The Expression of Angiogenesis Genes is Related to Immune Microenvironment and Prognosis of Lung Adenocarcinoma

Xue Wang  
Shandong University of Traditional Chinese Medicine

Xijian Liu  
Shandong University of Traditional Chinese Medicine

Lu Wang  
Shandong University of Traditional Chinese Medicine

Jiuwei Li  
Shandong University of Traditional Chinese Medicine

Ling Li  
Shandong University of Traditional Chinese Medicine

Yaxing Li  
Shandong University of Traditional Chinese Medicine

Hailiang Huang  
Shandong University of Traditional Chinese Medicine

Tao Han (✉ 60012002@sdutcm.edu.cn)  
Shandong University of Traditional Chinese Medicine

Research Article

Keywords: Lung adenocarcinoma, Angiogenesis, Immune microenvironment, Risk model, Survival prognosis

Posted Date: February 14th, 2023

DOI: https://doi.org/10.21203/rs.3.rs-2568517/v1

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

Objective: Lung adenocarcinoma (LC), the main type of non-small cell lung cancer, has a 5-year survival rate of only 14.6%. Tumor angiogenesis is the primary factor leading to the progression of LC. This study aimed to discuss the role of angiogenesis-related genes (ARGs) in the development and diagnosis of LC.

Methods: Clinical and transcriptomic data of LC patients were downloaded from TCGA and GEO databases and divided into training cohorts and validation cohorts. Combined with the ARGs of the Molecular Signatures Database, cluster analysis was performed to identify new cluster subgroups. Enrichment analyses were performed to elucidate the underlying mechanisms of subpopulation differences. MCPCounter, CIBERSORT and xCell analysis was used to determine the tumor immune microenvironment (TIM) and the immune status of identified subgroups. Lasso algorithm and multivariate Cox regression analysis were used to construct the prognostic risk model, and combined with the clinical information of patients with LC to verify the effectiveness of the risk model.

Results: We identified 2 cluster subgroups that could significantly predict differential survival based on LC survival prognostic genes and ARGs. Among them, cluster 2 showed a better prognosis and was associated with a high immune score, a high abundance of immune infiltrating cells, and a relatively high immune status. Enrichment analysis revealed that DEGs between the two subgroups were mainly enriched in angiogenesis and immune-related pathways. Combined with clinical features, higher risk scores were positively associated with LC worsening of disease progression, predicting poor survival. The validation cohort GSE68465 corroborates the validity of the risk model.

Conclusion: The abnormal expression of ARGs is closely related to the TIM of LC patients. The ARG risk model we constructed can be used to accurately predict the survival prognosis of LC.

Introduction

Lung adenocarcinoma (LC), as the main type of non-small cell lung cancer, has the pathological characteristics of relatively hidden onset and no special symptoms in the early stage. This makes some patients have missed the opportunity of operation when they are diagnosed, and the overall 5-year survival rate is only 14.6% (Nicholson et al., 2022). Although researchers have found that a variety of molecular biomarkers can assist in the prognosis of LC, their limitations and deficiencies restrict the progress of LC treatment (Calvayrac et al., 2017). Therefore, the construction of a new and effective prediction model will help to judge the prognosis and individualized treatment of patients with LC.

Tumor angiogenesis is the primary pathological factor in the process of tumor cell growth and metastasis, which is positively correlated with the degree of malignancy (Legg et al., 2008). Many studies have shown that tumor angiogenesis plays an important role in the process of tumor occurrence and development, which also promotes the research of antiangiogenesis (Lai et al., 2021). In 2004, antiangiogenic therapy was the first to break the dilemma of only chemotherapy for patients with advanced non-small cell lung cancer, making the median overall survival exceed 1 year for the first time.
After in-depth exploration of tumor angiogenesis, it is found that VEGF is the most important angiogenesis inducing factor in the tumor microenvironment, which combines with VEGF receptor on vascular endothelial cells to promote neoangiogenesis. Therefore, bevacizumab, which can inhibit VEGF and bind to its receptor, has been developed as a first-line treatment (Claesson-Welsh and Welsh, 2013). The clinical effectiveness also urges us to further explore the angiogenesis-related genes (ARGs) of LC.

