



OPEN

Comprehensive analysis of competitive endogenous RNA associated with immune infiltration in lung adenocarcinoma

Wenjie Chen^{1,5}, Wen Li^{1,5}, Zhenkun Liu^{1,5}, Guangzhi Ma^{1,2}, Yunfu Deng¹, Xiaogang Li¹, Zhu Wang³ & Qinghua Zhou^{1,4}✉

To identify the prognostic biomarker of the competitive endogenous RNA (ceRNA) and explore the tumor infiltrating immune cells (TIICs) which might be the potential prognostic factors in lung adenocarcinoma. In addition, we also try to explain the crosstalk between the ceRNA and TIICs to explore the molecular mechanisms involved in lung adenocarcinoma. The transcriptome data of lung adenocarcinoma were obtained from The Cancer Genome Atlas (TCGA) database, and the hypergeometric correlation of the differently expressed miRNA-lncRNA and miRNA-mRNA were analyzed based on the starBase. In addition, the Kaplan–Meier survival and Cox regression model analysis were used to identify the prognostic ceRNA network and TIICs. Correlation analysis was performed to analysis the correlation between the ceRNA network and TIICs. In the differently expressed RNAs between tumor and normal tissue, a total of 190 miRNAs, 224 lncRNAs and 3024 mRNAs were detected, and the constructed ceRNA network contained 5 lncRNAs, 92 mRNAs and 10 miRNAs. Then, six prognostic RNAs (*FKBP3*, *GPI*, *LOXL2*, *IL22RA1*, *GPR37*, and *has-miR-148a-3p*) were viewed as the key members for constructing the prognostic prediction model in the ceRNA network, and three kinds of TIICs (Monocytes, Macrophages M1, activated mast cells) were identified to be significantly related with the prognosis in lung adenocarcinoma. Correlation analysis suggested that the *FKBP3* was associated with Monocytes and Macrophages M1, and the *GPI* was obviously related with Monocytes and Macrophages M1. Besides, the *LOXL2* was associated with Monocytes and Activated mast cells, and the *IL22RA1* was significantly associated with Monocytes and Macrophages M1, while the *GPR37* and Macrophages M1 was closely related. The constructed ceRNA network and identified Monocytes, Macrophages M1 and activated Mast cells are all prognostic factors for lung adenocarcinoma. Moreover, the crosstalk between the ceRNA network and TIICs might be a potential molecular mechanism involved.

Lung cancer is one of the most frequently diagnosed malignancies worldwide, and it represents almost one-quarter of all cancer deaths¹. Lung cancer is categorised into adenocarcinoma (LUAD), squamous cell type and large cell type, and LUAD accounts for more than 40% of all lung cancer cases². For lung cancer patients, it is often difficult to diagnosed until an advanced disease stage, hence, the 5-year overall survival rate is still frustrating for all stages³. Besides, the conventional treatments, including surgery, radiotherapy, and chemotherapy, for metastatic lung cancer is remains modest, so it is vital to make a comprehensive understanding of the tumorigenesis and tumor progression and find out potential biomarkers to accurately predict the prognosis^{4,5}.

Over the years, it has been generally recognized that non-coding RNA (ncRNA) have transformed the tumorigenesis and progression landscape in serious tumors. Competitive endogenous RNA (ceRNA) network, composed of mRNAs, long non-coding RNAs (lncRNAs) and microRNAs (miRNAs), is a novel class of ncRNAs and plays an important role in the post-transcriptional regulation by modulate the oncogenes or tumor suppressor

¹Lung Cancer Center, West China Hospital, Sichuan University, Chengdu, China. ²Department of Thoracic Surgery, West China Hospital, Sichuan University, Chengdu, China. ³Laboratory of Molecular Diagnosis of Cancer, Cancer Center, West China Hospital, Sichuan University, Chengdu, China. ⁴Tianjin Key Laboratory of Lung Cancer Metastasis and Tumor Microenvironment, Tianjin Lung Cancer Institute, Tianjin Medical University General Hospital, Tianjin, China. ⁵These authors contributed equally: Wenjie Chen, Wen Li and Zhenkun Liu. ✉email: zhouqh135@163.com

genes and the interactions between protein and genes^{6–8}. Moreover, studies have recently demonstrated that the ceRNA regulatory networks is a prognostic indicator in different diseases, such as cardiovascular diseases, diabetic cataract, myocardial infarction, and some malignancies, including myeloid leukemia, bladder cancer and pancreatic adenocarcinoma^{9,10}.

Recently, studies have identified that the tumor immune microenvironment (TIME) plays a vital role in tumorigenesis, metastasis and could also critically influence therapeutic response. Tumor infiltrating immune cells (TIICs) which are regulated by different cytokines and chemokines consist the main regulators of tumor biology in the tumor microenvironment (TME)^{11,12}. Previous study shows that TIICs, especially the tumor-associated macrophage, are associated with oncological outcomes and drug response to Bacillus Calmette-Guerin (BCG) treatment in non-muscle-invasive bladder cancer¹³. Besides, it has also been shown that the TIICs could have effect on the ceRNA network by regulating the expression of non-coding RNA directly or indirectly¹⁴.

In the current study, we detected the different expression of ceRNA in TCGA lung adenocarcinoma and calculated the TIICs by CIBERSORT algorithm. Then, we built nomograms based on the prognosis model, which can be used to assess prognosis for lung adenocarcinoma patients. Moreover, the correlation analysis between the prognostic members in the ceRNA and prognostic TIICs was performed, our results might render to clarify the mechanism of the occurrence and development of lung adenocarcinoma combined with tumor immune microenvironment regulation.

Methods

Data collection and identification of differently expressed RNAs. We downloaded all the miRNA, lncRNA and mRNA expression profiles and relevant clinical data such as pathological factors, and the survival outcome of the lung adenocarcinoma cohort from The Cancer Genome Atlas (TCGA) (<https://portal.gdc.cancer.gov/>), including 497 cancer cases, 54 cases of adjacent normal tissues and 486 cases of relevant clinical data. Another independent validation sets GSE14814¹⁵ with 71 LUAD samples and common clinicopathological characteristics were obtained from the Gene Expression Omnibus (GEO) database (<https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE14814>).

The differently expressed RNAs were screened by using the R package ‘edgeR¹⁶’, $|\log(\text{fold change})| \geq 1.0$ and $p\text{-value} < 0.05$ were set as the cut-off. To visualize the differently expressed RNAs, we used the R package ‘ggplot2’.

Construction of the ceRNA network and functional enrichment analysis. The hypergeometric correlation of the differently expressed miRNA-lncRNA and miRNA-mRNA were analyzed based on the starBase¹⁷ (<http://starbase.sysu.edu.cn/>) information by using the R package ‘GDCRNATools¹⁸’, $p\text{-value} < 0.05$ were set as the cut-off. Then Cytoscape¹⁹ v3.8.2 was performed to visualize the ceRNA network of miRNA regulated the lncRNAs and mRNAs. The gene ontology (GO) analysis and the Kyoto Encyclopaedia of Genes and Genomes (KEGG) pathway enrichment analysis were performed by the R package ‘clusterProfiler²⁰’, the false discovery rate (FDR) and $p\text{-value} < 0.05$ was set as the cut-off.

Construction of prognostic prediction model. Univariate Cox hazards regression analysis was used to preliminary screen the candidate genes from the ceRNA network, $p\text{-value} < 0.05$ were set as the cut-off. Lasso regression analysis²¹ and multivariate Cox hazards regression analysis²² were performed to further verify the prognostic value, $p\text{-value} < 0.05$ were set as the cut-off. The multivariate Cox hazards regression coefficients were set as the β , and we calculated each case risk score with the formula: $\text{risk score} = \beta_{\text{gene}[1]} \times \text{Expression}_{\text{gene}[1]} + \beta_{\text{gene}[2]} \times \text{Expression}_{\text{gene}[2]} + \dots + \beta_{\text{gene}[n]} \times \text{Expression}_{\text{gene}[n]}$. Patients above the median risk score would be divided into the high-risk group, and the rest would be divided into the low-risk group. The ROC curves and K-M survival curves were utilised to evaluate the diagnostic efficacies. Finally, we constructed a nomogram to predict the prognosis of lung adenocarcinoma patients and the calibration curves were utilised to access the accuracy.

Estimation of the immune cells in the lung adenocarcinoma. CIBERSORT algorithm²³ was performed to estimate the component of the total immune cells in the lung adenocarcinoma and adjacent normal tissues, $p\text{-value} < 0.05$ were set as the cut-off. To visualize the composition of immune cells in the case, we used the R package ‘corplot’, ‘heatmap’ and ‘vioplot’.

