Identification and Validation of a Novel Prognostic Model of Inflammation-related Gene Signature of Lung Adenocarcinoma

Dayuan Luo
Central South University
Wei Feng
Central South University
Yunqian Ma
Central South University
Zhibin Jiang (506359153@qq.com)
Central South University

Research Article

Keywords:

Posted Date: February 15th, 2022

DOI: https://doi.org/10.21203/rs.3.rs-1323401/v1

Abstract

Previous literatures have suggested the importance of inflammatory response during lung adenocarcinoma (LUAD) development. This study aimed at exploring the inflammation-related genes and developing a prognostic signature for predicting the prognosis of LUAD. Survival-associated inflammation-related genes were identified by univariate Cox regression analysis in the dataset of The Cancer Genome Atlas (TCGA). The least absolute shrinkage and selection operator (LASSO) penalized COX regression model was used to derive a risk signature and divide samples into high-, medium- and low- risk group. Univariate and multivariate analyses suggested that the level of risk group was an independent prognostic factor of survival. Time-dependent receiver operating characteristic (ROC) curve indicated the AUC of 1-, 3- and 5- years of the risk signature was 0.715, 0.719, 0.699 respectively. A prognostic nomogram was constructed by integrating risk group and clinical features. The independent dataset GSE30129 of Gene Expression Omnibus (GEO) was used for verification. Furthermore, we performed Gene set enrichment analysis(GSEA) and discussed the differences in tumor mutation, tumor microenvironment among risk groups. Single Sample Gene Set Enrichment Analysis (ssGSEA) and CIBERSORT results suggested the status of immune cell infiltration was highly associated with risk groups. In a cohort of LUAD from The Cancer Immunome Atlas(TCIA) that predicted immune responses to CTLA-4 and PD-1/PD-L1 inhibitors, immunotherapy performed better in the low-risk group. Partial targeted drugs and chemotherapy drugs for lung cancer had higher drug sensitivity in the high-risk group. The level of risk group can provide some reference significance for the selection of immunotherapy and drug regimen for LUAD patients.

Introduction

Lung cancer has the highest mortality among 36 types of cancers and is the second most common cancer worldwide, which is divided into small cell lung cancer (SCLC) and non-small cell lung cancer (NSCLC) according to cell origin[1]. Adenocarcinoma is the most common type of all newly diagnosed NSCLC cases[2]. The inflammatory response of lung adenocarcinoma runs through all stages of lung tumor development and has an inseparable relationship with its prognosis[3]. Chronic inflammation may lead to tumorigenesis by inducing gene mutations, cell proliferation, anti-apoptosis, enhanced invasion, promotion of angiogenesis and secretion of immunosuppressive factors. Inflammatory cytokines and chemokines secreted during lung inflammation, such as tumor necrosis factor (TNF), transforming growth factor-β (TGF-β), Interleukin-1 (IL-1) and cyclooxygenase-2 (COX-2) have been shown to be related to the occurrence and development of lung cancer[4–6]. In addition, immune and microenvironmental disorders associated with inflammatory responses are also thought to be contributing factors to cancer. Tumor microenvironment (TME), which is composed of tumor cells, immune cells, stromal cells, inflammatory mediators and extracellular matrix, participates in the proliferation, drug resistance and metastatic growth of tumor cells together with immune disorders[7–8].

In view of the close relationship among inflammatory response, immune and tumor microenvironment, inflammation is an important part of lung tumor research. At present, there are many basic studies on lung tumor inflammation, but few clinical translational studies, especially few studies on inflammation-related genes as prognostic indicators. In this study, we analyzed 313 samples of LUAD from the TCGA and identified 54 prognostic inflammation-related gene. We constructed an accurate 8-gene predictive model and identified a prognostic risk signature. We explored the clinical efficacy of this model in LUAD patients, and comprehensively discussed the relationship between risk signature and tumor microenvironment, immune cell infiltration, tumor mutation, immunotherapy and drug sensitivity.

Result

Prognostic model of inflammation-related genes. The flow chart of this study was presented (Supple Fig. 1). 551 samples (54 none-tumor samples and 497 tumor samples) of LUAD were downloaded from TCGA. Excluding none-tumor samples and samples with incomplete information of survival time and survival status, 464 eligible samples were obtained. 254 inflammation-related genes were obtained from Molecular Signatures Database(MSigDB). R-package "survival" was used for univariate COX analysis to obtain the prognostic genes. As shown in the forest map (Fig. 1A), 54 inflammation-related genes associated with OS were obtained. In order to avoid the possibility of collinearity between gene expression levels, LASSO regression is used to perform variable screening and re-sampling 1000 times for these 54 genes, and selected the genes with more than 900 repetitions. The figure shows that the optimal number of variables is equal to 15(Figure. 1B-D). We further conducted multivariate Cox regression analysis on these 15 genes by stepwise method, and finally included 8 genes to construct a COX proportional hazard model (Figure. 1E).