Interestingly, we found that cancer angiogenesis is closely related to tumor immune microenvironment (TIM). TIM is a dynamic relationship network, which is composed of many types of cells (such as tumor cells, T cells, macrophages, etc.) and extracellular components (such as cytokines, growth factors, extracellular matrix, etc.). Studies have confirmed that the specificity of the TIM, hypoxia state, angiogenic factors, inflammatory factors and the state of immune cells can promote tumor angiogenesis (Jiang et al., 2020, Maeda-Otsuka et al., 2019). For example, in the hypoxic microenvironment of bladder cancer, endothelial cells promote angiogenesis and allow cells to grow to adapt to the complex TIM (Li et al., 2020). In patients with gastric cancer, angiogenesis is associated with weakened immune environment and worse survival (Oshi et al., 2021). Angiogenesis in pancreatic cancer may be related to the changes of immunosuppressive components induced by the interaction between epithelial cancer cells and stromal cells in the TIM (Ren et al., 2018). It can be seen that immune cells play a key role in cell reprogramming, which changes the original human immune homeostasis and gives tumor cells the ability to survive and develop (Mao et al., 2021). Therefore, TIM plays an important role in the process of disease. With the deepening of research, TIM is closely related to the pathogenesis of LC (Liu et al., 2021, Huang et al., 2021, Yang et al., 2022). Evaluating the immune microenvironment status of ARGs in LC can not only understand the internal mechanism of disease occurrence and development, but also promote the progress of immunotherapy and improve the prognosis of patients with LC.

Therefore, we comprehensively analyzed the ARGs related to the prognosis of LC, and constructed a risk model to explore the impact of angiogenesis on the immune status and survival prognosis of patients with LC. This work is expected to provide reference for cell angiogenesis and immune microenvironment of LC, and provide a new theoretical basis for antiangiogenesis and immunotherapy of patients with LC.

**Methods**

2.1 Data collection

The sequencing RNA data and clinical data related to LC in this study are from TCGA database and GEO database. Among them, LC samples with complete clinical information were included from TCGA as the training cohort, and 442 samples of GSE68465 were selected from GEO as the validation cohort for further analysis. The clinical characteristics of the two cohorts are shown in Table 1. 146 ARGs came from Molecular Signatures Database.

2.2 Enrichment analysis
Difference analysis was performed using "limma" to obtain differential genes between the 2 cluster groups. Get the latest KEGG Pathway gene annotations through the KEGG rest API, map genes to the background set. Enrichment analysis was performed using "clusterProfiler" (Wu et al., 2021) to obtain the results of gene set enrichment, combined with Metascape visualization. Additionally, GSEA (Subramanian et al., 2005) was used to identify changes in related pathways and molecular mechanisms between the 2 groups.

2.3 Clustering subpopulations and immune assessment

Combined with the prognostic survival genes of LC and Cox analysis, 25 genes related to angiogenesis and survival prognosis of LC were found. According to the transcriptional data of 25 ARGs, "ConsensusClusterPlus" was used for cluster analysis of LC samples (Wilkerson and Hayes, 2010). Immune infiltrating cells in malignant tumor tissues were evaluated using MCPCounter (Lu et al., 2022), CIBERSORT (Le et al., 2021), xCell (Aran et al., 2017), and immune score, stromal score and microenvironment score were calculated.

2.4 Construction of risk model

"Glmnet" was used to integrate survival time, survival status and gene expression data, and Lasso-Cox method was used for regression analysis. In addition, we also set up 10 fold cross validation to obtain the optimal model. We set the lambda value to 0.0118635943287256, and finally obtained 7 genes (ANGPTL4, SPHK1, HTATIP2, FGFBP1, NCL, PDGFB, COL4A2). The calculation formula of risk score is:

\[
\text{RiskScore} = 0.13674169804817 \times \text{ANGPTL4} + 0.11429283469428 \times \text{SPHK1} + 0.374982089648168 \times \text{HTATIP2} + 0.0570109248934948 \times \text{FGFBP1} + 0.345918302843882 \times \text{NCL} + 0.0554868166738347 \times \text{PDGFB} + 0.0620930183188786 \times \text{COL4A2}.
\]

2.5 Validation of risk model

We divided patients into high-risk groups and low-risk groups according to risk scores, and analyzed the clinical data of LC based on gender, primary tumor(T), lymph nodes(N), distant metastasis(M), disease stage, and survival prognosis. The validation queue GSE68465 was used to verify the effectiveness of the risk model again.