Correlation analysis. We performed Cox hazards regression analysis and Lasso regression analysis to identify the survival related immune cells, $p\text{-value} < 0.05$ were set as the cut-off. Then we constructed a nomogram to predict the prognosis of the lung adenocarcinoma patients. The ROC curves and calibration curves were utilised to access the accuracy. Correlation analysis was performed for each key gene in the ceRNA network and each survival related immune cell with the Pearson method.

Result

Identification of the differently expressed miRNAs, lncRNAs and mRNAs. All miRNA, lncRNA and mRNA profiles of lung adenocarcinoma were obtained from the TCGA. A total of 190 differently expressed miRNAs (122 up-regulated and 68 down-regulated) (Fig. 1A), 224 differently expressed lncRNAs (176 up-reg-

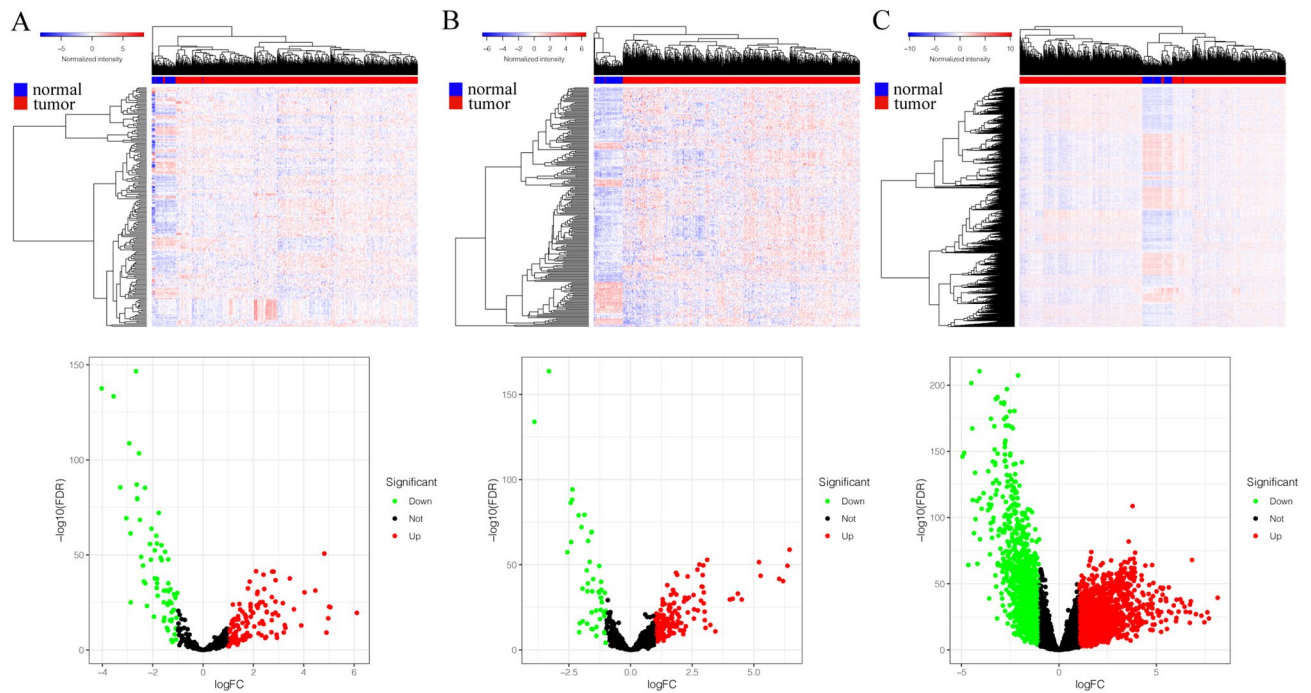


Figure 1. Differently expressed RNAs in TCGA lung adenocarcinoma. (A) The heatmap of the miRNAs expression in LUAD, and the volcano graph shows the distribution of differently expressed miRNAs. (B) The heatmap of the lncRNAs expression in LUAD, and the volcano graph shows the distribution of differently expressed lncRNAs. (C) The heatmap of the mRNAs expression in LUAD, and the volcano graph shows the distribution of differently expressed mRNAs. In the volcano graph, the X axis represents the fold changes of DEGs, and the Y axis represents the adjusted p-value, red dots present up-regulated genes and blue dots present down-regulated genes (\log fold change > 1.0 , p -value < 0.05).

ulated and 48 down-regulated) (Fig. 1B) and 3024 differently expressed mRNAs (1768 up-regulated and 1256 down-regulated) (Fig. 1C) between tumor and normal tissue were identified, $|\log$ FC > 1.0 and p -value < 0.05 set as the cut-off.

Construction of the ceRNA network and functional enrichment analysis. Based on the information of starBase database, the lncRNA-miRNA and mRNA-miRNA interaction of the differently expressed genes was confirmed and a ceRNA network contained 5 lncRNAs, 92 mRNAs and 10 miRNAs were constructed (Fig. 2A). The KEGG pathway analysis showed that the genes in the network were primarily clustered in the small cell lung cancer, Cell cycle, Epstein-Barr virus infection and Cellular senescence (Fig. 2B). For the biological process (BP), the genes in the network were primarily enriched in the positive regulation of DNA-dependent DNA replication, positive regulation of cell cycle process and cell cycle G1/S phase transition. For the cell component (CC), the genes were primarily enriched in complex of collagen trimers and basement membrane. For the molecular function (MF), the genes were mainly enriched in extracellular matrix structural constituent conferring tensile strength, guanyl-nucleotide exchange factor activity and Ras guanyl-nucleotide exchange factor activity (Fig. 2C).

Identification of key genes and construction of prognostic prediction model. All clinicopathological characteristics of patients in TCGA lung adenocarcinoma were summarized in Table 1. Univariate Cox hazards regression analysis detected 45 candidate genes from the ceRNA network (Supplementary Table) and the lasso regression analysis filtered 12 genes were reliable for prediction model, p -value < 0.05 set as the cut-off (Fig. 3A,B). Then multivariate Cox hazards regression analysis identified 6 genes (*FKBP3*, *GPI*, *LOXL2*, *IL22RA1*, *GPR37*, *has-miR-148a-3p*) were as the key members for constructing the prognostic prediction model in the network, p -value < 0.05 set as the cut-off (Fig. 3C). Based on the model, a nomogram was constructed and a good discrimination of the model was validated by the calibration curves (Fig. 3D,E). Kaplan-Meier curves validated these 6 key genes (Fig. 4A-F) and the model were significantly associated with overall survival, p -value < 0.05 set as the cut-off (Fig. 4G). The ROC curves showed that the area under curve (AUC) at 1 year was 0.677, at 3 year was 0.705, and at 5 years was 0.709 (Fig. 4H). We used GSE14814 lung adenocarcinoma