Validation of risk groups and genetic prognostic models. Regression coefficients of eight genes included in the prognostic model are shown (supple table. 1). The risk score of each sample is calculated as follows formula: Risk Score= 0.1241*CCL20 expression - 0.6192*CCR2 expression + 0.3660*GNAI3 expression + 0.1240*ITGA5 expression + 0.2050*NMI expression + 0.1330*PCDH7 expression + 0.430*PSEN1 expression - 0.3382*SLC11A2 expression. The two best cut-off values of risk score are equal to 1.05, 2.01 respectively, the tumor samples were divided into high-, medium- and low-risk group (Supple Fig. 2). A total of 289 tumor samples of GSE30129 from GEO were taken as the validation group, risk score and risk group were defined as the same way, the two best cutoff values of risk score are equal to 11.28 and 15.7(Supple Fig. 3). Survival analysis based on R-package “survival” showed that the overall survival differences were statistically significant among the three risk groups both in the two data sets(p-value<0.001). (Fig. 2A,E). R-package “timeROC” was used to draw time-dependent ROC curves. AUC at 1-, 3- and 5- years was equal to 0.715, 0.719, 0.699 in the cohort of TCGA(Fig. 2B), and 0.608, 0.631, 0.639 in GSE30129, which shows that the model has good predictive ability(Fig. 2F). Principal component analysis(PCA) and uniform manifold approximation and projection analysis(UMAP) indicate that all samples can be well clustered as risk groups(Fig. 2C,D,G,H). Univariate and multivariate analysis suggested that risk group can serve as an independent prognostic signature of LUAD both in TCGA cohort and GSE30129(Fig. 2L,Q,M,R). The survival state curve shows most of the survival cases are

concentrated in the low-risk group, while the death cases tend to be concentrated in the high-risk group, indicates there is a significant correlation between the risk score and the survival status of the patients, (Fig. 2I,J,N,O).

Clinical correlation of the prognostic index. The clinical baseline characteristics of TCGA cohort (Supple Table 2) and GSE30129 (Supple Table 3) are shown. The risk score was significantly correlated with the clinical features such as tumor stage (*p*-value < 0.001), T (*p*-value < 0.001) and N (*p*-value < 0.01) (Fig. 3A), the proportion of samples of later stage, T and N tend to be higher as the level of risk group raise up(Fig. 3B-D). The decision curve analysis shows that the net benefit rate of risk score is the highest at each threshold probability compared with other clinical characteristics, suggesting that it have practical application value (Fig. 3I). The ROC curve constructed by clinical characteristics showed that the AUC of risk score and tumor stage in predicting 1-year survival was 0.715 and 0716, which were close to each other. The AUC of 3-year and 5-year survival predicted by risk score was 0.719 and 0.699, higher than 0.688 and 0.644 of 3-year and 5-year survival predicted by tumor stage(Fig. 3E-G). The C-index of risk score was higher than that of other clinical characteristics in univariate regression model with bootstrap re-sampling (1000 times) method(Fig. 3H).

Construct and verify the prognostic nomogram. We integrated signatures including risk groups, T, M, N and age into the COX proportional hazard model to construct a nomogram(Fig. 4A). The TCGA data set and GSE30129 were included in 313 cases and 280 cases with complete clinical information respectively. The C-index of the nomogram in the TCGA cohort and GSE30129 are equal to 0.745 and 0.692 respectively. According to the calibration curves(Fig. 4B-C), the prediction probability of this nomogram is very close to the observation probability after 1000 simulations of the 1-, 3- and 5-years calibration curves drawn by bootstrap re-sampling method in the TCGA cohort and GSE30629.

Gene set enrichment analysis. Gene set enrichment analysis(GSEA) was carried out on the high- and low-risk group in TCGA by using GSEA software(version 4.1.0). In terms of KEGG pathways, the high-risk group mainly focused on cell cycle regulation, DNA replication, homologous chromosome, p53 signaling pathway, pentose phosphate pathway, etc. In the low-risk group, the main pathways were enriched in T and B cell receptor signal pathway, cell adhesion signal pathway, chemokine signal pathway, etc. In terms of GO functions, the high-risk group mainly focused on mitosis, epithelial polarization, protein regulation, desmosome, cadherin, etc. The low-risk group focused on adaptive immune response, B cell receptor signal transduction, lipid metabolism, immune-related T cell activity, T cell selectivity, cytokine receptor activity, immune checkpoint activity (Fig. 5), etc.

Tumor microenvironment and tumor mutation burden. Tumor microenvironment was quantified by score of stromal cells (stromal score) and score of immune cells (immune score). Sum of immune score and stromal score equals the tumor microenvironment (estimate score).Immune score, stromal score and estimate score were negatively correlated to risk score. Tumor purity was positively correlated with risk score (Fig. 6). Scatter plots of correlation coefficients showed that there was a positive correlation between tumor mutation burden (TMB) and risk score(Supple Fig. 4E). In the waterfall plot, the proportion

of samples with mutations in the high-risk group was significantly higher than that in the low-risk group, and missense mutations were the most common type of mutation. The most frequently mutated gene was TP53, which was also the gene with the most significant difference in mutation ratio among the high- and low- risk group, followed by TTN(Supple Fig. 4A-C). TP53 gene and KRAS gene have mutually exclusive mutation relationship(Supple Fig. 4D).