Results

3.1 Identification of class 2 cluster subtypes based on angiogenesis

Combined with the transcriptome analysis of LC and clinical data, we obtained 6822 genes related to survival and prognosis, and then performed venny analysis with angiogenesis genes to obtain 25 core ARGs. Next, the training cohort was clustered according to the 25 ARGs. When k = 2, the best intra group consistency is displayed (Fig. 1A-C). Moreover, we found that ARGs showed significant differential expression in two subtypes through heat map (Fig. 1D). In addition, through the survival analysis of the two subtypes, it was found that the prognostic survival rate of cluster C2 was higher than that of cluster
C1 \((P = 1.2e-5, \text{Fig. 1E})\). The above cluster analysis showed that ARGs effectively divided patients with LC into two subtypes, and there were differences in gene expression and prognostic survival rate.

3.2 Differential expression and enrichment analysis of 2 clusters

In order to explore the expression differences of the two clusters in patients with LC, we further analyzed the differentially expressed genes (DEGs) between the two clusters and enriched them to explore their biological processes and signal mechanisms. According to the standard of "\(P < 0.05\) and fold change > 1.5", 533 DEGs between the two clusters were screened, of which 265 were up-regulated and 268 were down-regulated (Fig. 2A). Go analysis showed that DEGs were enriched in biological processes related to angiogenesis, including positive regulation of cytokine production, vascular development, tissue morphogenesis, PID integrin1 pathway, etc (Fig. 2B). The protein-protein interaction network identified multiple clustering models, all involving lung cancer development and immune microenvironment, including constructive signaling by aberrant PI3K in cancer, extracellular matrix organization, positive regulation of epithelial growth factor receptor signaling pathway, mitotic cell cycle process, etc. (Fig. 2C, E). In addition, KEGG pathway analysis also showed that DEGs were involved in the signal pathways related to lung cancer angiogenesis and immunity, such as focal adhesion, ECM receptor interaction, PI3K Akt signal pathway, p53 signal pathway, IL-17 signaling pathway, etc (Fig. 2D, F). In addition, GSEA analyzed the differential signaling pathways between the two cluster subtypes. The results showed that DNA replication, cell cycle and p53 signaling pathways were highly expressed in group C1 (Fig. 2H). The above enrichment analysis shows that ARGs may be related to the formation of TIM and participate in the process of LC.

3.3 Different immune infiltration of 2 cluster subtypes

Since the 2 cluster subtypes were differentially enriched in terms of immunity, we further performed immunological analysis of the 2 cluster subtypes. First, according to the xCell algorithm, the immune, stroma and microenvironment score of cluster 2 with better survival prognosis of LC are higher than those of cluster 1 (Fig. 3A). Secondly, we analyzed immune cells by combining MCPCounter (Fig. 3B) and CIBERSORT methods (Fig. 3C). It was found that the aggregation of cluster 2 in T cells \((P = 4.70e-03)\), B lineage \((P = 1.20e-03)\), myoid dendritic cells \((P = 3.40e-06)\), neutrophils \((P = 1.70e-07)\), endothelial cells \((P = 9.10e-10)\), CD4 T cells \((P = 2.10e-04)\), monocytes \((P = 1.30e-06)\), dendritic cells \((P = 4.20e-05)\) and mass cells \((P = 1.60e-10)\) was significantly higher than that of cluster 1. This suggests that our clustering grouping by ARGs has significant differences in immune status between cluster 1 and cluster 2.

