



Two predicted models based on ceRNAs and immune cells in lung adenocarcinoma

Miaomiao Zhang^{1,*}, Peiyan Zheng^{2,*}, Yuan Wang¹ and Baoqing Sun²

¹The First Affiliated Hospital of Guangzhou Medical University, Guangzhou Institute of Respiratory Diseases, State Key Laboratory of Respiratory Disease, Guangzhou, China

²Department of Allergy and Clinical Immunology, Guangzhou Institute of Respiratory health, State Key Laboratory of Respiratory Disease, National Clinical Research Center of Respiratory Disease, First Affiliated Hospital of Guangzhou Medical University, State Key Laboratory of Respiratory Disease, Guangzhou, China

*These authors contributed equally to this work.

ABSTRACT

Background. It is well accepted that both competitive endogenous RNAs (ceRNAs) and immune microenvironment exert crucial roles in the tumor prognosis. The present study aimed to find prognostic ceRNAs and immune cells in lung adenocarcinoma (LUAD).

Materials and Methods. More specifically, we explored the associations of crucial ceRNAs with the immune microenvironment. The Cancer Genome Atlas (TCGA) database was employed to obtain expression profiles of ceRNAs and clinical data. CIBERSORT was utilized to quantify the proportion of 22 immune cells in LUAD.

Results. We constructed two cox regression models based on crucial ceRNAs and immune cells to predict prognosis in LUAD. Subsequently, seven ceRNAs and seven immune cells were involved in prognostic models. We validated both predicted models via an independent cohort GSE72094. Interestingly, both predicted models proved that the longer patients were smoking, the higher risk scores would be obtained. We further investigated the relationships between seven genes and immune/stromal scores via the ESTIMATE algorithm. The results indicated that CDC14A and HIF0 expression were significantly related to stromal scores/immune scores in LUAD. Moreover, based on the result of the ceRNA model, single-sample gene set enrichment analysis (ssGSEA) suggested that differences in immune status were evident between high- and low-risk groups.

Subjects Bioinformatics, Oncology, Respiratory Medicine

Keywords Lung adenocarcinoma, Immune microenvironment, Prognosis, ceRNA (competitive endogenous RNAs)

INTRODUCTION

Lung cancer is considered as the most common cause of survival-associated cancer globally (*Bray et al., 2018*). Besides, lung adenocarcinoma (LUAD) approximately accounts for 50% of all lung cancer types (*Imielinski et al., 2012*). Although the improvement of diagnosis and the comprehensive treatment in the lung cancer has been made, the 5-year overall survival rate for lung cancer is still approximately 18%, mainly due to invisible early symptoms

Submitted 12 October 2020
Accepted 8 February 2021
Published 23 March 2021

Corresponding authors
Yuan Wang,
YuanWangxz1992@126.com
Baoqing Sun,
sunbaoqing@vip.163.com

Academic editor
Santosh Patnaik

Additional Information and
Declarations can be found on
page 16

DOI 10.7717/peerj.11029

© Copyright
2021 Zhang et al.

Distributed under
Creative Commons CC-BY 4.0

OPEN ACCESS

(Siegel, Miller & Jemal, 2018). Thus, it is of great necessity to identify sensitive and accurate biomarkers in the early stage.

Competitive endogenous RNAs (ceRNAs), consisting of a group of regulatory RNA molecules, bind with the specific miRNAs and then affect the protein level (Karthi & Subramanian, 2014; Phelps et al., 2016; Shukla, Singh & Barik, 2011). Previous studies have shown that ceRNA regulatory networks play essential roles in tumor progression and migration (Bartel, 2009; Qi et al., 2015; Sanchez-Mejias & Tay, 2015; Tay, Rinn & Pandolfi, 2014). Besides, increasing studies have demonstrated that immune microenvironment (TIME) was also associated with the prognosis in various kinds of tumors, including LUAD (Lavin et al., 2017; Liu, Niu & Qiu, 2020). Stromal cells and immune cells constitute the central part of TIME. Stromal cells participate in the process of tumorigenesis and metastasis and are generally conducive to the movement of cancer cells (Ni et al., 2018). Additionally, various studies revealed that immune cells exhibit large effects on clinical outcomes in many cancers. Donnem et al. summarized that different types of immune cells affected the clinical stage and prognosis in non-small lung cancer (Donnem et al., 2016; Liu et al., 2020). As a result, to further investigate potential biomarkers based on immune cells and ceRNAs in LUAD, we constructed Cox proportional hazards models and predicted nomograms.

However, the mechanisms of ceRNAs in the tumor microenvironment still remain unclear in LUAD. It is well accepted that cancer stem cells (CSCs) are tumorigenic cells with self-renewal ability. Besides, CSCs promote tumor immune to escape via expressing a variety of immune factors (Bruttel & Wischhusen, 2014). In the current study, RNA stem scores and DNA stem scores were applied to assess tumor stem cell content (Malta et al., 2018). ESTIMATE algorithm was utilized to evaluate the stromal and immune cells in every sample (Yoshihara et al., 2013). Thus, we explored the correlation between crucial prognostic genes with RNAss/DNAss scores and stromal/immune scores. All in all, in this study, we comprehensively analyze the associations between crucial ceRNAs and immune infiltration landscape, crucial ceRNAs and tumor stem cells, which may offer an important perspective for early diagnosis and treatment of LUAD.

MATERIALS AND METHODS

Identification of differentially expressed ceRNAs

In the present study, expression profiles of LUAD and the corresponding clinical information were derived from The Cancer Genome Atlas (TCGA) database, including lncRNA, miRNA, and mRNA. The R package “DESeq2” was used to screen the differentially expressed mRNAs, miRNAs and lncRNAs (DEmRNA, DEmiRNA and DElncRNA) between normal and tumor samples (Love, Huber & Anders, 2014). Specifically, a false discovery rate (FDR) <0.05 and \log_2 |fold change| > 1 were set as cutoff criteria.

ceRNA network construction

Both lncRNA-miRNA and miRNA-mRNA interactions were predicted using miRcode and starBase (Chou et al., 2018; Jeggari, Marks & Larsson, 2012; Li et al., 2014). DEmiRNAs, DEmRNAs, and DE lncRNAs, which presented a significant difference in hypergeometric

distribution detection and correlation analysis ($P < 0.05$), were selected to construct a lncRNA-miRNA-mRNA regulatory network and then was visualized by Cytoscape 3.8.0 (Shannon *et al.*, 2003).

Construction of nomogram based on key ceRNAs

Univariate Cox analysis was employed to screen the survival-associated ceRNAs. Lasso regression analysis was applied to ensure that the Cox model was not overfitted. Eventually, the candidate ceRNAs were integrated to establish the multivariate cox regression models. Risk scores were described as following: risk score = $(\text{Exp RNA1} * \beta_1) + (\text{Exp RNA2} * \beta_2) + \dots + (\text{Exp RNAn} * \beta_n)$. It is worth noting that Exp stands for the expression level, and β denotes the regression coefficient. Based on the risk scores, patients with LUAD were divided into high-risk and low-risk groups. The corresponding Kaplan–Meier survival curves were carried out to reveal the overall survival (OS) within different groups. In addition, we also performed the ROC curve to evaluate the specificity and sensitivity of the model. Finally, according to the multivariate model result, we constructed a nomogram to predict patients' prognostic values. The calibration curve was applied to evaluate the accuracy of the nomogram.

Construction of nomogram based on immune cells

With a good deconvolution performance in gene expression profiles, CIBERSORT method could estimate the proportion of particular cell types (Newman *et al.*, 2015). In this study, CIBERSORT algorithm was utilized to estimate 22 tumour-infiltrating immune cells in LUAD by collating and calculating the genes expression in the tumor samples in the TCGA database. Samples with a CIBERSORT output of $p < 0.05$ presented that the predicted proportion of immune cells were correct.

The Wilcoxon test was conducted to examine the significant difference of immune cells between the tumor samples and healthy samples. We also performed multivariate Cox regression and produced the corresponding risk scores. Based on the multivariate Cox model results, we finally built a predicted nomogram. Similar to ceRNAs, patients were also stratified into high-risk and low-risk groups on the basis of the mean risk score. Subsequently, the prognostic model's sensitivity and specificity were examined by the ROC curve, and the calibration curve investigated the accuracy of the nomogram.

Validation of prognostic equations with GEO

To better elucidate the accuracy of both multivariate Cox models, an external validation dataset GSE72094 ($n = 442$) was applied in the present study. We selected this dataset because it contained most LUAD patients with clinical data in the Gene Expression Omnibus (GEO) database. Based on the median risk score generated from TCGA, we also classified patients with high- and low-risk patients. K-M and ROC curves were applied to evaluate the prognostic efficacy of multivariate models in TCGA.