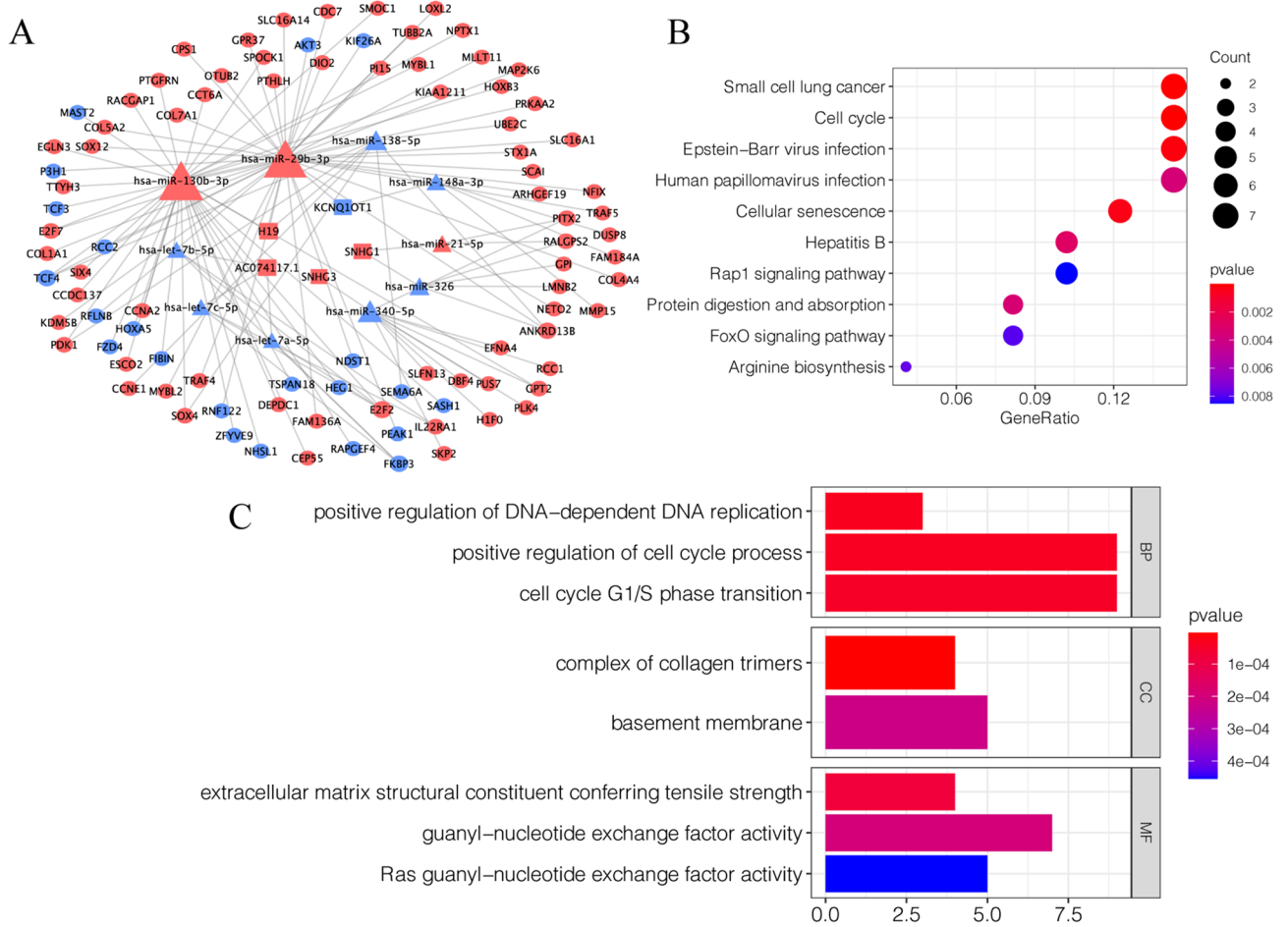


Figure 2. Construction of the ceRNA network and functional enrichment analysis. **(A)** The ceRNA network of the differently expressed RNAs in lung carcinoma. The triangle nodes indicate miRNAs; the rectangle nodes indicate lncRNAs; the ellipse nodes indicate mRNAs. The blue nodes indicate down-regulated genes; the red nodes indicate up-regulated genes. Node size indicates the number of the interactions. **(B)** The KEGG enrichment analysis of the ceRNA network genes. **(C)** The GO enrichment analysis of the ceRNA network genes.

data sets ($n = 71$) to evaluate the prognostic value of the model. The Kaplan–Meier curves validated the model were significantly associated with overall survival, p -value < 0.05 , and the AUC at 3 year was 0.915, at 5 year was 0.615 (Fig. 4G). Subgroup analysis classified by patients’ age, gender, smoking status or clinic stage were subsequently conducted, the results revealed that the high-risk group in this model were significantly associated with worse prognosis among all the subgroups (Fig. 5A–D). The expression of the 6 key genes (*FKBP3*, *GPI*, *LOXL2*, *IL22RA1*, *GPR37*, *has-miR-148a-3p*) in lung adenocarcinoma patients with different tumor stages, T stages, N stages and M stages (Fig. 5E–J).

Comprehensive analysis of immune cells and prognostic value in lung adenocarcinoma. The CIBERSORT algorithm estimated the component of immune cells in lung adenocarcinoma (Fig. 6A,B). Cox hazards regression analysis detected 3 immune cells (Monocytes, Macrophages M1, activated mast cells) were associated with the prognosis in lung adenocarcinoma (Table 2), and the lasso regression analysis validated the 3 immune cells were reliable for prediction as an immune cell model, p -value < 0.05 set as the cut-off (Fig. 7A,B). Based on the result, a nomogram was constructed (Fig. 7C). The ROC curves showed that the area under curve (AUC) at 1 year was 0.621, at 3 year was 0.604, and at 5 years was 0.585 (Fig. 7D). The discrimination of the 3 immune cells were validated by the calibration curves (Fig. 7E).

Age	Cases (%)
≤ 65	215 (44.2%)
> 65	226 (46.6%)
Unknown	45 (9.2%)
Gender	
Male	222 (47.44%)
Female	264 (56.41%)
Smoking history	
Smoked	284 (60.68%)
Non- smoked	190 (40.60%)
Unknow	12 (2.56%)
Clinical stage	
I	262 (55.98%)
II	112 (23.93%)
III	79 (16.88%)
IV	25 (5.34%)
Unknow	8 (1.71%)
T _stage	
T1	163 (34.83%)
T2	260 (55.56%)
T3	41 (8.76%)
T4	19 (4.06%)
Tx/Unknow	3 (0.64%)
N _stage	
N0	312 (66.67%)
N1	90 (19.23%)
N2	70 (14.96%)
N3	2 (0.43%)
Nx/Unknow	12 (2.56%)
M _stage	
M0	333 (71.15%)
M1	24 (5.13%)
Mx/Unknow	129 (27.56%)

Table 1. Clinicopathological characteristics of patients in TCGA lung carcinoma.

Correlation analysis between the key genes of ceRNA network and the prognostic immune cells.

The ROC analysis at 3 year performed for comparison the effect of the ceRNA network and the combination of TIICs. The results showed that the AUC for network and network+ TIICs is 0.705 and 0.706, respectively (Fig. 8A). Pearson correlation analysis indicated that the *FKBP3* was negatively associated with Monocytes ($R = -0.13$, $p < 0.05$) and positively associated with Macrophages M1 ($R = 0.12$, $p < 0.05$) (Fig. 8C). The *GPI* was negatively associated with Monocytes ($R = -0.18$, $p < 0.05$) and positively associated with Macrophages M1 ($R = 0.15$, $p < 0.05$) (Fig. 8D). The *LOXL2* was negatively associated with Monocytes ($R = -0.12$, $p < 0.05$) and positively associated with activated mast cells ($R = 0.11$, $p < 0.05$) (Fig. 8E). The *IL22RA1* was negatively associated with Monocytes ($R = -0.11$, $p < 0.05$) and positively associated with Macrophages M1 ($R = 0.21$, $p < 0.05$) (Fig. 8F). Moreover, the *GPR37* was positively associated with Macrophages M1 ($R = 0.12$, $p < 0.05$) (Fig. 8G). The correlation of all key genes and the prognostic immune cells were summarized in the heatmap (Fig. 8B).

Discussion

Lung cancer is a heterogeneous disease which is still a big challenge to treatment. Even a part of patients with early-stage tumors who received surgery ultimately develop recurrence²⁴. In the process of tumor occurrence and progression, molecular components play an important role and are generally considered as potential prognostic factors²⁵. Tumor infiltrating immune cells have been proved to be very important in the occurrence and development of tumors, and have shown the prognostic value of a variety of malignant tumors^{26–28}.