3.4 Construction of angiogenesis-related gene risk model

By clustering ARGs on LC, we found that there were significant differences between the two molecular subtypes, which led us to further explore. Combined with Lasso-Cox analysis to screen risk genes, 7 genes (ANGPTL4, SPHK1, HTATIP2, FGFBP1, NCL, PDGFB, COL4A2) were obtained to construct a risk model (Fig. 4A, B). This risk model can effectively divide LC patients into high-risk and low-risk groups based on risk scores. And according to the heat map, 7 genes were significantly highly expressed in the
high-risk group. In addition, the survival prognosis analysis showed that the overall survival rate of patients in the low-risk group was higher than that in the high-risk group (Fig. 4C). The ROC analysis also showed that the auc was 0.71 in 1 year, 0.73 in 3 years, and 0.69 in 5 years, indicating that the risk model has the ability to accurately predict (Fig. 4D). Combined with xCell, the cellular immune infiltration between the two groups was evaluated. We found that immune, stroma and microenvironment score were significantly higher in the low-risk group compared to the high-risk group (Fig. 4E). The above Lasso-Cox and risk immune analysis results indicated that the construction of a risk model of ARGs has the potential to predict the prognosis of patients with LC, and is significantly related to the immune microenvironment of patients with LC.

3.5 Clinical relevance of constructing risk models

We further analyzed the association between the constructed risk model and the clinical characteristics of LC. The clinical features involved the five most important factors of LC, including TNM, disease stage and survival prognosis. The results showed that the risk model scores were all significantly associated with the five clinical key factors of LC. Among them, in different genders, the risk score of male was significantly higher than that of female (Fig. 5A), and the risk score increased with the increase of TNM and disease stage (Fig. 5B-E). When we delved deeper into these 5 clinically critical factors in combination with survival prognosis, we found that prognosis also showed the same result, higher risk scores had worse prognosis (Fig. 5F-O). Among them, the sample size of the M1 phase of the distant metastasis clinical data is too small, resulting in meaningless results. This fully proves that our ARGs risk model has the predictive performance that is closely related to LC.

3.6 Validation cohort to test the risk model

In order to verify the validity of the constructed risk model, we verified the prognosis of 7 risk genes through the LC microarray gene chip GSE68465. As in the training cohort, patients with higher risk scores had poorer outcomes (Fig. 5P). We also conducted a validation analysis on the TIM that was closely related to the risk model score of the previous analysis. The results showed that the immune score, stroma score, and microenvironment score of the low-risk group were significantly higher than those of the high-risk group (Fig. 5Q), which were also consistent with the results of the training group. Through the analysis of the validation cohort, it was effectively verified that the ARG risk model we constructed was related to the survival prognosis and TIM of LC.

**Discussion**

The mortality rate of LC has been high, and one of the important reasons for preventing effective clinical treatment is that the TIM leads to local immunosuppression of the tumor, which leads to the further growth, infiltration and metastasis of tumor cells (Bonanno et al., 2019). An essential factor in the development of tumor cells is tumor angiogenesis, which has been recognized as a marker of solid tumor growth, invasion and metastasis (Parmar and Apte, 2021). More and more studies have shown that tumor angiogenesis is closely related to the TIM (Jang et al., 2022, Albini et al., 2018, Hu et al., 2022).
Therefore, we analyzed the ARGs and TIM of the intractable LC, and identified 2 cluster subtypes, which exhibit different angiogenesis and can effectively predict the survival prognosis of LC patients. Survival prognosis combined with immune analysis showed that cluster 2 with better prognosis in LC also had higher immune status and related immune, stromal and microenvironment scores. In addition, the enrichment analysis also showed that the DEGs between the two clusters were involved in pathways closely related to cellular angiogenesis, immunity and cancer. Based on this, we established a LC prognostic risk model of ARGs. In order to provide reference for the clinical treatment of patients with LC.

