The tumor environment immune phenotype of LUSC by genome-wide analysis

Yuansheng Zheng  
Fudan University

Guoshu Bi  
Fudan University

Yunyi Bian  
Fudan University

Ming Li  
Fudan University

Yiwei Huang  
Fudan University

Mengnan Zhao  
Fudan University

Zhencong Chen  
Fudan University

Cheng Zhan (czhan10@fudan.edu.cn)  
Fudan University  https://orcid.org/0000-0001-8745-9276

Wei Jiang  
Fudan University

Research

Keywords: Immune infiltration, Lung squamous carcinoma, Genome-wide, Tumor environment, Immune therapy, Random forest

Posted Date: May 20th, 2020

DOI: https://doi.org/10.21203/rs.3.rs-29041/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 International Immunopharmacology on July 1st, 2021. See the published version at https://doi.org/10.1016/j.intimp.2021.107564.
Abstract

Background

To compare the landscape of tumor microenvironment (TME) of lung squamous carcinoma (LUSC) in different immune pattern and explore potential factors on immune therapy and prognosis.

Method and Materials:

We have obtained the LUSC data from TCGA, GEO, and our department and classified them into 2 TME clusters by random forest model based on the infiltration pattern of 24 immune cell populations. Then we systemically compared the genomic significance, clinical characteristic and immune infiltration pattern in 2 TME clusters.

Results

Samples were divided into 2 TME clusters based on the relative abundance of 24 immune cells and a random forest classifier model was constructed. TME cluster B was a higher immune infiltration group with lower mutation load, richer co-infiltrate immune cells, up-regulated immune-related cytokines including immune checkpoint molecules, and higher active immune cells. TME cluster was also an independent predictor in prognosis (B vs A, p < 0.05) in patients from TCGA, GEO and our department.

Conclusions

Our study has described the microenvironment landscape of LUSC in different immune infiltration patterns and systemically analyzed genomic and clinical characteristics with distinct immunophenotypes, thus partly revealed the interaction between tumors and immune microenvironment, which may guide a more precise and personalized immune therapeutic strategy for LUSC patients.

Background

Lung cancer was one of the most common cancer worldwide (12% of all cancers), which caused enormous economic and medical burdens every year. About 80–85% of lung cancers are non-small cell carcinoma (NSCLC) and squamous cell carcinoma is one of the major subtypes of NSCLC (more than 30%). Strategies targeting the immune responses have demonstrated significant antitumor activity across a range of solid tumors, including NSCLC. Several immune therapy drugs including immune checkpoint inhibitors such as pembrolizumab and nivolumab have been widely applied in the treatment of patients with LUSC and showed excellent results.
The tumor microenvironment (TME) could significantly impact the response of immunotherapy. Microenvironment-mediated drug resistance can be induced by soluble factors secreted by a tumor or stromal cells, while different immune cells can also improve or obstruct therapeutic efficacy and may vary in their activation status within the TME. Therefore, evaluating the TME status can help us to predict the potential response to immunotherapy.

In our study, we collected 728 LUSC patients and classified them according to the immune cells’ infiltration abundance in the tumor samples calculated by GSEA. Then we integrally compared the genomic signatures, clinical characteristics and immune microenvironment including immune checkpoints factors between patients with different immune cells’ infiltration abundance and seek valuable potential biomarkers to predict the response to immunotherapy. We hope this research will offer some important insights into the complex relationship between the heterogeneity of intratumoral immune cells, tumor molecular subtypes, and disease progression in LUSC.

Method And Materials

Acquisition and preprocessing of LUSC data sets

Discovery cohort: Gene expression data of LUSC patients (FPKM normalized) and corresponding clinical information of The Cancer Genome Atlas (TCGA) were downloaded from the UCSC Xena browser (GDC hub: https://gdc.xenahubs.net)\textsuperscript{7}. The data with unclear prognosis information including outcome status and survival time was removed.

Validation cohort one: Gene Expression Omnibus (GEO) data of LUSC patient’s data was composed by the datasets including GSE3141, GSE8894, GSE19188, GSE29013, GSE30219, GSE37745, GSE43580, GSE50081 and GSE115457 from the same chip platform (Affymetrix Human Genome U133 Plus 2.0 Array) after systematic screening\textsuperscript{8–15}. The expression data and clinical data were downloaded from the GEO database (http://www.ncbi.nlm.nih.gov/geo). The probe sets of Affymetrix Human Genome U133 Plus 2.0 chips were annotated to gene names based on the annotation platforms GPL570. We have eliminated the batch effect due to the heterogeneity among different studies by the COMBAT empirical Bayes method in the gsva package and performed the data normalization with the limma package\textsuperscript{16,17}.

Validation cohort two

44 samples were obtained from patients with lung squamous-cell carcinoma who underwent surgical resection between July 2012 and December 2012 at Zhongshan Hospital, Fudan University. Every patient was provided informed consent to conduct genomic studies in accordance with the ethical principles of the Declaration of Helsinki\textsuperscript{18}. The study was approved by the ethical committees of Zhongshan Hospital (No. 2011–219(2)). All pulmonary resections were performed by experienced thoracic surgeons and qualified pathologists to confirm the diagnosis of LUSC. Patients with metastasis, history of chemoradiotherapy or immunotherapy, and unclear survival information were excluded. Annual follow-up
was performed by call or clinic. RNA sequencing for all tumor samples was performed using Illumina Hiseq 2500 and BGI-500RNAseq platforms.

**Classification based on microenvironment immune cell abundance**

We have used the gene signatures proposed by Bindea et al. to detect the relative abundances of the 24 immune cell populations by GSVA algorithm\(^1\), which consist of 585 genes representing 24 microenvironment cell subsets from both innate and adaptive immunity (Supplement Table 1). GSVA algorithm was an unsupervised gene set enrichment method that computes an enrichment score by integrating the collective expression of a given gene set relative to the other genes in the sample\(^2\). A matrix containing the enrichment score ranging from −1 to 1 for each cell type in each tumor sample was produced by GSVA function, representing the relative abundances of the immune cell populations used for further analyses.

