Using machine learning modeling to explore new immune-related prognostic markers in non-small cell lung cancer

Jiasheng Xu  
The Second Affiliated Hospital of Nanchang University

Kaili Liao  
the second Affiliated Hospital of Nanchang university

Han Nie  
The second affiliated Hospital of Nanchang university

xiaozhong wang (✉ xiaozhongwangncu@163.com)  
the Second Affiliated Hospital of nanchang university

Research

Keywords: Machine learning, model building, immune-related prognostic markers, immunohistochemistry

Posted Date: May 23rd, 2020

DOI: https://doi.org/10.21203/rs.3.rs-30017/v1

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

Version of Record: A version of this preprint was published at Frontiers in Oncology on October 30th, 2020. See the published version at https://doi.org/10.3389/fonc.2020.550002.
Abstract

BACKGROUND

To find new immune-related prognostic markers for non-small cell lung cancer (NSCLC)

METHODS

We found suitable data chip (GSE14814) related to NSCLC in geo database. The non-small cell lung cancer observation (NSCLC-OBS) group was evaluated for immunity, and the NSCLC-OBS were divided into high and low groups for differential gene screening according to the score of immune evaluation. A single factor COX regression analysis was performed to select the genes related to prognosis. A prognostic model was constructed by machine learning, and the Receiver Operating Characteristic (ROC) model was analyzed to test whether the model has a test efficacy for prognosis, and then test the association between the selected prognostic genes and the patient's prognosis. A chip-in-chip non-small cell lung cancer chemotherapy (NSCLC-ACT) sample was used as a validation dataset for the same validation and prognostic analysis of the model. The relative infiltration scores of 24 immune cells in NSCLC-ACT patients were compared with those of high and low risk groups. The coexpression genes of hub genes were obtained by pearson analysis and gene enrichment, function enrichment and protein interaction analysis were carried out and the correlation between prognostic genes and immune checkpoints was further analyzed. The tumor samples of patients with different clinical stages were detected by immunohistochemistry and the expression difference of prognostic genes in tumor tissues of patients with different stages was compared.

RESULTS

By screening, we found that LYN\textsuperscript{C3}, COPG2IT1, HLA.DQA1, TNFRSF17 is closely related to prognosis. After machine learning we found that the immune prognosis model constructed from these 5 genes was ROC analyzed, and the AUC values were greater than 0.9 at three time periods of 1, 3, and 5 years; the total survival period of the low-risk group containing these 5 hubgene was significantly better than that of the high-risk group. The Kaplan–Meier curve showed that the increase of COPG2IT1, HLA.DQA1 expression and the decrease of LYN\textsuperscript{C3}, TNFRSF17 expression were significantly related to the shortening of survival time. The results of prognosis analysis and ROC analysis in ACT samples were consistent with those of OBS groups. Hubgene was most expressed in fibroblasts, but there was no significant difference in immune infiltration in the high and low risk groups in 24 immune cells. The coexpression genes are mainly involved B cell receptor signaling pathway and mainly enriched in biological processes such as apoptotic cell clearance, Intestinal immune network for IgA Production. Prognostic key genes are highly correlated with PDCD1, PDCD1LG2, LAG3, CTLA4 immune checkpoints (p < 0.05). The immunohistochemical results showed that the expression of COPG2IT1 and HLA.DQA1 in stage III
increased significantly and the expression of LYN·C3 and TNFRSF17 in stage III decreased significantly compared with that of stage I. The experimental results are consistent with the previous analysis.

CONCLUSION

LYN·C3·COPG2IT1·HLA.DQA1·TNFRSF17 may be a new immune marker to judge the prognosis of patients with non-small cell lung cancer.