Correlation analysis between prognostic model and immune. In order to further explore the correlation between risk score and immune status, 16 kinds of immune cells and related immune functions in all samples were scored by ssGSEA. Most of the estimated infiltration abundances of immune cells in the high-risk group were down-regulated with statistical significance (*p*-value<0.05) (Fig. 7A), including B cells、dendritic cells (DCs)、immature dendritic cells (iDCs)、activated dendritic cells (aDCs)、plasmacytoid dendritic cells (pDCs)、CD8+ T cells、macrophages、mast cells、neutrophils、T helper cells、T follicular helper cell (Tfh)、Type 1 T help Cells (Th1 cells)、tumor infiltrating lymphocyte (TIL)、regulatory T cells (Treg), both risk score and TMB were negatively correlated with abundance of multiple types of immune cells (Fig. 7D). In terms of immune function, the activity scores of high-risk group were significantly lower in check-point、cytolytic activity、Human lymphocyte antigen (HLA)、inflammation-promoting、T cell co-inhibition、T cell co-stimulation、type II IFN response(Fig. 7C). CIBERSORT method was used to calculate the relative content of 22 types of immune cells in each sample. Similar to the result of ssGSEA, the relative content of B memory cells, resting CD4 T memory cells, monocytes, M0 macrophages, resting dendritic cells, mast cells were higher in the samples in the low-risk group (Supple Fig. 5). Correlation analysis showed that the expression values of most immune checkpoint related genes were up-regulated in the low-risk group, including PD-CD1 and BTLA, while TNFSF9 and CD276 were down-regulated in the low-risk group(*p*-value<0.05)(Fig. 7B). Previous studies have summarized the immune infiltration patterns of various cancer types, which can be defined as six types, including: C1(wound healing), C2 (INF-γ dominant), C3 (inflammatory), C4(lymphocyte depleted), C5 (immunologically quiet), C6(TGF-β dominant). In this study, the ensemble average of risk score of the sample with type C1 (WoundHealing) was the highest, and that of the sample with type C3 (Inflammatory) was the lowest(Fig. 7E).

Immunotherapy effect and drug sensitivity analysis. There was no statistical difference in IPS between the high-risk and low-risk groups among samples that predicted a negative response in both immunotherapy regimens(Fig. 8A). Among tumor samples that were predicted to have a positive immune response to a single regimen of PD-1/PD-L1 (Fig. 8B) or CTLA4 inhibitors (Fig. 8C) and a positive immune response to both regimens (Fig. 8D), the ensemble average of immunophenoscore (IPS) in the low-risk group were higher than those in the high-risk group, and the differences were statistically significant. The IC50 values of commonly used LUAD targeted drugs and chemotherapy drugs were calculated to analyze the drug sensitivity of 464 samples from TCGA. The IC50 values of cisplatin (Fig. 8E), docetaxel (Fig. 8F), erlotinib (Fig. 8G) and gefitinib (Fig. 8H) were significantly increased in the low-risk group in TCGA, suggesting higher sensitivity of patients in the high-risk group to use these drugs.

Discussion

In this study, we obtained inflammatory genes associated with prognosis by univariate COX analysis, and constructed an 8-gene predictive model by LASSO-penalized COX regression. A prognostic risk signature was generated based on eight inflammation-related genes, which can divide the samples into high-, medium- and low-risk group. The risk group could serve as an independent predictor for lung cancer prognosis, and participate in the construction of nomogram, which was successfully validated in an independent GEO dataset, indicating the general applicability of this signature. Among these eight genes, CCL20 is expressed by macrophages, T cells and B cells and is responsible for the chemotactic attraction of immature dendritic cells, effector/memory T cells and B cells. It is up-regulated in tumor tissues and negatively correlated with the prognosis of lung cancer, thus it may be a potential therapeutic target[9]. CCR2 is one of the four chemokine receptors expressed by monocytes, it is related to the migration of inflammatory sites and plays a leading role in promoting the recruitment of monocytes with tumorigenic and metastatic activity[10–11]. ITGA5 is a member of the integrin α chain family, which participates in a variety of biological functions, including cell proliferation, differentiation, adhesion, survival and apoptosis[12]. It may also be involved in the PI3K/Akt signaling pathway that mediates inflammation, apoptosis and reactive oxygen species production of pulmonary endothelial cells[13–15]. PCDH7 is a cell surface receptor protein that is highly expressed in NSCLC, inducing cell transformation and promoting tumor growth in vitro and in vivo. The MAPK pathway plays a central role in PCDH7-mediated tumorigenesis[16]. In addition, the deletion of PCDH7 has been shown to increase the sensitivity of mutant lung cancer cells with KRAS mutation to Mek and Erk inhibitors, thus the treatment of PCDH7 may have a synergistic effect with EGFR or MAPK inhibitors[17]. NMI is an important component of a transcription factor complex that allows for the continuous activation of telomerase in breast and ovarian cancer, and participates in the regulation of bradykinin BDKRB2 and MAPK/ERK pathways, thus mediating tumor progression and metastasis[18]. GNAI3 is a potential tumor suppressor, which inhibits GNAI2-mediated MDSC amplification and Colitis-Associated tumorigenesis by negatively regulating IL6 signaling pathway[19]. SLC11A2, a member of the SLC11 family, also known as DMT1, has been shown to control the iron pool in the cytoplasm, thus affecting the step of drug activation or the level of hydroxyl radicals in cells[20–21]. PSEN1 is a recognized gene associated with Alzheimer's disease, which can induce the intramembrane division of Notch receptors and then activate Notch signaling pathways. A study shows that its abnormal expression may activate downstream Notch1 signaling pathways and be related to the development of lung cancer[22–24].

The mRNAs associated with high-risk group was enriched in common tumor-related pathways and functions, such as cell cycle, DNA replication, chromosome recombination, and cell desmosomes. While the mRNAs associated with low-risk group was enriched in T and B cell receptor signal pathways, chemokine signal pathways and so on, it also enriched in immune and inflammatory functions such as immune response, cytokine activity, regulation of T and B cells etc. The results suggest that low-risk group samples may be more active in immune and inflammation-related biological processes. Finally, the high-risk group had higher tumor purity but lower stroma score and immune score, consistent with our previous analysis.

