

Mine TCGA Database for Tumor Microenvironment-Related Genes of Prognostic value in lung squamous cell carcinoma

Xiao Chen

1.Shandong University Qilu Hospital 2.Tai'an City Central Hospital

Rui Li

Shandong University Qilu Hospital

Yun-Hong Yin

Shandong University Qilu Hospital

Xiao Liu

Shandong University Qilu Hospital

Xi-Jia Zhou

Shandong University Qilu Hospital

Yi Qing Qu (✉ quyiqing@sdu.edu.cn)

Shandong University Qilu Hospital <https://orcid.org/0000-0002-9538-7601>

Research

Keywords: TCGA, LUSC, tumor microenvironment, TME, TME-related genes, prognosis.

Posted Date: May 13th, 2020

DOI: <https://doi.org/10.21203/rs.3.rs-28076/v1>

License: © ⓘ This work is licensed under a Creative Commons Attribution 4.0 International License. [Read Full License](#)

Abstract

Background: Tumor microenvironment (TME) plays a significant role in the development of cancer. However, the roles of TME in lung squamous cell carcinoma (LUSC) are not well studied. In our study, we aimed to identify differentially expressed tumor microenvironment-related genes as biomarker for predicting the prognosis of LUSC.

Methods: We combined The Cancer Genome Atlas (TCGA) and Estimation of Stromal and Immune cells in Malignant Tumor tissue using Expression data (ESTIMATE) datasets to identify differentially expressed genes in lung squamous cell carcinoma microenvironment. Then, functional enrichment analysis and protein-protein interaction (PPI) network were conducted. The top six genes in the PPI network were regarded as tumor microenvironment-related hub genes. Finally, the relationship between hub genes and tumor-infiltrating immune cells was deciphered using TIMER.

Results: Our study revealed that immune and stromal scores are associated with specific clinicopathologic variables in LUSC. These variables include gender, age, distant metastasis and prognosis. In addition, a total of 874 upregulated and 72 downregulated genes were identified. Functional enrichment analysis demonstrated a correlation between DEGs and the tumor microenvironment, tumor immune cells differentiation and activation. C3AR1, CSF1R, CCL2, CCR1, TYROBP, CD14 were selected as the hub genes. A positive correlation was obtained between the expression of hub genes and the abundance of six immune cells.

Conclusions: The results of the present study showed that ESTIMATE algorithm-based stromal and immune scores may be a reference indicator of cancer prognosis. We identified five TME-related genes, which could be used to predict the prognosis of LUSC patients.

Background

Lung cancer is the deadliest malignant cancer and the leading cause of cancer death in the world with about 2.09 million new cases and 1.76 million deaths each year [1, 2]. According to histological examination, approximately 85% of lung cancer are non-small cell lung cancer (NSCLC), of which lung adenocarcinoma (LUAD) and lung squamous cell carcinoma (LUSC) are the most common subtypes [3, 4]. In different types of lung cancer, LUSC has the second highest incidence of cell carcinoma and mortality, which have been increasing in recent years [5]. Platinum-based dual chemotherapy has been the cornerstone of first-line treatment in patients with advanced NSCLC, however, this pattern changed in 2016 with the outcome of the iconic KEYNOTE-024 trial, which improved overall survival (OS) by using pembrolizumab [6, 7]. Targeted therapies against some mutated genes, such as EGFR, ALK, ROS1, HER2, AKT1, RET, BRAF have proved to bring better clinical outcomes to adenocarcinoma patients, but the prognostic biomarker for LUSC patients is little known [8, 9]. Therefore, efforts towards the identification of LUSC are still needed, which are important for the development of targeted therapies.