Background

Non-small cell lung cancer is the most common type of lung cancer, accounting for about 80% -85% of the total lung cancer. [1] According to statistics from the American Cancer Society, lung cancer accounted for 26% of all cancer deaths in women and 29% of all male deaths in the united states in 2012[2], lung cancer is still the number one killer of human cancer-related deaths. In the past few decades, with various surgical methods including surgical resection, chemotherapy, radiation therapy and molecular targeted therapy, the survival rate of patients diagnosed with NSCLC has significantly improved [7]. The emergence of gene-targeted therapies such as: EGFR( epidermal growth factor receptor) gene changes and EML4-ALK( spinous cortex microtubule-associated protein-like 4 and anaplastic lymphoma kinase) gene rearrangements had promoted the development of TKI targeted therapy. And tyrosine kinase inhibitors and crizotinib had came to be the important targeted therapy [3–6]. But the current effect is still unsatisfactory, and the 5-year overall survival rate of lung cancer is still less than 15% [8, 9]. In recent years, immunotherapy has been developed and increasingly used for patients with lung cancer, for example: PD-L1 in overexpressed in many tumor cells including lung cancer cells ,and plays an important role in regulating the immune response of tumor cells [10–12]. Currently, there are several clinical trials involving FDA-approved immune checkpoint inhibitors. These trials attack by blocking PD-L1/PD-1 signaling pathway tumor cells expressing PD-L1. However, the current research on prognostic genes related to immune checkpoints is very scarce. Therefore, finding immune prognostic markers in NSCLC is of great significance for the prediction and targeted treatment of NSCLC.

In this study, we processed the NSCLC data set GSE14814 in the geo database. The NSCLC-OBS group was evaluated for immunity, and then divided into high and low risk groups for differential gene analysis and single factor COX regression analysis, Receiver Operating Characteristic (ROC) analysis and survival analysis. LYN, C3, COPG2IT1, HLA.DQA1, TNFRSF17 were closely related to the survival of NSCLC. These five genes were modeled and verified in the NSCLC-ACT group in GSE14814. The results showed that the overall survival and survival rate of the high-risk group was significantly lower than those of the low-risk group. Subsequently, we used immunohistochemistry to detect the expression of these five prognostic genes in tumor tissues of 100 patients with different clinical stages. The results of immunohistochemistry are consistent with the conclusions of the model analysis. Thus, we determined that LYN, C3, COPG2IT1, HLA. DQA1, TNFRSF17 are important immune prognostic markers of NSCLC.
2 Materials And Methods

2.1 Data acquisition and immune score

The gene expression profile data set GSE14814 was obtained from the National Biotechnology Information Center (NCBI) GEO (gene expression Omnibus) database, which contained 62 samples of untreated non-small cell lung cancer patients and 62 non-small cells after chemotherapy samples from patients with cell-cell lung cancer. Samples of 62 untreated patients with non-small cell lung cancer were evaluated for tumor purity scores and immune scores using the Estimate software, and the distribution of immune score was plotted using the ggstatsplot software package in R language. The NSCLC-OBS patients were divided into high and low groups based on the immune scores, using the R language limma package to compare the two groups to obtain differentially expressed genes. Use the "ggplot2" "heatmap" package in the R software package to draw the volcano map and heat map of the difference results.

2.2 Screening of prognostic related genes and GSEA analysis

For the clinical information of patients with differential genes combined, the unidimensional COX regression analysis was used to reduce the dimension for the first time, and then the second dimension reduction was carried out by 1000 minimum depth method in random forest (A method of machine learning, the purpose is to obtain prognostic value. The largest variable has more accurate results than the conventional multi-factor COX hazard ratio model. The model kernel is a non-linear model, which is closer to the biological reality than the linear model kernel of multi-factor COX regression.), To obtain the marker with the highest prognostic value gene. In order to obtain a wider range of random forest explanatory variables, and due to the inaccurate biological fitting of single-factor COX regression, the first dimension reduction result was selected with a P value of 0.1, and the obtained variables were included in the second dimension reduction analysis. In order to obtain a more stable learning result, the tree of the random forest was set to 1000. The result of the learning data also outputs the risk score value of each patient for the next step to show the effect of risk stratification.

Perform GSEA analysis of prognostic marker, and use the GO and KEGG data sets on the GSEA official website for functional enrichment and pathway enrichment. Significant enrichment criteria: the absolute value of NES is greater than 1, and the NOMp value is less than 0.05.

2.3 Prediction model construction and evaluation

According to the previous calculation results of COX, GSEA, and random forest model, the immune prognosis model construction and risk assessment were performed for the patients in the OBS group, and the risk scores were ranked. The survminer package in R language was used to find the best cutting point, and the patients were divided into high and low risk groups for subsequent evaluation of the model efficiency. The expression of five key prognostic genes in the high-risk group and the low-risk group was analyzed and compared, and a heat map was drawn. Based on the results of the high and low risk
grouping, the prognostic difference of the K-M survival curve compared with the high and low risk group is plotted, and then the ROC analysis was performed to verify the above results.