Immune cell infiltration was an important regulator of tumor progression, our study further found that there was a significant correlation between risk score and immune status. The results of ssGSEA showed that the estimated infiltration abundances of most immune cells and activity scores of immune functions were higher in the low-risk group than in the high-risk group. The relative content of each type of immune cell in a single sample obtained using CIBERSORT method also showed a higher degree of immune infiltration in the low-risk group. The results suggest that the low-risk score may represent a more immunoreactive microenvironment and the high-risk score represents an immunosuppressed microenvironment, which may be one of the reasons for the poor prognosis of patients in the high-risk group. Previous literature suggest that inflammatory mediators and cytokines produced in the process of inflammatory response can establish gradients to recruit or reject immune cell subsets, and ultimately promote the formation of immunosuppressive microenvironment and immune escape, thus accelerating tumor metastasis[25–26]. Therefore, some studies have pointed out that the evaluation of tumor immune cell infiltration in tumor microenvironment suggests that "hot" or inflammatory tumors tend to have a better therapeutic response[27]. In addition, according to the six pan-cancer immunophenotypes summarized in previous studies, type C3 (inflammatory) is characterized by immune control and immune balance, and has the best prognosis among the six types, while type C1 (wound healing) is characterized by increased expression of angiogenic genes and high tumor cell proliferation rate, suggesting a poor prognosis. In this study, we found that the risk score of type C3 was the lowest, while that of type C1 was the highest, which was consistent with the difference in prognosis predicted by our model.

Tumor mutation burden (TMB) is defined as the total number of somatic / acquired mutations in each coding region of the tumor genome (Mut/Mb)[28]. In this study, risk scores were positively correlated with TMB, with a higher proportion of mutation events occurring in tumor samples in the high-risk group. Most of the mutant genes are mainly missense mutations, which may produce new antigens recognized by the host immune system and lead to anti-tumor immune response. tumors with high mutation burden may produce more new antigens, thus increasing the possibility of immune recognition and tumor cell killing[29]. However, due to the instability and heterogeneity of TMB in tissues and the lack of unified standard detection methods, the effect of immunotherapy could not be accurately predicted[30–32].

We found significant differences in the expression of multiple immune checkpoint genes in different risk groups, suggesting that immunotherapy effect may be associated with risk group, and different immune checkpoint inhibitors may suit to different risk groups. We further analyzed data from TCIA for evaluating LUAD immune response. The analysis showed that the ensemble averages of IPS in the low-risk group with positive immunoresponse to PD-L1/PD-1 inhibitors or CTLA-4 inhibitors, as well as both PD-L1/PD-1 inhibitors and CTLA-4 inhibitors, was higher than that in the high-risk group. Patients in the low-risk group may have a better immune response to immunotherapy regiments, and the risk group may have the potential to screen LUAD patients who are immunogenic and more responsive to ICIs. In addition, current studies have shown that inflammation may be associated with tumor multidrug resistance, and the common mechanisms of tumor drug resistance, such as the overexpression of membrane-anchored MDR transporters, can be directly affected by inflammation and inflammatory mediators[33]. Taking a single

gene as an example, previous studies have shown that the increased expression of inflammation-related gene COX-2 is related to the level of MDR protein, which can strongly interfere with the results of chemotherapy in cancer patient, suggesting that inflammatory genes may be used as a predictor of tumor treatment efficacy[34]. In this study, tyrosine kinase inhibitors (TKIs) with epidermal growth factor receptor (EGFR) as target genes, such as gefitinib and erlotinib, and commonly used chemotherapy drugs, such as cisplatin and docetaxel, were more sensitive in high-risk group of patients. This result has certain reference significance for anti-tumor drug strategy, but it still needs to be further verified in prospective studies with large samples. Exploring the relationship between inflammation, immunity and tumor drug resistance may become a new direction for anti-tumor treatment population screening and regimen selection.

Conclusion

In conclusion, our study elaborated the prognostic significance of important inflammation-related genes in patients with lung adenocarcinoma, and constructed a prognostic model composed of eight genes, which can accurately predict the prognosis of patients with lung adenocarcinoma. This study will help us further understand the role of inflammation-related genes in influencing cellular pathways, immune cell infiltration, immune checkpoint gene expression, tumor mutation, tumor microenvironment and antitumor drug selection. This study is helpful to guide more effective immunotherapy strategies for lung cancer.

Materials And Method

Download gene expression data and clinical information. RNA-seq gene expression profiles of LUAD patients were downloaded from the TCGA (https://portal.gdc.cancer.gov/), including fragments per kilobase per million(FPKM). The workflow type is HTSeq-FPKM. R-package “limma” was used to average the repeated data of RNA expression profile. Clinical data of LUAD samples were retrieved and downloaded from TCGA database, including gender, age, tumor stage and survival information. Tumor samples were distinguished from normal samples by TCGA ID. In addition, GSE30129 was downloaded from GEO (https://www.ncbi.nlm.nih.gov/geo/)for verification, the download content includes the soft matrix file of RNA-seq profile and the annotated file “GPL570”. Inflammation-related genes were obtained from MSigDB on GSEA website (https://www.gsea-msigdb.org/gsea/msigdb/index.jsp). Data related to PD-1 /PD-L1 and CTLA-4 immunotherapy in LUAD cohort were obtained from TCIA (https://tcia.at/home). All data obtained from TCGA,GEO,GSEA, MSigDB and TCIA are public, the involved human data in these databases adhered to relevant ethical guidelines. Our study based on public source and the acquisition process follows the rules and guidelines of the official website, so there are no ethical issues and other conflicts of interests.