**Cluster analysis and construction of the random forest classifier model**

Samples were grouped based on the immune cells infiltration pattern by an unsupervised clustering method based on Euclidean distance and Ward’s linkage and the optimal number of “TME clusters” was determined based on the percentage of the variance of the data using the ConsensusClusterPlus package with 1000 repeats\(^2\)\(^,\)\(^2\)\(^2\). Then we calculated the correlation matrix by Spearman's test. The TME cluster classifier model was established using the random forest algorithm by randomForest package in R, a machine learning dimension reduction strategy based on the construction of thousands of classification or regression trees\(^2\)\(^3\)\(^,\)\(^2\)\(^4\). Patients in the TCGA database were randomly assigned to training and testing groups at a ratio of 7:3\(^2\)\(^5\).

**Differentially expressed genes and functional analysis**

The expressed genes in different TME clusters were evaluated by the limma package in R and the threshold for identification of DEGs was set as \(p < 0.05\) and \(\log FC (\text{fold change}) > 0.5\)\(^2\)\(^6\). Functional enrichment analysis of DEGs was performed by the clusterProfiler package in R. Gene Ontology (GO) and Kyoto Encyclopedia of Genes and Genomes (KEGG) terms were completed in R and the cutoff was adjusted \(p < 0.01\) and false discovery rate (FDR) < 0.05. Single-sample GSEA analysis was carried out using the GSVA Bioconductor package according to the gene signatures associated with tumor development and immune response\(^2\)\(^0\). We also identified pathways that were up- or downregulated in TME clusters with GSEA. Gene sets were obtained from the MSigDB database of the Broad Institute, and c5.bp.symbols gene sets were selected to perform quantification of pathway activity\(^2\)\(^7\).

**Comparison of somatic mutations and copy number variation among TME clusters**
Somatic mutation and copy number variation data were obtained from the Xena browser. We have compared the somatic mutations and copy number variation in TME clusters by Kruskal–Wallis test and the results were shown with the oncoplot function in the maftools package.

**Statistical analysis**

Categorical variables were compared using Fisher’s exact test and Pearson’s c2 test and continuous variables were compared using Student’s t-test and the Wilcoxon test. Kaplan–Meier survival estimates were generated and visualized by the ggplot2 package. Multivariable Cox regression analyses were used to test the independent prognostic value of the immune clusters using the R package survival and the coxph function. All analyses were performed in R version 3.6.1. Unless otherwise stated, results were considered statistically significant, if p-value < 0.05. The p-values were all two-sided.

**Result**

**The landscape of the microenvironment phenotype in LUSC**

Our study included 493 LUSC patients in TCGA, 370 LUSC patients in GEO and 44 LUSC patients in our hospital. The design of the study was shown in supplement Fig. 1A. Patients from TCGA were divided into the discovery cohort while the patients from GEO and our hospital were divided into the validation cohort. We used the GSVA enrichment score of the 24 immune cell populations to define the relevant abundance in each tumor sample to describe the immune infiltration phenotype in LUSC patients (supplement Table 2). We employed the unsupervised consensus clustering base on the GSVA score of the 24 cell populations by the ConsensusClusterPlus package in R, and after k-means clustering, we found the best separation was a dichotomy (supplement Fig. 1B). We divided the LUSC patients from TCGA into two heterogeneous clusters named TME cluster A and TME cluster B, which meant the different immune infiltration patterns of the 24 immune cells. (shown in Fig. 1A). The TME cluster A had 243 patients while the TME cluster B covered 250 patients. The difference between 2 TME clusters was the immune infiltration pattern. We also found the different 5-year OS and MST between two TME clusters (Fig. 1B, A vs. B, 43.9% vs. 52.2%, p = 0.028) (MST, A vs. B, 49.0 months vs. 61.9 months). TME cluster A was marked with blue and the B group was marked by red, which meant the different immune cell infiltration abundance. Red meant higher infiltration of the innate and adaptive immune response cell including pDCs, Tgd cells, macrophages, B cells, T cells, cytotoxic cells, and Treg cells. There was a significant difference between the two TME clusters groups (Fig. 1C) and the relevance between every immune cell and prognosis was shown in Fig. 1D. We found there was a significant association between most immune infiltration cells and survival. According to the baseline analysis (Fig. 1E), we also found only sex was associated with immune cell infiltration (p = 0.017) while the age, stage, and location were not related to immune cell infiltration. Male patients had higher immune cell infiltration after logistic analysis (p = 0.046).

**Construction and validation of the random forest classifier model**
We have established a new classification method through random decision forests based on Bootstrap aggregating and random subspace method to sort patients with LUSC based on the different immune infiltration relative abundant. We divided randomly the TCGA patients into 2 cohorts including 345 patients in train cohorts and 148 patients in the test cohort. And the relative abundance of 24 immune cell populations was input variate while the unsupervised TME cluster (Fig. 1A and supplement Fig. 1C) was set as the response factor. The most suitable mtry and ntree were selected as 4 and 3000 and we use the random forest model to verify the training and internal testing cohorts. The accuracy rate was 100% and 99.3%, so we could recognize it was a stable model and the separation of the scatter diagram also showed the validity of the model. We could find there was nearly no overlapping area in the scatter diagram (supplement Fig. 5A). Therefore, we could use the random forest model to verify the patients in GEO and our hospital. The classified results of the GEO cohort and our hospital were shown in Fig. 2A-2B. And we found the immune cell infiltration in GEO and hospital patients were similar to patients in TCGA, which also certified the validity of the random forest model again. The TME cluster B in the GEO cohort had more immune cell infiltration (shown in Fig. 2C) and the TME cluster B had a higher 5 years OS rate (shown in Fig. 2E). Patients in our hospital were also classified into 2 groups by the model and the TME cluster B had more immune infiltration relative abundant (Fig. 2D), but there was no survival difference between 2 groups in our hospital (Fig. 2F, 78.3% vs. 84.0%, p = 0.640), which might be due to the lack of samples.