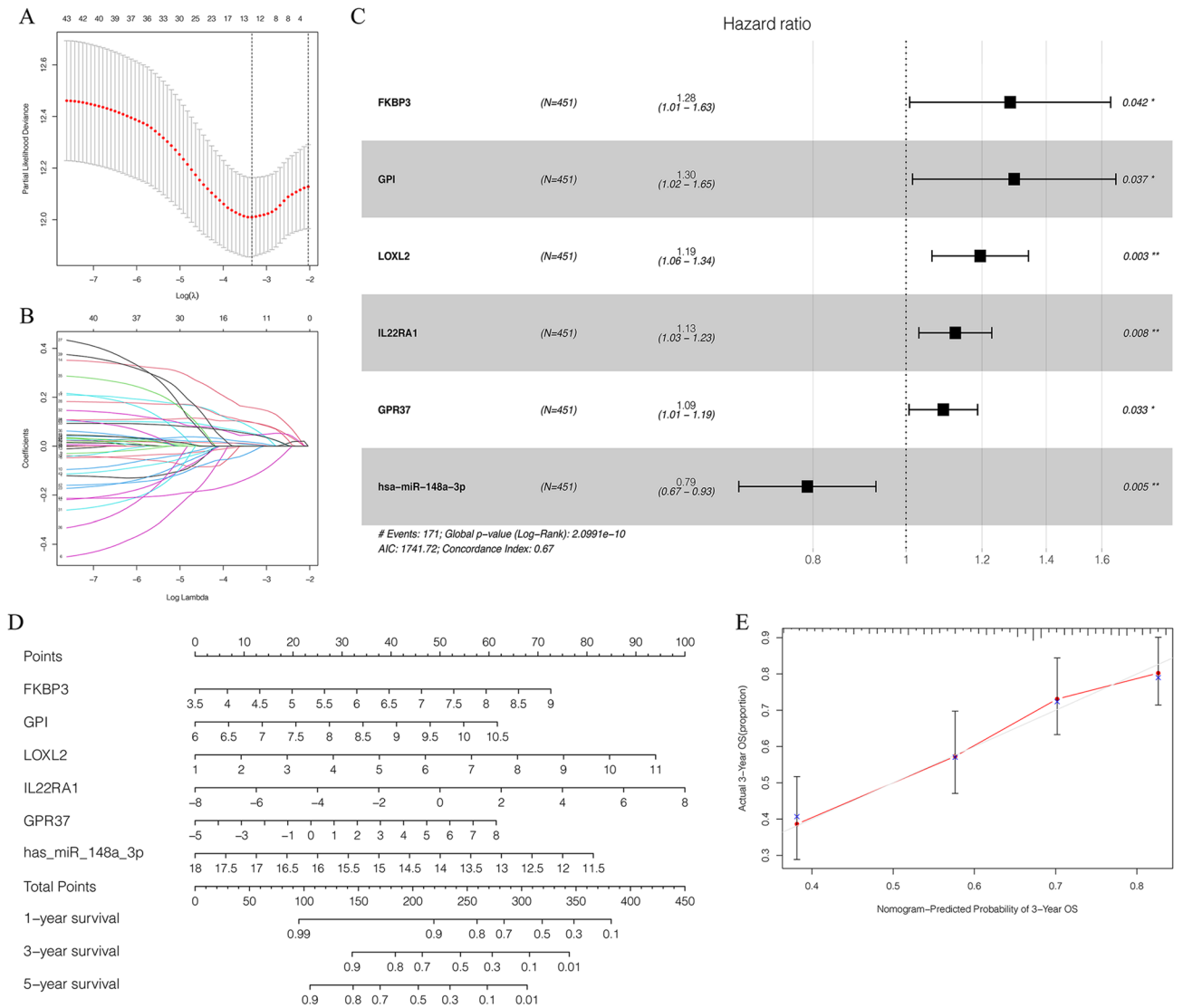


Figure 3. Identification of prognostic members in ceRNA network. (A,B) The Lasso regression suggested that 12 genes were essential for prognostic prediction model. (C) The Multivariate Cox regression analysis identified six genes (*FKBP3*, *GPI*, *LOXL2*, *IL22RA1*, *GPR37*, *has-miR-148a-3p*) were as the key gene for constructing the prognostic prediction model, p -value < 0.05. (D) The nomogram based on the key genes model, and (E) the calibration curves validated the discrimination.

Generally, the ceRNA regulatory networks include the lncRNA, miRNA, circular RNA and mRNA²⁹. Previous studies have shown that the non-coding RNAs play a vital role in cancer occurrence and prognosis, and it has also been generally recognized that the interactions between lncRNA, miRNA and mRNA could regulate the expression levels of mRNAs and the affect the core protein signals, inducing the changes in physiological functions of cells¹⁴. In this study, we detected the differently expressed miRNAs, lncRNAs and mRNAs in the TCGA lung adenocarcinoma for the first and constructed a ceRNA network from these differently expressed genes.

To identify the key biomarker for prognosis, we performed Cox hazards regression analysis, Lasso regression analysis and survival analysis. Then, *FKBP3*, *GPI*, *LOXL2*, *IL22RA1*, *GPR37* and *has-miR-148a-3p* were identified as the prognostic biomarkers that were significantly associated with overall survival of LUAD patients. In previous studies, these 6 key genes have been confirmed to play an important regulatory role in the progression of tumors. Zhu et al. have provided evidence that *FKBP3* plays an essential role in proliferation and cell cycle progression of NSCLC and might be valued as a prognostic marker in NSCLC³⁰. *IL22RA1* is a receptor of interleukin 22, has been reported working as a signal transducer and activator of STAT3 signaling in the malignant transformation, it plays an important role in promoting tumor growth and metastasis and inhibiting cell apoptosis³¹⁻³³. Besides, many studies have shown that the abnormal expression of *LOXL2* in a variety of cancers is related to epithelial-mesenchymal transition, metastasis, poor prognosis, resistance to radiotherapy and chemotherapy, and tumor progression³⁴. Consistent with our result, a previous study demonstrated that *LOXL2* levels are positively

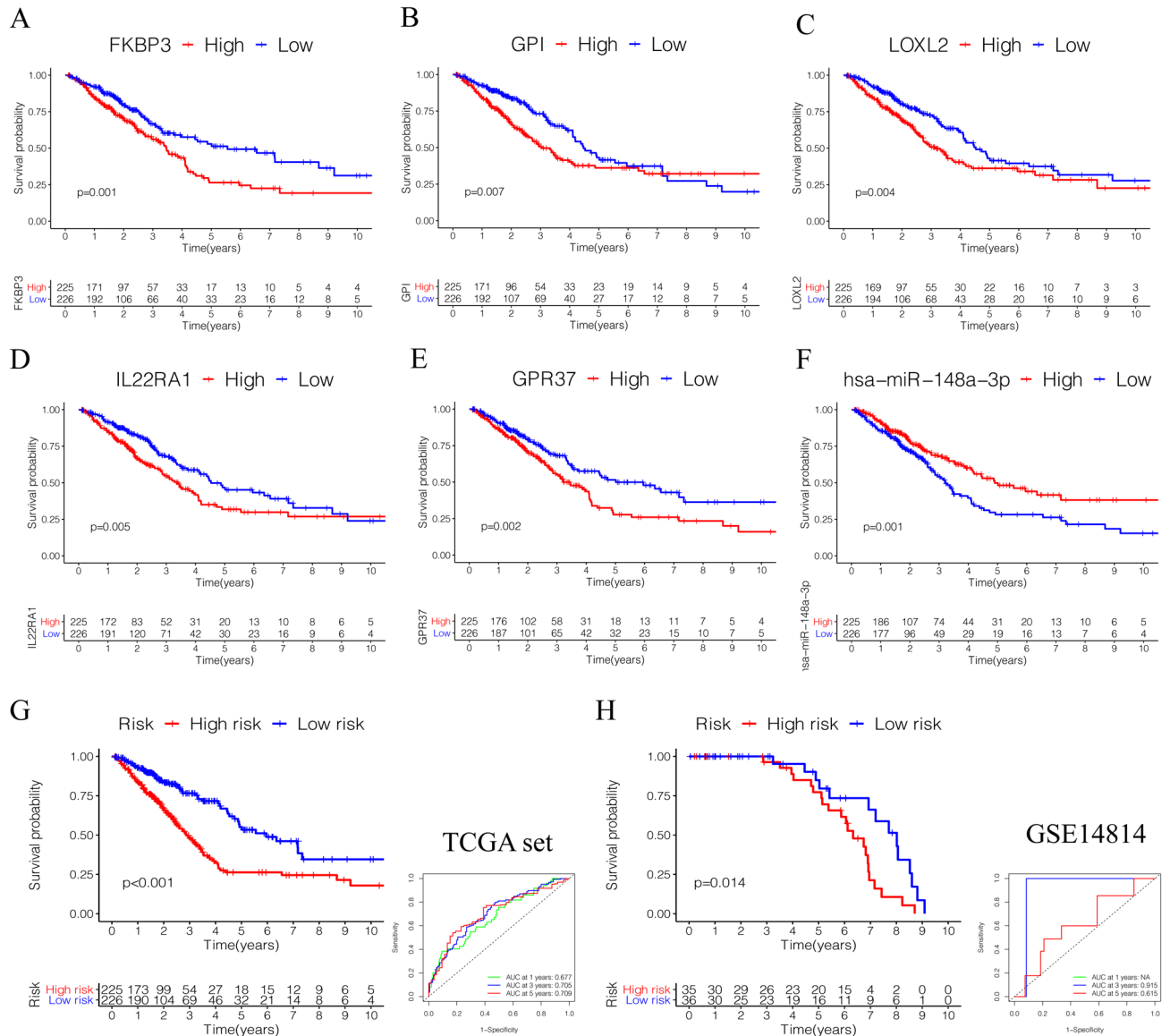


Figure 4. Survival analysis of prognostic genes in ceRNA network. (A–F) The Kaplan–Meier curves validated the 6 key genes (*FKBP3*, *GPI*, *LOXL2*, *IL22RA1*, *GPR37*, *has-miR-148a-3p*) were significantly associated with overall survival, p -value < 0.05 . (G) The Kaplan–Meier overall survival curves for risk score groups. Red curves represent high risk group and blue curves represent low risk group, p -value < 0.05 . (H) The ROC curves showed that the area under curve (AUC) at 1 year was 0.677, at 3 year was 0.705, and at 5 years was 0.709.

associated with poor prognosis in NSCLC patients. In addition, *GPR37* also has been shown as a poor prognostic factor in LUAD patients with TP53 mutation combined EGFR mutation³⁵.