The risk model of ARGs we constructed included 7 risk genes, ANGPTL4, SPHK1, HTATIP2, FGFBP1, NCL, PDGFB, COL4A2. Angiopoietin like 4 (ANGPTL4) is a secreted protein that can promote the survival and development of vascular endothelial cells. Studies have shown that ANGPTL4 is abnormally expressed in a variety of cancers, promoting angiogenesis, leading to tumor growth and metastasis (Yi et al., 2013, Chen et al., 2020, Liao et al., 2017). Among them, in the process of breast cancer lung metastasis, ANGPTL4 can destroy the connection between vascular endothelial cells, increase the permeability of pulmonary capillaries, accelerate the angiogenesis of breast cancer cells, and lead to a high lung metastasis rate (Padua et al., 2008). Sphingosine kinase 1 (SPHK1), as another risk gene in the risk prognosis model, is a sphingosine kinase, which is involved in the angiogenesis, microenvironment and metastasis of lung cancer (Shen et al., 2021), gastric cancer (Yin et al., 2019), breast cancer (Nagahashi et al., 2018). It can promote tumor angiogenesis, so that new capillaries can deliver necessary energy supply materials to tumor tissues, and promote tumor metastasis (Wang et al., 2020). Previous studies have shown that the abnormally high expression of SPHK1 in non-small cell lung cancer activates signal translator and activator of transcription 3, thereby increasing the proliferation and migration of non-small cell lung cancer cells.

What is worth discussing is the risk gene HIV-1 Tat interactive protein 2 (HTATIP2), also known as TAT interacting protein 30 (TIP30). We have found that it has both anti-cancer and anti-cancer effects through existing research reports. For example, Li et al. Detected and compared gastric cancer samples with normal gastric tissues and found that the expression of HTATIP2 decreased significantly in gastric cancer tissues (Li et al., 2009). In contrast, the findings of Zhang et al. in prostate cancer are just the opposite. They analyzed HTATIP2 by immunohistochemistry and found that its overexpression can promote the metastasis of prostate cancer cells, resulting in poor prognosis (Zhang et al., 2008). Combined with our analysis in the risk model, with the increase of the risk factor and the number of deaths in LC patients, the expression of HTATIP2 gene also gradually increased, suggesting that HTATIP2 may, like p53, not only have a tumor suppressor effect, but also very likely exist Potential cancer risk (Kim et al., 2019).

Fibroblast growth factor-binding protein 1 (FGFBP1), as an initiating factor of angiogenesis, can bind to fibroblast growth factor 1 and participate in the process of cell growth together (Tomaszewski et al., 2011). Nucleolin (NCL), an eukaryotic nucleolar phosphoprotein, is aberrantly expressed in a variety of cancers and has received extensive attention (Miranda et al., 2021, Thongchot et al., 2022, Lee et al., 2021). High expression of NCL can not only be used as an indicator of enhanced mitotic activity, but also has anti-apoptotic effects, increasing the proliferation and invasion of tumor cells (Bywater et al., 2013). More importantly, it is up-regulated in angiogenic endothelial cells, promoting tumor angiogenesis (Tuteja
Platelet derived growth factor subunit B (PDGFB) is a protein composed of platelet-derived growth factor and vascular endothelial growth factor. It is an important component of cell development and is closely related to the formation of tumor microenvironment. The latest study explores its role in pancreatic cancer by constructing platelet PDGFB knockout mice (Zhang et al., 2022). The results show that the deletion of PDGFB has a good effect on the tumor microenvironment, can significantly reduce the content of tumor-associated fibroblasts and extracellular matrix, and reduce the further deterioration of the tumor. The high expression of PDGFB and signal transducer and activator of transcription 3 will increase the microvascular proliferation of glioma, and the expression of CD31, a detection marker of endothelial cell proliferation, will also increase significantly (Doucette et al., 2012). Collagen type IV alpha 2 chain (COL4A2), one of the encoded proteins of type IV collagen, can be related to the physiological process of the lung. Specifically, COL4A2 can regulate the growth and development of alveoli, and its abnormal expression will affect the balance of lung epithelial progenitors and differentiated cells, and fibroblasts will also be affected, which will lead to the integrity of alveolar growth and abnormal differentiation (Loscertales et al., 2016).

