# microbial

# Predictable regulation of survival by intratumoral microbeimmune crosstalk in patients with lung adenocarcinoma

Shuo Shi<sup>1</sup>, Yuwen Chu<sup>2,3</sup>, Haiyan Liu<sup>4,5</sup>, Lan Yu<sup>6,7,8</sup>, Dejun Sun<sup>8,9</sup>, Jialiang Yang<sup>2,3,5</sup>, Geng Tian<sup>2,3</sup>, Lei Ji<sup>2,3</sup>, Cong Zhang<sup>10,\*</sup> and Xinxin Lu<sup>11,\*</sup>

<sup>1</sup> The First Affiliated Hospital of Guangxi Medical University, Nanning 530021, Guangxi, China.

<sup>2</sup> Geneis Beijing Co., Ltd., Beijing 100102, China.

<sup>3</sup> Qingdao Geneis Institute of Big Data Mining and Precision Medicine, Qingdao 266000, Shandong, China.

<sup>4</sup> College of Information Engineering, Changsha Medical University, Changsha 410219, Hunan, China.

<sup>5</sup> Academician Workstation, Changsha Medical University, Changsha 410219, Hunan, China.

<sup>6</sup> Clinical Medical Research Center, Inner Mongolian People's Hospital, No. 20, Zhaowuda Road, Hohhot, Inner Mongolia, China.

<sup>7</sup> Inner Mongolia Key Laboratory of Gene Regulation of The Metabolic Disease, Inner Mongolian People's Hospital, No. 20, Zhaowuda Road, Hohhot, Inner Mongolia, China.

<sup>8</sup> Inner Mongolia Academy of Medical Sciences, Inner Mongolian People's Hospital, No. 20, Zhaowuda Road, Hohhot, Inner Mongolia, China.

<sup>9</sup> Pulmonary and Critical Care Medicine, Inner Mongolian People's Hospital, No. 20, Zhaowuda Road, Saihan District, Hohhot, Inner Mongolia, China.

<sup>10</sup> Hospital of Chengdu University of Traditional Chinese Medicine/No. 39, 12th Bridge Road, Jinniu District, Chengdu City, Sichuan Province, 610072, China.

<sup>11</sup> Nanjing Medical University Affiliated Cancer Hospital & Jiangsu Cancer Hospital & Jiangsu Institute of Cancer Research. \* Corresponding Authors:

<u>Cong Zhang</u>, Hospital of Chengdu University of Traditional Chinese Medicine/No. 39, 12th Bridge Road, Jinniu District, Chengdu City, Sichuan Province, 610072, China; E-mail: zhangcong1989@cdutcm.edu.cn

Xinxin Lu, Nanjing Medical University Affiliated Cancer Hospital & Jiangsu Cancer Hospital & Jiangsu Institute of Cancer Research; E-mail: luxinxinde@163.com

ABSTRACT Intratumoral microbiota can regulate the tumor immune microenvironment (TIME) and mediate tumor prognosis by promoting inflammatory response or inhibiting anti-tumor effects. Recent studies have elucidated the potential role of local tumor microbiota in the development and progression of lung adenocarcinoma (LUAD). However, whether intratumoral microbes are involved in the TIME that mediates the prognosis of LUAD remains unknown. Here, we obtained the matched tumor microbiome and host transcriptome and survival data of 478 patients with LUAD in The Cancer Genome Atlas (TCGA). Machine learning models based on immune cell marker genes can predict 1- to 5-year survival with relative accuracy. Patients were stratified into high- and low-survival-risk groups based on immune cell marker genes, with significant differences in intratumoral microbial communities. Specifically, patients in the high-risk group had significantly higher alpha diversity (p < 0.05) and were characterized by an enrichment of lung cancerrelated genera such as Streptococcus. However, network analysis highlighted a more active pattern of dominant bacteria and immune cell crosstalk in TIME in the low-risk group compared to the high-risk group. Our study demonstrated that intratumoral microbiota-immune crosstalk was strongly associated with prognosis in LUAD patients, which would provide new targets for the development of precise therapeutic strategies.

doi: 10.15698/mic2024.02.813 Received originally: 31.08.2023; In revised form: 09.01.2024, Accepted 16.01.2024, Published 19.02.2024.

*Keywords*: lung adenocarcinoma, intratumoral microbiota, tumor microenvironment, immune cell, prognosis.

#### Abbreviations:

AM – alveolar macrophage, LC – lung cancer, LUAD – lung adenocarcinoma, NSCLC – non-small cell LC, PC – plasma cell, TCGA – The Cancer Genome Atlas, TIME – tumor immune microenvironment.

#### INTRODUCTION

Lung cancer (LC) is one of the most common malignancies and a leading cause of disease-related death around the world [1]. Histopathological differences divide LC into nonsmall cell lung cancer (NSCLC) and small cell lung cancer (SCLC). Lung adenocarcinoma (LUAD) is the most prominent cytological type of NSCLC, accounting for approximately 40% of LC cases [2]. Although multimodal treatment strategies including immunotherapy, targeted therapy, chemoradiotherapy, and surgical resection have made great progress in recent decades [3, 4], the 5-year survival rate for patients with LC remains below 20% [5, 6]. Therefore, it is urgent to clarify the pathogenesis, diagnostic biomarkers, and therapeutic targets of LC to facilitate the diagnosis and treatment of LC.

The tumor immune microenvironment (TIME) largely determines the prognosis and the effect of immunotherapy of patients with cancer [7-10]. The composition of the tumor microenvironment varies by tumor type, but signature features include immune cells, stromal cells, blood vessels, and extracellular matrix, and are generally recognized as active agents of cancer progression [11]. Tumors are infiltrated by a variety of adaptive and innate immune cells that can perform both pro- and anti-tumor functions [12]. Song et al. developed a seven-gene prognostic signature based on nature killer cell marker genes in The Cancer Genome Atlas (TCGA) LUAD cohort, and its ability to predict prognosis has been well validated in different cohorts [13]. One study quantitatively analyzed the immune cell infiltration across 32 cancer types and observed considerable heterogeneity in the prognostic correlation of these cells across different cancer types, and in particular, established an immune-cell characteristic score model for LUAD that had a favorable prognostic performance [14]. Although the effects of immune infiltration on cancer treatment and prognosis have been extensively studied [7, 9, 10, 15], the factors that influence immune infiltration and the contributing factors to the individual heterogeneity of TIME have been largely unknown.