Tumor microenvironment (TME) not only is an attractive target to develop novel strategies for cancers but also plays a key role during tumor initiation, progression and metastasis [10]. There are many kinds of cells in TME including various cell types (endothelial cells, immune cells, etc.) and extracellular cells (cytokines, growth factors, extracellular matrix, etc.) that are surrounding tumor cells and nourished by vascular networks [11]. Immune cells and stromal cells play a critical role in tumor message transfer, immune surveillance and tumor niche information [12]. The prospect of effective immunotherapies for cancer is now becoming a clinical reality [13]. Immune score as a prognostic factor has become a part of tumor diagnosis and prognosis [14]. Study has predicted prognosis and reflect the immune microenvironment of LUAD [15]. The ESTIMATE (Estimation of Stromal and Immune cells in Malignant Tumor tissues using Expression data) algorithm is a new method for predicting tumor purity, and the presence of infiltrating stromal/immune cells in tumor tissues by analysis gene expression data [16]. The ESTIMATE algorithm has been applied to estimate the diagnostic and prognostic values of immune and stromal cells in many tumor types, including renal, thyroid cancer and glioblastoma [17–19]. However, the proportion of stromal/immune cells in LUSC has not been explored using the ESTIMATE algorithm.

The Cancer Genome Atlas (TCGA) is an open project, contain genomic, epigenomic, transcriptomic, and proteomic data [20]. In this study, RNA expression patterns and clinical data including gender, age, tumor stage and survival state of LUSC patients were extracted from the TCGA database. The relationship between the ESTIMATE scores and clinical data was explored. Meanwhile, the relationship between tumor microenvironment related genes and the outcomes of LUSC patients was also examined. Besides, the hub genes were identified and their correlation with immune cell infiltration were evaluated.

Methods

Microarray data

The gene expression profiles of lung squamous cell carcinoma (LUSC) patients were extracted from the TCGA database (<https://portal.gdc.cancer.gov/>), including 502 LUSC tissues and 49 non-LUSC tissues. FPKM data was downloaded for differential analysis. Extracted clinical data included age, gender, tumor stage, survival state, and outcome. After the data were downloaded and transposed, the Estimation of Stromal and Immune cells in Malignant Tumor tissue using Expression data (ESTIMATE) algorithm was performed, and immune and stromal scores were calculated [16].

Identification of differentially expressed genes

Evidence for differential gene expression was analyzed using the “limma R” package (<http://www.bioconductor.org/packages/release/bioc/html/limma.html>) [21]. Differentially expressed genes (DEGs) were screened and obtained using the cut-off values of fold change (log FC) > 1 and adjusted the false discovery rate (FDR) to a P value < 0.05. Based on the cut-off values, patients were divided into low- and high-stromal/immune score groups. Group comparisons of stromal/immune scores between different clinical indexes were performed by R package. P value ≥ 0.05 was considered statistically significantly different.

Heatmaps and clustering analysis

Clustering analysis and the construction of immune and stromal heatmaps were done using the pheatmap R package(<https://cran.r-project.org/web/packages/pheatmap/index.html>).

Enrichment analysis of DEGs

The DEGs were filtered with R software using the VennDiagram package (<https://cran.r-project.org/web/packages/VennDiagram/index.html>).

The overlapping DEGs would be used for further analysis. Gene Ontology (GO) analysis and Kyoto Encyclopedia of Genes and Genomes (KEGG) pathways were used to perform the functional annotation of DEGs. GO analysis was done for assessment of biological processes (BP), molecular functions (MF), and cellular components (CC). GO, and KEGG pathway analyses results were processed by the “clusterProfiler”(<http://www.bioconductor.org/packages/release/bioc/html/clusterProfiler.html>), “enrichplot”(<http://www.bioconductor.org/packages/release/bi>) and “ggplot2” packages. A P value of less than 0.05 was used regarded as statistically significant.

Identification of hub genes

The PPI network of overlapping DEGs would be obtained from STRING (<https://string-db.org/> ver11.0, update Jan 2019) [22]. Active interaction sources were set as follows: Textmining, Experiments, Databases, Co-expression, Neighborhood, Gene Fusion and Co-occurrence. The required minimum interaction score was set at 0.4. The barplot were generated by the R software (<https://www.r-project.org/ver 3.6.2>). The top six DEGs were selected as the tumor microenvironment-related hub genes.