Estimation of tumor stemness and microenvironment

Stemness scores of "RNAss" and "DNAss" referred to the result based on mRNA and DNA-methylation respectively. The larger stemness scores were, the higher likelihood

more stem cells were infiltrating. ESTIMATE algorithm could speculate the level of stromal cell, immune cells and tumor purity, respectively (Yoshihara *et al.*, 2013). Correspondingly, stromal/immune/Estimate scores were significantly associated with stromal cells, immune cells and tumor purity, respectively. We utilized the “ESTIMATE” package to assess immune, stromal scores, and the sum of both in individual patients. The higher scores, the higher proportion of corresponding cells were. Additionally, the correlation of hub ceRNAs with above scores was examined by Spearman analysis. The ssGSEA method could apply expressed traits of immune cell population to individual tumor samples (Barbie *et al.*, 2009). Based upon the results of ssGSEA, we calculated infiltrating level of immune cells and immune-associated functions in LUAD samples by using the “gsva” package (Hänzelmann, Castelo & Guinney, 2013). Furthermore, we compared the differences in these immune data sets between the high-risk group and the low-risk group.

Statistical analysis

All statistical analyses were accomplished with R software 4.0.2. The Wilcoxon rank sum tests were conducted for comparisons between two groups. Only a two-sided *P*-value <0.05 was recognized as statistically significant.

RESULTS

Clinicopathological characteristics of LUAD clinicopathological characteristics in the TCGA database and GSE72094 are displayed in Tables 1 and 2, respectively. In addition, complete clinical information for each patient is listed in Table S1.

Identification of DEmRNAs, DE miRNAs and DE lncRNAs

Figure 1 presents our workflow for the bioinformatic analysis. Among the whole expression profiles, totally 1645 upregulated DEmRNAs, and 1,344 downregulated DEmRNA were screened out (Figs. 2A and 2B). Furthermore, we also identified 163 upregulated DE lncRNA, 52 downregulated DE lncRNA (Figs. 2C and 2D), 111 upregulated DE miRNA, and 87 downregulated DE miRNA (Fig. 3B). Specifically, the top 20 significantly upregulated- and downregulated- miRNAs were shown in Fig. 3A.

Construct Nomogram of ceRNAs

To investigate the endogenously competitive relationships between the lncRNA and mRNA, we constructed an intersected ceRNA network. Seven lncRNAs (H19, AC074117.1, FBXL19-AS1, MAGI2-AS3, SNHG3, PVT1 and SNHG1), 94mRNA, and 15miRNA were involved in the network (Table S2). Hsa-miR-130b-3p and hsa-miR-29b-3p were the top 2 regulated miRNA in the ceRNA network (Fig. 4). In addition, the initial univariate Cox regression (Fig. S1) and Lasso regression analysis were further conducted to investigate the key biomarkers (Figs. 5B and 5C). Seven survival-associated biomarkers were finally involved in a new multiple Cox regression analysis (Fig. 5A). According to the result, all biomarkers belonged to protein-coding RNAs. Consequently, the corresponding risk score was generated (Table 3): risk score = [Exp CDC14A * (-0.281888)]+(Exp LOXL2*0.114969) + (Exp CCT6A * 0.174244)+(Exp E2F7 * 0.187164)+(Exp GPR37*0.099737)+(Exp H1F0 * 0.197590)+(Exp SMOC1 * 0.068385).

Table 1 Clinicopathological characteristics baseline in LUAD in TCGA.

Characteristics	Number	(%)
Total	522	100%
Age		
	>65	50.2%
	<=65	46.2%
	Unknown	3.6%
Gender		
	Female	53.6%
	Male	46.4%
Vital status		
	Dead	36.0%
	Alive	64.0%
Clinical stage		
	I	53.4%
	II	21.8%
	III	16.3%
	IV	5.0%
	Unknown	1.5%
T stage		
	T1	33.0%
	T2	53.8%
	T3	9.0%
	T4	3.6%
	Unknown	0.6%
M stage		
	M0	67.6%
	M1	4.8%
	Unknown	27.6%
N stage		
	N0	64.2%
	N1	18.8%
	N2	14.4%
	N3	0.4%
	Unknown	0.2%
Smoking status		
	<=20	8.6%
	>20	26.6%
	Unknown	64.80%

Notes.

LUAD, lung adenocarcinoma.

As shown in [Fig. 5D](#), the prognostic score could be a promising indicator to distinguish the LUAD patients according to the individual prognosis ($P = 5.701e-8$). The area under ROC reflected the predictive efficacy of the prognostic signature (AUC of 3-year

Table 2 Clinicopathological characteristics baseline in LUAD in GSE72094.

Characteristics		Number	(%)
Total		442	100%
Age	>65	294	50.2%
	<=65	127	46.2%
	Unknown	21	3.6%
Gender	Female	280	53.6%
	Male	242	46.4%
Smoking status	Ever	335	75.8%
	Never	33	7.5%
	Unknown	74	16.7%
Stage	I	265	60.0%
	II	69	15.6%
	III	63	14.3%
	IV	17	3.9%
	Unknown	28	6.3%
Vital status	Alive	298	67.4%
	Dead	122	27.6%
	Unknown	22	5.0%

survival:0.772) (Fig. 5E). By the GEO validation set, the AUC of gene signature at 3-year was 0.711 (Fig. S2A). Additionally, K-M analysis further confirmed the difference between high- and low-risk groups ($P = 9.963e-05$) (Fig. S2B). Eventually, we assessed the prognosis of patients with LUAD by a nomogram (Fig. 6A), and the accuracy of the nomogram was confirmed by the calibration curve (Fig. 6C). As shown in Fig. 7A, based on this prognostic model, the risk score in the group with fewer smoking years (≤ 20) was lower than the group with more smoking years (> 20) ($P = 0.025$).

Construct nomogram of immune cells

By applying the CIBERSORT, we estimated the percentages of 22 immune cell types in LUAD (Figs. S3B and S3C) (Chen *et al.*, 2018). The violin plot indicated that 13 immune cell types showed a significant difference in percentages of immune cells between tumor samples and healthy samples by the Wilcoxon test (Fig. S3D). The correlation analysis of immune cells could be found in Fig. S3A. Subsequently, after the analysis of univariate cox regression and lasso regression, seven immune cells were taken as prognostic biomarkers and eventually integrated into a new multivariate model (Figs. 8A, 8B and 8C). The formula was presented as following (Table 4): risk score = [Exp CD8 *(-3.425649)]+(Exp Tregs*9.175062)+[Exp Monocytes*(-5.318492)]+(Exp M1 *6.134551)+(Exp Dendritic cells activated*5.539130)+(Exp Mast cells activated

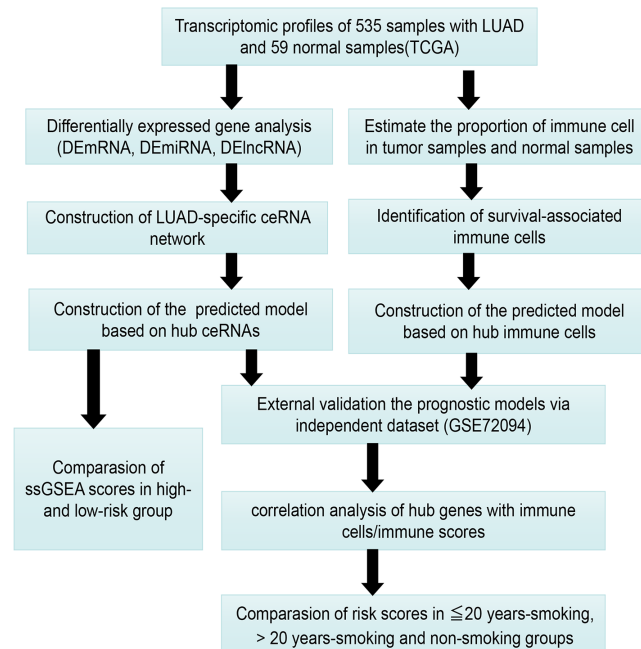


Figure 1 The workflow for bioinformatic analysis.

Full-size DOI: [10.7717/peerj.11029/fig-1](https://doi.org/10.7717/peerj.11029/fig-1)

*24.611110)+(Exp Eosinophils*37.351722). The K-M analysis results showed a significant difference in overall survival between high- and low- risk groups ($P = 3.342e-05$) (Fig. 8D). The ROC curve (AUC of 3-years: 0.766) demonstrated that the multivariate model was recognized as a fair potential to monitor the prognostic efficacy (Fig. 8E). In the external validation set, the model also carried out adequate predicted capacity. The AUC of immune cell signature was 0.738 at 3-years (Fig. S2C). K-M curves reflected that the low-risk group had longer overall survival than the high-risk group ($P = 2.922e-04$) (Fig. S2D). A predicted nomogram was also performed, and the discrimination was conducted to test the calibration quality (Figs. 6B and 6D). Similarly, we could also found that the longer smoking history was related to the higher the risk scores based on the predicted model of immune cells in smoking groups ($P = 0.0046$) (Fig. 7B). The correlations of key prognostic biomarkers in the two models illustrate that the expression of E2F7 is positively related to Macrophages M1 ($R = 0.4$) (Fig. S4).