**DEGs and functional annotation**

The 5-year survival difference in different tumor microenvironment phenotypes may be related to the DEGs. So, we have examined the expression profiles from TCGA and we found 1828 significant different genes including 301 up-regulated genes in TME cluster A and 1527 up-regulated genes in TME cluster B (supplement Table 3). The result was shown in Fig. 3A and we simultaneously labeled the top 10 up-regulated or down-regulated genes. The KEGG and GO enrichment analysis was operated by clusterProfiler in R. We sought out the enrichment pathways on immune cell activation and adhesion in TME cluster B by KEGG and GO analysis. The enrichment pathways in TME cluster B included the cytokine-cytokine receptor interaction pathway and other metabolism signal pathways (Fig. 3B-3C). The above DEGs enrichment results were consistent with the immune infiltration pattern in TME cluster B. At the same time, the GSEA algorithm was performed to further verify the enrichment pathways and showed similar results (supplement Fig. 3). Besides, we also have utilized the ssGSEA to compare the different patterns in those TME subgroups and the results showed that immune-related pathways were up-regulated in the TME cluster B group, which was consistent with the KEGG/GO result (supplement Table 4). Several important signal pathways including EMT, TGF-β, RAS and so on were also up-regulated in TME cluster B, which meant those signal pathways may play an important role in immune infiltration (supplement Fig. 4A). The DEGs and functional annotation analysis of data from GEO and our hospital was shown in supplement Fig. 4B-4G.

**Immune microenvironment traits associated with TME clusters**
The difference of 2 TME clusters was immune infiltration pattern, so we have compared the expression level of the immune-related genes, cytokines, and miRNA of LUSC patients in TCGA to explore the tumor environment difference characteristics. Pervious a POPLAR trial had mentioned a 7-gene signature including (CD8A, GZMA, GZMB, IFNγ, EOMES, CXCL9, CXCL10, and TBX21) which was associated with activated T cells, immune cytolytic activity, and interferon-γ expression. Another article quantified the cytolytic activity of the local immune infiltrate and devised a simple and quantitative measure of immune cytolytic activity ('CYT') based on transcript levels of two key cytolytic effectors, granzyme A (GZMA) and perforin (PRF1), which are dramatically upregulated upon CD8 + T cell activation. In our study, the 8 genes associated with the cytotoxic function was up-regulated in TME cluster B (p < 0.05), which meant the TME B had higher immune cytotoxic (Fig. 4A). Besides, we also found some genes related to innate immunity including MYD88, TLR3 and so on were up-regulated in TME cluster B (p < 0.05) but the MAVS gene was down-regulated (Fig. 4B). Meanwhile, the TME cluster B had a more MHC-associated antigen molecular expression than cluster A. We still compared the miRNA between in 2 clusters (shown in supplement Fig. 5D). According to the heat map, we found the TME cluster B had more immune cells including activated cells and suppressor cells such as Treg cells or iDCs. Therefore, we have explored the immune activation level by calculating the ratio of CD8 + T cells/ Treg cells and aDC/iDC. The results were shown that the TME cluster B group had a higher ratio, which meant the TME cluster B had a more activated immune environment including effect cells and antigen-presenting cells (Fig. 4C). The expression of immunoregulation factors in 2 clusters was shown in Fig. 4D-4E. We have compared 15 immune checkpointing molecules and found 12 higher-expression molecules in the cluster B group (p < 0.001, Fig. 4D). There were also 14 significant higher-expression co-stimulated factors in cluster B. (p < 0.001, Fig. 4E). Important immune checkpoints molecules such as CTLA4, PDCD1 (PD-1), LAG3, and CD274 (PD-L1) were up-regulated in cluster B (Fig. 4F). We could also observe the similar trend of immune-related factors of patients from GEO and our hospital even though some immune molecules were no different (supplement Fig. 6). We also have built a correlation between immune checkpoints (Fig. 4G).

**Tumor genomic alterations associated with TME clusters**

Patients in different immune infiltration patterns had a different genomic mutation, which could cause survival differences. Some studies suggested there was a significant relation to immunotherapy effects and genomic mutation. We have compared the mutation load and copy number variation (CNV) between two TME clusters and found the TME cluster B had higher mutation load (243.97 vs 257.20, p = 0.023) and CNV (1629 vs 2419, p < 0.001), which suggested the number of somatic mutations may play an important role in immune infiltration (Fig. 5A and supplement Fig. 5B-5C).

Then we have explored further every mutation gene in two TME clusters (Fig. 5B and supplement Table 5) and found 1205 significant genes (adjusted p < 0.05). Some studies have reported the cancer-related genes of LUSC such as EGFR, FGFR1, PIK3CA, DDR2, MET, BRAF, TP63, SOX2 ZNRF2 and so on. In our study, we found the significant related genes included EGFR (10.25% vs. 20.61% adjusted p = 0.024), FGFR1 (21.31% vs. 33.05% adjusted p = 0.031), PIK3CA (50.01% vs. 63.37% adjusted p = 0.014), TP63 (39.76% vs. 55.6% adjusted p = 0.006) and SOX2 (42.22% vs. 57.15% adjusted p = 0.009). The TME
cluster B had higher mutation rates of cancer-related genes, which could indicate the correlation between genomic mutation and immune infiltration. However, there was no association between mutation load and cytotoxic (Fig. 4G) so it needed to explore further.

**Prognostic value of TME clusters**

We have compared the survival rates of patients between different TME clusters. Patients in TME cluster B from TCGA had a better 5-year OS (A vs. B, 43.9% vs. 52.2%, \( p = 0.028 \)) and the TME cluster B group from GEO also had a higher 5-year survival (A vs. B, 33.1% vs. 46.4% \( p = 0.0073 \)). There was no significant survival difference between different TME cluster patients in our hospital (A vs. B, 78.3% vs. 84.0%, \( p = 0.640 \)). The cox multivariable analysis also showed the TME cluster B had a better prognosis in patients from TCGA (B vs. A, HR = 0.673, \( p = 0.006 \)) and GEO (B vs. A, HR = 0.662 \( p = 0.005 \)). We have built the predictive model of prognosis by nomogram and the c-index of the model was 0.654 and 0.731, which suggested the TME cluster was a valuable predictive factor (supplement Fig. 2).