The TIME provides a friendly niche for the presence of a wide range of microbes, and tissue-specific intracellular microbes have been identified in most human tumors [16]. The lungs of healthy individuals have long been considered sterile, but with the maturity of second-generation sequencing technology, the diversity of the lung microbiota and its relationship to lung disease and LC has been confirmed [17, 18]. Over the past decade, microbial communities have been implicated in the initiation, progression, metastasis, and response to treatment of a variety of cancers [19-22]. Recent studies have shown that microbes exist in tumor cells and immune cells, indicating that these microbes can affect the status of tumor immune microenvironment [23-25]. Studies have shown differences in the lung microbiome between patients with LC and those with benign lung disease, and that certain bacteria may have the potential to predict LC [26]. During the development of lung cancer, the number and species of commensal microorganisms in the lung changed, which promoted the proliferation and function of resident immune cells in the lung, furthermore, it promotes the development of LC through its effect on inflammatory reaction [27]. However, whether the microbiome in tumor tissue is related to the TIME and



**FIGURE 1: Overview of the analysis pipeline.** The tumor microbiome abundance of LUAD was annotated by Poore *et al.* from RNA-Seq data and the matched host gene expression was downloaded from The Cancer Genome Atlas. Immune cell marker genes were used to build machine learning models to predict patient survival. COX regression analysis based on immune cell marker genes stratified patients into high- and low-risk. The intratumoral microbiota, the tumor immune microenvironment, and their crosstalk between the high-and low-risk groups were further explored.

30

prognosis of LUAD remains unclear. Besides, the pattern of microbe-immune cell crosstalk in the TIME of LUAD and its prognostic implications need exploration. The wholetranscriptome sequencing data provided by TCGA offers a good opportunity to explore the crosstalk between the intratumoral microbiota and the TIME, which can be guantified based on host gene expression. Here, we identified immune cell marker genes in LC tissues and correlated them with the prognosis of patients with LUAD, and compared the intratumoral microbiota of high- and low-risk patients, as well as crosstalk patterns with immune cells in the TIME. We found that the immune cell marker genebased machine learning model can predict the survival of patients with LUAD accurately. The intratumoral microbiota differed significantly between high- and low-risk patients, and there was variation in the crosstalk pattern between the microbial components and immune cells in the TIME.

#### RESULTS

#### **Pipeline of this study**

The workflow of this study is shown in **Figure 1**. To illuminate the intratumoral microbiota in LUAD, we revisited and obtained the intratumoral microbial profiles in multiple cancer types, which were processed by Poore *et al.* using sequencing data in TCGA [28]. The LUAD samples in TCGA consist of 478 RNA sequencing (RNA-seq) data from the primary tumor of 478 patients. In addition, we also obtained the host gene expression of these patients with LUAD in TCGA which matched with the tumor microbiome.

We next downloaded cell marker genes from CellMarker2 database and selected lung tissue of LUAD to get marker genes in lung cells. By filtering out genes unrelated to immunity, we obtained 297 immune cell marker genes in lung tissue associated with LUAD. These genes corresponded mainly to 33 types of immune cells (**Figure 2A**). Among these, T cells had the most marker genes, reaching 31, followed by macrophages and cancer stem cells. Other cell types such as effector T cells, naive B cells, and alveolar macrophages (AM) etc. had only three marker genes. These genes will be used to predict the survival time of patients with LUAD.

#### Immune cell marker genes show strong power in prediction of LUAD survival

478 patients with LUAD with their gene expression have been obtained from TCGA. First, we dichotomized patients based on survival time of one to five years, respectively. Five-fold cross validation was implemented to verify the



FIGURE 2: Survival prediction model based on immune cell marker genes using machine learning. (A) The number of marker genes corresponding to specific immune cell types. (B) ROC curves of 1- to 5-year survival prediction by five-fold cross validation of six machine learning algorithms.

accuracy of the six machine learning algorithms. Specifically, we classified all patients into five groups. Four of five group samples were used to train the model and the remaining one was used to test the model. After repeating this process five times, each group has been tested once and trained four times. As shown in **Figure 2B**, the prediction accuracy fluctuated slightly with the different survival time as the grouping threshold, and the difference of prediction accuracy of different algorithms was also very small. In some models, such as GB (gradient boosting), the mean AUC (area under the curve) for predicting one-year survival was up to 0.72, which means it is effective to predict the survival time of patients with LUAD through 297 immune cell marker genes.

## Immune-related activities are associated with survival in LUAD patients

Since the TIME plays an important role in the development of LUAD [29, 30], we next explored the impact of differential expression of immune cell marker genes on the TIME of patients with LUAD. First, we performed univariate COX regression on 297 immune cell marker genes and identified 84 genes that were significantly associated with patient survival. Multivariate COX regression analyses were then performed based on these survival-related genes. The regression coefficients of these prognostic genes were obtained and the risk score of each patient was calculated based on the expression levels and coefficients of each gene. Figure 3A shows the survival curves of high-risk and low-risk. A P-value of less than 0.001 indicates that these prognosis-related immune cell marker genes could significantly distinguish the survival time of patients (Figure 3A). Furthermore, we examined the ten genes most significantly associated with survival and found that they corresponded to immune cell types such as macrophages and regulatory T cells (Figure 3B).

We next characterized the immune cell infiltration of patients in high-risk and low-risk. CIBERSORT was used to quantify the abundance of 22 types of immune cells in the TIME (Figure 3C). Among them, M2 macrophages had the highest abundance, followed by CD4+ T cells and CD8+ T cells. Moreover, we identified six types of immune cells that significantly differed between the high-risk group and the low-risk group, including M0 macrophages, B cell plasma, and myeloid dendritic cells (Figure 3D).

