells [5]. Yu *et al* found that lncRNAs can be used as biomarkers to mark different stages of cancer immunity to adjust tumor immunity [10]. In recent

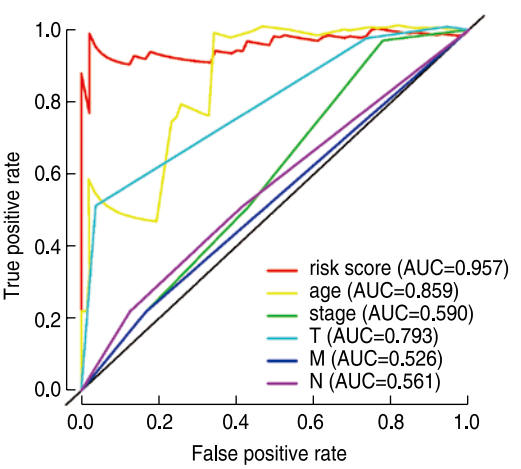

Fig. 2 Operating characteristic (ROC) curve of clinical parameters in rectal adenocarcinomas

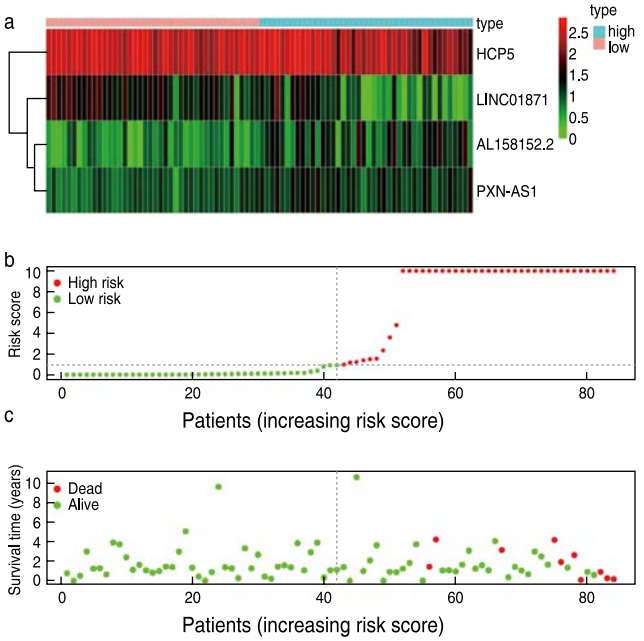

Fig. 3 Heat map (a), risk score (b) and scatter plots of survival time (c) of immune-related lncRNAs in rectal adenocarcinoma

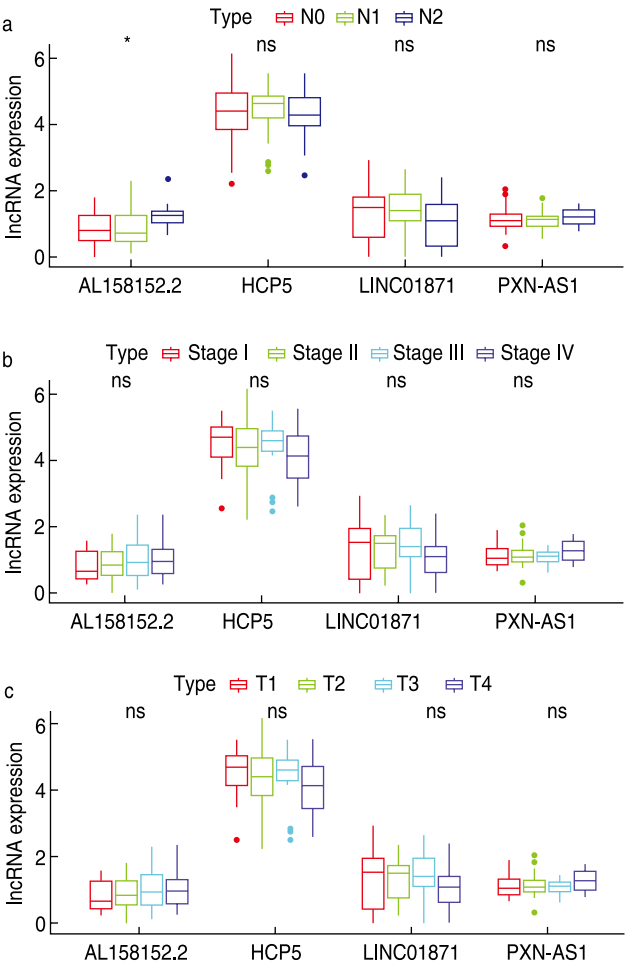

Fig. 4 The lymph node stage was positively correlated with the expression of AL158152.2, not the expressions of HCP5, LINC01871 and PXN−AS1 (a), there were no correlations between T stage, clinical stage of rectal adenocarcinomas and 4 key immune-related lncRNAs (b and c)