**Discussion**

In recent decades, studies focused on the immune landscape have obtained a lot of attention in cancer biological and clinical research. Immune markers such as CD3 + or CD8 + have suggested the predictive value of prognosis in several cancer types\(^ {30} \). The immune landscape is an important factor in the immune response in cancer. Therefore, to explore genomic and clinical characteristics in different immune infiltration pattern of LUSC patients, we collected 728 LUSC patients from common datasets and our institution and divided them into 2 groups named “TME cluster” based on the abundant of immune cells by GSEA method. Then we systemically compared the genomic significance, clinical characteristic and immune infiltration pattern in 2 TME clusters and found TME cluster was an important predictive factor in prognosis.

Several studies have suggested that the immune infiltration status in the tumor samples was significantly correlated with the prognosis in breast cancer, renal cancer, and neck cancer\(^ {31-33} \). In our results, we also found that TME cluster A, a subgroup with lower immune cell abundant, was associated with poor survival. Some studies have reported the abundance of immune cells in tumor tissue could affect the response to immune checkpoints inhibitors such as nivolumab, which could explain the non-ideal OS in some advanced NSCLC patients with immune therapy. However, it also has been reported the recruitment of Tregs could promote the immune escape of tumors in NSCLC\(^ {34} \), which might be conflictive to our founding. Therefore, we speculated the co-infiltration of Treg cells and the effect of T cells in cluster B could maintain the immune balance\(^ {35,36} \). The anti-tumor effect of the immune microenvironment was decided by the ratio of effect cells and suppressor cells. We have calculated the ratio of CD8 + T cell/Treg cell and aDC/iDC and found the TME cluster B had higher activated immune cells.

The co-expression immune-related genes of defining affect T cells and IFN-\( \gamma \) expression including CD8A, GZMA, GZMB, IFNy, EOMES, CXCL9, CXCL10, and TBX21 were both up-regulated in TME cluster B, which was consistent with the T cells infiltration pattern. It has been reported the high-expression of immune-
related genes could increase the response rate to immunotherapy in NSCLC cancer and improve the overall survival, which might suggest TME cluster B was an immunotherapy-sensitive subgroup. At the same time, the cytolytic activity score which was associated with OS as some studies reported based on the PRF1 and GZMA which reflected the anti-tumor effect also was higher in cluster B. Sumana Narayanan et al. suggested a higher CYT score was associated with higher expression of immune checkpoint molecules and higher mutation load in colon cancer. It was further verified in LUSC that cluster B had a higher mutation load in our study. Tumor mutation burden was an important predictive factor in immune therapy. Several significant mutation DNA repair genes including TP53, ATR, and LIG1 which were considered to be associated with the effect of PD-1 inhibitors were both up-regulated in the TME B cluster. Key targeted gene aberrations in LUSC such as PI3K, ROS and EGFR mutation were more common in TME cluster B, which means LUSC patients with those mutations may also be suitable for immunotherapy. Besides, it has been reported higher CNV could increase the expression of the cancer-related genes, which was consistent with our study.

The low immune infiltration pattern was considered being associated with distinct tumor immune escape mechanisms including reduced expression of major histocompatibility complex (MHC) class I, adhesion and costimulatory molecules, loss of antigens, and increased expression of immunosuppressive components such as HLA-G, HLA-E, and PD-L1 and of other immunosuppressive factors such as cytokine and metabolites that contribute to the escape from immune recognition. The expression of adhesion (CADM1) and costimulatory molecules (CD80, CD86) were increased in cluster B while the immune checkpoint molecules such as PD-L1 and CTLA-4 were up-regulated in cluster B, which was coincident with GO analysis. TGF-β, another factor that has been identified as a key driver of tumor plasticity, was up-regulated in cluster B, which might be an important therapy site.

Rachel Rosenthal et al suggested the diverse TME impact upon neoantigen presentation, as well as the tumor-specific mechanisms leading to immune escape. The disruption of tumor antigen presentation was a significant pathway to promote immune escape. Antigen presentation cells (APC) such as dendritic cells and macrophages were up-regulated in TME cluster B and the ratio of activated DC/inhibited DC was also higher in cluster B, so, we speculated the cluster with lower infiltration pattern could lead to immune escape via the down-regulation of active APC. Some specific genes in the APM process including PSMB5, PSMB6, PSMB7, PSMB8, PSMB9, PSMB10, TAP1, TAP2, ERAP2, CANX, CALR, PDIA3, TAPBP, B2M, HLA-A, HLA-B, and HLA-C could predict the efficiency of this antigen-processing and presenting steps based on previous studies. Therefore, identification of such genes might lead to further insight into the complex interaction between tumor cells and the immune system and thus facilitate the development of personalized immune therapeutic regimens and enhance the response rate of immune checkpoint inhibitors in LUSC patients.

There were several limitations to our study. First, accessible gene-expression data of patients who received immunotherapy is insufficient currently, so we can't validate our findings in patients who received immunotherapy. Second, most of the tumor samples from TCGA and GEO were single-loci. Given
the spatial heterogeneity of intratumor immunoreactivity, the exploration of the multi-foci sample is still warranted. Besides, the study was a retrospective descriptive study. In summary, by machine learning methods and multi-omics profiling, our large-cohort study still described a comprehensive landscape of LUSC immune infiltration patterns and integrated biomarkers associated with distinct immunophenotypes, thus explored the interaction between tumors and immune microenvironment, which may guide a more precise and personalized immune therapeutic strategy for LUSC patients.