Since we used 297 genes to predict the survival time of patients with LUAD and the experiments results showed relatively high precision, these 297 genes should be significantly associated with the prognosis of LUAD in function. Therefore, we next explored the GO terms of these genes. Enrichment analysis showed that these genes were significantly associated with 1359 GO terms (adjusted P < 0.05). The GO terms can be divided into three classes: 1220 biological process, 59 cellular components, 80 molecular function. **Figure 3E** shows the five GO terms with the largest number of genes in each class. These genes enriched processes are associated with immune-related activities such as regulation of cell-cell adhesion, regulation of T cell activation, and cytokine receptor binding. **Figure 3F-H** showed

the most significantly enriched GO terms in each class, along with the corresponding genes. CD74 is a receptor for the cytokine macrophage migration inhibitor [31], and Kashima et al. reported that CD74 is a novel gene that plays a key role in the drug-resistant state [32]. FOXP3 is a member of the forkhead transcription factor family, which is primarily expressed in a subset of CD4 + T cells and plays an inhibitory role in the immune system [33]. Yang et al. reported that FOXP3 can act as a co-activator of the Wnt-bcatenin signaling pathway, inducing epithelialmesenchymal transition and tumor growth and metastasis in NSCLC [34]. Takanami found that CCR7 may be involved in the development of lymph node metastasis in NSCLC [35].

#### Intratumoral microbiota differentiation between highand low-survival-risk patients

Recent studies have shown that the intratumoral microbiota plays a key role in theTIME [23, 36], so we next explored whether there are differences in the intratumoral microbiota between high- and low-survival-risk patients. Proteobacteria was the most abundant phylum and Pseudomonas was the most abundant genus (Figure 4A). There was significant difference in alpha-diversity between high- and low-survival-risk patients (P < 0.05). For instance, the microbial abundance of the low-risk group was significantly higher that of the high-risk group (Figure 4B, P = 0.039), while the Shannon (Figure 4C, P = 0.017) and Simpson indices (Figure 4D, P = 0.044) of the low-risk group were significantly lower than that of the high-risk group. Moreover, beta-diversity analysis showed that intratumoral microbial profiles were significantly different between the low-risk and high-risk group (Figure 4E, P = 0.02), and beta-diversity was more dissimilar among individuals in the low-risk group (Figure 4F, P < 0.001). Based on linear discriminant analysis effect size (LEfSe) analysis, at the phylum level, six phyla were enriched in the high-risk group and two phyla were enriched in the low-risk group (Figure 4G). At the genus level, 17 genera were enriched in the high-risk group and one genus was enriched in the low-risk group (Figure 4H).

# Different microbiota-immune crosstalk patterns in the TIME between the high- and low-risk group

We next explored whether microbiota-immune cell crosstalk in the TIME differed between the high- and low-risk group. Considering the predominance of dominant bacteria in the community, we performed Spearman correlation analysis for the top 50 genera in relative abundance and 22 types of immune cells. **Figures 5A** and **B** show only the microbe-immune cell pairs that were significantly associated (p < 0.05). The results of network analysis showed that the low-risk group presented more active microbe-immune crosstalk than the high-risk group (**Figure 5A-C**). The highrisk network had 134 edges, including 75 positive correlations and 59 negative correlations, while the low-risk network had 161 edges, including 89 positive correlations and 72 negative correlations (**Figure 5C**). In addition, more nodes and higher average degree of nodes in the low-risk



FIGURE 3: Tumor immune microenvironment and related functions are associated with survival in LUAD patients. (A) Survival curve in different risk score groups obtained by COX regression analysis. (B) Sankey plot showing the correlation between immune cells and the ten genes most associated with survival. (C) Relative abundance of the immune cell components in each patient. (D) Boxplot showing the differences in the abundance of immune cells between the high- and low-risk group. Wilcoxon test was used to perform the statistical test. (E) Five GO terms with the largest number of genes in each class. (F-H) The most significantly enriched GO terms in each class, along with the corresponding genes.



FIGURE 4: The intratumoral microbial profile was significantly different between the high- and low-risk group. (A) Relative abundance of the intratumoral microbes at the genus level in each patient. Boxplot showing the difference in (B) microbial richness, (C) the Shannon and (D) Simpson indices between the high- and low-risk group. Wilcoxon test was used to perform the statistical test. (E) PCoA based on the Bray-Curtis dissimilarity matrix showing the difference in intratumoral microbial community composition between the high- and low-risk group. (F) Boxplot showing the difference in Bray-Curtis dissimilarity index between the high- and low-risk group. Significantly different microbes in abundance between the high- and low-risk group at the (G) phylum and (H) genus level.

group network than in the high-risk group indicated more complex and robust microbe-immune crosstalk pattern in the low-risk group.

To further explore the relationship between specific microbes and specific cells, we screened the significantly correlated microbial-immune pairs with absolute correlation coefficients greater than 0.2 (Figure 5D-E). We ob-

served some common and significant microbe-immune cell associations in the high-risk and low-risk groups. For instance, in both groups, *Lachnoclostridium*, *Acinetobacter*, and *Paenibacillus* were positively correlated with M1 macrophages. *Aeromonas* and *Vibrio* were negatively correlated with regulatory T cells. However, we still identified multiple correlations between microbes and cell types which were specific for the survival risk group. Most notably, memory CD4+ T cells were positively associated with more than a dozen bacteria in the low-risk group, compared with just two in the high-risk group. The role of memory T cells in LC has been extensively studied [37, 38]. One study reported that when tissue-resident memory T cells are present in tumors, they act together to attack the cancer cells and protect the host [38]. Our results show that the interaction between memory T cells and intratumoral bacteria in TIME is different in patients with LUAD with different survival risks.





Mast cell activated

high- and (B) low-risk groups. Only edges with p < 0.05 were shown in the figure. The size of the node indicates the number of nodes connected to it in the network. The solid yellow line and the dotted gray line indicate positive and negative correlations, respectively. (C) Comparison of parameters of microbe-immune interaction network in high- and low-risk group. On the basis of Figure 5a and b, microbial-immune cell relationship pairs with an absolute value of correlation coefficient greater than 0.2 were screened, further resulting in network plots of (D) high- and (E) low-risk groups.

T cell CD4+ mer

#### DISCUSSION

Recent studies have identified the presence of intratumoral microbiota in various non-gastrointestinal tumors, including LUAD. However, the role of the intratumoral microbiota in the prognosis of LUAD remains largely unknown. In this study, survival of patients with LUAD could be distinguished based on immune cell marker genes. The intratumoral microbiota varied between high- and low-survivalrisk patients using these immune cell marker genes. Moreover, the intratumoral microbiota-immune cell crosstalk pattern were found to be different between these two groups, which may contribute to the prognosis of patients with LUAD.