Overall survival Analysis

The prognostic value of hub genes was evaluated using Gene Expression Profiling Interactive Analysis (GEPIA(<http://gepia.cancer-pku.cn/>)). GEPIA is a newly developed interactive web server for cancer and normal gene expression profiling and analyses[23]. The P value ≤ 0.05 was set as the cut-off value.

The Human Protein Atlas

All the data in the Human Protein Atlas(<https://www.proteinatlas.org/>) are open and composed of six independent parts, the Tissue Atlas shows the distribution of protein in all major tissues and organs of the human body[24]. We analyzed the hub gene protein expression in LUSC via immunohistochemistry (IHC) analysis.

Hub genes and tumor-infiltrating immune cells

The TIMER (<https://cistrome.shinyapps.io/timer/>) is a web resource for systematical evaluations of the clinical impact of different immune cells in diverse cancer types. The relationship between hub genes and tumor-infiltrating immune cells (B cell, CD4 T cell, CD8 T cell, neutrophil, macrophage and dendritic cell) was deciphered using TIMER [25].

Results

Patient characteristics

The data on gene expression profiles and complete clinicopathologic variables of 484 LUSC case were extracted from the TCGA database in Jan 2020. The detail demographic and baseline characteristic of 484 LUSC patients are described in Table 1.

Table 1
The baseline characteristics of all patients.

Variables	Subtype	Count	Percent(%)
Age(years)	<60	87	17.98
	≥ 60	397	82.02
Gender	Female	126	26.03
	Male	358	73.97
Stage	I	238	49.17
	II	157	32.44
	III	82	16.94
	IV	7	1.45
T(Tumor)	T1	110	22.73
	T2	284	58.68
	T3	68	14.05
	T4	22	4.54
N(Node)	N0	308	63.64
	N1	127	26.24
	N2	38	7.85
	N3	5	1.03
	Nx	6	1.24
M(Metastasis)	M0	401	82.85
	M1	7	1.45
	Mx	76	15.70
Total		484	100

Table 2
30 pathways with significant differences in the KEGG database.

ID	Description	p-value	p.adjust	q-value	Count
hsa04640	Hematopoietic cell lineage	4.66E-29	1.21E-26	9.52E-27	45
hsa04061	Viral protein interaction with cytokine and cytokine receptor	1.76E-25	2.28E-23	1.80E-23	42
hsa05150	Staphylococcus aureus infection	3.06E-25	2.64E-23	2.08E-23	41
hsa04060	Cytokine-cytokine receptor interaction	2.32E-24	1.50E-22	1.18E-22	70
hsa04514	Cell adhesion molecules (CAMs)	1.83E-21	9.46E-20	7.46E-20	46
hsa04672	Intestinal immune network for IgA production	2.08E-19	8.99E-18	7.09E-18	26
hsa04145	Phagosome	1.22E-16	3.98E-15	3.14E-15	41
hsa04062	Chemokine signaling pathway	1.23E-16	3.98E-15	3.14E-15	46
hsa05340	Primary immunodeficiency	4.08E-15	1.17E-13	9.26E-14	20
hsa05323	Rheumatoid arthritis	8.83E-15	2.28E-13	1.80E-13	30
hsa05310	Asthma	9.69E-15	2.28E-13	1.80E-13	18
hsa05140	Leishmaniasis	1.93E-14	4.16E-13	3.28E-13	27
hsa04610	Complement and coagulation cascades	3.05E-13	6.08E-12	4.80E-12	27
hsa05330	Allograft rejection	1.06E-12	1.95E-11	1.54E-11	18
hsa05322	Systemic lupus erythematosus	1.68E-12	2.90E-11	2.29E-11	33
hsa05332	Graft-versus-host disease	5.36E-12	8.67E-11	6.83E-11	18
hsa05320	Autoimmune thyroid disease	1.04E-11	1.58E-10	1.25E-10	20
hsa04380	Osteoclast differentiation	1.60E-11	2.31E-10	1.82E-10	31
hsa05321	Inflammatory bowel disease (IBD)	9.72E-11	1.32E-09	1.04E-09	21
hsa04658	Th1 and Th2 cell differentiation	1.08E-10	1.40E-09	1.10E-09	25
hsa04659	Th17 cell differentiation	1.23E-10	1.52E-09	1.20E-09	27
hsa05152	Tuberculosis	1.35E-10	1.59E-09	1.25E-09	36
hsa04940	Type I diabetes mellitus	1.60E-10	1.80E-09	1.42E-09	17
hsa05144	Malaria	2.86E-10	3.08E-09	2.43E-09	18
hsa05416	Viral myocarditis	1.13E-09	1.18E-08	9.27E-09	19
hsa04662	B cell receptor signaling pathway	1.90E-09	1.89E-08	1.49E-08	22
hsa05133	Pertussis	2.48E-09	2.37E-08	1.87E-08	21
hsa04612	Antigen processing and presentation	4.15E-09	3.84E-08	3.03E-08	21
hsa05145	Toxoplasmosis	9.43E-09	8.42E-08	6.64E-08	25
hsa04064	NF-kappa B signaling pathway	1.42E-07	1.22E-06	9.64E-07	22