Next, we estimated tumor-infiltrating immune cells through the CIBERSORT algorithm, and used Cox hazards regression analysis and lasso regression analysis to identify prognostic immune cells. Our results show that the Monocytes, Macrophages M1 and Activated mast cells identified were significantly related with survival. Previous studies proved that Monocytes and macrophages play an important role in the immune system, Monocyte activation and then macrophage differentiation is an important step for immune responses³⁶. In the early stages of tumor progression, Monocyte’s recruitment can be found in many types of cancer, and Monocytes induce direct killing of malignant cells through cytokine-mediated cell death and phagocytosis^{37,38}. Macrophages M1 is one major type cell subtype of tumor-associated macrophages (TAMs) which are an important part of

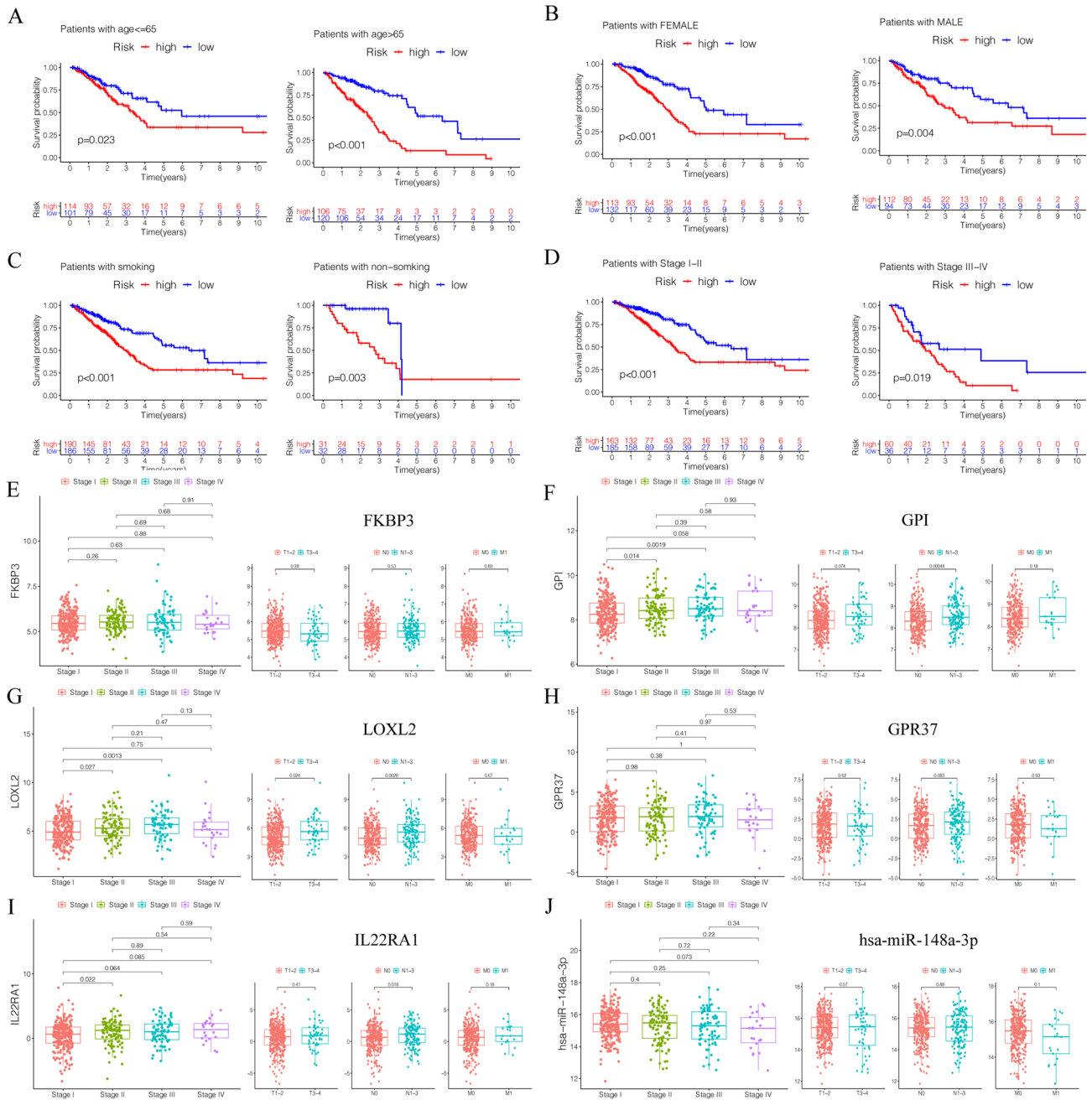


Figure 5. Subgroup analysis classified by patients’ age (A), gender (B), smoking status (C) and clinic stage (D) were conducted. (E–J) The expression of the 6 key genes (*FKBP3*, *GPI*, *LOXL2*, *IL22RA1*, *GPR37*, *has-miR-148a-3p*) in lung adenocarcinoma patients with different tumor stages, T stages, N stages and M stages.

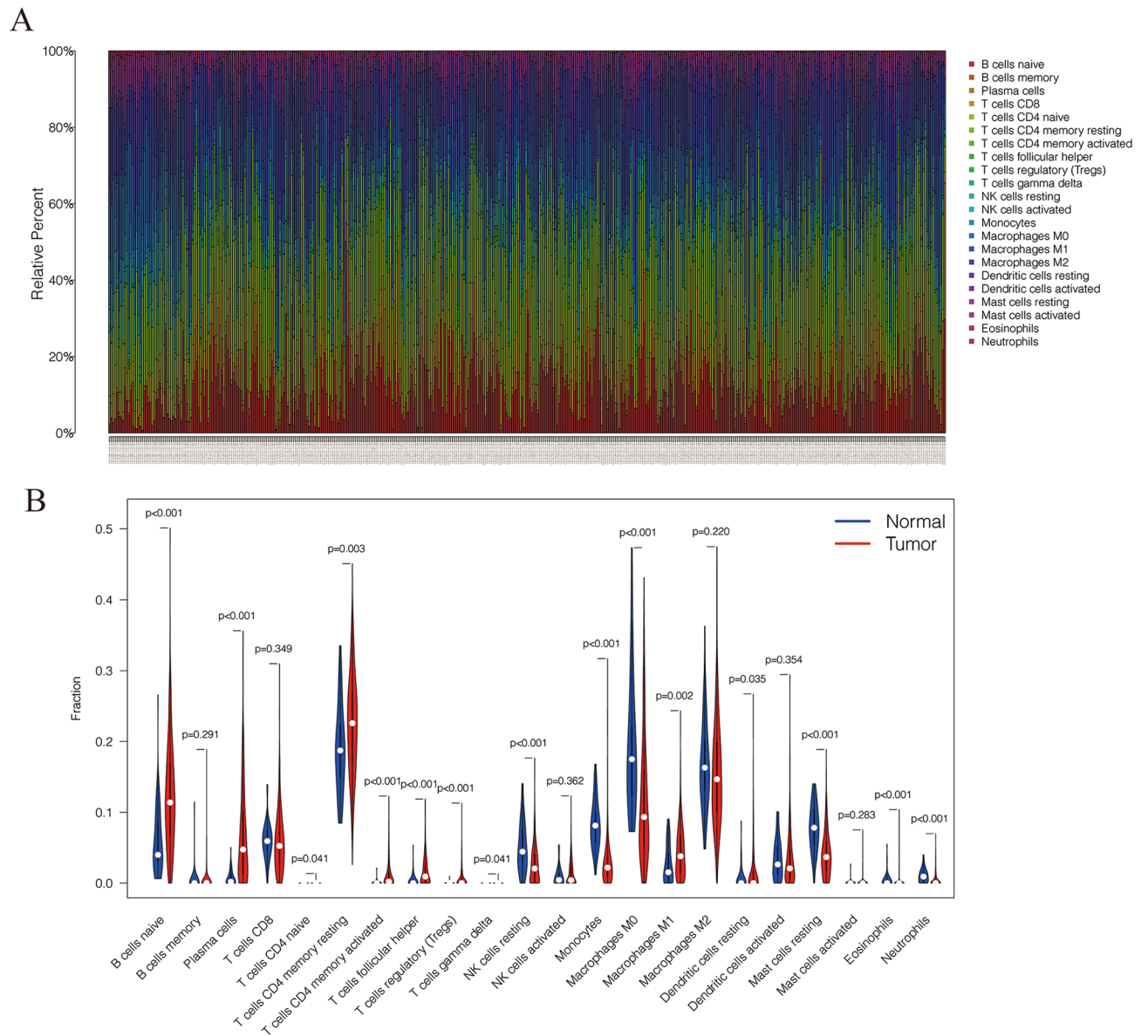


Figure 6. The composition of immune cells in lung adenocarcinoma. **(A)** The component of immune cells in the TCGA-LUAD cases. **(B)** The violin plot shows the levels of immune cells in the tumor group (red) and the normal tissue group (blue).