2.4 Verification of prediction models

The constructed prediction model was applied to a sample of NSCLC-ACT patients, score calculations were performed on ACT patients. The expression of five key prognostic genes in the high-risk group and the low-risk group was analyzed and compared, and a heat map was drawn. K-M survival analysis and ROC calculation model accuracy were performed respectively.

2.5 Invasion of Hubgene in immune cells and immune targets

The GSVA package was used to score the relative infiltration of 24 immune cells in NSCLC-ACT patients, and the NSCLC-ACT patients were divided into high and low risk groups according to the risk grouping results of the random forest model. The infiltration of 24 immune cells in high and low risk groups was compared Hubgene co-expressed genes were also calculated based on pearson (screening method $P < 0.05$, sorted by R value from small to large). Using R software (3.6.1) cluster analysis package cluster Profiler to function the hubgene and its co-expressed genes Enrichment and pathway enrichment analysis. Analyze and calculate the significant levels of enrichment of 5 prognostic genes and their co-expressed genes in each signal pathway through a hypergeometric distribution exact test to assess the signal pathways that are significantly affected ($P < 0.05$). The protein database and the cytoscape software were used to analyze the protein interactions of the Hub gene co-expressed genes. In order to further evaluate the relationship between immune factors at the non-cellular level and hubgene, the correlation analysis between hubgene and immune checkpoints in the model was performed, and immune targets that were highly correlated with hubgene were screened according to the pearson correlation coefficient.

2.6 Experimental verification of prognostic genes

100 patients diagnosed with stage I and stage III non-small cell lung cancer were selected from our hospital and were divided into two groups. The selected patients had underwent surgery. The general conditions of the two groups of patients are shown in Table 1. All patients were diagnosed by pathological examination, and were diagnosed by two senior pathologists. Tissue chips were taken from their tumor tissues. Tissue chip was dewaxed for 20 min in xylene, and then in fresh xylene. Repeat 1 time. Soak the dewaxed chip in 100% ethanol for 5 min twice, and soak each of 95% ethanol, 80% ethanol, and distilled water for 5 min. Basic antigen repair solution (Tris-E DTA, pH = 9) was used. Use a pressure cooker to boil, place the chip, time it for 2 min, and naturally cool to room temperature. Incubate for 10 min at room temperature in the dark with 3% H2O2, block with normal sheep serum working solution for 30 minutes at room temperature. Then add primary antibody RP215, 4 °C overnight, HRP-labeled secondary antibody at room temperature for 30 min. DAB coloration, hematoxylin counterstaining. Microscopic observation, randomly selected 5 high-power fields, and two pathologists independently read the film. Cytoplasmic staining score uses four intensity levels (0: negative, 1: weak
positive, 2: moderate positive, 3: strong positive) and percentage of positive cells (0: 0%, 1: 1%-5%, 2: 6%-25%, 3: 26%-50%, 4: 51%-100%). The final score is product of intensity grade and positive cell rate (percentage) grade, 0–3 for low expression and 4–9 for high expression.

Table 1

<table>
<thead>
<tr>
<th>group</th>
<th>gender</th>
<th>age</th>
<th>cancer type</th>
<th>stage</th>
<th>TNM</th>
<th>Radiotherapy or Chemotherapy</th>
</tr>
</thead>
<tbody>
<tr>
<td>stage I lung cancer group</td>
<td>26</td>
<td>24</td>
<td>52.8±1.7</td>
<td>I</td>
<td>T1N0M0</td>
<td>none</td>
</tr>
<tr>
<td></td>
<td>female</td>
<td></td>
<td>22 cases of LUSC and 28 cases of LUAD</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>stage III lung cancer group</td>
<td>27</td>
<td>23</td>
<td>54.6±3.2</td>
<td>IIIA</td>
<td>T1N2M0</td>
<td>none</td>
</tr>
<tr>
<td></td>
<td>female</td>
<td></td>
<td>24 cases of LUSC and 26 cases of LUAD</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

| P value | > 0.05 | > 0.05 | > 0.05 | / |

2.7 Statistical Analysis

Data analysis was performed using SPSS 20.0 statistical software. Comparison of gene expression levels in different tissues was performed using the χ² test and independent sample t test, and clinical pathological characteristics were analyzed using the χ² test. Survival analysis was performed using Kaplan-Meier method. P < 0.05 was defined as a statistically significant difference.