Obtain prognostic inflammation-related genes and construct genetic prognostic model. Univariate Cox regression analysis was used to obtain inflammation-related genes associated with overall survival (OS). The Lasso regression algorithm in R-package "glmnet" was used to establish the penalty coefficient and selection variables, and the ten-fold cross-validation method was used to determine the penalty

coefficient (λ) of the regression model. By applying the penalty coefficient to shrink the regression coefficient of prognosis-related inflammatory genes, most of the independent variable coefficients were reduced to zero, and only a relatively small number of genes with non-zero weight were retained. After obtaining the optimal number of variables, the multivariate Cox regression analysis was further carried out by stepwise forward and backward regression methods to minimize the risk of overfitting and collinearity.

Risk groups and genetic prognostic model validation. The model gene's regression coefficient is multiplied by the gene's expression value and summed up to give each patient's risk score. The formula for calculating the risk score is as follows: Risk Score = (Gene 1 Expression × Coefficient) + (Gene 2 Expression ×Coefficient) +. .. + (Gene n Expression ×Coefficient). "Gene n Expression" represents the expression values of genes included in the regression model, “Coefficient” represents the regression coefficient of the gene. X-tile software is used to obtain the best cut-off value of risk score[35]. The software uses Kaplan-Meier method and log-rank test to analyze and compare the survival differences among risk groups under different cut-off values. When the *p*-value is minimum, the risk score is the best cut-off value, which can divide samples into different risk groups. PCA and UMAP were performed using the function " prcomp " and R-package "UMAP" in R software. R-package "Survminer" was used for survival analysis, Kaplan-Meier method was used to evaluate the survival difference of patients in different risk groups, log-rank test was used for inter-group comparison, the result was considered to be statistically significant when *p*-value < 0.05. The "time ROC" R packet is used to construct a time-dependent receiver operating characteristic (ROC) curve to measure the prediction performance of the model. When the area under the curve (AUC) is greater than 0.6, it is considered to have prediction ability. Univariate and multivariate Cox analyses were used to identify the prognostic significance of risk score and clinical characteristics. Bootstrap re-sampling method was used to calculate the C-index of risk score and clinical characteristics in the univariate regression model to compare the difference in their predictive ability. The risk signature and prognostic model performance were verified using independent data set GSE30129 from GEO.

Construct and verify the prognostic nomogram. The prognostic nomogram was constructed by using the R-package "rms" by integrating risk group and clinical characteristics. Calibration curves and C-index were used to evaluate the nomogram performance. We used bootstrap re-sampling method to re-sample for 1000 times and draw calibration curves to reflect the consistency between the predicted probability and the observed probability in a visual form. When the predicted probability is equal to the observed probability, the calibration curve generated will coincide with the 45° diagonal line emitted along the diagonals in the graph. External validation of Nomogram is performed using the dataset GSE30129 from GEO.

Gene set enrichment analysis. The pathways and functions enrichment analysis for the mRNAs associated with the high or low risk was carried out using c2.cp.kegg.v7.4.symbols.gmt and c5.go.v7.4.symbols.gmt as gene sets database at 1,000 random sample permutations using Gene Set

Enrichment Analysis(GSEA) software (version 4.1).The enrichment functions or pathways were statistically significant when *p*-value<0.05.

Tumor microenvironment, tumor mutation burden, immune infiltration. We quantified the levels of infiltration of immune cells and stromal cells in different risk groups by immune score and stromal score, combined with tumor purity score, to compare differences in tumor microenvironment. To explore the correlation between risk groups and TMB, we analyzed the available somatic mutation data from LUAD cohort in TCGA, which was downloaded in MAF format and was analyzed using R-package “maftools”. We used single-sample gene set enrichment analysis (ssGSEA) to calculate the scores of 16 immune cells and 13 immune-related pathways in each sample using a "GSVA" R-package. CIBERSORT method is an excellent tool for assessing immune cell infiltration and can be used to assess the relative abundances of 22 types of immune cells in single sample. These two methods were combined to compare the differences in immune cell infiltration degree or functional activity between different risk groups. Furthermore, we compared the difference in the ensemble average of risk score between different immunophenotypes, which were identified by previous literature[36]. The immunophenotyping data of pan-cancer samples were obtained from the NCI Genomic Data Commons official website (GDC, https://portal.gdc.cancer.gov), and were intermixed with LUAD samples of TCGA. According to the immune checkpoint-related genes summarized in previous literature, the differences in the ensemble average of expression values of immune checkpoint-related genes among different risk groups were compared. Analysis of variance (ANOVA)was used for comparison between different groups. Spearman analysis was used for correlation analysis(the correlation between TMB and risk score as an example). The results considered to have statistical significance when *p*-value<0.05.