years, increasing evidence has shown that immune-related lncRNAs can be used as prognostic biomarkers for malignant tumors. At present, a prognostic model based on immune-related lncRNAs has been successfully constructed in a few malignant tumors, including breast cancer, head and neck squamous cell carcinoma, glioma, pancreatic cancer, and renal clear cell carcinoma [11–16]. There are many prognostic factors in rectal cancer, among which are common clinical related factors, such as surgical methods and R0 resection [3]; however, there are few reports on the prognostic role of immune-related lncRNAs. In this study, we aimed to construct a prognostic risk model of rectal adenocarcinoma based on immune-related lncRNAs and further analyze and obtain prognostic markers.

In the present study, a prognostic model was constructed by analyzing rectal adenocarcinoma samples from TCGA database. Risk score = (-4.084) * expression LINC01871 + (3.112) * expression AL158152.2 + (7.616) * expression PXN-AS1 + (-0.867) * expression HCP5. It had a high accuracy. LINC01871 is a protective immune-related lncRNA, while PXN-AS1 and AL158152.2 are harmful prognostic markers. At present, there are few reports on the prognostic immune-related lncRNAs in rectal adenocarcinomas. Tao *et al* found that NKILA, an immune-related lncRNA encoded by a gene on chromosome 20q13, was expressed at low levels in various human tumors, such as breast, lung, and rectal cancers. NKILA inhibits proliferation, migration, and invasion of rectal cancer cells by inhibiting NF-κB signaling, which is related to clinical progress and prognosis [17]. Zhao *et al* obtained a prognostic risk model of rectal cancer consisting of five lncRNAs (AC079789.1, AC106900.2, AL121987.1, AP004609.1, and LINC02163), in which AC106900.2 and LINC02163 are immune-related lncRNAs, but their functions are still unknown [7]. He *et al* found that HCP5, EPB41L4A-AS1, SNHG12, and LINC00649 are significantly related to the occurrence and prognosis of colorectal cancer through the competitive endogenous RNA network mediated by lncRNA, among which HCP5 and SNHG12 are immune-related lncRNAs. Studies have shown that SNHG12 can increase the expression of cell cycle-related proteins and inhibit the expression of caspase-3 in colorectal cancer and human osteosarcoma. In addition, silencing the expression of SNHG12 inhibited the proliferation of triple-negative breast cancer cells. It was found that SNHG12, as miR-199a/b-5p, regulates

the expression of MLK3 in hepatocellular carcinoma and affects the activation of the NF-κB pathway [18]. Therefore, SNHG12 may be a potential biomarker. The lncRNA HLA complex P5 (HCP5) is located at 6P21.33, which is homologous to the retrovirus gene sequence [19]. HCP5, which is considered a susceptibility gene site for HCV-related liver cancer, is downregulated in ovarian cancer,. HCP5 targets miR-139-5p and inhibits the expression of miR-139-5p. The miR-139-5p/ZEB1/Wnt signaling pathway is involved in the occurrence and development of EMT in CRC [20]. It was reported that HCP5 is highly expressed in glioma tissues and can promote the proliferation, migration, and invasion of glioma cells, inhibit apoptosis, and promote malignant biological behavior of glioma cells [18]. The lncRNA PXN-AS1-L is upregulated in hepatocellular carcinoma, nasopharyngeal carcinoma, lung cancer, and glioma, and promotes tumor occurrence by upregulating PXN [16, 18, 20–22]. In addition, we know nothing about the function and mechanism of LINC01871 and AL158152.2. However, the conclusion of this study was based on TCGA database, and we lacked domestic data to verify the prediction model and markers of immune-related lncRNAs in rectal adenocarcinomas.

In summary, we constructed a prognostic model of rectal adenocarcinomas based on the expression levels of four immune-related lncRNAs (LINC01871, PXN-AS1, HCP5, and AL158152.2) by analyzing TCGA database and immune-related gene sets, which have high prediction accuracy. Two negative prognostic biomarkers (PXN-AS1 and AL158152.2) and positive prognostic biomarker (LINC01871) were identified. However, the mechanism of action in rectal adenocarcinomas needs to be further explored.

Conflicts of interest

The authors indicated no potential conflicts of interest.