Evaluation of Immune Scores and Stromal Scores

We downloaded the complete gene expression profiles and clinicopathologic information of all 484 LUSC patients from the TCGA database. Based on the ESTIMATE algorithm, stromal scores ranged from - 2230.11 to 1845.98, immune scores ranged from - 1188.16 to 3455.61, while the ESTIMATE scores ranged from - 2993.00 to 4961.47. Female LUSC patients had higher immune scores ($P = 2.886e-04$) and ESTIMATE scores ($P = 0.007$) than male LUSC patients (Fig. 1a). Moreover, immune scores ($P = 0.009$), stromal scores ($P = 0.017$), and ESTIMATE scores ($P = 0.007$) increased with an increase in age (Fig. 1b). LUSC patients with distant metastases had higher immune scores (Fig. 1c, $P = 0.88$, median scores 1034.73 VS 1016.71), stromal scores (Fig. 1c, $P = 0.307$, median scores 56.57 VS -95.23), and ESTIMATE scores (Fig. 1c, $P = 0.666$, median scores 1253.27 VS 948.14), although statistically not significant. Besides, Stage IV cases have higher stromal scores (median scores - 25.95 VS -65.06) and ESTIMATE scores (median scores 1038.32 VS 956.59) compared with stage I, although statistically not significant (Fig. 1d). These results indicate that immune scores, stromal scores and ESTIMATE scores are associated with specific clinicopathologic variables.

Kaplan-Meier survival curves were plotted to find out the potential correlation of overall survival with immune, stromal, and ESTIMATE scores. Firstly, 484 cases of LUSC patients were divided into high and low score groups based on their scores. Our results of ESTIMATE scores demonstrate a more prolonged overall survival of LUSC patients in the low score group compared with the high score group (Fig. 2a, median 539 d VS 587 d, $P = 0.05$ in log-rank test). Consistently, stromal (Fig. 2b, median 513.5 d VS 585 d, $P = 0.066$) and immune (Fig. 2c, median 535 d VS 581 d, $P = 0.219$) score results showed longer

median overall survival of LUSC patients in the lower score group compared to the higher score group, although the differences were not statistically significant.

Comparison of gene expression profiles with immune scores and stromal scores in LUSC

We compared the data of 484 LUSC patients extracted from the TCGA database to reveal the correlation of gene expression profiles in immune scores and stromal scores. The stromal score based heatmaps are presented in Fig. 2d, which demonstrated that 1195 and 136 genes were up-downregulated in the high score than low score groups, respectively (fold change > 1, $P < 0.05$). The immune score based heatmaps are presented in Fig. 2e, which showed 1187 and 176 up-downregulated genes in the high than low score groups, respectively (fold change > 1, $P < 0.05$). Moreover, Venn diagrams (Fig. 2f,g) showed that 874 genes were commonly upregulated, and 72 genes were commonly downregulated.