Immunotherapy effect and drug sensitivity. Immunophenoscore (IPS) is a score calculated quantitatively based on MHC molecules, immune regulators, effector cells and suppressor cells. It is a good predictor of the response of immune checkpoint inhibitors (ICIs). IPS scores were calculated based on representative cell type gene expression, with scores ranging from 0 to 10. The IPS of each LUAD patient was obtained from TCIA (https://tcia.at/home)[37–38]. The "pRRophetic" algorithm in R software was used to estimate the IC50 value of commonly used targeted drugs and chemotherapy drugs for lung cancer, and the drug sensitivity differences among different risk groups were compared. ANOVA is the statistical method for comparison between different risk groups, the results considered to have statistical significance when *p*-value<0.05.

Declarations

Acknowledgements

The authors express sincere gratitude to the TCGA, GEO, GSEA and TCIA databases for the availability of data.

Author contributions

DY.L. collected the data and performed the analyses. DY.L. and W.F. validated and interpreted the results. DY.L., ZB.J. and W.F. conceived the project and designed the work flow, YQ.M. translated and edited the article. The manuscript was written with contributions from all authors.

Competing interests

The authors declare that they have no conflicts of interest.

Funding

This work was supported by Natural Science Foundation of Hunan Province (No. 2021JJ40919), Hunan Science and Health Union foundation (No. 2018JJ6053). The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

References

1. Sung, H. *et al.* Global Cancer Statistics 2020: GLOBOCAN Estimates of Incidence and Mortality Worldwide for 36 Cancers in 185 Countries. *CA: a cancer journal for clinicians* **71**, 209-249, doi:10.3322/caac.21660 (2021).
2. Travis, W. D. *et al.* The 2015 World Health Organization Classification of Lung Tumors: Impact of Genetic, Clinical and Radiologic Advances Since the 2004 Classification. *J Thorac Oncol* **10**, 1243-1260, doi:10.1097/jto.0000000000000630 (2015).
3. Budisan, L. *et al.* Links between Infections, Lung Cancer, and the Immune System. *International journal of molecular sciences* **22**, doi:10.3390/ijms22179394 (2021).
4. Gong, K. *et al.* Tumor necrosis factor in lung cancer: Complex roles in biology and resistance to treatment. *Neoplasia (New York, N.Y.)* **23**, 189-196, doi:10.1016/j.neo.2020.12.006 (2021).
5. Ding, K., Yi, M., Li, L. & Zhang, Y. Interleukin polymorphisms and protein levels associated with lung cancer susceptibility and phenotypes. *Expert review of clinical immunology* **17**, 1029-1040, doi:10.1080/1744666x.2021.1952072 (2021).
6. Li, J. *et al.* COX-2 rs5275 and rs689466 polymorphism and risk of lung cancer: A PRISMA-compliant meta-analysis. *Medicine* **97**, e11859, doi:10.1097/md.0000000000011859 (2018).
7. Son, B. *et al.* The role of tumor microenvironment in therapeutic resistance. *Oncotarget* **8**, 3933-3945, doi:10.18632/oncotarget.13907 (2017).
8. Giannone, G. *et al.* Immuno-Metabolism and Microenvironment in Cancer: Key Players for Immunotherapy. *International journal of molecular sciences* **21**, doi:10.3390/ijms21124414 (2020).
9. Wang, G. Z. *et al.* Tobacco smoke induces production of chemokine CCL20 to promote lung cancer. *Cancer letters* **363**, 60-70, doi:10.1016/j.canlet.2015.04.005 (2015).
10. Marcuzzi, E., Angioni, R., Molon, B. & Calì, B. Chemokines and Chemokine Receptors: Orchestrating Tumor Metastasization. *International journal of molecular sciences* **20**, doi:10.3390/ijms20010096 (2018).