In the experiment of predicting the survival of patients with LUAD by selected 297 immune cell marker genes, the machine learning model obtained the result of AUC = 0.72, which is relatively accurate, but there is still room for improvement. These immune cell-specific genes correspond to a wide variety of immune cells, not all of which are involved in the development of LUAD. Therefore, more mechanistic studies are needed to investigate and identify specific immune cells that may influence the progress of LUAD, or to identify tumor immune infiltration characteristics in patients with different prognostic risks. Therefore, predictive models based on marker genes of specific immune cells that regulate the development of LUAD through a well-defined mechanism or function can greatly improve the accuracy of patient prognosis prediction. Altogether, only by accurately identifying and classifying immune cell marker genes related to pathogenesis and treatment can achieve personalized and precise treatment.

The diversity of TIME profiles in patients with LUAD has been highlighted in previous reports, proving that it could serve as a hallmark for LUAD development [39-42]. Shinohara et al. conducted a single-sample gene set enrichment analysis of TIME-related gene sets to develop a new scoring system (TIME score), the TIME score captures the intricate interactions between tumor proliferation, anti-tumor immunity and immunosuppression, which may be useful in predicting the prognosis or selecting treatment strategies in patients with LC [42]. Taniguchi et al. revealed that AMs promote the proliferation of cancer cells [43]. Under tumor-containing conditions, the expression of statin  $\beta A$  (IN-HBA) in lung AMs is up-regulated, thus promoting tumor proliferation and forming a "vicious cycle" in in vivo tumor environment. We found that the abundance of M0 macrophages was significantly higher in the high-risk group compared to the low-risk group (p < 0.0001), consistent with previous reports. In addition, B and plasma cells (PCs) were found to be more abundant in the high-risk group. By a comprehensive analysis of 50,000 tumor-infiltrating B and PCs, Hao et al. found that memory B cells and PCs were highly enriched and highly differentiated in tumor tissues, and PC were significantly increased in smokers with distinct differentiation trajectories [44]. Furthermore, one study showed that memory T cells in lung tumors predicted good outcomes for patients, and that patients with high levels of these cells in their tumors were 34% less likely to die [38]. Consistently, we found that tissue-resident memory CD4+

T cells were more enriched in the low-risk group compared with the high-risk group.

We found that the intratumoral microbiome profiles were significantly different between the high- and low-risk group. Interestingly, a large number of genera were significantly enriched in the high-risk group, including multiple LC-related pathogens, such as Streptococcus, Escherichia, and Klebsiella. Li et al. reported that LC cells infected with Streptococcus pneumoniae formed larger tumors in mice compared to untreated LC cells, and their abundance was associated with survival [45]. LC surgery is prone to serious infectious complications caused by Gram-negative bacteria such as Escherichia coli, which may reduce long-term survival after discharge through cancer recurrence and metastasis [46]. Klebsiella expression is more pronounced in lung squamous-cell carcinoma than in LUAD, however, we still found significantly enriched Klebsiella in high-risk LUAD patients, suggesting their potential significance in LUAD prognosis [47]. Microbiota-immune crosstalk in the TIME may contribute to the heterogeneity of outcomes in patients with LUAD. Jin et al. reported that LC alters the number and type of microbes in the lung and activates the immune system, creating an inflammatory environment for LC and ultimately promoting the development of LC [27]. Although intratumoral microbial alpha diversity was significantly lower in low-risk patients than in high-risk patients, we identified more complex and close microbe-immune interactions in low-risk patients. Predictive models that combine immune cell marker genes with their associated intratumoral microbiota may further improve performance in predicting patient survival. In addition, our results suggest that targeting specific microbes within tumors can modify tumor immune infiltration by exploiting the association of microbes with immune components. Future multicenter studies with larger cohorts will be needed to determine the TIME characteristics that are most favorable to patient outcomes for LUAD.

There were several limitations in this study. A major limitation was that our study on the interactions between intratumoral microbiota and immune cells were only based on a single TCGA dataset, and lacked external independent verification. Moreover, the causal relationship and specific mechanisms between intratumoral microbiota and immune and LUAD prognosis require rigorous experimental verification. Another limitation was that the tumor microbiome abundance was obtained by Kraken pipeline from RNA sequencing data. Therefore, it is necessary to validate our results by other microbial detection methods, such as metagenomic sequencing or PCR analysis.

In conclusion, this study advances the understanding of the relationship between intratumoral microbe-immune crosstalk and prognosis in patients with LUAD. Although components in the TIME have emerged as potential targets for lung cancer immunotherapy, our study suggests that ignoring the important role of intratumoral microbiota in the TIME may not enable all patients to benefit from immunotherapy. Future development of emerging immunotherapy strategies for LUAD will require perturbation of the microbe-immune cell crosstalk pattern in the TIME to achieve truly individualized precision-targeted therapies.

#### MATERIALS AND METHODS

#### Data acquisition

The intratumoral microbiome data used in this study and the metadata were downloaded from a previous work conducted by Poore *et al.*, and are available at ftp://ftp.microbio. me/pub/cancer\_microbiome\_analysis/. Poore et al. developed a Kraken TCGA microbial-detection pipeline, which uses an ultrafast Kraken algorithm to map sequence readings that do not align with the human reference genome to known bacterial, viral, and archaea microbial genomes [28]. The bacterial abundance data in tumor tissue of patients with LUAD was used in this study. The decontamination process was detailed in the original paper [28]. The overall survival time and survival status of the samples were collected from UCSC Xena (http://xena.ucsc.edu/). The quantification of host gene expression by **RNA-Seq** were downloaded from https://portal.gdc.cancer.gov/.

#### Selection of immune cell marker genes

The immune cell marker gene must satisfy two conditions: first, it must be the immune cell marker gene in the lung, and second, all the immune cell marker genes must be related to LUAD. By defining LUAD and immune cells, we obtained a total of 297 immune cell marker genes from the CellMarker 2.0 [48] database.