Besides, the Univariate and multivariate survival analyses revealed that risk scores in the two models were both independent prognostic factors (Fig. S5).

Correlation of hub ceRNAs expression with tumor stemness and microenvironment

As shown in Fig. 9, CCT6A, H1F0 and E2F7 were positively related to RNAss and DNAss ($R = 0.22$ to 0.53). A significant negative association was detected between CDC14A and RNAss ($R = -0.25$, $P = 7.5e-8$). Nevertheless, we did not find a significant correlation between LOXL2, GPR37 and SMOC1 with tumor stemness. In addition, H1F0 displayed a negative association with stromal scores, immune scores and estimate scores ($R = -0.34$

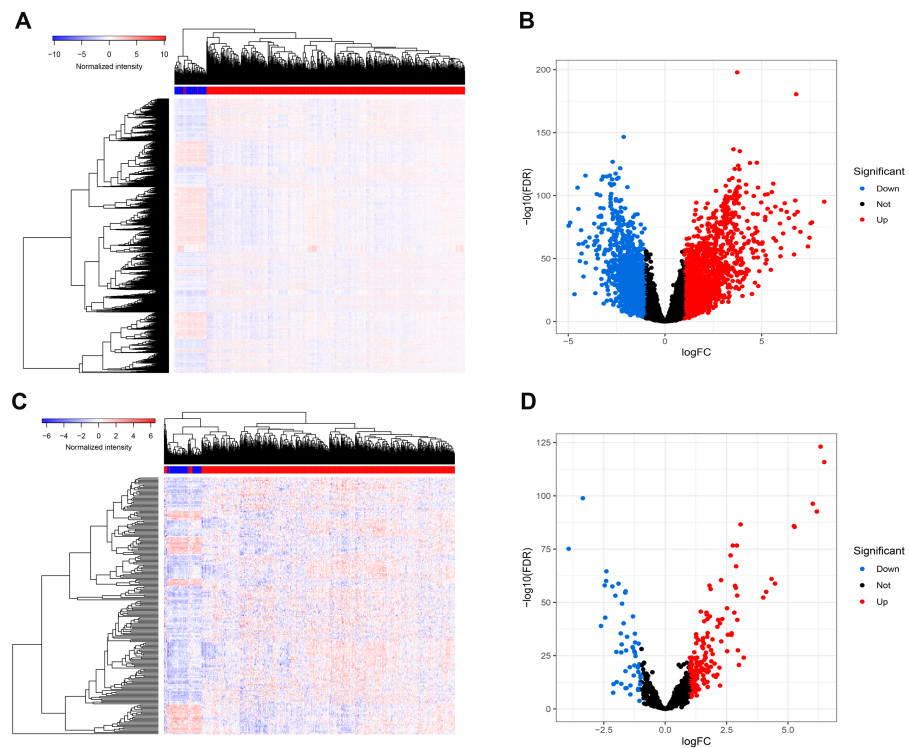


Figure 2 Differentially expressed mRNAs and lncRNA. Heatmap (A) and volcano plot (B) of DEmRNA. Heatmap (C) and volcano plot (D) of DELncRNA. Red dots and green dots represent up-regulated, down-regulated differentially expressed genes, respectively. DEmRNA and DELncRNA represent differentially expressed mRNA and lncRNA, respectively.

Full-size DOI: 10.7717/peerj.11029/fig-2

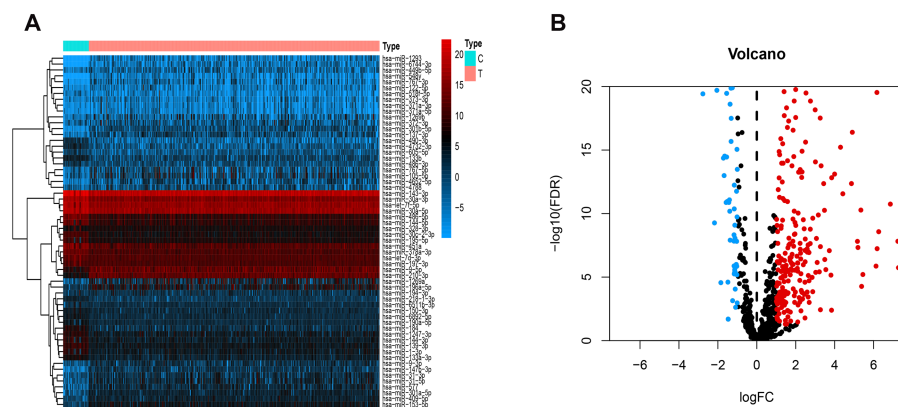


Figure 3 Differentially expressed miRNAs. Heatmaps of the top 20 differentially upregulated and the top 20 differentially downregulated miRNAs (A). Volcano plot (B) of DEmiRNAs. Red dots and green dots represent up-regulated, down-regulated differentially expressed genes, respectively. DEmiRNAs, differentially expressed miRNAs.

Full-size DOI: 10.7717/peerj.11029/fig-3

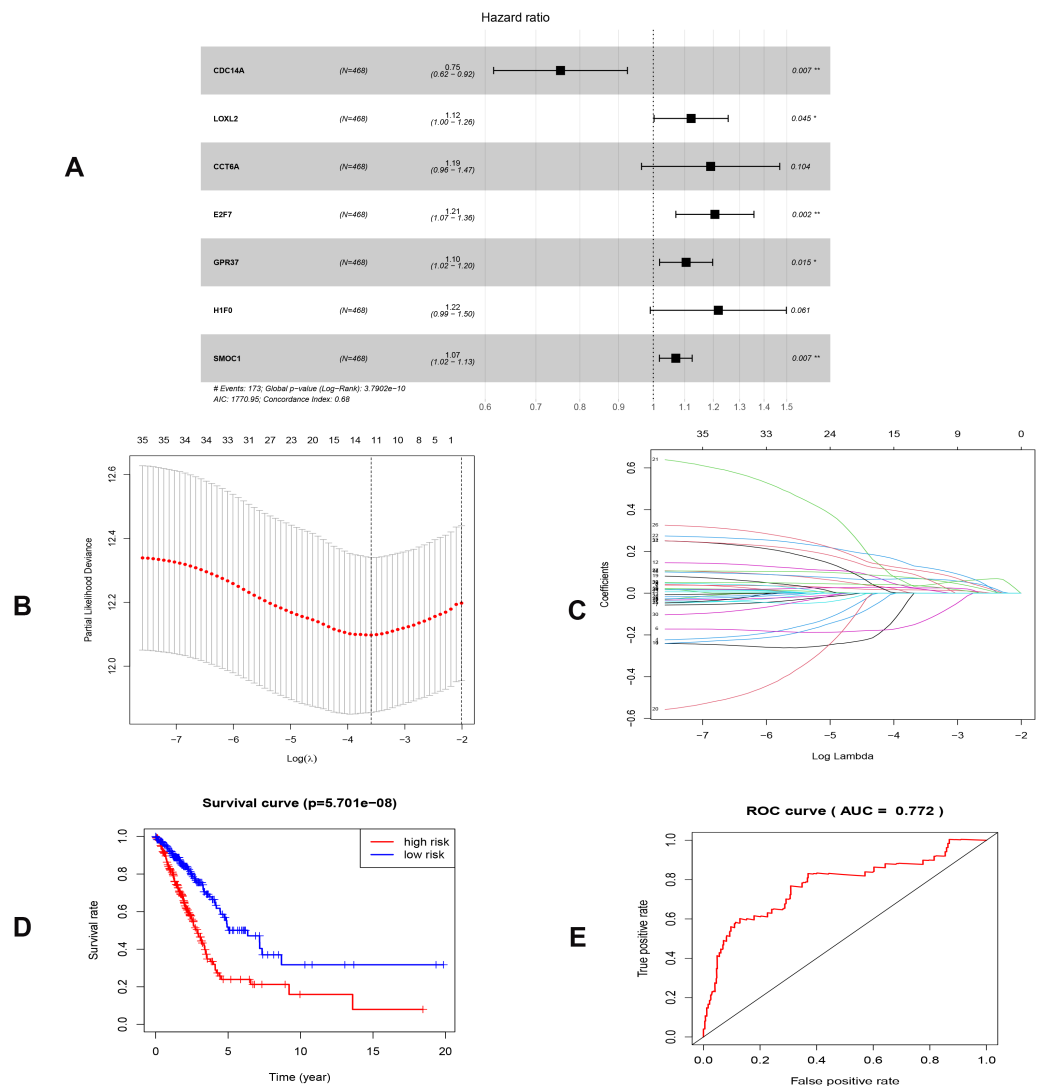


Figure 5 Construction of the prognostic model involving the key RNAs of the ceRNA network in LUAD. (A), multivariate Cox regression. (B), (C) lasso regression. (D) the Kaplan–Meier survival curve. (E) the ROC curve. ROC, the Receiver Operating Characteristic. LUAD, lung adenocarcinoma.