3 Results

3.1 NSCLC-OBS sample grouping and differential gene screening

We divided 62 NSCLC-OBS samples into high and low score groups according to their tumor purity and immune scores using the estimate software and the "ggstatsplot" package in R. The violin distribution map of the immune scores (Fig. 1A), and the T-test results showed the high and low score groups. There were significant differences in tumor heterogeneity (p < 0.01) (Fig. 1B). A differential gene analysis was performed on 62 samples of the high and low score groups, and the results showed that there were 38 differential genes. The R software package was used to draw the volcanic map of the difference results (Fig. 1C) and heatmap (Fig. 1D).

3.2 Screening and GSEA analysis of prognostic related genes

A single factor COX regression analysis was performed on the differential genes based on the patient's clinical information, and a total of 95 variables were obtained. After including them in the secondary dimension reduction analysis, we obtained LYN, C3, COPG2IT1, HLA.DQA1, TNFRSF17 (Fig. 1E) The five hubgene, COPG2IT1, HLA.DQA1 expression and LYN, C3, and TNFRSF17 expressions that were most relevant to the prognosis of patients will reduce the survival time of patients (Fig. 2A-E). The results of
GSEA for each prognostic relevant marker were shown in Fig. 2F-I. As we can see that the GO analysis results of prognostic genes are mainly enriched in ADAPTIVE_IMMUNE_RESPONSE, and the KEGG mainly in CHEMOKINE: HEDGEHOG and JAK signaling pathways.

3.3 Construction and evaluation of prediction models

According to the previous calculation results, we identified LYN, C3, COPG2IT1, HLA.DQA1, TNFRSF17 as genes that affect immune prognosis. According to the difference in the expression of these 5 genes, we divided 62 NSCLC-OBS samples into high and low risk groups (Fig. 3A-B). Compared with the low-risk group, the gene expression of COPG2IT1 and HLA.DQA1 increased and the gene expression of LYN, C3, and TNFRSF17 decreased in the high-risk group (Fig. 3C). The K-M survival curve results showed that the survival time and survival rate of patients identified as high-risk group were significantly lower than those of low-risk group (Fig. 3D). Using ROC to calculate the model, we found that at 1, 3, The auc values of the prediction models for these three time periods over 5 years are all greater than 0.9 (Fig. 3E), which proved that this prediction model has a high accuracy in the NSCLC-OBS sample.

3.4 Verification of prediction models

This prediction model was applied to 62 NSCLC-ACT samples. The results showed that the high and low-risk groups divided by the model had significant differences (Fig. 4A-B). Compared with the low-risk group, the gene expression of COPG2IT1 and HLA.DQA1 increased and the gene expression of LYN, C3, and TNFRSF17 decreased in the high-risk group (Fig. 4C). The survival time and survival rate of patients in the high-risk group were significantly lower than those in the low-risk group. (Fig. 4D), the ROC curve shows that the AUC value of the prediction model in the three time periods of 1, 3, and 5 years is greater than 0.9 (Fig. 4E), which proved that this prediction model has high accuracy.

3.5 Model Hubgene infiltration and immune targets in immune cells

The NSCLC-ACT patient samples were evaluated to obtain the relative infiltration scores of 24 immune cells in the NSCLC-ACT patient group. According to the prediction model, the ACT patients were divided into high and low risk groups. After comparing the 24 types of immune cell infiltration in the high and low risk groups, we found that hubgene of the prediction model was most expressed in fibroblasts, but there was no significant difference in immune infiltration between the high and low risk groups in 24 immune cells (Fig. 5A).