Functional enrichment analysis of DEGs

Functional enrichment analyses by applying the clusterProfiler R package were conducted to determine the functions of the DEGs. Figure 3a shows the top 10 biological processes (BP), the top 10 cellular components (CC) and the top 10 molecular functions (MF) identified following GO analyses. The ten enriched BPs include T cell activation, leukocyte migration, leukocyte proliferation, leukocyte cell-cell adhesion, lymphocyte proliferation, mononuclear cell proliferation, regulation of lymphocyte activation, positive regulation of cell activation, regulation of T cell activation, positive regulation of cytokine production. All the above BPs are involved in immune cells differentiation and activation. The ten enriched CCs include external side of plasma membrane, secretory granule membrane, collagen-containing extracellular matrix, MHC class II protein complex, tertiary granule, tertiary granule membrane, MHC protein complex, ficolin-1-rich granule membrane, clathrin-coated endocytic vesicle, specific granule. All the above BPs are involved in extracellular matrix and membrane. The ten enriched MFs include carbohydrate binding cytokine receptor activity, cytokine binding, immunoglobulin binding, glycosaminoglycan binding, chemokine activity, chemokine receptor binding, cytokine activity, CCR chemokine receptor binding, heparin binding. Also, the above MFs are associated with surface receptor activity and protein binding. Notably, KEGG analysis (Fig. 3b) suggested that most DEG-related pathway were significantly linked to immune response.

PPI network and identification of hub genes

The present study used the online STRING tool to explore protein-protein interaction (PPI) networks among significant target genes. The functional network contained 153 nodes and 281 edges (Fig. 3c). According to the PPI network, the top six significant genes (C3AR1, CSF1R, CCL2, CCR1, TYROBP, CD14) were selected as hub genes (Fig. 3d).

The roles of hub genes in overall survival in LUSC

This study plotted Kaplan-Meier survival curves to identify the underlying functions of hub genes in the overall survival of LUSC patients. The log-rank test of six hub genes showed remarkably predict poor overall survival in LUSC except TYROBP (Fig. 4a, $P < 0.05$). And all six genes expression was significant (Fig. 4b, $P < 0.05$).

Immunohistochemistry of the five hub genes

Immunohistochemistry of the five hub genes were collected based on the Human Protein Atlas. The protein levels of 4 genes were higher in tumor tissues compared with normal tissues. Immunohistochemistry of CCR1 were not found. (Fig. 5a) Protein levels of C3AR1 in normal tissue and tumor tissue. (Fig. 5b) Protein levels of CSF1R in

normal tissue and tumor tissue. (Fig. 5c) Protein levels of CCL2 in normal tissue and tumor tissue. (Fig. 5d) Proteins level of CD14 in normal tissue and tumor tissue. **The correlation between hub genes and immune cell infiltration**

The present study further analyzed the correlation between hub genes and immune cell infiltration. Our results revealed positive correlations between hub genes expression and immune cell infiltration (C3AR1 Fig. 6a, CSF1R Fig. 6b, CCL2 Fig. 6c, CCR1 Fig. 6d, CD14 Fig. 6e. Details are given in the Table 3.