Full-size [DOI: 10.7717/peerj.11029/fig-5](https://doi.org/10.7717/peerj.11029/fig-5)

Differences in ssGSEA scores in TCGA/GEO cohort

To further elucidate the immune status in risk- and low groups of ceRNA signatures, we performed the ssGSEA algorithm to estimate the immune infiltrating between two groups (Fig. S9). Of note, low-risk group were associated with higher scores of aDCs, B_cells, CD8+ T_cells, iDCs, Mast_cells, Neutrophils, pDCs, T_helper_cells and TIL (Figs. 10A and 10B). Similarly, in the lower-risk group, the scores of Check-point, HLA, T_cell_co-stimulation and Type_II_IFN_Reponse in still higher in the low-risk group. Additionally, the differences above were confirmed by the GEO cohort (Figs 10C and 10D).

Table 3 Multivariate Cox proportional hazards regression model based on ceRNAs in LUAD.

Genes	Coef	HR	95% CI	P value
CDC14A	-0.282	0.754	(0.616–0.924)	0.007
LOXL2	0.115	1.122	(1.002–1.256)	0.045
CCT6A	0.174	1.190	(0.965–1.468)	0.104
E2F7	0.187	1.206	(1.071–1.358)	0.002
GPR37	0.100	1.105	(1.019–1.198)	0.015
H1F0	0.198	1.218	(0.991–1.498)	0.061
SMOC1	0.068	1.071	(1.019–1.125)	0.007

Notes.

LUAD, lung adenocarcinoma; HR, hazard ratio; CI, confidence interval.

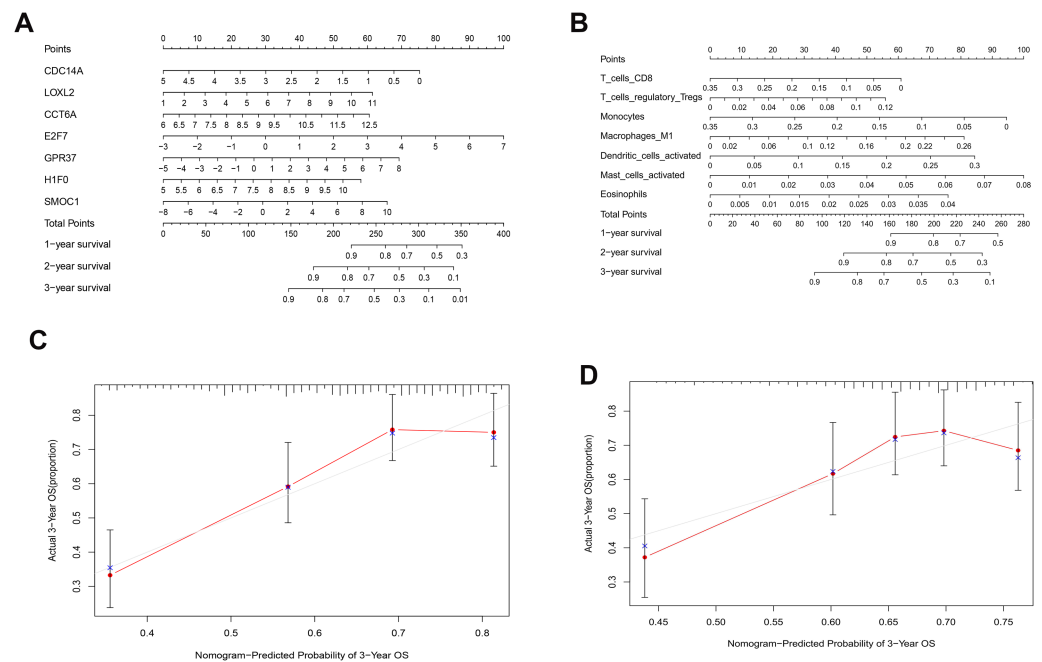


Figure 6 Nomograms based on key ceRNAs and immune cells. Nomograms predicting 1-, 2- and 3-year overall survival of ceRNAs signatures (A) and immune cells signatures (B). Calibration curves of the nomogram at the 3-year of ceRNAs (C) signatures and immune cells signatures (D).

Full-size [DOI: 10.7717/peerj.11029/fig-6](https://doi.org/10.7717/peerj.11029/fig-6)

DISCUSSION

With the reduced overall survival rate, lung adenocarcinoma is the most common pathological type since most patients were diagnosed at an advanced stage. Currently, there is a general consensus that both molecular and cellular characteristics play an imposing role in oncogenesis and metastasis, and thus they were suggested as underlying prognostic signatures. However, the integrative analysis of ceRNAs and microenvironments has yet to be fully explored. Previously, there were several studies that explored potential biomarkers of diagnosis and prognosis in LUAD via constructing the ceRNA model. Sui et al. constructed lncRNA-related network of LUAD with 29 lncRNAs, 24 miRNAs

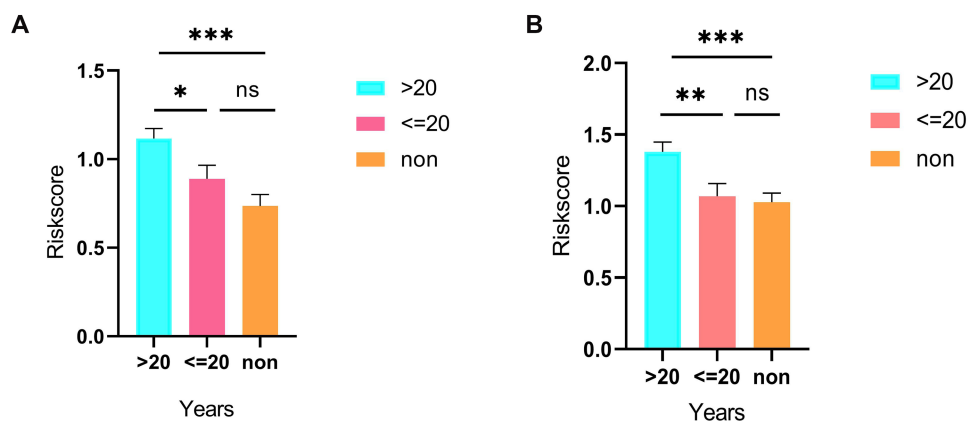


Figure 7 Differences in risk scores between >20-smoking years, ≤20-smoking years and non-smoking based on ceRNAs (A) and immune cells models (B).

Full-size DOI: [10.7717/peerj.11029/fig-7](https://doi.org/10.7717/peerj.11029/fig-7)

and 72mRNA (Sui et al., 2016). Then, they investigated the relevance of lncRNA and clinical characteristics. Cao et al. (2020) established a ceRNA network consisting of 23 lncRNAs, 117mRNAs and 22miRNAs in lung cancer. Cao et al. also proposed a 6-lncRNA prognostic model. Li et al. (2018) constructed three ceRNA networks and a 32-ceRNA prognostic signature. However, they did not further explore the correlation between immune microenvironment and ceRNAs comprehensively. In the present study, we not only presented nomogram prediction models based on ceRNAs and immune cells, but also further investigated the relationships between crucial ceRNAs and immune microenvironment. GSE72094 dataset proved the feasibility of predicted equations.

Cigarette smoking is an established risk factor of lung cancer, which accounts for the leading cause of death worldwide (Ettinger et al., 2019). In recent years, the prevalence of LUAD has increased and smoking patients with LUAD constitute a certain proportion (Zhang et al., 2015).