According to pearson calculation, 100 genes co-expressed with LYN, C3, COPG2IT1, HLA.DQA1, and TNFRSF17 were obtained (p < 0.05, sorted from small to large). Functional enrichment and pathway enrichment analysis of the five hubgenes and 100 co-expressed genes showed that the co-expressed genes are mainly involved in B cell receptor signaling pathway. And these genes were mainly enriched in biological processes such as Apoptotic cell clearance, Leishmaniasis, Hematopoietic cell lineage and Intestinal immune network for IgA production (Fig. 5B-C). Protein interaction analysis of the co-expressed genes of the hub gene yielded three significantly associated protein interaction networks (Figs. 5D-F).
Through the correlation analysis of 5 hubgene and immune checkpoints in the prediction model, we found that LYN, C3, COPG2IT1, HLA.DQA1, TNFRSF17 were highly correlated with PDCD1, PDCD1LG2, LAG3, CTLA4 immune checkpoints ($p < 0.05$) (Fig. 6).

3.6 Experimental verification of prognostic related genes

Except for the staging of lung cancer, there was no significant difference in general information between the two groups of patients ($P < 0.05$). The expression levels of five prognostic related genes in the tumor tissues of the two groups of patients are shown in Table 2. The results of immunohistochemical analysis showed that compared with the stage I lung cancer group, the expression levels of COPG2IT1 and HLA.DQA1 in the stage III lung cancer group were significantly increased, and the expression levels of LYN, C3 and TNFRSF17 were significantly decreased (Fig. 7). The results of the immunohistochemical experiments are consistent with the conclusions obtained in the previous analysis.

| Table 2 |
|-----------------|-----------------|-----------------|--------|
| Gene            | stage I lung cancer group | stage III lung cancer group | $P$ value |
| LYN             | 9.3 ± 0.62       | 2.5 ± 0.36       | < 0.01 |
| C3              | 8.9 ± 0.37       | 2.1 ± 0.35       | < 0.01 |
| COPG2IT1        | 3.6 ± 0.59       | 7.3 ± 0.26       | < 0.05 |
| HLA.DQA1        | 4.3 ± 0.87       | 10.5 ± 1.12      | < 0.05 |
| TNFRSF17        | 7.8 ± 0.86       | 3.2 ± 0.36       | < 0.05 |

4 Discussion

Non-small cell lung cancer (NSCLC) accounts for about 80% -85% of all lung cancers, and the main pathological types include adenocarcinoma and squamous cell carcinoma. For the most part, the treatment of NSCLC depends on the stage at which it is treated. Patient's surgery resection rate is about 25%[13]. However, many patients are still at risk of recurrence of lung cancer after surgical resection. Among patients with NSCLC after surgical resection, the 5-year survival rate of patients in stage I exceeds 70%, but patients in stage IIIA, the 5-year survival rate is only 25% [13–16]. Despite the improvement of surgical methods, the combination of radiotherapy and chemotherapy, and the emergence of molecular targeted therapy, the overall 5-year survival rate of lung cancer remains less than 15%. In recent years, emerging immunotherapy performs well in a variety of cancers, including lung cancer. Blocking the PD-L1/PD-1 signaling pathway to attack tumor cells expressing PD-L1 is the current mainstream method. Currently, there are only four types of PD-1/PD-L1 blockers in the clinic. and it is now preferred to use immunotherapies alone or in combination with chemotherapy. Although the anti-cancer activity of PD-1 and PD-L1 inhibitors is exciting, this type of immunotherapy is not effective for all patients, and meta-analyses indicate that a higher risk of rash, thyroid dysfunction, pruritus, pneumonia, and colitis in
patients treated with PD-1/PD-L1 inhibitors[17–19]. Large amounts of long-term use of PD-1/PD-L1 inhibitors can easily lead to adverse effects such as the expansion of the MDM2 family or distortion of EGFR and resistance to PD-1/PD-L1 inhibitors[20]. Therefore, predicting and finding more biomarkers that may be related to immune prognosis is important for non-small cell lung cancer significance.