**Conclusions**

Our study has described the microenvironment landscape of LUSC in different immune infiltration patterns and systemically analyzed genomic and clinical characteristics with distinct immunophenotypes, thus partly revealed the interaction between tumors and immune microenvironment, which may guide a more precise and personalized immune therapeutic strategy for LUSC patients. The immune infiltration pattern could be a prognosis predictor in LUSC.
| Characteristic  | TME cluster A (250) | TME cluster B (243) | p-value |
|----------------|---------------------|---------------------|---------|
| Age (%)        |                     |                     | 0.072   |
| <60            | 51 (21.0)           | 39 (15.6)           |         |
| 60–70          | 109 (44.9)          | 102 (40.8)          |         |
| >70            | 83 (34.2)           | 109 (43.6)          |         |
| Sex (%)        |                     |                     | 0.017   |
| male           | 192 (79.0)          | 173 (69.2)          |         |
| female         | 51 (21.0)           | 77 (30.8)           |         |
| Location (%)   |                     |                     | 0.357   |
| central        | 68 (28.0)           | 76 (30.4)           |         |
| peripheral     | 51 (21.0)           | 40 (16.0)           |         |
| unknown        | 124 (51.0)          | 134 (53.6)          |         |
| Stage (%)      |                     |                     | 0.917   |
| Stage I        | 116 (47.7)          | 125 (50.0)          |         |
| Stage II       | 79 (32.5)           | 81 (32.4)           |         |
| Stage III      | 44 (18.1)           | 41 (16.4)           |         |
| Stage IV       | 4 (1.6)             | 3 (1.2)             |         |
| T stage (%)    |                     |                     | 0.481   |
| T1             | 49 (20.2)           | 65 (26.0)           |         |
| T2             | 147 (60.5)          | 139 (55.6)          |         |
| T3             | 36 (14.8)           | 34 (13.6)           |         |
| T4             | 11 (4.5)            | 12 (4.8)            |         |
| N stage (%)    |                     |                     | 0.75    |
| N0             | 157 (64.6)          | 159 (63.6)          |         |
| N1             | 63 (25.9)           | 64 (25.6)           |         |
| N2             | 20 (8.2)            | 20 (8.0)            |         |
| N3             | 2 (0.8)             | 3 (1.2)             |         |
| NX             | 1 (0.4)             | 4 (1.6)             |         |
Table 2
The result of UVA and MVA Cox proportional hazards regression in TCGA data

| Variable (TCGA) | Univariable | | Multivariable | |
|-----------------|-------------|-----------------|-----------------|
|                 | HR (95%CI)  | P               | HR (95%CI)      | P               |
| Age             |             |                 |                 |                 |
| ≤ 60            |             |                 |                 |                 |
| 60 ~ 70         | 1.068 (0.808–1.411) | 0.646          | 1.346 (0.864–2.097) | 0.190          |
| ≥ 70            | 1.037 (0.786–1.368) | 0.796          | 1.380 (0.888–2.144) | 0.152          |
| Sex             |             |                 |                 |                 |
| Female          |             |                 |                 |                 |
| Male            | 1.204 (0.872–1.663) | 0.260          | 1.221 (0.877–1.670) | 0.236          |
| Stage           |             |                 |                 |                 |
| Stage I         |             |                 |                 |                 |
| Stage II        | 0.976 (0.724–1.316) | 0.873          | 1.031 (0.745–1.428) | 0.852          |
| Stage III       | 1.564 (1.119–2.185) | 0.009          | 1.578 (1.079–2.309) | 0.019          |
| Stage IV        | 2.749 (1.127–6.705) | 0.026          | 3.710 (1.479–9.306) | 0.005          |
| Smoking         |             |                 |                 |                 |
| No              |             |                 |                 |                 |
| Yes             | 1.267 (0.947–1.695) | 0.111          | 0.489 (0.210 ~ 1.136) | 0.332          |
| Reformed        | 0.758 (0.570–1.008) | 0.057          | 0.655 (0.279 ~ 1.540) | 0.096          |
| Radiation therapy |         |                 |                 |                 |
| No              |             |                 |                 |                 |
| Yes             | 1.014 (0.645–1.594) | 0.952          | 1.197 (0.731 ~ 1.958) | 0.475          |
| Unknown         | 2.714 (1.963–3.752) | < 0.001       | 2.878 (2.053 ~ 4.034) | < 0.001       |
| TME cluster     |             |                 |                 |                 |
| Cluster A       |             |                 |                 |                 |
| Cluster B       | 0.732 (0.555–0.967) | 0.028          | 0.673 (0.507 ~ 0.893) | 0.006          |
Table 3
The result of UVA and MVA Cox proportional hazards regression in GEO data

| Variable (GEO) | Univariable | Multivariable |
|----------------|-------------|---------------|
|                | HR (95%CI)  | P             |
|                |             |               |
|                | HR (95%CI)  | P             |
| Age            |             |               |
| ≤ 60           | —           | —             |
| 60 ~ 70        | 0.891 (0.609–1.304) | 0.552 | 0.938 (0.639–1.378) | 0.746 |
| ≥ 70           | 1.045 (0.706–1.546) | 0.826 | 1.155 (0.777–1.715) | 0.476 |
| Unknown        | 0.866 (0.568 ~ 1.319) | 0.503 | 1.302 (0.723 ~ 2.343) | 0.379 |
| Sex            |             |               |
| Female         | —           | —             |
| Male           | 1.448 (0.978–2.412) | 0.064 | 1.277 (0.851–1.917) | 0.237 |
| Unknown        | 1.045 (0.617 ~ 1.769) | 0.871 | 0.705 (0.327 ~ 1.523) | 0.374 |
| TME cluster    |             |               |
| Cluster A      | —           | —             |
| Cluster B      | 0.688 (0.523 ~ 0.906) | 0.007 | 0.662 (0.495~0.886) | 0.005 |

Supplemental Figure Legends

Supplement Fig. 1. A. The flow chart of our study. B. Consensus matrixes of all GC cohorts for each k (k = 2–4), displaying the clustering stability using 1000 iterations of hierarchical clustering. C. The correlation coefficient between the 24 types of immune cells in LUSC patients from TCGA.

Supplement Fig. 2. Nomogram to predict the OS in LUSC patients from TCGA (A) and GEO (C). The validation of the predictive model based on the TCGA data (B) and GEO data (D).

Supplement Fig. 3. A-B, The cnetplot showing the DEGs and functional annotation pathways in the KEGG database (A) and GO database (B). C-F, GSEA showing the significantly enriched pathways in different TME clusters.