Besides, we stratified TCGA patients with smoking history into three groups (smoking years >20, smoking years ≤20 and non-smoking) and calculated risk scores in each sample via both predicted formulas, respectively. Not unexpectedly, both predicted formulas confirmed that the more extended smoking group (smoking years >20) have a higher risk than the shorter smoking group (smoking years ≤20). Thus, it could be concluded that smoking was an adverse factor for prognosis in both models. Risk scores in the >20 years-smoking group were increased when compared with non-smoking groups in both models. However, there were no significant differences between ≤ the 20 years-smoking group and the non-smoking group. It was related to that the patients with less duration of smoking were more frequently female. E2F7, a member of EF transcription factors, was known as a transcriptional repressor that influences the prognosis in a lot of tumors, such as breast cancer, endometrial carcinoma, gallbladder cancer, and colon cancer. Previous studies have proved that miR-30a-5p prevents tumor migration and metastasis by targeting E2F7 in gallbladder cancer (Ye et al., 2018). Liu et al. (2018) have revealed that E2F7 was reported significantly overexpressed in ER-positive breast cancer than healthy breast tissues

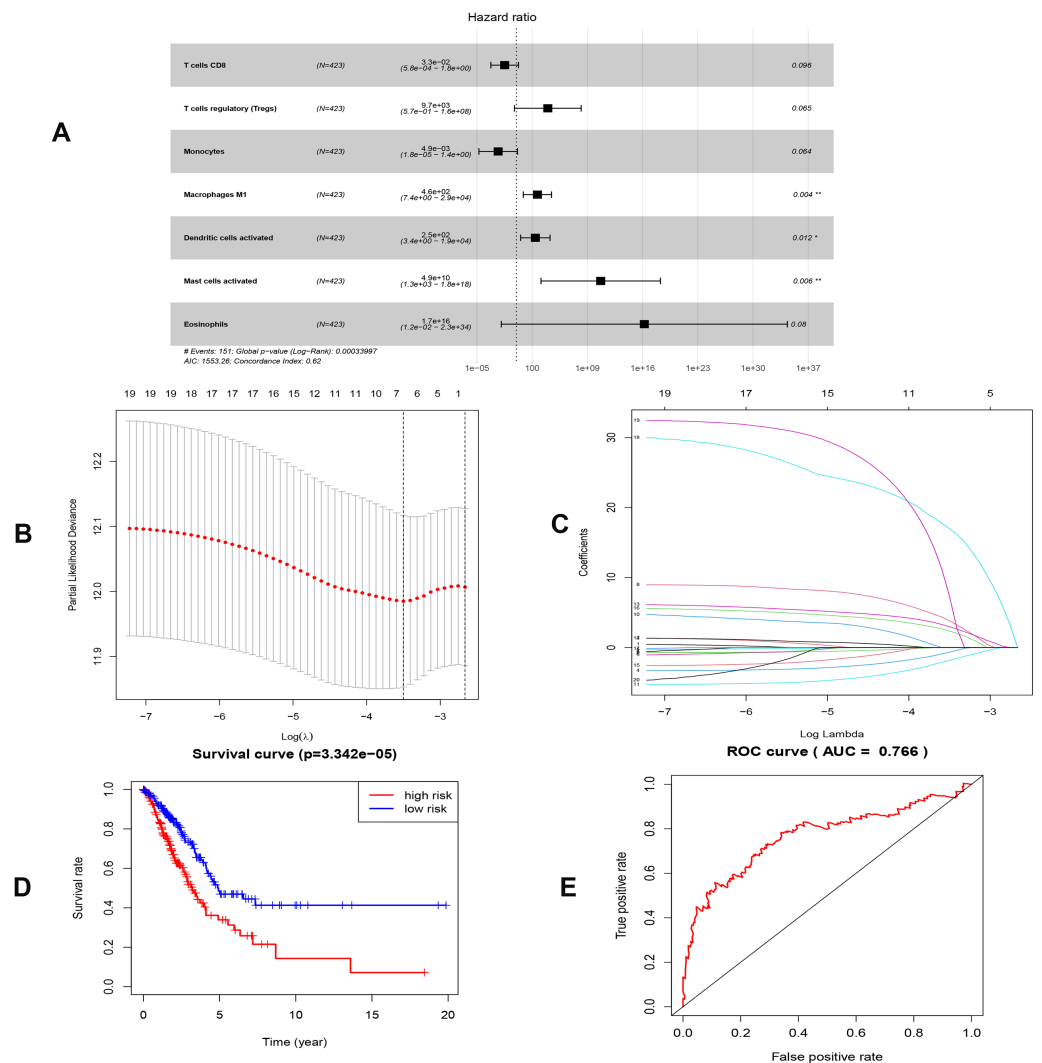


Figure 8 Construction of the prognostic model based on key immune cells in LUAD. (A), multivariate Cox regression. (B), (C) lasso regression. (D) the Kaplan–Meier survival curve. (E) the ROC curve. ROC, the Receiver Operating Characteristic. LUAD, lung adenocarcinoma.

Full-size DOI: 10.7717/peerj.11029/fig-8

and then led to tamoxifen resistance in breast cancer cells. Additionally, overexpressed E2F7 could be suggested as a significant biomarker to identify the high-risk and low-risk groups of patients with lung cancer, which is also in consistence with our report (*Sun et al., 2018*).

SMOC1 is a cancer-associated protein, identified as an extracellular glycoprotein of the SPARC-related modular calcium-binding protein family. SMOC1 was overexpressed in brain cancer, including oligodendrogliomas, astrocytomas, and glioblastomas (*Brellier et al., 2011; Fackler et al., 2011*). However, till now, there is no enough evidence for the impact of SMOC1 in LUAD reported in the literature. Moreover, the current study may provide a study direction for researches of SMOC1 in the LUAD. Additionally, CCTA6 is reported to encode a molecular chaperone and play critical roles in damaged proteins

Table 4 Multivariate Cox proportional hazards regression model based on immune cells in LUAD.

Immune cells	Coef	HR	95% CI	P value
T cells CD8	-3.426	0.033	(5.767e-04-1.835)	0.096
T cells regulatory (Tregs)	9.175	9653.370	(0.571-1.631e+08)	0.065
Monocytes	-5.318	0.005	(1.778e-05-1.351)	0.064
Macrophages M1	6.135	461.532	(7.447-2.860e+04)	0.004
Dendritic cells activated	5.539	254.456	(3.446-1.879e+04)	0.012
Mast cells activated	24.611	4.881e+10	(1.320e+03-1.805e+18)	0.006
Eosinophils	37.352	1.66589E+16	(0.012-2.282e+24)	0.080

Notes.

LUAD, lung adenocarcinoma; Coef, correlation coefficients; HR, hazrd ratio.

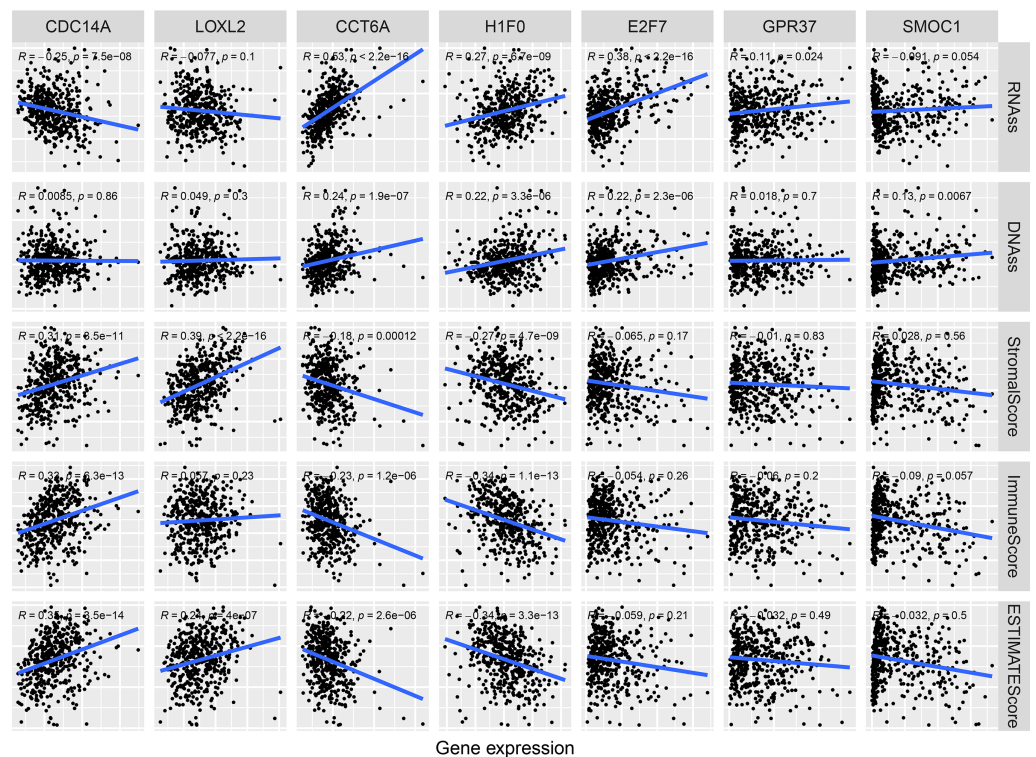
Cancer: LUAD

Figure 9 Correlations between crucial ceRNAs expression and stem cell index, crucial ceRNAs expression and the stromal/immune/ESTIMATE score in LUAD. RNAss and DNAss represent RNA-based stemness score, DNA-based stemness score, respectively. The sum of stromal score and immune score is ESTIMATE Score.

Full-size DOI: 10.7717/peerj.11029/fig-9

repair, cytoskeletal organization, and the cell cycle (Van Hove et al., 2015; Yam et al., 2008). Ying et al. (2017) found that CCTA6 served as a SMAD2-binding protein, which could inhibit the function of SMAD2, and promote metastasis.