### Quantification of immune cells in the TIME of patients with $\ensuremath{\mathsf{LUAD}}$

CIBERSORT [49] was used to perform the immune cell analysis based on the TCGA LUAD gene expression data. We converted FPKM (fragments per kilobase of transcript per million fragments mapped) value to TPM (transcripts per million) value because TPM can correct the batch effect, so the sum of FPKM is a fixed value.

#### Survival time prediction method

Six machine learning models were used to predict the one to five-year survival of patients with LUAD. These models were implemented using the scikit-learn library in Python. Bagging (Bootstrap Aggregating) is an ensemble learning method that reduces model variance and improves prediction stability and accuracy by combining multiple decision trees. The fundamental idea is to randomly select multiple sub-samples from the training dataset using bootstrap sampling and train multiple base learners on these sub-samples. Finally, the predictions of these base learners were aggregated to obtain the final ensemble prediction.

LGBM, XGBoost (XGB), and GB (Gradient Boosting) are three machine learning algorithms based on Gradient Boosting Trees. The formula for the Gradient Boosting Tree algorithm is as follows:

$$F_0(x) = argmin_{\gamma} \sum_{i=1}^n L(y_i, \gamma)$$

 $L(y_i, \gamma)$  is the loss function,  $y_i$  is the true label of the first sample of the training data, and  $\gamma$  is the initial predicted value of the model. For the m-round iteration, a new decision tree model  $h_m(x)$  is constructed on the basis of the previous round model, with the goal of reducing the loss function  $r_{im}$ :

$$r_{im} = \frac{\partial L(y_i, F_{m-1}(x_i))}{\partial F_{m-1}(x_i)}$$

The newly constructed model  $h_m(x)$  is weighted with the previous one  $F_{m-1}(x)$  to get the updated model  $F_{m-1}(x)$ :

 $\mathbf{F}_{\mathbf{m}}(\mathbf{x}) = \mathbf{F}_{\mathbf{m}-1}(\mathbf{x}) + \eta * h_{\mathbf{m}}(\mathbf{x})$ 

Among  $\eta$  is the learning rate, which can control the weight of each model. Iterative updates are repeated until a predetermined number of iterations is reached.

LGBM, XGBoost, and GB are three gradient boosting tree algorithms that optimize and enhance the gradient boosting algorithm, thereby improving model efficiency and prediction performance. On the other hand, Adaboost is a specialized implementation of gradient boosting trees. It iteratively trains a series of weak classifiers (usually decision trees) and calculates weights for each weak classifier based on its error rate, resulting in a strong classifier.

The advantages of Adaboost lie in its ability to effectively enhance classifier accuracy and handle complex problems. As an ensemble learning method, Random Forest predicts patient survival by constructing multiple decision trees, each trained on random samples of data and features. The predictions of multiple trees are then combined to obtain the final survival prediction. To evaluate the model performance and prevent overfitting, we employed 5-fold cross-validation. The average concordance index obtained from the 5-fold cross-validation served as the evaluation metric for the models.

#### Microbial diversity analysis

The microbial alpha diversity was measured by the Shannon and Simpson indices, and was calculated by the "vegdist" function in R package "vegan". The microbial beta diversity was measured by the Bray-Curtis dissimilarity matrix. Linear discriminant analysis Effect Size (LEfSe) was used to identify the significantly different microbes in relative abundance, with a linear discriminant analysis (LDA) score greater than 2 as the threshold.

#### Construction of microbe-immune cell crosstalk network

To investigate the interactions between intratumoral microbes and immune cells, we conducted a correlation analysis on 22 types of immune cells and the top 50 microbes in relative abundance at the genus level. The "psych" package in R was used to perform the Spearman correlation analysis and calculate the correlation coefficients and p-values. First, to explore the overall properties of microbe-immune interaction networks in high-risk and low-risk groups, significant relationships with p-values < 0.05 were selected to construct the microbeimmune interaction network. Gephi was used to visualize the network. Furthermore, to identify the interaction of a particular microbe with a particular cell, pairs with an absolute value of correlation coefficient greater than 0.2 and a p value less than 0.05 were screened for further network construction.

#### Statistical analysis

All statistical calculations were conducted using R software (Version 4.2.1). Differences between two groups were compared using Wilcoxon rank sum test. Correlations between immune cells and intratumoral microbes were calculated using Spearman's correlation analysis. Survival curves were performed using the Kaplan–Meier (KM) method, and the significance was determined by the log-rank test. Univariate Cox regression analysis was used to calculate the significance of

the association between immune cell marker genes and prognosis in patients with LUAD. The Python package "Sklearn" and library "matplotlib" was used to plot receiver operating characteristic (ROC) curves and obtain the area under the curve (AUC). p < 0.05 was considered statistically significant.

#### **Data availability**

The intratumoral microbiome abundance data and metadata are available at <u>ftp://ftp.microbio.me/pub/cancer</u><u>microbiome\_analysis/</u>. The host gene expression data is available at <u>https://portal.gdc.cancer.gov/</u>.

#### ACKNOWLEDGEMENTS

Funding: Science and Technology Planning Project of Inner Mongolia (No. 2020GG0084); National Natural Science Foundation of China (No. 81960449). We appreciate the computing resources provided by Geneis Beijing Co., Ltd.

#### REFERENCES

1. Barta JA, Powell CA, Wisnivesky JP (2019). Global Epidemiology of Lung Cancer. Ann Glob Health 85(1): 8. doi: 10.5334/aogh.2419

2. Sung H, Ferlay J, Siegel RL, Laversanne M, Soerjomataram I, Jemal A, Bray F (**2021**). Global Cancer Statistics 2020: GLOBOCAN Estimates of Incidence and Mortality Worldwide for 36 Cancers in 185 Countries. **CA Cancer J Clin** 71(3): 209-249. doi: 10.3322/caac.21660

3. Ferrer I, Zugazagoitia J, Herbertz S, John W, Paz-Ares L, Schmid-Bindert G (2018). KRAS-Mutant non-small cell lung cancer: From biology to therapy. Lung Cancer 124: 53-64. doi: 10.1016/j.lungcan.2018.07.013