11. Hao, Q., Vadgama, J. V. & Wang, P. CCL2/CCR2 signaling in cancer pathogenesis. *Cell communication and signaling : CCS* **18**, 82, doi:10.1186/s12964-020-00589-8 (2020).
12. De Marco, R., Tolomelli, A., Juaristi, E. & Gentilucci, L. Integrin Ligands with α/β-Hybrid Peptide Structure: Design, Bioactivity, and Conformational Aspects. *Medicinal research reviews* **36**, 389-424, doi:10.1002/med.21383 (2016).
13. Larzabal, L. *et al.* TMPRSS4 regulates levels of integrin α5 in NSCLC through miR-205 activity to promote metastasis. *British journal of cancer* **110**, 764-774, doi:10.1038/bjc.2013.761 (2014).
14. Jiang, H. H. *et al.* Cell wall mannoprotein of Candida albicans polarizes macrophages and affects proliferation and apoptosis through activation of the Akt signal pathway. *International immunopharmacology* **72**, 308-321, doi:10.1016/j.intimp.2019.03.032 (2019).
15. Xu, F. & Zhou, F. Inhibition of microRNA-92a ameliorates lipopolysaccharide-induced endothelial barrier dysfunction by targeting ITGA5 through the PI3K/Akt signaling pathway in human pulmonary microvascular endothelial cells. *International immunopharmacology* **78**, 106060, doi:10.1016/j.intimp.2019.106060 (2020).
16. Zhou, X. *et al.* Modulation of Mutant Kras(G12D) -Driven Lung Tumorigenesis In Vivo by Gain or Loss of PCDH7 Function. *Molecular cancer research : MCR* **17**, 594-603, doi:10.1158/1541-7786.mcr-18-0739 (2019).
17. Zhou, X. *et al.* PROTOCADHERIN 7 Acts through SET and PP2A to Potentiate MAPK Signaling by EGFR and KRAS during Lung Tumorigenesis. *Cancer Res* **77**, 187-197, doi:10.1158/0008-5472.can-16-1267-t (2017).
18. Zhao, J. *et al.* NMI promotes hepatocellular carcinoma progression via BDKRB2 and MAPK/ERK pathway. *Oncotarget* **8**, 12174-12185, doi:10.18632/oncotarget.14556 (2017).
19. Li, Z. W. *et al.* GNAI1 and GNAI3 Reduce Colitis-Associated Tumorigenesis in Mice by Blocking IL6 Signaling and Down-regulating Expression of GNAI2. *Gastroenterology* **156**, 2297-2312, doi:10.1053/j.gastro.2019.02.040 (2019).
20. Montalbetti, N., Simonin, A., Kovacs, G. & Hediger, M. A. Mammalian iron transporters: families SLC11 and SLC40. *Molecular aspects of medicine* **34**, 270-287, doi:10.1016/j.mam.2013.01.002 (2013).
21. Girardi, E. *et al.* A widespread role for SLC transmembrane transporters in resistance to cytotoxic drugs. *Nature chemical biology* **16**, 469-478, doi:10.1038/s41589-020-0483-3 (2020).
22. R, A. A. Risk factors for Alzheimer's disease. *Folia neuropathologica* **57**, 87-105, doi:10.5114/fn.2019.85929 (2019).
23. Azimi, M., Le, T. T. & Brown, N. L. Presenilin gene function and Notch signaling feedback regulation in the developing mouse lens. *Differentiation; research in biological diversity* **102**, 40-52, doi:10.1016/j.diff.2018.07.003 (2018).
24. Guo, H. *et al.* A novel lncRNA, loc107985872, promotes lung adenocarcinoma progression via the notch1 signaling pathway with exposure to traffic-originated PM2.5 organic extract. *Environmental pollution (Barking, Essex : 1987)* **266**, 115307, doi:10.1016/j.envpol.2020.115307 (2020).

25. Mukaida, N. & Baba, T. Chemokines in tumor development and progression. *Experimental cell research* **318**, 95-102, doi:10.1016/j.yexcr.2011.10.012 (2012).
26. Wood, S. L., Pernemalm, M., Crosbie, P. A. & Whetton, A. D. The role of the tumor-microenvironment in lung cancer-metastasis and its relationship to potential therapeutic targets. *Cancer treatment reviews* **40**, 558-566, doi:10.1016/j.ctrv.2013.10.001 (2014).
27. Thommen, D. S. *et al.* A transcriptionally and functionally distinct PD-1(+) CD8(+) T cell pool with predictive potential in non-small-cell lung cancer treated with PD-1 blockade. *Nature medicine* **24**, 994-1004, doi:10.1038/s41591-018-0057-z (2018).
28. Sha, D. *et al.* Tumor Mutational Burden as a Predictive Biomarker in Solid Tumors. *Cancer discovery* **10**, 1808-1825, doi:10.1158/2159-8290.cd-20-0522 (2020).
29. Sholl, L. M. *et al.* The Promises and Challenges of Tumor Mutation Burden as an Immunotherapy Biomarker: A Perspective from the International Association for the Study of Lung Cancer Pathology Committee. *J Thorac Oncol* **15**, 1409-1424, doi:10.1016/j.jtho.2020.05.019 (2020).
30. Büttner, R. *et al.* Implementing TMB measurement in clinical practice: considerations on assay requirements. *ESMO open* **4**, e000442, doi:10.1136/esmoopen-2018-000442 (2019).
31. Chan, T. A. *et al.* Development of tumor mutation burden as an immunotherapy biomarker: utility for the oncology clinic. *Ann Oncol* **30**, 44-56, doi:10.1093/annonc/mdy495 (2019).
32. Bravaccini, S., Bronte, G. & Ulivi, P. TMB in NSCLC: A Broken Dream? *International journal of molecular sciences* **22**, doi:10.3390/ijms22126536 (2021).
33. Mirzaei, S. A., Dinmohammadi, F., Alizadeh, A. & Elahian, F. Inflammatory pathway interactions and cancer multidrug resistance regulation. *Life sciences* **235**, 116825, doi:10.1016/j.lfs.2019.116825 (2019).
34. Yu, T., Lao, X. & Zheng, H. Influencing COX-2 Activity by COX Related Pathways in Inflammation and Cancer. *Mini reviews in medicinal chemistry* **16**, 1230-1243, doi:10.2174/1389557516666160505115743 (2016).
35. Camp, R. L., Dolled-Filhart, M. & Rimm, D. L. X-tile: a new bio-informatics tool for biomarker assessment and outcome-based cut-point optimization. *Clinical cancer research : an official journal of the American Association for Cancer Research* **10**, 7252-7259, doi:10.1158/1078-0432.ccr-04-0713 (2004).
36. Thorsson, V. *et al.* The Immune Landscape of Cancer. *Immunity* **48**, 812-830.e814, doi:10.1016/j.immuni.2018.03.023 (2018).
37. Charoentong, P. *et al.* Pan-cancer Immunogenomic Analyses Reveal Genotype-Immunophenotype Relationships and Predictors of Response to Checkpoint Blockade. *Cell reports* **18**, 248-262, doi:10.1016/j.celrep.2016.12.019 (2017).
38. Hakimi, A. A. *et al.* An Integrated Metabolic Atlas of Clear Cell Renal Cell Carcinoma. *Cancer cell* **29**, 104-116, doi:10.1016/j.ccell.2015.12.004 (2016).