CDC14A, which is of considerable significance to regulate the actin, was reported downregulated in many tumors and then reduced CDC14A was associated with the poor

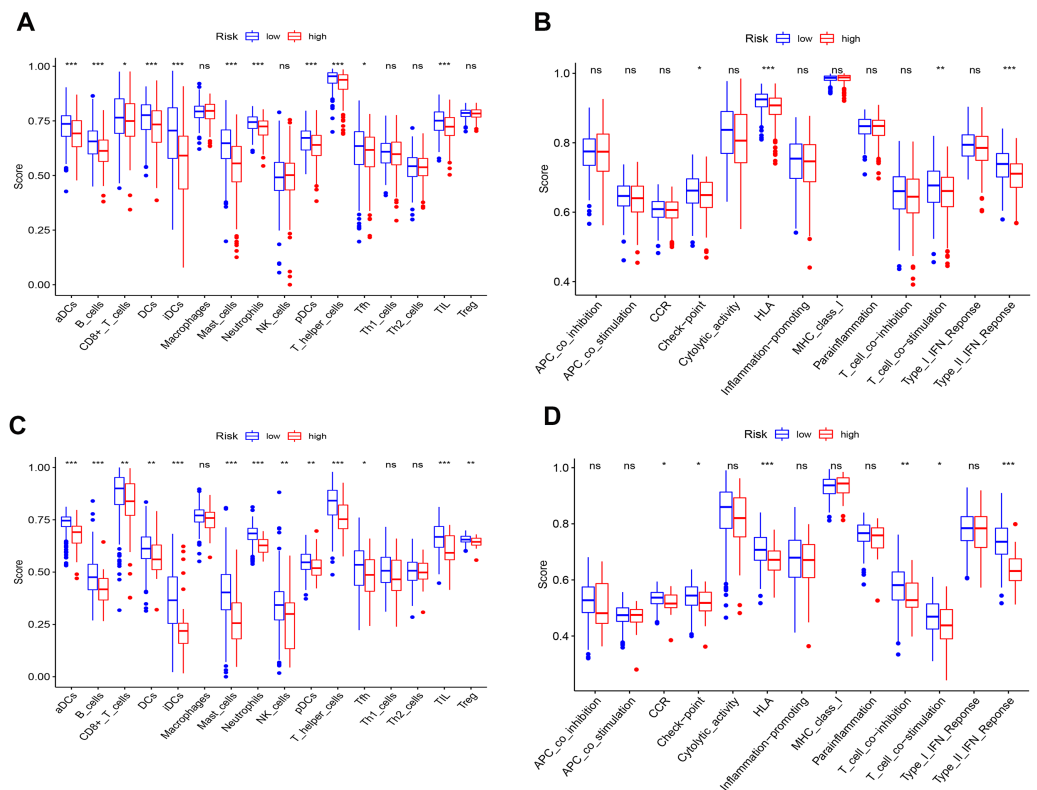


Figure 10 Differences of ssGSEA scores in high- and low-risk groups in TCGA (A–B) cohort and GEO (C–D) cohort. (A & C) refer to the difference in immune cells. (B & D) refer to the difference in immune-associated functions * $p < 0.05$, ** $p < 0.01$, and *** $p < 0.001$.

Full-size [DOI: 10.7717/peerj.11029/fig-10](https://doi.org/10.7717/peerj.11029/fig-10)

clinical outcomes (Chen et al., 2016). Our results indicated that CDC14A was a protective biomarker in LUAD (HR=0.75, $P = 0.0066$), which corresponded well with previous studies. Similarly, based on the correlation with stemness scores and immune scores, we could also draw inferences that CDC14A suppressed tumor progression.

As one of H1 histone genes, H1F0 (known as H1.0) is heterogeneous in many tumors (Di Liegro, Schiera & Liegro, 2018). Knockdown of H1.0 influenced the differentiation of embryonic stem cells (Terme et al., 2011). Recent studies revealed that H1F0 were prognostic indicators in many cancers, such as breast cancer, liver cancer, and kidney cancer (Torres et al., 2016). In the current study, the relationships of H1F0 expression with stemness scores and Estamine scores suggested that H1F0 might promote LUAD.

T cells regulatory (Tregs) account for 5%–10% of peripheral CD4+ T cells in humans. Increasing evidence demonstrated that the immunosuppression mediated by Tregs is one of the essential mechanisms of immune evasion in tumors (Dunn, Old & Schreiber, 2004; Shevach, 2002; Zou, 2005). Mast cells are increased in several tumors, and their accumulation was associated with a low survival rate in many cancers, such as pancreatic adenocarcinoma (Strouch et al., 2010) and melanoma (Ribatti et al., 2003). Macrophages M1 was historically recognized as the proinflammatory subgroup,

expressing a series of chemokines and consequently play antitumor roles (*Hao et al., 2012*). Controversially, our results suggested that M1 was highly expressed in LUAD (*Fig. 8*). Actually, heterogeneity of environment signals could influence TAM development, and comprehensive nomenclatures of TAM have consequently been suggested (*Ostuni et al., 2015*).

Additionally, our study found that there were obvious differences in immune status between high- and low- risk groups in both TCGA and GEO cohorts. This indicated that decreased risk scores of ceRNAs were related to the enhancement of antitumor immunity. Furthermore, we could say that diminished immune function is responsible for poor prognosis in LUAD patients. However, there are inevitably some disadvantages in our study that should be taken into consideration. Firstly, considering the limit data from the TCGA and GEO, it could result in large analysis deviation and be verified by other large cohorts. Secondly, our follow-up experiment will explore the molecular mechanisms of prognostic signatures in the present study.

CONCLUSION

To conclude, the present study firstly provided a combined analysis of ceRNA and immune cells and then established nomograms to predict the prognosis in LUAD reliably. Based on the present study result, we will carry out the biological experiments to validate key ceRNAs and mechanisms of immune cells in the future.

ACKNOWLEDGEMENTS

The authors would like to thank the TCGA, GEO, TIMER, GEPIA, HPA, UCLAN and TISIDB databases.

ADDITIONAL INFORMATION AND DECLARATIONS

Funding

The authors received no funding for this work.

Competing Interests

The authors declare there are no competing interests.

Author Contributions

- Miaomiao Zhang conceived and designed the experiments, performed the experiments, analyzed the data, prepared figures and/or tables, authored or reviewed drafts of the paper, and approved the final draft.
- Peiyan Zheng performed the experiments, analyzed the data, prepared figures and/or tables, authored or reviewed drafts of the paper, and approved the final draft.
- Yuan Wang conceived and designed the experiments, analyzed the data, authored or reviewed drafts of the paper, and approved the final draft.
- Baoqing Sun conceived and designed the experiments, authored or reviewed drafts of the paper, and approved the final draft.

Data Availability

The following information was supplied regarding data availability:

All raw data are available at TCGA and GEO. TCGA IDs and sample GEO accession numbers are available in the [Supplemental Files](#).

Supplemental Information

Supplemental information for this article can be found online at <http://dx.doi.org/10.7717/peerj.11029#supplemental-information>.

REFERENCES

- Barbie DA, Tamayo P, Boehm JS, Kim SY, Moody SE, Dunn IF, Schinzel AC, Sandy P, Meylan E, Scholl C, Fröhling S, Chan EM, Sos ML, Michel K, Mermel C, Silver SJ, Weir BA, Reiling JH, Sheng Q, Gupta PB, Wadlow RC, Le H, Hoersch S, Wittner BS, Ramaswamy S, Livingston DM, Sabatini DM, Meyerson M, Thomas RK, Lander ES, Mesirov JP, Root DE, Gilliland DG, Jacks T, Hahn WC. 2009. Systematic RNA interference reveals that oncogenic KRAS-driven cancers require TBK1. *Nature* 462:108–112 DOI 10.1038/nature08460.
- Bartel DP. 2009. MicroRNAs: target recognition and regulatory functions. *Cell* 136:215–233 DOI 10.1016/j.cell.2009.01.002.
- Bray F, Ferlay J, Soerjomataram I, Siegel RL, Torre LA, Jemal A. 2018. Global cancer statistics 2018: GLOBOCAN estimates of incidence and mortality worldwide for 36 cancers in 185 countries. *CA: A Cancer Journal for Clinicians* 68:394–424 DOI 10.3322/caac.21492.
- Brellier F, Ruggiero S, Zwolanek D, Martina E, Hess D, Brown-Luedi M, Hartmann U, Koch M, Merlo A, Lino M, Chiquet-Ehrismann R. 2011. SMOC1 is a tenascin-C interacting protein over-expressed in brain tumors. *Matrix Biology* 30:225–233 DOI 10.1016/j.matbio.2011.02.001.
- Bruttel VS, Wischhusen J. 2014. Cancer stem cell immunology: key to understanding tumorigenesis and tumor immune escape? *Frontiers in Immunology* 5:360 DOI 10.3389/fimmu.2014.00360.
- Cao Q, Dong Z, Liu S, An G, Yan B, Lei L. 2020. Construction of a metastasis-associated ceRNA network reveals a prognostic signature in lung cancer. *Cancer Cell International* 20:208 DOI 10.1186/s12935-020-01295-8.
- Chen B, Khodadoust MS, Liu CL, Newman AM, Alizadeh AA. 2018. Profiling tumor infiltrating immune cells with CIBERSORT. *Methods in Molecular Biology* 1711:243–259 DOI 10.1007/978-1-4939-7493-1_12.
- Chen NP, Uddin B, Voit R, Schiebel E. 2016. Human phosphatase CDC14A is recruited to the cell leading edge to regulate cell migration and adhesion. *Proceedings of the National Academy of Sciences of the United States of America* 113:990–995 DOI 10.1073/pnas.1515605113.
- Chou CH, Shrestha S, Yang CD, Chang NW, Lin YL, Liao KW, Huang WC, Sun TH, Tu SJ, Lee WH, Chiew MY, Tai CS, Wei TY, Tsai TR, Huang HT, Wang CY, Wu HY,