We performed immune assessment on untreated NSCLC patient samples and grouped them for differential gene analysis, single-factor COX regression analysis, Receiver Operating Characteristic (ROC) analysis and survival analysis. We identified LYN, C3, COPG2IT1, HLA.DQA1, TNFRSF17, these five genes may be important immune prognostic genes in NSCLC. We modeled these five genes and applied this model to chemotherapy-treated NSCLC samples, and the results showed that the survival and survival of patients identified as high risk group were significantly lower than those of low-risk group, and ROC analysis of the model shows that AUC values are greater than 0.9, which proves that the model has great credibility. Some literatures show that LYN in allergic airway inflammation can reduce inflammatory cell infiltration and levels of IL-13 and IL-4 and downregulate allergen-induced airway inflammation[21]. And C3, HLA.DQA1 has been reported in many reports with glomerular nephropathy, idiopathic membranous nephropathy, steroid-sensitive nephrotic syndrome in children, and other immune system diseases[22–25]. It is worth noting that B-cell maturation antigen (BCMA) is a transmembrane glycoprotein in TNFRSF17, and current targeted immunotherapy for BCMA has been used in clinical trials of multiple myeloma with satisfactory results[26–28]. Immune checkpoint molecules (immune checkpoint) is a regulatory molecule in the immune system that plays a suppressive role in the immune system and are essential for maintaining autotolerance, preventing autoimmune reactions and minimize tissue damage by controlling the time and intensity of the immune response. The expression of immune checkpoint molecules on immune cells will inhibit the function of immune cells and prevent the body from producing effective anti-tumor immune response, tumor formation immune escape. PDCD1, PDCD1LG2 are the key targets in the treatment of widely used PD-1 / PD-L1 inhibitors and plays an important role in adolescent idiopathic arthritis, diffuse large B-cell lymphoma, head and neck squamous cell carcinoma, colorectal cancer and other diseases[29–33]. LAG3 is the third clinically targeted inhibitory receptor pathway, it can increase the expression of CD4 +, CD25- and promote T cell dysfunction in tumor microenvironment[34–35], and significantly expressed in immunohistochemistry of solid tumors including pancreatic cancer, gastric cancer, colorectal cancer, melanoma, urogenital tract cancer, etc[36] It is worth noting that CTLA4, LAG3 and PD1 seem to interact[37–38]. The combination of anti-PD-1-CTLA4 in the treatment of prostate cancer is gratifying[39], meanwhile, antagonizing LAG3 and PD1 can enhance Tumor-specific cellular response and induce tumor rejection[40, 41]. Correlation analysis of the five hubgene genes and immune checkpoints, LYN, C3, COPG2IT1, HLA.DQA1, TNFRSF17 highly correlated with PDCD1, PDCD1LG2, LAG3, CTLA4 immune checkpoints. This indicates that these five hubgene genes may affect the disease progression of NSCLC patients by regulating immune checkpoints expression. Co-expression analysis of these five hubgene genes yields 100 co-expressed genes (p < 0.05). Performing functional enrichment and pathway enrichment analysis of hubgene and its co-expressed genes, we found that the GO prognostic gene were mainly enriched in ADAPTIVE_IMMUNE_RESPONSE, and KEGG was mainly enriched in CHEMOKINE, HEDGEHOG and JAK signal pathways. Co-expressed genes are mainly involved in the B cell
receptor signaling pathway, and are mainly enriched in Apoptotic cell clearance, Leishmaniasis Hematopoietic cell lineage, Intestinal immune network for IgA Production and other biological processes. Among them, CHEMOKINE is an emerging family of chemokine cytokines, which shows a wide range of functions, such as regulating steady-state leukocytes flow and development, and activating the innate immune system. Inappropriate balance of chemokine synthesis or chemokine receptor expression can lead to a variety of pathological diseases. Some drugs containing chemokine-derived peptides may exert antitumor activity in lung cancer, prostate cancer, colon cancer, melanoma and breast cancer by affecting the tumor microenvironment[42]. And ADAPTIVE_IMMUNE_RESPONSE, B cell receptor signaling pathway, and adaptive cell clearance all play important roles in the process of tumorigenesis and development.

5 Conclusion

In conclusion, our research indicates that LYN, C3, COPG2ITL, HLA.DQAL, and TNFRSFL17 are potential prognostic markers for non-small cell lung cancer. The results of immunohistochemical experiments of patients' pathological tissues are consistent with this conclusion. These prognosis-related genes are mainly enriched in B cell receptor signaling pathways and are highly related to PDCD1, PDCD1LG2, LAG3, and CTLA4 immune checkpoints; this suggests that immunotherapy may improve the prognosis of non-small cell lung cancer patients by regulating these prognosis-related genes.