References

1. Bray F, Ferlay J, Soerjomataram I, *et al.* Global cancer statistics 2018: GLOBOCAN estimates of incidence and mortality worldwide for 36 cancers in 185 countries. CA Cancer J Clin, 2018, 68: 394–424.
2. Arnold M, Sierra MS, Laversanne M, *et al.* Global patterns and trends in colorectal cancer incidence and mortality. Gut, 2017, 66: 683–691.
3. Khalfallah M, Dougaz W, Jerraya H, *et al.* Prognostic factors in rectal cancer: where is the evidence? Tunis Med, 2017, 95: 79–86.
4. Geng H, Tan XD. Functional diversity of long non-coding RNAs in immune regulation. Genes Dis, 2016, 3: 72–81.
5. Li Y, Jiang T, Zhou W, *et al.* Pan-cancer characterization of immune-related lncRNAs identifies potential oncogenic biomarkers. Nat Commun, 2020, 11: 1000.
6. Zhang Z, Wang S, Ji D, *et al.* Construction of a ceRNA network reveals potential lncRNA biomarkers in rectal adenocarcinoma. Oncol Rep, 2018, 39: 2101–2113.
7. Zhao K, Wang M, Kang H, *et al.* A prognostic five long-noncoding RNA signature for patients with rectal cancer. J Cell Biochem, 2019 Nov. 10.
8. Liu H, He YK, Li QL, *et al.* Expression and clinical significance of lncRNA AC010145.4 in patients with rectal cancer. Chin J Cancer Prev Treat (Chinese), 2018, 25: 1558–1561.
9. Atianand MK, Caffrey DR, Fitzgerald KA. Immunobiology of long noncoding RNAs. Annu Rev Immunol, 2017, 35: 177–198.
10. Yu WD, Wang H, He QF, *et al.* Long noncoding RNAs in cancer-immunity cycle. J Cell Physiol, 2018, 233: 6518–6523.
11. Khadirnaikar S, Kumar P, Pandi SN, *et al.* Immune associated LncRNAs identify novel prognostic subtypes of renal clear cell carcinoma. Mol Carcinog, 2019, 58: 544–553.
12. Shen Y, Peng X, Shen C. Identification and validation of immune-related lncRNA prognostic signature for breast cancer. Genomics, 2020, 112: 2640–2646.
13. 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: 14916–14927.
14. Wang W, Zhao Z, Yang F, *et al.* An immune-related lncRNA signature for patients with anaplastic gliomas. J Neurooncol, 2018, 136: 263–271.
15. Zhou M, Zhang Z, Zhao H, *et al.* An immune-related six-incRNA signature to improve prognosis prediction of glioblastoma multiforme. Mol Neurobiol, 2018, 55: 3684–3697.
16. Wu S, Dai X, Xie D. Identification and validation of an immune-related RNA signature to predict survival of patients with head and neck squamous cell carcinoma. Front Genet, 2019, 10: 1252.
17. Tao F, Xu Y, Yang D, *et al.* LncRNA NKILA correlates with the malignant status and serves as a tumor-suppressive role in rectal cancer. J Cell Biochem, 2018, 119: 9809–9816.
18. Jiang L, Wang R, Fang L, *et al.* HCP5 is a SMAD3-responsive long non-coding RNA that promotes lung adenocarcinoma metastasis via miR-203/SNAI axis. Theranostics, 2019, 9: 2460–2474.
19. Yang C, Sun J, Liu W, *et al.* Long noncoding RNA HCP5 contributes to epithelial-mesenchymal transition in colorectal cancer through ZEB1 activation and interacting with miR-139-5p. Am J Transl Res, 2019, 11: 953–963.
20. Zhang Z, Peng Z, Cao J, *et al.* Long noncoding RNA PXN-AS1-L promotes non-small cell lung cancer progression via regulating PXN. Cancer Cell Int, 2019, 19: 20.
21. Jia X, Niu P, Xie C, *et al.* Long noncoding RNA PXN-AS1-L promotes the malignancy of nasopharyngeal carcinoma cells via upregulation of SAPCD2. Cancer Med, 2019, 8: 4278–4291.
22. Shang F, Du SW, Ma XL. Upregulation of lncRNA PXN-AS1-L is associated with unfavorable prognosis in patients suffering from glioma. Eur Rev Med Pharmacol Sci, 2019, 23: 8950–8955.

DOI 10.1007/s10330-020-0472-2
Cite this article as: Jian DN, Cheng Y, Zhang J, *et al*. Construction and validation of an immune-related lncRNA prognostic model for rectal adenocarcinomas. Oncol Transl Med, 2021, 7: 130–135.