- Ho SY, Chen PR, Chuang CH, Hsieh PJ, Wu YS, Chen WL, Li MJ, Wu YC, Huang XY, Ng FL, Buddhakosai W, Huang PC, Lan KC, Huang CY, Weng SL, Cheng YN, Liang C, Hsu WL, Huang HD. 2018. miRTarBase update 2018: a resource for experimentally validated microRNA-target interactions. *Nucleic Acids Research* 46:D296–D302 DOI 10.1093/nar/gkx1067.
- Di Liegro CM, Schiera G, Di Liegro I. 2018. H1.0 linker histone as an epigenetic regulator of cell proliferation and differentiation. *Gene* 9(6):310 DOI 10.3390/genes9060310.
- Donnem T, Kilvaer TK, Andersen S, Richardsen E, Paulsen EE, Hald SM, Al-Saad S, Brustugun OT, Helland A, Lund-Iversen M, Solberg S, Gronberg BH, Wahl SG, Helgeland L, Fløtten O, Pohl M, Al-Shibli K, Sandanger TM, Pezzella F, Busund LT, Bremnes RM. 2016. Strategies for clinical implementation of TNM-immunoscore in resected nonsmall-cell lung cancer. *Annals of Oncology* 27:225–232 DOI 10.1093/annonc/mdv560.
- Dunn GP, Old LJ, Schreiber RD. 2004. The immunobiology of cancer immunosurveillance and immunoediting. *Immunity* 21:137–148 DOI 10.1016/j.immuni.2004.07.017.
- Ettinger DS, Wood DE, Aggarwal C, Aisner DL, Akerley W, Bauman JR, Bharat A, Bruno DS, Chang JY, Chirieac LR, D'Amico TA, Dilling TJ, Dobelbower M, Gettinger S, Govindan R, Gubens MA, Hennon M, Horn L, Lackner RP, Lanuti M, Leal TA, Lin J, Loo Jr BW, Martins RG, Otterson GA, Patel SP, Reckamp KL, Riely GJ, Schild SE, Shapiro TA, Stevenson J, Swanson SJ, Tauer KW, Yang SC, Gregory K, Hughes M. 2019. NCCN guidelines insights: non-small cell lung cancer, Version 1.2020. *Journal of the National Comprehensive Cancer Network* 17:1464–1472 DOI 10.6004/jnccn.2019.0059.
- Fackler MJ, Umbricht CB, Williams D, Argani P, Cruz LA, Merino VF, Teo WW, Zhang Z, Huang P, Visvanathan K, Marks J, Ethier S, Gray JW, Wolff AC, Cope LM, Sukumar S. 2011. Genome-wide methylation analysis identifies genes specific to breast cancer hormone receptor status and risk of recurrence. *Cancer Research* 71:6195–6207 DOI 10.1158/0008-5472.can-11-1630.
- Hänzelmann S, Castelo R, Guinney J. 2013. GSEA: gene set variation analysis for microarray and RNA-seq data. *BMC Bioinformatics* 14:7 DOI 10.1186/1471-2105-14-7.
- Hao NB, Lü MH, Fan YH, Cao YL, Zhang ZR, Yang SM. 2012. Macrophages in tumor microenvironments and the progression of tumors. *Clinical & Developmental Immunology* 2012:948098 DOI 10.1155/2012/948098.
- Imielinski M, Berger AH, Hammerman PS, Hernandez B, Pugh TJ, Hodis E, Cho J, Suh J, Capelletti M, Sivachenko A, Sougnez C, Auclair D, Lawrence MS, Stojanov P, Cibulskis K, Choi K, De Waal L, Sharifnia T, Brooks A, Greulich H, Banerji S, Zander T, Seidel D, Leenders F, Ansén S, Ludwig C, Engel-Riedel W, Stoelben E, Wolf J, Goparju C, Thompson K, Winckler W, Kwiatkowski D, Johnson BE, Jänne PA, Miller VA, Pao W, Travis WD, Pass HI, Gabriel SB, Lander ES, Thomas RK, Garraway LA, Getz G, Meyerson M. 2012. Mapping the hallmarks

- of lung adenocarcinoma with massively parallel sequencing. *Cell* **150**:1107–1120 DOI [10.1016/j.cell.2012.08.029](https://doi.org/10.1016/j.cell.2012.08.029).
- Jeggari A, Marks DS, Larsson E. 2012. miRcode: a map of putative microRNA target sites in the long non-coding transcriptome. *Bioinformatics* **28**:2062–2063 DOI [10.1093/bioinformatics/bts344](https://doi.org/10.1093/bioinformatics/bts344).
- Kartha RV, Subramanian S. 2014. Competing endogenous RNAs (ceRNAs): new entrants to the intricacies of gene regulation. *Frontiers in Genetics* **5**:8–8 DOI [10.3389/fgene.2014.00008](https://doi.org/10.3389/fgene.2014.00008).
- Lavin Y, Kobayashi S, Leader A, Amir ED, Elefant N, Bigenwald C, Remark R, Sweeney R, Becker CD, Levine JH, Meinhof K, Chow A, Kim-Shulze S, Wolf A, Medaglia C, Li H, Rytlewski JA, Emerson RO, Solovyov A, Greenbaum BD, Sanders C, Vignali M, Beasley MB, Flores R, Gnjatich S, Pe'er D, Rahman A, Amit I, Merad M. 2017. Innate immune landscape in early lung adenocarcinoma by paired single-cell analyses. *Cell* **169**:750–765 DOI [10.1016/j.cell.2017.04.014](https://doi.org/10.1016/j.cell.2017.04.014).
- Li J-H, Liu S, Zhou H, Qu L-H, Yang J-H. 2014. starBase v2.0: decoding miRNA-ceRNA, miRNA-ncRNA and protein-RNA interaction networks from large-scale CLIP-Seq data. *Nucleic Acids Research* **42**:D92–D97 DOI [10.1093/nar/gkt1248](https://doi.org/10.1093/nar/gkt1248).
- Li X, Li B, Ran P, Wang L. 2018. Identification of ceRNA network based on a RNA-seq shows prognostic lncRNA biomarkers in human lung adenocarcinoma. *Oncology Letters* **16**:5697–5708 DOI [10.3892/ol.2018.9336](https://doi.org/10.3892/ol.2018.9336).
- Liu J, Li X, Wang M, Xiao G, Yang G, Wang H, Li Y, Sun X, Qin S, Du N, Ren H, Pang Y. 2018. A miR-26a/E2F7 feedback loop contributes to tamoxifen resistance in ER-positive breast cancer. *International Journal of Oncology* **53**:1601–1612 DOI [10.3892/ijo.2018.4492](https://doi.org/10.3892/ijo.2018.4492).
- Liu X, Niu X, Qiu Z. 2020. A five-gene signature based on stromal/immune scores in the tumor microenvironment and its clinical implications for liver cancer. *DNA and Cell Biology* **39**:1621–1638 DOI [10.1089/dna.2020.5512](https://doi.org/10.1089/dna.2020.5512).
- Liu Z, Wan Y, Qiu Y, Qi X, Yang M, Huang J, Zhang Q. 2020. Development and validation of a novel immune-related prognostic model in lung squamous cell carcinoma. *International Journal of Medical Sciences* **17**:1393–1405 DOI [10.7150/ijms.47301](https://doi.org/10.7150/ijms.47301).
- Love MI, Huber W, Anders S. 2014. Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. *Genome Biology* **15**:550 DOI [10.1186/s13059-014-0550-8](https://doi.org/10.1186/s13059-014-0550-8).
- Malta TM, Sokolov A, Gentles AJ, Burzykowski T, Poisson L, Weinstein JN, Kamińska B, Huelsken J, Omberg L, Gevaert O, Colaprico A, Czerwińska P, Mazurek S, Mishra L, Heyn H, Krasnitz A, Godwin AK, Lazar AJ, Stuart JM, Hoadley KA, Laird PW, Noushmehr H, Wiznerowicz M. 2018. Machine learning identifies stemness features associated with oncogenic dedifferentiation. *Cell* **173**:338–354 DOI [10.1016/j.cell.2018.03.034](https://doi.org/10.1016/j.cell.2018.03.034).
- Newman AM, Liu CL, Green MR, Gentles AJ, Feng W, Xu Y, Hoang CD, Diehn M, Alizadeh AA. 2015. Robust enumeration of cell subsets from tissue expression profiles. *Nature Methods* **12**:453–457 DOI [10.1038/nmeth.3337](https://doi.org/10.1038/nmeth.3337).