Immune cell	HR	95%CI	p-value
Monocytes	6.31E-04	(1.02E-06—3.88E-01)	0.0246
Macrophages M1	1.09E+02	(1.67E+00—7.09E+03)	0.0276
Activated mast cells	4.03E+09	(1.02E+02—1.58E+17)	0.0132
B cells naive	4.71E-01	(5.56E-02—3.99E+00)	0.4902
B cells memory	1.13E+02	(1.66E-01—7.70E+04)	0.1555
Plasma cells	4.15E-01	(3.27E-02—5.27E+00)	0.4975
T cells CD8	1.36E-01	(2.51E-03—7.41E+00)	0.3284
Memory resting T cells CD4	3.41E-01	(3.11E-02—3.74E+00)	0.3787
Memory activated T cells CD4	5.61E-01	(4.51E-05—6.96E+03)	0.9042
T cells follicular helper	6.33E+00	(4.08E-03—9.81E+03)	0.6225
T cells regulatory (Tregs)	1.35E+03	(1.31E-01—1.40E+07)	0.1263
Resting NK cells	3.11E+00	(5.66E-03—1.71E+03)	0.7246
Activated NK cells	2.46E+02	(3.26E-02—1.86E+06)	0.2268
Macrophages M0	1.92E+00	(3.23E-01—1.14E+01)	0.4728
Macrophages M2	1.32E+00	(1.71E-01—1.02E+01)	0.7891
Resting dendritic cells	8.08E-02	(4.03E-04—1.62E+01)	0.3523
Activated dendritic cells	1.82E+01	(2.33E-01—1.42E+03)	0.1920
Resting mast cells	6.40E-01	(2.57E-03—1.60E+02)	0.8741
Eosinophils	1.37E+13	(1.53E-07—1.23E+33)	0.1968
Neutrophils	6.16E+01	(6.90E-07—5.50E+09)	0.6591

Table 2. Univariate cox hazards regression analysis of immune cells. *HR* hazards regression, *CI* confidence interval.

the tumor microenvironment, and the infiltration of TAMs is closely related to the poor survival of patients with solid tumors^{39,40}. Cytokines secreted by TAMs induce anti-apoptotic programs in cancer cells, such as *IL-6* derived from TAMs that mediate the resistance of solid tumors to many chemotherapies via activating TGF- β pathway⁴¹⁻⁴³. Although Mast cells participate in the maintenance of healthy lungs through innate immunity and adaptive immunity⁴⁴, there is evidence that mast cell activation leads to the proliferation of non-small cell lung cancer cells⁴⁵.

According to previous study, Kim et. al. proved that the *IL-22* could promote the infiltrating immune cells, including macrophages, and then regulate the inflammatory processes¹⁴. Our correlation analysis revealed that the correlation between *IL22RA1* with macrophage was significantly obvious suggesting that the *IL22RA1* might be an immune-related gene in lung carcinoma. In addition, the correlation analysis revealed that the *GPR37* were also associated with immune cells. Therefore, our results support that the crosstalk between the ceRNA network and tumor-infiltrating immune cells might play a vital role in the formation and progression of lung adenocarcinoma.

In conclusion, through comprehensive analysis, we constructed a ceRNA network and identified *FKBP3*, *GPI*, *LOXL2*, *IL22RA1*, *GPR37* and *has-miR-148a-3p* as the prognostic members of the network, and then identified Monocytes, Macrophages M1 and activated Mast cells as the prognostic tumor-infiltrating immune cells, thus constructing two nomograms that can be used as indicators for evaluating the survival of lung adenocarcinoma patients. Moreover, our results also suggest the crosstalk between the ceRNA network and tumor-infiltrating immune cells to explore the molecular mechanisms involved in lung adenocarcinoma.

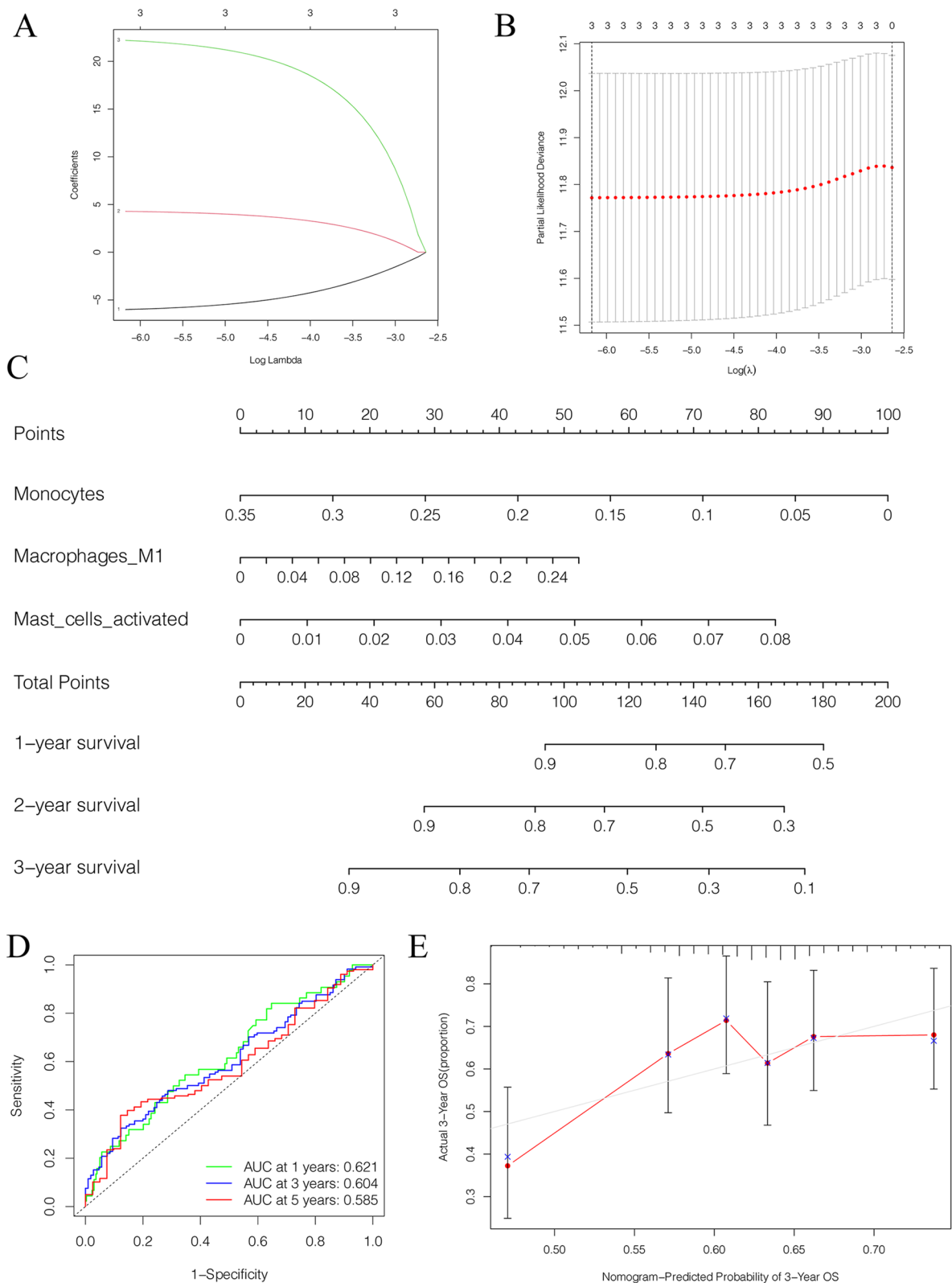


Figure 7. Identification of prognostic value of immune cells in lung adenocarcinoma. (A,B) The Lasso regression suggested that 3 immune cells were essential for prognostic prediction model. (C) The nomogram based on the immune cells model. (D) The ROC curves showed that the area under curve at 1 year was 0.621, at 3 year was 0.604, and at 5 years was 0.585. (E) The calibration curves validated the discrimination.