4. Reck M, Rabe KF (**2017**). Precision Diagnosis and Treatment for Advanced Non-Small-Cell Lung Cancer. **N Engl J Med** 377(9): 849-861. doi: 10.1056/NEJMra1703413

5. de Groot PM, Wu CC, Carter BW, Munden RF (2018). The epidemiology of lung cancer. Transl Lung Cancer Res 7(3): 220-233. doi: 10.21037/tlcr.2018.05.06

6. Herbst RS, Morgensztern D, Boshoff C (2018). The biology and management of non-small cell lung cancer. Nature 553(7689): 446-454. doi: 10.1038/nature25183

7. Thorsson V, Gibbs DL, Brown SD, Wolf D, Bortone DS, Ou Yang TH, Porta-Pardo E, Gao GF, Plaisier CL, Eddy JA, Ziv E, Culhane AC, Paull EO, Sivakumar IKA, Gentles AJ, Malhotra R, Farshidfar F, Colaprico A, Parker JS, Mose LE, Vo NS, Liu J, Liu Y, Rader J, Dhankani V, Reynolds SM, Bowlby R, Califano A, Cherniack AD, Anastassiou D, et al. (**2018**). The Immune Landscape of Cancer. **Immunity** 48(4): 812-830 e814. doi: 10.1016/j.immuni.2018.03.023

8. Bagaev A, Kotlov N, Nomie K, Svekolkin V, Gafurov A, Isaeva O, Osokin N, Kozlov I, Frenkel F, Gancharova O, Almog N, Tsiper M, Ataullakhanov R, Fowler N (**2021**). Conserved pan-cancer microenvironment subtypes predict response to immunotherapy. **Cancer Cell** 39(6): 845-865 e847. doi: 10.1016/j.ccell.2021.04.014

9. Desbois M, Udyavar AR, Ryner L, Kozlowski C, Guan Y, Durrbaum M, Lu S, Fortin JP, Koeppen H, Ziai J, Chang CW, Keerthivasan S, Plante M, Bourgon R, Bais C, Hegde P, Daemen A, Turley S, Wang Y (2020). Integrated digital pathology and transcriptome analysis identifies molecular mediators of T-cell exclusion in ovarian cancer. Nat Commun 11(1): 5583. doi: 10.1038/s41467-020-19408-2

10. He Y, Jiang Z, Chen C, Wang X (2018). Classification of triplenegative breast cancers based on Immunogenomic profiling. J Exp Clin Cancer Res 37(1): 327. doi: 10.1186/s13046-018-1002-1

#### **CONFLICT OF INTEREST**

The authors declare that they have no known competing financial interests or personal relationships.

#### COPYRIGHT

© 2024 Shi *et al.* This is an open-access article released under the terms of the Creative Commons Attribution (CC BY) license, which allows the unrestricted use, distribution, and reproduction in any medium, provided the original author and source are acknowledged.

Please cite this article as: Shuo Shi, Yuwen Chu, Haiyan Liu, Lan Yu, Dejun Sun, Jialiang Yang, Geng Tian, Lei Jixx, Cong Zhang and Xinxin Lu (**2024**). Predictable regulation of survival by intratumoral microbe-immune crosstalk in patients with lung adenocarcinoma. **Microbial Cell** 11: 29-40. doi: 10.15698/mic2024.02.813

11. Xiao Y, Yu D (**2021**). Tumor microenvironment as a therapeutic target in cancer. **Pharmacol Ther 221**: 107753. doi: 10.1016/j.pharmthera.2020.107753

12. Anderson NM, Simon MC (2020). The tumor microenvironment. Curr Biol 30(16): R921-R925. doi: 10.1016/j.cub.2020.06.081

13. Song P, Li W, Guo L, Ying J, Gao S, He J (**2022**). Identification and Validation of a Novel Signature Based on NK Cell Marker Genes to Predict Prognosis and Immunotherapy Response in Lung Adenocarcinoma by Integrated Analysis of Single-Cell and Bulk RNA-Sequencing. **Front Immunol** 13: 850745. doi: 10.3389/fimmu.2022.850745

14. Zuo S, Wei M, Wang S, Dong J, Wei J (**2020**). Pan-Cancer Analysis of Immune Cell Infiltration Identifies a Prognostic Immune-Cell Characteristic Score (ICCS) in Lung Adenocarcinoma. **Front Immunol** 11: 1218. doi: 10.3389/fimmu.2020.01218

15. Liang L, Yu J, Li J, Li N, Liu J, Xiu L, Zeng J, Wang T, Wu L (**2021**). Integration of scRNA-Seq and Bulk RNA-Seq to Analyse the Heterogeneity of Ovarian Cancer Immune Cells and Establish a Molecular Risk Model. **Front Oncol** 11: 711020. doi: 10.3389/fonc.2021.711020

16. Nejman D, Livyatan I, Fuks G, Gavert N, Zwang Y, Geller LT, Rotter-Maskowitz A, Weiser R, Mallel G, Gigi E, Meltser A, Douglas GM, Kamer I, Gopalakrishnan V, Dadosh T, Levin-Zaidman S, Avnet S, Atlan T, Cooper ZA, Arora R, Cogdill AP, Khan MAW, Ologun G, Bussi Y, Weinberger A, Lotan-Pompan M, Golani O, Perry G, Rokah M, Bahar-Shany K, et al. (**2020**). The human tumor microbiome is composed of tumor type-specific intracellular bacteria. **Science** 368(6494): 973-980. doi: 10.1126/science.aay9189

17. Dickson RP, Huffnagle GB (2015). The Lung Microbiome: New Principles for Respiratory Bacteriology in Health and Disease. PLoS Pathog 11(7): e1004923. doi: 10.1371/journal.ppat.1004923

18. Dickson RP, Martinez FJ, Huffnagle GB (2014). The role of the microbiome in exacerbations of chronic lung diseases. Lancet 384(9944): 691-702. doi: 10.1016/S0140-6736(14)61136-3

19. Mouradov D, Greenfield P, Li S, In EJ, Storey C, Sakthianandeswaren A, Georgeson P, Buchanan DD, Ward RL, Hawkins NJ, Skinner I, Jones IT, Gibbs P, Ma C, Liew YJ, Fung KYC, Sieber OM (**2023**). Oncomicrobial Community Profiling Identifies Clinicomolecular and Prognostic Subtypes of Colorectal Cancer. **Gastroenterology** 165(1): 104-120. doi: 10.1053/j.gastro.2023.03.205