Abbreviations

NSCLC
non-small cell lung cancer

NSCLC-OBS
non-small cell lung cancer observation

ROC
Receiver Operating Characteristic

NSCLC-ACT
A chip-in-chip non-small cell lung cancer chemotherapy

COX
cyclo-oxygen-ase

LYN
v-yes-1 Yamaguchi sarcoma viral related oncogene homolog

COPG2IT1
COPG2IT1 Antigen

HLA-DQA1
Human leucocyte antigen HLA-DQA1

TNFRSF17
Tumor necrosis factor receptor superfamily member 17, a protein that in humans is encoded by the TNFRSF17 gene.
**PDCD1**  
Programmed cell death protein 1, also known as pd-1 and CD279 (differentiated cluster 279), is a protein on the cell surface that down-regulates the immune system's response to human cells.

**PDCD1LG2**  
programmed cell death 1 ligand 2

**LAG3**  
Lymphocyte-activation protein 3, belongs to Ig superfamily and contains 4 extracellular Ig-like domains. The LAG3 gene contains 8 exons. The sequence data, exon/intron organization, and chromosomal localization all indicate a close relationship of LAG3 to CD4.

**CTLA4**  
CTLA4 or ctla-4 (cytotoxic t-lymphocyte associated protein 4), also known as CD152 (differentiated cluster 152), is a protein receptor that acts as an immune checkpoint and down-regulates the immune response.

---

**Declarations**

**Acknowledgements:**
None

**Data Availability:**
All datas are available. Please contact us to access if it is needed.

**Ethical approval and consent to participate:**
This study was carried out in accordance with the recommendations of the Ethics Committee of the First Affiliated Hospital of Nanchang University. The protocol was approved by the Ethics Committee of the First Affiliated Hospital of Nanchang University. All subjects gave written informed consent in accordance with the Declaration of Helsinki.

**Funding:** The study did not accept any funding.

**Author contribution statement:**
JX: research design and drafting the manuscript
HN: Experiment implementation
KL: literature search and Experiment implementation
XW: review and revision of the manuscript and writing guidance

**Disclosure statement**
Competing interests: There are no conflicts of interest in this study.

References


31. Masugi Yohei
Ogino Shuji
10.1158/2326-6066.CIR-17-0122


33. Franzen Alina
10.18632/oncotarget.23080


35. Solinas Cinzia
10.3390/cancers11081213

36. Park Joon Oh., Park Young Suk., Kang Won Ki., Lim Ho Yeong., Kim Kyoung-Mee., Kim Seung Tae
10.1097/CJI.0000000000000283

37. 10.1126/sciimmunol.aan3864


Figures
Figure 1

A: Violin distribution of immune scores for OBS samples; B: Comparison of tumor heterogeneity between high and low score groups; C: Volcano map of differential genes in OBS samples; D: Heat map of differential genes in OBS samples E: Single factor regression analysis of differential genes to screen for key genes for prognosis
Figure 2

A-E: Prognostic analysis of 5 key prognostic genes; F: GO results of GSEA analysis of 5 key prognostic genes; G: KEGG results of GSEA analysis of 5 key prognostic genes
Figure 3

A: The abscissa is the number of patients in OBS group, and the high and low risk groups are divided by the risk score; B: The abscissa is the number of patients, and the division of high-score and low-risk groups is verified by survival; C: Heatmap of the expression of 5 key prognostic genes in high-risk and low-risk patients in OBS group; D: Comparison of survival analysis between high-risk and low-risk patients; E: ROC analysis test results of model sensitivity and specificity
Figure 4

A: The abscissa is the number of patients in ACT group, and the high and low risk groups are divided by the risk score; B: The abscissa is the number of patients, and the division of high-score and low-risk groups is verified by survival; C: Heatmap of the expression of 5 key prognostic genes in high-risk and low-risk patients in ACT group; D: Comparison of survival analysis between high-risk and low-risk patients; E: ROC analysis test results of model sensitivity and specificity
Figure 5

A: Hubgene infiltration in 24 immune cells in high-risk and low-risk groups of ACT patients; B: Functional enrichment results of co-expressed genes of 5 key prognostic genes; C: Pathway enrichment results of co-expressed genes for 5 key prognostic genes; D-F: Protein-protein interaction network of co-expressed genes of 5 key prognostic genes
Figure 6

Correlation analysis results of 5 key prognostic genes with PDCD1, PDCD1LG2, LAG3 and CTLA4 immune checkpoints
Figure 7

Expression levels of 5 key prognostic genes in tumor tissues of two groups of patients with NSCLC.