Figures

Figure 1

Univariate cox, LASSO and multivariate Cox regression analysis for overall survival related inflammation-related genes.

(A) Forest plots showing the prognostic value detection of inflammation-related genes, in which the HRs, corresponding 95% confidence intervals, and *p*-values are displayed. **(B)** LASSO regression analysis was used to calculate the coefficient of inflammation-related genes. **(C)** Fifteen genes were selected as active covariates to determine the prognostic value after ten-fold cross-validation for the LASSO model. **(D)** Forest plots showing the fifteen genes of the LASSO model. **(E)** Forest plots showing the Eight genes which were selected by stepwise forward and backward regression methods for the COX proportional hazard model.

Figure 2

Eight-gene risk signature can predict the prognosis in the TCGA cohort and GSE30129.

(A, E) Kaplan-Meier curves of overall survival probability of risk groups in TCGA **(A)** and GSE30129 **(E)**. **(B, F)** Kaplan-Meier overall survival rate curves for the three groups in TCGA **(B)** and GSE30129 **(F)**. **(C, G)**PCA in the TCGA cohort **(C)** and GSE30129 **(G)**. **(D, H)** Umap analysis in the TCGA cohort **(D)** and GSE30129 **(H)**. **(I, N)** Risk score distribution in the TCGA cohort **(I)** and GSE30129 **(N)**. **(J, O)** Survival time and survival status distribution in the TCGA cohort **(J)** and GSE30129 **(O)**. **(K, P)** Heatmap showing the expression level for eight inflammation-related genes among the risk groups in TCGA **(K)** and GSE30129 **(P)**. **(L, Q)** Univariate Cox regression analysis of the association between clinical characteristics, the risk score, and patient overall survival in the TCGA cohort **(L)** and GSE30129 **(Q)**. **(M, R)** Multivariate Cox regression analysis of the association between clinical characteristics, the risk score, and patient overall survival in the TCGA cohort **(M)** and GSE30129 **(R)**.

Figure 3

Correlations between the risk signature and clinical features.

(A) The heatmap and clinical features of three risk groups. **(B, C , D)** Bar plot of correlation between risk score and N **(B)**, Stage **(C)**, T **(D)** of LUAD. **(E, F, G)** ROC curves for 1-year **(E)**, 3-year **(F)**, 5-year **(G)** survival prediction and clinical characteristics. **(H)**Time-related concordance index of risk score and clinical characteristics. **(I)** The decision curve analysis of risk score and clinical characteristics. * *p*-value < 0.05, ** *p*-value < 0.01, *** *p*-value < 0.001.

Figure 4

Construction and verification of the Nomogram.

(A) Nomogram of TCGA cohorts based on the risk groups and clinical features. **(B)** 1-, 3-, 5-year calibration curve for verification of nomogram in the TCGA cohort. **(C)** 1-, 3-, 5-year calibration curve for verification of nomogram in GSE30129.

Figure 5

Gene Set Enrichment Analysis of the mRNAs associated with the high or low risk group in TCGA.

(A) Top enriched GO functions in the low-risk group. **(B)** Top enriched GO functions in the high-risk group. **(C)** Top enriched KEGG pathways in the high-risk group. **(D)** Top enriched KEGG pathways in the low-risk group. The names of enriched KEGG pathways or GO functions are listed on the right side(*p*-value 0.05).

Figure 6

Correlation analysis between prognostic model and tumor microenvironment.

(A) Correlation analysis between stromal score and risk score. **(B)** Correlation analysis between immune cell score and risk score. **(C, D, E, F)** Differences in stromal score **(C)**, immune scores **(D)**, tumor purity **(E)**, estimate score **(F)** between risk groups. * *p*-value < 0.05, ** *p*-value < 0.01, *** *p*-value < 0.001.

Figure 7

Correlation analysis between prognostic model and immune.

(A) Differences in 16 immune infiltration cells between high- and low-risk groups. **(B)** Differences in gene expression level of immune checkpoint-related genes between high- and low-risk patients. **(C)** Differences in 13 immune-related function between high- and low-risk groups. **(D)** Relationship among estimated infiltration abundances of immune cells, risk score and TMB. **(E)** Differences in risk scores of different immune subtypes. * *p*-value < 0.05, ** *p*-value < 0.01, *** *p*-value < 0.001.

Figure 8

Immunotherapy effect and drug sensitivity analysis.

(A, B, C, D) Differences in Immunophenoscore (IPS) among different risk groups in the four situations: negative immunoresponse to both PD-L1/PD-1 inhibitors and CTLA-4 inhibitors**(A)**; positive immunoresponse to PD-L1/PD-1 inhibitors **(B)**; positive immunoresponse to CTLA-4 inhibitors **(C)**; positive immunoresponse to both PD-L1/PD-1 inhibitors and CTLA-4 inhibitors**(D). (E, F, G, H)** Differences in sensitivity of cisplatin **(E)**, docetaxel **(F)**, erlotinib **(G)**, gefitinib **(H)** between high and low risk groups. The hollow diamond pattern in the violin diagram represents the IPS ensemble average level.

Supplementary Files

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

- SupplementaryFigure1.jpg
- SupplementaryFigure2.jpg
- SupplementaryFigure3.jpg
- SupplementaryFigure4.jpg
- SupplementaryFigure5.jpg
- SupplementaryTables.docx