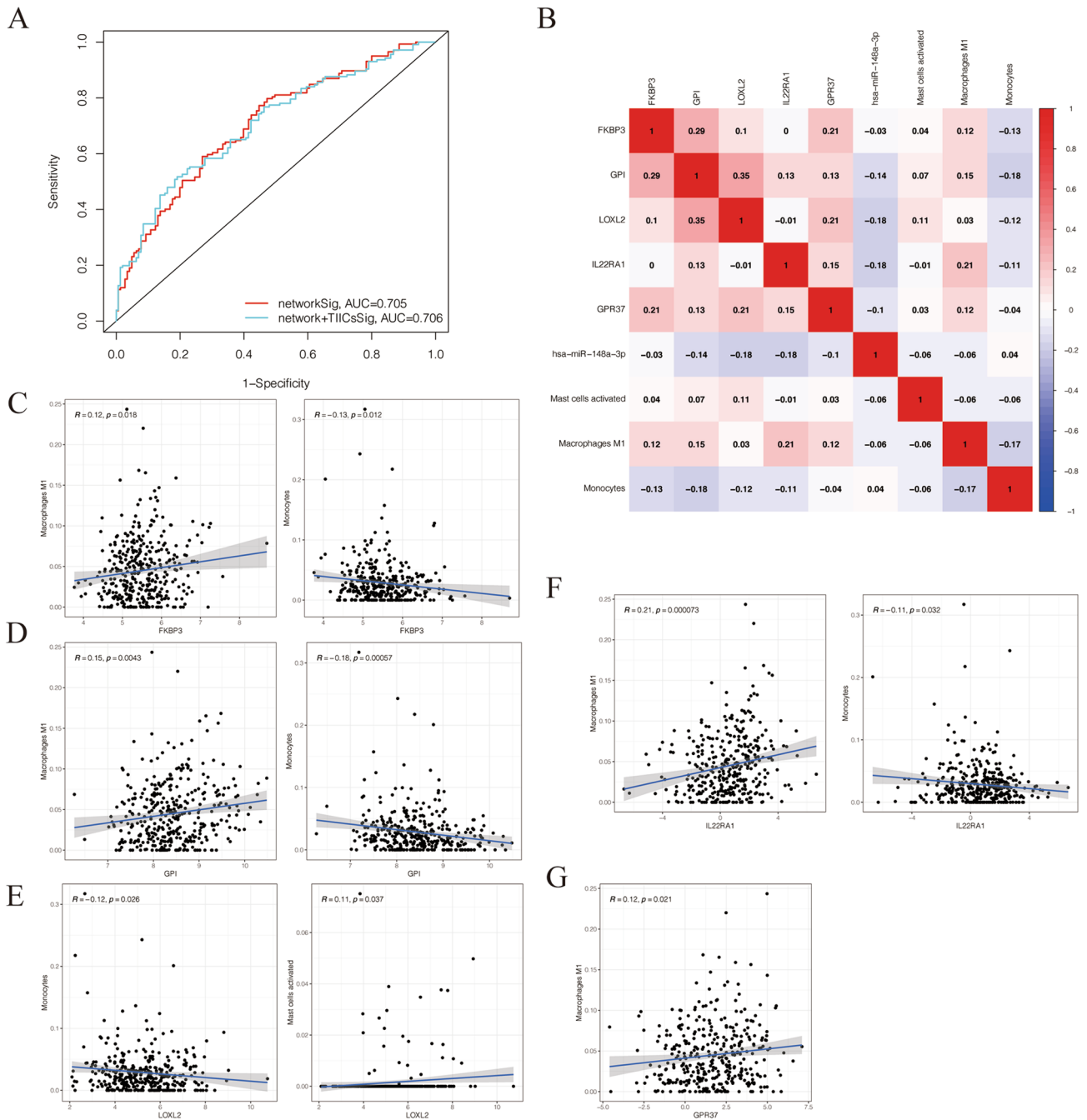


Figure 8. Correlation analysis between the key genes of ceRNA network and the prognostic immune cells. **(A)** The ROC curves of the ceRNA network signatures and ceRNA network combined TIICs signatures. **(B)** The co-expression heatmap of all key genes and the prognostic immune cells. **(C)** The *FKBP3* was associated with Monocytes and Macrophages M1. **(D)** The *GPI* was associated with Monocytes and Macrophages M1. **(E)** The *LOXL2* was negatively associated with Monocytes and Activated mast cells. **(F)** The *IL22RA1* was related with Monocytes and Macrophages M1. **(G)** The *GPR37* was associated with Macrophages M1. R, correlation coefficients. *p*-value < 0.05.

Data availability

All the data used in this study were downloaded from The Cancer Genome Atlas portal (TCGA, <https://portal.gdc.cancer.gov/>).