20. Zhao LY, Mei JX, Yu G, Lei L, Zhang WH, Liu K, Chen XL, Kolat D, Yang K, Hu JK (**2023**). Role of the gut microbiota in anticancer therapy: from molecular mechanisms to clinical applications. **Signal Transduct Target Ther** 8(1): 201. doi: 10.1038/s41392-023-01406-7

21. Yuan X, Wang Z, Li C, Lv K, Tian G, Tang M, Ji L, Yang J (**2022**). Bacterial biomarkers capable of identifying recurrence or metastasis carry disease severity information for lung cancer. **Front Microbiol 13**: 1007831. doi: 10.3389/fmicb.2022.1007831

22. Banerjee S, Wei Z, Tian T, Bose D, Shih NNC, Feldman MD, Khoury T, De Michele A, Robertson ES (**2021**). Prognostic correlations with the microbiome of breast cancer subtypes. **Cell Death Dis** 12(9): 831. doi: 10.1038/s41419-021-04092-x

23. Ma J, Huang L, Hu D, Zeng S, Han Y, Shen H (**2021**). The role of the tumor microbe microenvironment in the tumor immune microenvironment: bystander, activator, or inhibitor? J Exp Clin Cancer Res 40(1): 327. doi: 10.1186/s13046-021-02128-w

24. Quail DF, Joyce JA (**2013**). Microenvironmental regulation of tumor progression and metastasis. **Nat Med** 19(11): 1423-1437. doi: 10.1038/nm.3394

25. Cullin N, Azevedo Antunes C, Straussman R, Stein-Thoeringer CK, Elinav E (**2021**). Microbiome and cancer. **Cancer Cell** 39(10): 1317-1341. doi: 10.1016/j.ccell.2021.08.006

26. Cheng C, Wang Z, Wang J, Ding C, Sun C, Liu P, Xu X, Liu Y, Chen B, Gu B (**2020**). Characterization of the lung microbiome and exploration of potential bacterial biomarkers for lung cancer. **Transl Lung Cancer Res** 9(3): 693-704. doi: 10.21037/tlcr-19-590

27. Jin C, Lagoudas GK, Zhao C, Bullman S, Bhutkar A, Hu B, Ameh S, Sandel D, Liang XS, Mazzilli S, Whary MT, Meyerson M, Germain R, Blainey PC, Fox JG, Jacks T (**2019**). Commensal Microbiota Promote Lung Cancer Development via gammadelta T Cells. **Cell** 176(5): 998-1013 e1016. doi: 10.1016/j.cell.2018.12.040

28. Poore GD, Kopylova E, Zhu Q, Carpenter C, Fraraccio S, Wandro S, Kosciolek T, Janssen S, Metcalf J, Song SJ, Kanbar J, Miller-Montgomery S, Heaton R, McKay R, Patel SP, Swafford AD, Knight R (**2020**). Microbiome analyses of blood and tissues suggest cancer diagnostic approach. **Nature** 579(7800): 567-574. doi: 10.1038/s41586-020-2095-1

29. Sorin M, Rezanejad M, Karimi E, Fiset B, Desharnais L, Perus LJM, Milette S, Yu MW, Maritan SM, Dore S, Pichette E, Enlow W, Gagne A, Wei Y, Orain M, Manem VSK, Rayes R, Siegel PM, Camilleri-Broet S, Fiset PO, Desmeules P, Spicer JD, Quail DF, Joubert P, Walsh LA (**2023**). Single-cell spatial landscapes of the lung tumour immune microenvironment. **Nature** 614(7948): 548-554. doi: 10.1038/s41586-022-05672-3

30. Zhang Y, Yang M, Ng DM, Haleem M, Yi T, Hu S, Zhu H, Zhao G, Liao Q (2020). Multi-omics Data Analyses Construct TME and Identify the Immune-Related Prognosis Signatures in Human LUAD. Mol Ther Nucleic Acids 21: 860-873. doi: 10.1016/j.omtn.2020.07.024

31. David K, Friedlander G, Pellegrino B, Radomir L, Lewinsky H, Leng L, Bucala R, Becker-Herman S, Shachar I (**2022**). CD74 as a regulator of transcription in normal B cells. **Cell Rep** 41(5): 111572. doi: 10.1016/j.celrep.2022.111572

32. Kashima Y, Shibahara D, Suzuki A, Muto K, Kobayashi IS, Plotnick D, Udagawa H, Izumi H, Shibata Y, Tanaka K, Fujii M, Ohashi A, Seki M, Goto K, Tsuchihara K, Suzuki Y, Kobayashi SS (2021). Single-Cell Analyses Reveal Diverse Mechanisms of Resistance to EGFR Tyrosine Kinase Inhibitors in Lung Cancer. Cancer Res 81(18): 4835-4848. doi: 10.1158/0008-5472.CAN-20-2811

33. Kim CH (2009). FOXP3 and its role in the immune system. Adv Exp Med Biol 665: 17-29. doi:  $10.1007/978\text{-}1\text{-}4419\text{-}1599\text{-}3\_2$ 

34. Yang S, Liu Y, Li MY, Ng CSH, Yang SL, Wang S, Zou C, Dong Y, Du J, Long X, Liu LZ, Wan IYP, Mok T, Underwood MJ, Chen GG (**2017**). FOXP3 promotes tumor growth and metastasis by activating Wnt/beta-catenin signaling pathway and EMT in non-small cell lung cancer. **Mol Cancer** 16(1): 124. doi: 10.1186/s12943-017-0700-1

35. Takanami I (**2003**). Overexpression of CCR7 mRNA in nonsmall cell lung cancer: correlation with lymph node metastasis. **Int J Cancer** 105(2): 186-189. doi: 10.1002/ijc.11063

36. Byrd DA, Fan W, Greathouse KL, Wu MC, Xie H, Wang X (2023). The intratumor microbiome is associated with microsatellite instability. J Natl Cancer Inst 115(8):989-993. doi: 10.1093/jnci/djad083