Supplement Fig. 4. A, The heatmap showing corresponds pathways signatures between different TME clusters based on ssGSEA. B-D, The DEGs (B), GO (C) and KEGG (D) analyses of DEGs between different TME clusters from GEO. E-G, The DEGs (E), GO (F) and KEGG (G) analyses of DEGs between different TME clusters from our hospital.
Supplement Fig. 5. A, The scatter diagram showing the analysis of the Principal component between different TME cluster A (blue) and TME cluster B (red). B-C, The genomic alterations of TME cluster A (B) and B (C). D, The volcano plot showing the expressed miRNAs between different TME clusters.

Supplement Fig. 6. The expression profiling of the 8 immune-related genes in the POPLAR study and cytolytic activity (CYT) score from GEO (A) and our hospital (G). The relative expression level of molecules potentially involved in the initiation of innate immunity (left) and MHC-I/II antigen-presenting process (right) in GEO data (B) and our hospital (H). Violin plots showing the CD8 + T cells/Treg ratio and aDC/iDC ratio between two TME clusters from GEO (C) and our hospital (I). The relative expression level of immune co-inhibitors and co-stimulators in GEO (D-E) and our hospital (J-K). The box plots showing the expression pattern of the four important immune checkpoint molecules in TME clusters A and B in GEO (F) and our hospital (L). The statistical difference of two TME clusters was compared through the Wilcoxon test. *, p < 0.05; **, p < 0.01; ***, p < 0.001; ****, p < 0.0001.

Abbreviations

LUSC
Lung Squamous Carcinoma
NSCLC
Non-Small Cell Lung Cancer
TME
Tumor Microenvironment
DEGs
Differentially Expressed Genes
CNV
Copy Number Variation
OS
Overall Survival

Declarations

Ethics approval and consent to participate: Every patient was provided informed consent and the study was approved by the ethical committees of Zhongshan Hospital (No. 2011–219(2)).

Consent for publication: All authors read and approved the final manuscript.

Availability of data and materials: The data that support the findings of this study are available from the corresponding author, upon reasonable request.

Competing interests: The authors declare that they have nothing to disclose concerning commercial support.
**Funding:** This work was supported by the Research and Development Fund of Zhongshan Hospital, Fudan University (Grant number 2018ZSLC28 and 2016ZSFZ56).

**Authors' contributions**

(I) Conception and design: Yuansheng Zheng, Guoshu Bi, Yunyi Bian, Cheng Zhan, and Qun Wang

(II) Administrative support: Yuansheng Zheng, Guoshu Bi, Yunyi Bian, and Qun Wang

(III) Provision of study materials: Yuansheng Zheng, Guoshu Bi, Yunyi Bian, Ming Li, and Mengnan Zhao

(IV) Collection and assembly of data: Yuansheng Zheng, Guoshu Bi, Yunyi Bian, Ming Li, Zhencong Chen, and Mengnan Zhao

(V) Data analysis and interpretation: Yuansheng Zheng, Guoshu Bi, Yunyi Bian, and Yiwei Huang

(VI) Manuscript writing: All authors

(VII) Final approval of manuscript: All authors

**Acknowledgments:** We thank the Zhongshan Hospital, Fudan University for supporting the study and the International Science Editing (http://www.internationalscienceediting.com) for editing this manuscript.

**References**

1. Siegel RL, Miller KD, Jemal A. Cancer statistics, 2020. CA Cancer J Clin. 2020 2020-01-01;70(1):7–30.
2. Molina JR, Yang P, Cassivi SD, Schild SE, Adjei AA. Non-small cell lung cancer: epidemiology, risk factors, treatment, and survivorship. MAYO CLIN PROC. 2008 2008-05-01;83(5):584 – 94.
3. Lu T, Yang X, Huang Y, et al. Trends in the incidence, treatment, and survival of patients with lung cancer in the last four decades. CANCER MANAG RES. 2019 2019-01-20;11:943 – 53.
4. Reck M, Rodriguez-Abreu D, Robinson AG, et al. Pembrolizumab versus Chemotherapy for PD-L1-Positive Non-Small-Cell Lung Cancer. N Engl J Med. 2016;2016-11-10(19):1823–33. 375().
5. Gettinger S, Rizvi NA, Chow LQ, et al. Nivolumab Monotherapy for First-Line Treatment of Advanced Non-Small-Cell Lung Cancer. J CLIN ONCOL. 2016 2016-09-01;34(25):2980–7.
6. Fehrenbacher L, Spira A, Ballinger M, et al. Atezolizumab versus docetaxel for patients with previously treated non-small-cell lung cancer (POPLAR): a multicentre, open-label, phase 2 randomised controlled trial. LANCET. 2016 2016-04-30;387(10030):1837–46.
7. Tang Z, Li C, Kang B, Gao G, Li C, Zhang Z. GEPIA: a web server for cancer and normal gene expression profiling and interactive analyses. NUCLEIC ACIDS RES. 2017 2017-01-01;45(W1):W98-102.
8. Bild AH, Yao G, Chang JT, et al. Oncogenic pathway signatures in human cancers as a guide to targeted therapies. NATURE. 2006 2006-01-19;439(7074):353–7.
9. Hou J, Aerts J, den Hamer B, et al. Gene expression-based classification of non-small cell lung carcinomas and survival prediction. PLOS ONE. 2010 2010-04-22;5(4):e10312.
10. Xie Y, Xiao G, Coombes KR, et al. Robust gene expression signature from formalin-fixed paraffin-embedded samples predicts prognosis of non-small-cell lung cancer patients. CLIN CANCER RES. 2011;2011-09-01(17):5705–14. 17(.
11. Rousseaux S, Debernardi A, Jacquiau B, et al. Ectopic activation of germline and placental genes identifies aggressive metastasis-prone lung cancers. SCI TRANSL MED. 2013 2013-05-22;5(186):166r-186r.
12. Botling J, Edlund K, Lohr M, et al. Biomarker discovery in non-small cell lung cancer: integrating gene expression profiling, meta-analysis, and tissue microarray validation. CLIN CANCER RES. 2013 2013-01-01;19(1):194–204.
13. Jabs V, Edlund K, Konig H, et al. Integrative analysis of genome-wide gene copy number changes and gene expression in non-small cell lung cancer. PLOS ONE. 2017 2017-01-01;12(11):e187246.
14. Der SD, Sykes J, Pintilie M, et al. Validation of a histology-independent prognostic gene signature for early-stage, non-small-cell lung cancer including stage IA patients. J THORAC ONCOL. 2014;2014-01-01(1):59–64. 9(.
15. Weber R, Meister M, Muley T, et al. Pathways regulating the expression of the immunomodulatory protein glycolycin in nonsmall cell lung cancer. INT J ONCOL. 2019 2019-02-01;54(2):515–26.
16. Johnson WE, Li C, Rabinovic A. Adjusting batch effects in microarray expression data using empirical Bayes methods. BIOSTATISTICS. 2007 2007-01-01;8(1):118–27.
17. Ritchie ME, Phipson B, Wu D, et al. limma powers differential expression analyses for RNA-sequencing and microarray studies. NUCLEIC ACIDS RES. 2015 2015-04-20;43(7):e47.
18. Zhan C, Zhang Y, Ma J, et al. Identification of reference genes for qRT-PCR in human lung squamous-cell carcinoma by RNA-SEq. Acta Biochim Biophys Sin (Shanghai). 2014;46(4):330–7. 2014-04-01.
19. Bindea G, Mlecnik B, Tosolini M, et al. Spatiotemporal dynamics of intratumoral immune cells reveal the immune landscape in human cancer. IMMUNITY. 2013 2013-10-17;39(4):782–95.
20. Hanzelmann S, Castelo R, Guinney J. GSVA: gene set variation analysis for microarray and RNA-seq data. BMC BIOINFORMATICS. 2013 2013-01-16;14:7.
21. Hartigan JA, Wong MA. Algorithm AS 136: A K-Means Clustering Algorithm. Journal of the Royal Statistical Society. Series C. (Applied Statistics). 1979 1979-01-01;28(1):100–8.
22. Monti S, Tamayo P, Mesirov J, Golub T. Consensus Clustering. A Resampling-Based Method for Class Discovery and Visualization of Gene Expression Microarray Data. MACH LEARN. 2003 2003-01-01;52(1):91–118.
23. Breiman L. Random, Forests. MACH LEARN. 2001 2001-01-01;45(1):5–32.
24. Xie Y, Xiao G, Coombes KR, et al. Robust gene expression signature from formalin-fixed paraffin-embedded samples predicts prognosis of non-small-cell lung cancer patients. CLIN CANCER RES. 2011;2011-09-01(17):5705–14. 17(.
25. Bi G, Chen Z, Yang X, et al. Identification and validation of tumor environment phenotypes in lung adenocarcinoma by integrative genome-scale analysis. Cancer Immunol Immunother. 2020 2020-03-18.