Received: 16 January 2021; Accepted: 13 May 2021

Published online: 26 May 2021

References

1. Siegel, R. L., Miller, K. D. & Jemal, A. Cancer statistics, 2020. *CA Cancer J. Clin.* **70**, 7–30. <https://doi.org/10.3322/caac.21590> (2020).
2. Herbst, R. S., Morgensztern, D. & Boshoff, C. The biology and management of non-small cell lung cancer. *Nature* **553**, 446–454. <https://doi.org/10.1038/nature25183> (2018).
3. Tanoue, L. T., Tanner, N. T., Gould, M. K. & Silvestri, G. A. Lung cancer screening. *Am. J. Respir. Crit. Care Med.* **191**, 19–33. <https://doi.org/10.1164/rccm.201410-1777CI> (2015).
4. Duma, N., Santana-Davila, R. & Molina, J. R. Non-small cell lung cancer: Epidemiology, screening, diagnosis, and treatment. *Mayo Clin. Proc.* **94**, 1623–1640. <https://doi.org/10.1016/j.mayocp.2019.01.013> (2019).
5. Hirsch, F. R. *et al.* Lung cancer: Current therapies and new targeted treatments. *Lancet* **389**, 299–311. [https://doi.org/10.1016/S0140-6736\(16\)30958-8](https://doi.org/10.1016/S0140-6736(16)30958-8) (2017).
6. Guo, L. L. *et al.* Competing endogenous RNA networks and gastric cancer. *World J. Gastroenterol.* **21**, 11680–11687. <https://doi.org/10.3748/wjg.v21.i41.11680> (2015).
7. Tay, Y., Rinn, J. & Pandolfi, P. P. The multilayered complexity of ceRNA crosstalk and competition. *Nature* **505**, 344–352. <https://doi.org/10.1038/nature12986> (2014).
8. Karreth, F. A. & Pandolfi, P. P. ceRNA cross-talk in cancer: when ce-bling rivalries go awry. *Cancer Discov.* **3**, 1113–1121. <https://doi.org/10.1158/2159-8290.CD-13-0202> (2013).
9. Weng, W. *et al.* Identification of a competing endogenous RNA network associated with prognosis of pancreatic adenocarcinoma. *Cancer Cell Int.* **20**, 231. <https://doi.org/10.1186/s12935-020-01243-6> (2020).
10. Emilsson, V. *et al.* Genetics of gene expression and its effect on disease. *Nature* **452**, 423–428. <https://doi.org/10.1038/nature06758> (2008).
11. Singh, R., Mishra, M. K. & Aggarwal, H. Inflammation, immunity, and cancer. *Mediators Inflamm.* **2017**, 6027305. <https://doi.org/10.1155/2017/6027305> (2017).
12. Grivennikov, S. I., Greten, F. R. & Karin, M. Immunity, inflammation, and cancer. *Cell* **140**, 883–899. <https://doi.org/10.1016/j.cell.2010.01.025> (2010).
13. Kardoust Parizi, M., Shariat, S. F., Margulis, V., Mori, K. & Lotan, Y. Value of tumour-infiltrating immune cells in predicting response to intravesical BCG in patients with non-muscle-invasive bladder cancer: A systematic review and meta-analysis. *BJU Int.* <https://doi.org/10.1111/bju.15276> (2020).
14. Raziq, K. *et al.* Competitive endogenous network of lncRNA, miRNA, and mRNA in the chemoresistance of gastrointestinal tract adenocarcinomas. *Biomed. Pharmacother.* **130**, 110570. <https://doi.org/10.1016/j.biopha.2020.110570> (2020).
15. Zhu, C. Q. *et al.* Prognostic and predictive gene signature for adjuvant chemotherapy in resected non-small-cell lung cancer. *J. Clin. Oncol.* **28**, 4417–4424. <https://doi.org/10.1200/JCO.2009.26.4325> (2010).
16. Anders, S. & Huber, W. Differential expression analysis for sequence count data. *Genome Biol.* **11**, R106. <https://doi.org/10.1186/gb-2010-11-10-r106> (2010).
17. Li, J. H., Liu, S., Zhou, H., Qu, L. H. & Yang, J. H. starBase v2.0: decoding miRNA-ceRNA, miRNA-ncRNA and protein-RNA interaction networks from large-scale CLIP-Seq data. *Nucleic Acids Res.* **42**, D92–97. <https://doi.org/10.1093/nar/gkt1248> (2014).
18. Li, R. *et al.* GDCRNATools: An R/Bioconductor package for integrative analysis of lncRNA, miRNA and mRNA data in GDC. *Bioinformatics* **34**, 2515–2517. <https://doi.org/10.1093/bioinformatics/bty124> (2018).
19. Shannon, P. *et al.* Cytoscape: A software environment for integrated models of biomolecular interaction networks. *Genome Res.* **13**, 2498–2504. <https://doi.org/10.1101/gr.1239303> (2003).
20. Yu, G., Wang, L. G., Han, Y. & He, Q. Y. clusterProfiler: An R package for comparing biological themes among gene clusters. *OMICS* **16**, 284–287. <https://doi.org/10.1089/omi.2011.0118> (2012).
21. Colombani, C. *et al.* Application of Bayesian least absolute shrinkage and selection operator (LASSO) and BayesCpi methods for genomic selection in French Holstein and Montbeliarde breeds. *J. Dairy Sci.* **96**, 575–591. <https://doi.org/10.3168/jds.2011-5225> (2013).
22. Cox, D. R. Regression models and life-tables. *J. Roy. Stat. Soc. Ser. B (Methodol.)* **34**, 187–202 (1972).
23. Newman, A. M. *et al.* Robust enumeration of cell subsets from tissue expression profiles. *Nat Methods* **12**, 453–457. <https://doi.org/10.1038/nmeth.3337> (2015).
24. Sata, Y. *et al.* Keys to successful induction chemoradiotherapy followed by surgery for stage III/N2 non-small cell lung cancer. *Surg. Today* **49**, 547–555. <https://doi.org/10.1007/s00595-019-1766-8> (2019).
25. Zancanato, F., Cordenonsi, M. & Piccolo, S. YAP and TAZ: A signalling hub of the tumour microenvironment. *Nat. Rev. Cancer* **19**, 454–464. <https://doi.org/10.1038/s41568-019-0168-y> (2019).
26. Tsou, P., Katayama, H., Ostrin, E. J. & Hanash, S. M. The emerging role of B cells in tumor immunity. *Cancer Res.* **76**, 5597–5601. <https://doi.org/10.1158/0008-5472.CAN-16-0431> (2016).
27. Takeuchi, Y. & Nishikawa, H. Roles of regulatory T cells in cancer immunity. *Int. Immunol.* **28**, 401–409. <https://doi.org/10.1093/intimm/dxw025> (2016).
28. Jochems, C. & Schlom, J. Tumor-infiltrating immune cells and prognosis: The potential link between conventional cancer therapy and immunity. *Exp. Biol. Med. (Maywood)* **236**, 567–579. <https://doi.org/10.1258/ebm.2011.011007> (2011).
29. Polisenio, L. *et al.* A coding-independent function of gene and pseudogene mRNAs regulates tumour biology. *Nature* **465**, 1033–1038. <https://doi.org/10.1038/nature09144> (2010).
30. Zhu, W. *et al.* FKBP3 promotes proliferation of non-small cell lung cancer cells through regulating Sp1/HDAC2/p27. *Theranostics* **7**, 3078–3089. <https://doi.org/10.7150/thno.18067> (2017).
31. Cui, X. *et al.* IL22 furthers malignant transformation of rat mesenchymal stem cells, possibly in association with IL22RA1/STAT3 signaling. *Oncol. Rep.* **41**, 2148–2158. <https://doi.org/10.3892/or.2019.7007> (2019).
32. He, W. *et al.* IL22RA1/STAT3 signaling promotes stemness and tumorigenicity in pancreatic cancer. *Cancer Res.* **78**, 3293–3305. <https://doi.org/10.1158/0008-5472.CAN-17-3131> (2018).
33. Trejejo-Nunez, G., Elsegeiny, W., Conboy, P., Chen, K. & Kolls, J. K. Critical role of IL-22/IL22-RA1 signaling in pneumococcal pneumonia. *J. Immunol.* **197**, 1877–1883. <https://doi.org/10.4049/jimmunol.1600528> (2016).
34. Wen, B., Xu, L. Y. & Li, E. M. LOXL2 in cancer: Regulation, downstream effectors and novel roles. *Biochim. Biophys. Acta Rev. Cancer* **1874**, 188435. <https://doi.org/10.1016/j.bbcan.2020.188435> (2020).
35. Wang, F. *et al.* Prognostic value of TP53 co-mutation status combined with EGFR mutation in patients with lung adenocarcinoma. *J. Cancer Res. Clin. Oncol.* **146**, 2851–2859. <https://doi.org/10.1007/s00432-020-03340-5> (2020).

36. Ivanova, E. A. & Orekhov, A. N. Monocyte activation in immunopathology: Cellular test for development of diagnostics and therapy. *J. Immunol. Res.* **2016**, 4789279. <https://doi.org/10.1155/2016/4789279> (2016).
37. Olingy, C. E., Dinh, H. Q. & Hedrick, C. C. Monocyte heterogeneity and functions in cancer. *J. Leukoc. Biol.* **106**, 309–322. <https://doi.org/10.1002/JLB.4RI0818-311R> (2019).
38. Woods, J. A. & Davis, J. M. Exercise, monocyte/macrophage function, and cancer. *Med. Sci. Sports Exerc.* **26**, 147–156. <https://doi.org/10.1249/00005768-199402000-00004> (1994).
39. Singhal, S. *et al.* Human tumor-associated monocytes/macrophages and their regulation of T cell responses in early-stage lung cancer. *Sci. Transl. Med.* <https://doi.org/10.1126/scitranslmed.aat1500> (2019).
40. Yang, L. & Zhang, Y. Tumor-associated macrophages: From basic research to clinical application. *J. Hematol. Oncol.* **10**, 58. <https://doi.org/10.1186/s13045-017-0430-2> (2017).
41. Kong, L. *et al.* Deletion of interleukin-6 in monocytes/macrophages suppresses the initiation of hepatocellular carcinoma in mice. *J. Exp. Clin. Cancer Res.* **35**, 131. <https://doi.org/10.1186/s13046-016-0412-1> (2016).
42. Xu, X., Ye, J., Huang, C., Yan, Y. & Li, J. M2 macrophage-derived IL6 mediates resistance of breast cancer cells to hedgehog inhibition. *Toxicol. Appl. Pharmacol.* **364**, 77–82. <https://doi.org/10.1016/j.taap.2018.12.013> (2019).
43. Yin, Y. *et al.* The immune-microenvironment confers chemoresistance of colorectal cancer through macrophage-derived IL6. *Clin. Cancer Res.* **23**, 7375–7387. <https://doi.org/10.1158/1078-0432.CCR-17-1283> (2017).
44. Virk, H., Arthur, G. & Bradding, P. Mast cells and their activation in lung disease. *Transl. Res.* **174**, 60–76. <https://doi.org/10.1016/j.trsl.2016.01.005> (2016).
45. Stoyanov, E., Uddin, M., Mankuta, D., Dubinett, S. M. & Levi-Schaffer, F. Mast cells and histamine enhance the proliferation of non-small cell lung cancer cells. *Lung Cancer* **75**, 38–44. <https://doi.org/10.1016/j.lungcan.2011.05.029> (2012).

Author contributions

W.J.-C. and Q.H.-Z. conceived and designed the study. W.J.-C. and W.-L. wrote the manuscript. X.G.-L. and W.-Z. collected the data. W.J.-C. and Z.K.-L. analyzed the data. G.Z.-M. and Y.F.-D. reviewed the draft. W.J.-C. prepared Figures and Tables. All authors have read and agreed to the published version of the manuscript.

Funding

This work was funded by the National Natural Science Foundation of China (Grant No. 81572288), the National Key Research and Development Plan of China (Grant No. 2016YEE0103400).

Competing interests

The authors declare no competing interests.

Additional information

Supplementary Information The online version contains supplementary material available at <https://doi.org/10.1038/s41598-021-90755-w>.

Correspondence and requests for materials should be addressed to Q.z.

Reprints and permissions information is available at www.nature.com/reprints.

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.



Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.

© The Author(s) 2021