- Ni SJ, Zhao LQ, Wang XF, Wu ZH, Hua RX, Wan CH, Zhang JY, Zhang XW, Huang MZ, Gan L, Sun HL, Dimri GP, Guo WJ. 2018. CBX7 regulates stem cell-like properties of gastric cancer cells via p16 and AKT-NF- κ B-miR-21 pathways. *Journal of Hematology & Oncology* 11:17 DOI 10.1186/s13045-018-0562-z.
- Ostuni R, Kratochvill F, Murray PJ, Natoli G. 2015. Macrophages and cancer: from mechanisms to therapeutic implications. *Trends in Immunology* 36:229–239 DOI 10.1016/j.it.2015.02.004.
- Phelps M, Coss C, Wang H, Cook M. 2016. Registered report: coding-independent regulation of the tumor suppressor PTEN by competing endogenous mRNAs. *Elife* 5:e12470 DOI 10.7554/eLife.12470.
- Qi X, Zhang DH, Wu N, Xiao JH, Wang X, Ma W. 2015. ceRNA in cancer: possible functions and clinical implications. *Journal of Medical Genetics* 52:710–718 DOI 10.1136/jmedgenet-2015-103334.
- Ribatti D, Ennas MG, Vacca A, Ferrel F, Nico B, Orru S, Sirigu P. 2003. Tumor vascularity and tryptase-positive mast cells correlate with a poor prognosis in melanoma. *European Journal of Clinical Investigation* 33:420–425 DOI 10.1046/j.1365-2362.2003.01152.x.
- Sanchez-Mejias A, Tay Y. 2015. Competing endogenous RNA networks: tying the essential knots for cancer biology and therapeutics. *Journal of Hematology & Oncology* 8:30 DOI 10.1186/s13045-015-0129-1.
- Shannon P, Markiel A, Ozier O, Baliga NS, Wang JT, Ramage D, Amin N, Schwikowski B, Ideker T. 2003. Cytoscape: a software environment for integrated models of biomolecular interaction networks. *Genome Research* 13:2498–2504 DOI 10.1101/gr.1239303.
- Shevach EM. 2002. CD4+ CD25+ suppressor T cells: more questions than answers. *Nature Reviews Immunology* 2:389–400 DOI 10.1038/nri821.
- Shukla GC, Singh J, Barik S. 2011. MicroRNAs: processing, maturation, target recognition and regulatory functions. *Molecular and Cellular Pharmacology* 3:83–92.
- Siegel RL, Miller KD, Jemal A. 2018. Cancer statistics, 2018. *CA: A Cancer Journal for Clinicians* 68:7–30 DOI 10.3322/caac.21442.
- Strouch MJ, Cheon EC, Salabat MR, Krantz SB, Gounaris E, Melstrom LG, Dangi-Garimella S, Wang E, Munshi HG, Khazaie K, Bentrem DJ. 2010. Crosstalk between mast cells and pancreatic cancer cells contributes to pancreatic tumor progression. *Clinical Cancer Research* 16:2257–2265 DOI 10.1158/1078-0432.ccr-09-1230.
- Sui J, Li YH, Zhang YQ, Li CY, Shen X, Yao WZ, Peng H, Hong WW, Yin LH, Pu YP, Liang GY. 2016. Integrated analysis of long non-coding RNA-associated ceRNA network reveals potential lncRNA biomarkers in human lung adenocarcinoma. *International Journal of Oncology* 49:2023–2036 DOI 10.3892/ijo.2016.3716.
- Sun CC, Zhou Q, Hu W, Li SJ, Zhang F, Chen ZL, Li G, Bi ZY, Bi YY, Gong FY, Bo T, Yuan ZP, Hu WD, Zhan BT, Zhang Q, Tang QZ, Li DJ. 2018. Transcriptional E2F1/2/5/8 as potential targets and transcriptional E2F3/6/7 as new

- biomarkers for the prognosis of human lung carcinoma. *Aging* **10**:973–987 DOI [10.18632/aging.101441](https://doi.org/10.18632/aging.101441).
- Tay Y, Rinn J, Pandolfi P. 2014.** The multilayered complexity of ceRNA crosstalk and competition. *Nature* **505**:344–352 DOI [10.1038/nature12986](https://doi.org/10.1038/nature12986).
- Terme JM, Sesé B, Millán-Ariño L, Mayor R, Izpisúa Belmonte JC, Barrero MJ, Jordan A. 2011.** Histone H1 variants are differentially expressed and incorporated into chromatin during differentiation and reprogramming to pluripotency. *Journal of Biological Chemistry* **286**:35347–35357 DOI [10.1074/jbc.M111.281923](https://doi.org/10.1074/jbc.M111.281923).
- Torres CM, Biran A, Burney MJ, Patel H, Henser-Brownhill T, Cohen AS, Li Y, Ben-Hamo R, Nye E, Spencer-Dene B, Chakravarty P, Efroni S, Matthews N, Misteli T, Meshorer E, Scaffidi P. 2016.** The linker histone H1.0 generates epigenetic and functional intratumor heterogeneity. *Science* **353**:aaf1644 DOI [10.1126/science.aaf1644](https://doi.org/10.1126/science.aaf1644).
- Van Hove I, Verslegers M, Hu TT, Carden M, Arckens L, Moons L. 2015.** A proteomic approach to understand MMP-3-driven developmental processes in the postnatal cerebellum: chaperonin CCT6A and MAP kinase as contributing factors. *Developmental Neurobiology* **75**:1033–1048 DOI [10.1002/dneu.22272](https://doi.org/10.1002/dneu.22272).
- Yam AY, Xia Y, Lin HT, Burlingame A, Gerstein M, Frydman J. 2008.** Defining the TRiC/CCT interactome links chaperonin function to stabilization of newly made proteins with complex topologies. *Nature Structural & Molecular Biology* **15**:1255–1262 DOI [10.1038/nsmb.1515](https://doi.org/10.1038/nsmb.1515).
- Ye YY, Mei JW, Xiang SS, Li HF, Ma Q, Song XL, Wang Z, Zhang YC, Liu YC, Jin YP, Hu YP, Jiang L, Liu FT, Zhang YJ, Hao YJ, Liu YB. 2018.** MicroRNA-30a-5p inhibits gallbladder cancer cell proliferation, migration and metastasis by targeting E2F7. *Cell Death & Disease* **9**:410 DOI [10.1038/s41419-018-0444-x](https://doi.org/10.1038/s41419-018-0444-x).
- Ying Z, Tian H, Li Y, Lian R, Li W, Wu S, Zhang HZ, Wu J, Liu L, Song J, Guan H, Cai J, Zhu X, Li J, Li M. 2017.** CCT6A suppresses SMAD2 and promotes prometastatic TGF- β signaling. *Journal of Clinical Investigation* **127**:1725–1740 DOI [10.1172/jci90439](https://doi.org/10.1172/jci90439).
- Yoshihara K, Shahmoradgoli M, Martínez E, Vegesna R, Kim H, Torres-Garcia W, Treviño V, Shen H, Laird PW, Levine DA, Carter SL, Getz G, Stemke-Hale K, Mills GB, Verhaak RG. 2013.** Inferring tumour purity and stromal and immune cell admixture from expression data. *Nature Communications* **4**:2612 DOI [10.1038/ncomms3612](https://doi.org/10.1038/ncomms3612).
- Zhang L, Li M, Wu N, Chen Y. 2015.** Time trends in epidemiologic characteristics and imaging features of lung adenocarcinoma: a population study of 21,113 cases in China. *PLOS ONE* **10**:e0136727 DOI [10.1371/journal.pone.0136727](https://doi.org/10.1371/journal.pone.0136727).
- Zou W. 2005.** Immunosuppressive networks in the tumour environment and their therapeutic relevance. *Nature Reviews Cancer* **5**:263–274 DOI [10.1038/nrc1586](https://doi.org/10.1038/nrc1586).