37. Weeden CE, Gayevskiy V, Marceaux C, Batey D, Tan T, Yokote K, Ribera NT, Clatch A, Christo S, Teh CE, Mitchell AJ, Trussart M, Rankin L, Obers A, McDonald JA, Sutherland KD, Sharma VJ, Starkey G, D'Costa R, Antippa P, Leong T, Steinfort D, Irving L, Swanton C, Gordon CL, Mackay LK, Speed TP, Gray DHD, Asselin-Labat ML (2023). Early immune pressure initiated by tissue-resident memory T cells sculpts tumor evolution in non-small cell lung cancer. Cancer Cell 41(5): 837-852 e836. doi: 10.1016/j.ccell.2023.03.019

38. Ganesan AP, Clarke J, Wood O, Garrido-Martin EM, Chee SJ, Mellows T, Samaniego-Castruita D, Singh D, Seumois G, Alzetani A, Woo E, Friedmann PS, King EV, Thomas GJ, Sanchez-Elsner T, Vijayanand P, Ottensmeier CH (**2017**). Tissue-resident memory features are linked to the magnitude of cytotoxic T cell responses in human lung cancer. **Nat Immunol** 18(8): 940-950. doi: 10.1038/ni.3775

39. Bischoff P, Trinks A, Obermayer B, Pett JP, Wiederspahn J, Uhlitz F, Liang X, Lehmann A, Jurmeister P, Elsner A, Dziodzio T, Ruckert JC, Neudecker J, Falk C, Beule D, Sers C, Morkel M, Horst D, Bluthgen N, Klauschen F (**2021**). Single-cell RNA sequencing reveals distinct tumor microenvironmental patterns in lung adenocarcinoma. **Oncogene** 40(50): 6748-6758. doi: 10.1038/s41388-021-02054-3

40. Xing X, Yang F, Huang Q, Guo H, Li J, Qiu M, Bai F, Wang J (**2021**). Decoding the multicellular ecosystem of lung adenocarcinoma manifested as pulmonary subsolid nodules by single-cell RNA sequencing. **Sci Adv** 7(5): eabd9738. doi: 10.1126/sciadv.abd9738

41. Yang M, Lin C, Wang Y, Chen K, Zhang H, Li W (**2022**). Identification of a cytokine-dominated immunosuppressive class in squamous cell lung carcinoma with implications for immunotherapy resistance. **Genome Med** 14(1): 72. doi: 10.1186/s13073-022-01079-x

42. Shinohara S, Takahashi Y, Komuro H, Matsui T, Sugita Y, Demachi-Okamura A, Muraoka D, Takahara H, Nakada T, Sakakura N, Masago K, Miyai M, Nishida R, Shomura S, Shigematsu Y, Hatooka S, Sasano H, Watanabe F, Adachi K, Fujinaga K, Kaneda S, Takao M, Ohtsuka T, Yamaguchi R, Kuroda H, Matsushita H (**2022**). New evaluation of the tumor immune microenvironment of non-small cell lung cancer and its association with prognosis. J Immunother Cancer 10(4): e003765. doi: 10.1136/jitc-2021-003765

43. Taniguchi S, Matsui T, Kimura K, Funaki S, Miyamoto Y, Uchida Y, Sudo T, Kikuta J, Hara T, Motooka D, Liu YC, Okuzaki D, Morii E, Emoto N, Shintani Y, Ishii M (**2023**). In vivo induction of activin A-producing alveolar macrophages supports the progression of lung cell carcinoma. **Nat Commun** 14(1): 143. doi: 10.1038/s41467-022-35701-8

44. Hao D, Han G, Sinjab A, Gomez-Bolanos LI, Lazcano R, Serrano A, Hernandez SD, Dai E, Cao X, Hu J, Dang M, Wang R, Chu Y, Song X, Zhang J, Parra ER, Wargo JA, Swisher SG, Cascone T, Sepesi B, Futreal AP, Li M, Dubinett SM, Fujimoto J, Solis Soto LM, Wistuba, II, Stevenson CS, Spira A, Shalapour S, Kadara H, et al. (2022). The Single-Cell Immunogenomic Landscape of B and Plasma Cells in Early-Stage Lung Adenocarcinoma. **Cancer Discov** 12(11): 2626-2645. doi: 10.1158/2159-8290.CD-21-1658

45. Li N, Zhou H, Holden VK, Deepak J, Dhilipkannah P, Todd NW, Stass SA, Jiang F (**2023**). Streptococcus pneumoniae promotes lung cancer development and progression. **iScience** 26(2): 105923. doi: 10.1016/j.isci.2022.105923

46. Chow SC, Gowing SD, Cools-Lartigue JJ, Chen CB, Berube J, Yoon HW, Chan CH, Rousseau MC, Bourdeau F, Giannias B, Roussel L, Qureshi ST, Rousseau S, Ferri LE (**2015**). Gram negative bacteria increase non-small cell lung cancer metastasis via Toll-like receptor 4 activation and mitogen-activated protein kinase phosphorylation. Int J Cancer 136(6): 1341-1350. doi: 10.1002/ijc.29111

47. Greathouse KL, White JR, Vargas AJ, Bliskovsky VV, Beck JA, von Muhlinen N, Polley EC, Bowman ED, Khan MA, Robles AI, Cooks T, Ryan BM, Padgett N, Dzutsev AH, Trinchieri G, Pineda MA, Bilke S, Meltzer PS, Hokenstad AN, Stickrod TM, Walther-Antonio MR, Earl JP, Mell JC, Krol JE, Balashov SV, Bhat AS, Ehrlich GD, Valm A, Deming C, Conlan S, et al. (**2018**). Interaction between the microbiome and TP53 in human lung cancer. **Genome Biol** 19(1): 123. doi: 10.1186/s13059-018-1501-6

48. Hu C, Li T, Xu Y, Zhang X, Li F, Bai J, Chen J, Jiang W, Yang K, Ou Q, Li X, Wang P, Zhang Y (**2023**). CellMarker 2.0: an updated database of manually curated cell markers in human/mouse and web tools based on scRNA-seq data. **Nucleic Acids Res** 51(D1): D870-D876. doi: 10.1093/nar/gkac947

49. 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. **Nat Methods** 12(5): 453-457. doi: 10.1038/nmeth.3337