26. Yu G, Wang LG, Han Y, He QY. clusterProfiler: an R package for comparing biological themes among gene clusters. OMICS. 2012 2012-05-01;16(5):284–7.

27. Benjamini Y, Yekutieli D. The Control of the False Discovery Rate in Multiple Testing under Dependency. The Annals of Statistics. 2001 2001-01-01;29(4):1165–88.

28. Spigel DR, Chaft JE, Gettinger S, et al. FIR: Efficacy, Safety, and Biomarker Analysis of a Phase II Open-Label Study of Atezolizumab in PD-L1-Selected Patients With NSCLC. J THORAC ONCOL. 2018 2018-11-01;13(11):1733–42.

29. Horn L, Spigel DR, Gettinger SN, et al. Clinical activity, safety and predictive biomarkers of the engineered antibody MPDL3280A (anti-PDL1) in non-small cell lung cancer (NSCLC): update from a phase Ia study. J CLIN ONCOL. 2015 2015-01-01;33(15_suppl):8029.

30. Peng D, Wang L, Li H, et al. An immune infiltration signature to predict the overall survival of patients with colon cancer. IUBMB LIFE. 2019 2019-11-01;71(11):1760–70.

31. Karn T, Jiang T, Hatzis C, et al. Association Between Genomic Metrics and Immune Infiltration in Triple-Negative Breast Cancer. JAMA ONCOL. 2017 2017-12-01;3(12):1707–11.

32. Oliva M, Spreafico A, Taberna M, et al. Immune biomarkers of response to immune-checkpoint inhibitors in head and neck squamous cell carcinoma. ANN ONCOL. 2019 2019-01-01;30(1):57–67.

33. Senbabaoglu Y, Gejman RS, Winer AG, et al. Erratum to: Tumor immune microenvironment characterization in clear cell renal cell carcinoma identifies prognostic and immunotherapeutically relevant messenger RNA signatures. GENOME BIOL. 2017;18(1):46. 2017-03-01.

34. Lv M, Xu Y, Tang R, et al. miR141-CXCL1-CXCR2 signaling-induced Treg recruitment regulates metastases and survival of non-small cell lung cancer. MOL CANCER THER. 2014 2014-12-01;13(12):3152–62.

35. Koyama K, Kagamu H, Miura S, et al. Reciprocal CD4 + T-cell balance of effector CD62Llow CD4 + and CD62LhighCD25 + CD4 + regulatory T cells in small cell lung cancer reflects disease stage. CLIN CANCER RES. 2008 2008-11-01;14(21):6770–9.

36. Vigano S, Perreau M, Pantaleo G, Harari A. Positive and negative regulation of cellular immune responses in physiologic conditions and diseases. Clin Dev Immunol. 2012 2012-01-20;2012:485781.

37. Fruh M, Peters S. Genomic Features of Response to Combination Immunotherapy in Lung Cancer. CANCER CELL. 2018 2018-05-14;33(5):791–3.

38. Duffy MJ, Crown J. Biomarkers for Predicting Response to Immunotherapy with Immune Checkpoint Inhibitors in Cancer Patients. CLIN CHEM. 2019 2019-10-01;65(10):1228–38.

39. Fehrenbacher L, Spira A, Ballinger M, et al. Atezolizumab versus docetaxel for patients with previously treated non-small-cell lung cancer (POPLAR): a multicentre, open-label, phase 2 randomised controlled trial. LANCET. 2016 2016-04-30;387(10030):1837–46.
40. Narayanan S, Kawaguchi T, Yan L, Peng X, Qi Q, Takabe K. Cytolytic Activity Score to Assess Anticancer Immunity in Colorectal Cancer. ANN SURG ONCOL. 2018 2018-08-01;25(8):2323–31.