The seven genes in the risk model were discussed in depth above, and it was found that their abnormal expression was closely related to tumor angiogenesis and immune microenvironment. This is exactly in line with our analysis. We found that in the process of constructing a risk model, cluster 2 with better survival prognosis of LC has a higher degree of immune cell infiltration than cluster 1. Moreover, through the analysis of the DEGs between the two cluster subtypes, it was found that they are involved in biological processes such as vascular development, cytokine production and tissue morphology, as well as PI3K-Akt signaling pathway, focal adhesion, ECM-receptor interaction, IL-17 signaling pathway, p53 signaling pathway and other important pathways for LC proliferation, invasion and metastasis. Taken together, we can reasonably infer that dysregulated angiogenesis may lead to impaired tumor microenvironment, leading to poor prognosis in LC. Therefore, we further analyzed the clinical characteristics of LC patients based on the constructed risk model.

Tumor angiogenesis is the primary pathway for tumor cell growth, invasion and metastasis. Tumor angiogenesis is positively correlated with tumor malignancy. We analyzed the clinical characteristics and ARGs risk scores of patients with LC and found that with the increase of risk score, the TNM of LC patients gradually increased, and they had poor survival prognosis. This also suggests that the higher the degree of ARGs mutation, the further deterioration of LC patients. Finally, we verified by LC microarray chip, and also proved the clinical validity of this risk model. In conclusion, our study clarifies the correlation of angiogenesis to LC and its immune microenvironment, which will provide a good prognostic basis for the clinical diagnosis of LC and help patients with LC receive better targeted treatment strategy.

Abbreviations

ARGs: Angiogenesis-related genes
DEGs: Differentially expressed genes

LC: Lung adenocarcinoma

TIM: Tumor immune microenvironment

**Declarations**

*Ethics approval and consent to participate:* Not applicable

*Consent for publication:* Not applicable

**Availability of data and materials:**


**Competing interests:**

The authors declare that they have no competing interests.

**Funding:**

1. Key Research and Development Program of Shandong (No. 2016CYJS08A01-6)

2. Research innovation team project on the mechanism and efficacy evaluation of Jingfang in the treatment of major diseases (No. 220316)

**Authors' contributions:**

Xue Wang and Xijian Liu are responsible for the design of the research ideas of the article. Lu Wang, Jiuwei Li and Ling Li analyzed and corrected the original data of the article. Yaxing Li conducted data processing and built a risk model. Hailiang Huang and Tao Han wrote articles and rechecked the data.

**Acknowledgments:** Not applicable

**Footnotes:** Not applicable

**References**


Tables

Table 1 is available in the Supplementary Files section.

Figures
Figure 1

Clustering and differential expression. (A-C) represents the effective construction of 2 cluster subgroups when $k = 2$, (D) shows the differential expression heat map of angiogenesis related genes between 2 cluster subgroups. (E) 2 cluster subgroups predict the survival and prognosis of patients with lung adenocarcinoma.
Figure 2

Differential expression gene (DEGs) analysis and enrichment analysis. (A) represents the DEGs of 2 clusters, red represents high expression and green represents low expression, (B, C, E) GO analysis and clustering of DEGs, (D, F) the signal pathway involved in DEGs, (G) analyzed the DEGs among 2 cluster through GSVA, (H) pathway analysis in GSEA.
Figure 3

Immunoassay of 2 clusters. (A) comparison of immune, stroma and microenvironment, (B, C) immune cell kurtosis was evaluated by MCPCounter and CIBERSORT.
Figure 4

Construction of risk model. (A) LASSO analysis with minimal lambda, (B) risk gene expression with risk score, survival status of lung adenocarcinoma patients, (C) survival prediction of lung adenocarcinoma patients by risk score, blue represents low risk score, red represents high risk score. (D) time-dependent ROC curves for risk model. (E) immunity differences between high and low risk.
Correlations between risk scores and clinical characteristics. (A-E) represent the differences in risk scores for different genders, TNM and disease stages, respectively. (F-O) Analysis of survival prognosis in patients with lung adenocarcinoma according to clinical characteristics. Shows an elevated risk score, which predicts a poorer prognosis. (P, Q) represent the validation of the risk model by GSE68465.

Supplementary Files

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

- Table1.pdf