41. Narayanan S, Kawaguchi T, Yan L, Peng X, Qi Q, Takabe K. Cytolytic Activity Score to Assess Anticancer Immunity in Colorectal Cancer. ANN SURG ONCOL. 2018 2018-08-01;25(8):2323–31.

42. Chae YK, Anker JF, Bais P, Namburi S, Giles FJ, Chuang JH. Mutations in DNA repair genes are associated with increased neo-antigen load and activated T cell infiltration in lung adenocarcinoma. Oncotarget. 2018 2018-01-30;9(8):7949–60.

43. Rousseaux S, Debernardi A, Jacquiau B, et al. Ectopic activation of germline and placental genes identifies aggressive metastasis-prone lung cancers. SCI TRANSL MED. 2013 2013-05-22;5(186):166r-186r.

44. Steven A, Seliger B. The Role of Immune Escape and Immune Cell Infiltration in Breast Cancer. Breast Care (Basel). 2018 2018-03-01;13(1):16–21.

45. Botling J, Edlund K, Lohr M, et al. Biomarker discovery in non-small cell lung cancer: integrating gene expression profiling, meta-analysis, and tissue microarray validation. CLIN CANCER RES. 2013 2013-01-01;19(1):194–204.

46. Horn LA, Riskin J, Hempel HA, et al. Simultaneous inhibition of CXCR1/2, TGF-beta, and PD-L1 remodels the tumor and its microenvironment to drive antitumor immunity. J IMMUNOTHER CANCER. 2020 2020-03-01;8(1).

47. Weber R, Meister M, Muley T, et al. Pathways regulating the expression of the immunomodulatory protein glycodelin in nonsmall cell lung cancer. INT J ONCOL. 2019 2019-02-01;54(2):515–26.

48. Rosenthal R, Cadieux EL, Salgado R, et al. Neoantigen-directed immune escape in lung cancer evolution. NATURE. 2019 2019-03-01;567(7749):479–85.

Figures
Figure 1

TME clusters of LUSC in the TCGA cohort. A. Unsupervised clustering of tumor microenvironment immune cells for 494 LUSC patients from TCGA. Clinicopathological information including location, status, sex, age, stage, as well as TME clusters, are shown in annotations above. B, Kaplan-Meier curves for OS stratified by TME clusters in the TCGA cohort. C. The fraction of immune cells in TMEcluster A and B in patients from TCGA. Within each group, the scattered dots represent the GSVA score of immune cells. The thick lines in the boxes represent the median value. The bottom and top of the boxes are the 25th and 75th percentiles (interquartile range). The whiskers encompass 1.5 times the interquartile range. The statistical difference of two TME clusters was compared through the Wilcoxon test. *, p<0.05; **, p<0.01;
***, p<0.001; ****, p<0.0001. D. The forest plot showing the prognostic value of each type of immune cell with a corresponding hazard ratio (HR) and 95% confidence interval (95% CI). E. The Sankey diagram describing the distribution of baseline information in TMEcluster A and B.

Figure 2

The random forest classifier model validating data from GEO and our institution. A-B, The relative abundance of 24 types of immune cells in LUSC patients from GEO (A) and our institution (B). C-D, The fraction of immune cells in TMEcluster A and B in patients from GEO (C) and our institution (D). Within each group, the scattered dots represent the GSVA score of immune cells. The thick lines in the boxes represent the median value. The bottom and top of the boxes are the 25th and 75th percentiles. The
whiskers encompass 1.5 times the interquartile range. The statistical difference of two TME clusters was compared through the Wilcoxon test. *, p<0.05; **, p<0.01; ***, p<0.001; ****, p<0.0001. E-F, Kaplan-Meier curves for overall survival (OS) stratified by TME clusters in GEO (E) and our institution (F).

Figure 3

Differentially expressed genes (DEGs) and functional annotation between different TME clusters. A, The volcano plot showing the DEGs between TME clusters A and B. B-E, Functional enrichment analyses of the DEGs including GO (B, D) and KEGG (C, E).
Figure 4

Immune microenvironment characteristics between different TME clusters in patients from TCGA. A, Violin plots showing the expression profiling of the 8 immune-related genes in the POPLAR study and cytolytic activity (CYT) score. The differences between every two groups were compared through the Wilcoxon test. P values indicated. B, The relative expression level of molecules potentially involved in the initiation of innate immunity (left) and MHC-I/II antigen-presenting process (right). C, Violin plots showing
the CD8+ T cells/Treg ratio and aDC/iDC ratio in the two TME clusters. D-E, The relative expression level of immune co-inhibitors (D) and co-stimulators (E). F, The box plots showing the expression pattern of the four important immune checkpoint molecules in TME clusters A and B. Within each group, the thick lines in the boxes represent the median value. The bottom and top of the boxes are the 25th and 75th percentiles. The whiskers encompass 1.5 times the interquartile range. The statistical difference of two TME clusters was compared through the Wilcoxon test. *, p<0.05; **, p<0.01; ***, p<0.001; ****, p<0.0001. G, Correlation coefficient between tumor immunogenicity, immune infiltration, immune checkpoint molecules, and mutation load.

Figure 5

Tumor Genomic Alterations associated with TME clusters. A. Violin plots showing the mutation load and copy number variation in the two clusters. The differences between the two clusters were compared through the Wilcoxon test. P values indicated. B. The waterfall plots summarizing the genomic alterations including somatic mutations and copy number variations in the two TME clusters. The type of alterations was annotated by different colors (bottom).

Supplementary Files

This is a list of supplementary files associated with this preprint. Click to download.
• supplementTable3.csv
• supplementfig5.tif
• supplementfig4.tif
• supplementfig3.tif
• supplementfig2.tif
• supplementFig1.tif
• supplementTable5.csv
• supplementfig6.tif
• supplementTable1.csv
• supplementTable4.csv
• supplementTable2.xlsx