Table 3
The correlation between hub genes and immune cell infiltration.

variable		C3AR1	CCL2	CCR1	CD14	CSF1R
B Cell	cor	0.365651381	0.238504235	0.385981383	0.264846997	0.34936269
	p	2.23E-16	1.57E-07	3.23E-18	5.12E-09	5.38E-15
CD8 + T Cell	cor	0.579654708	0.378469615	0.570719289	0.475588796	0.548176303
	p	5.52E-44	1.27E-17	2.12E-42	3.52E-28	1.31E-38
CD4 + T Cell	cor	0.270780268	0.217909542	0.286747243	0.261777733	0.410674865
	p	1.99E-09	1.63E-06	1.92E-10	6.97E-09	9.44E-21
Macrophage	cor	0.637486498	0.316436969	0.582452639	0.539590977	0.650202114
	p	9.21E-56	1.49E-12	1.14E-44	2.16E-37	1.18E-58
Neutrophil	cor	0.628934582	0.518157654	0.659862947	0.568640283	0.636583372
	p	8.81E-54	4.65E-34	7.99E-61	4.01E-42	1.90E-55
Dendritic Cell	cor	0.812257106	0.548273796	0.823855115	0.76262123	0.858242191
	p	1.43E-112	1.51E-38	1.85E-118	2.40E-91	8.02E-139

Discussion

Worldwide, lung cancer remains a highly lethal disease and the most common cause of cancer death [26]. NSCLC remains a major cause of cancer-related mortality in both men and women with survival rate of 19.3%[27]. In American, the rate of decline in lung cancer mortality has accelerated, with men dropping from 3% a year during 2008 through 2013 to 5% during 2013 through 2018 and women from 2% to nearly 4% ,spurring the biggest one-year drop in overall cancer mortality of 2.2% from 2016 to 2017. However, lung cancer still has more death in 2017 than breast, prostate, colorectal and brain cancers combined [2]. LUSC and LUAD are the two major subtypes of lung cancer, in which LUSC progresses faster than LUAD of the same stage [28]. Tumor-related prognosis or predictive markers in the immune microenvironment significantly changed the treatment prospects of lung cancer and significantly improved the prognosis of some patients, especially those with adenocarcinoma [29–33].

The tumor microenvironment consists of a complex tissue system through which immune and stromal cells can regulate anti-tumor immunogenicity. Recently, many studies have shown prognostic markers in tumor microenvironment for tumor prognosis. Zeng et al provides potential TME-related biomarker for the therapy and prognosis of RCC[34]. Zeng et al describe the comprehensive features of gastric cancer TME characteristics and provide new strategies for cancer treatment[35]. A present study found that many elements of TME other than tumor epithelial cells influence the progression of HNSCC[36]. Another study highlights the favorable prognostic impact of an immune-inflamed TME in LUAD [37]. There is also studies analyzed the immune microenvironment of stage I LUAD and found that the high expression of IL-7R and the lower expression of IL-12Rβ2 were all independent factors of poor prognosis[38]. Yue et al identified three TME-related DEGs signature which could be used to predict the prognosis of LUAD patients[39]. Chen et al revealed the differentiated regulation of immune-response related genes between LUAD and LUSC[28]. However, genome changes in squamous cell lung cancer have not been fully characterized and no molecular targeted drugs have been specifically developed for the treatment [40]. Moreover, prognostic biomarkers related to TME in LUSC is still lacking. In this study, we explore differentially expressed TME-related genes and reveal prognostic biomarkers for predicting the diagnosis and prognosis of LUSC.

The present study extracted data from the TCGA database to explore TME-related genes that affect the overall survival and tumorigenesis of LUSC. Subsequently, we identified prognostic and therapeutic biomarkers for LUSC. This study first assessed the relationship between clinicopathological variables and immune and stromal scores in LUSC. Our results suggest that immune score and stromal score are associated with specific clinicopathological variables (age, gender, tumor stage). In addition, immune and stromal score showed great potential in prognostic prediction with LUSC patients.

After distinguishing the high and low immune/stromal scores groups, we identified 946 differentially expressed genes (DEGs). The functional enrichment analyses demonstrated that these DEGs are involved in immune response, immune cells differentiation and activation, extracellular matrix and membrane. For example, GO analysis revealed that DEGs enriched T cell activation, leukocyte migration, leukocyte proliferation, leukocyte cell-cell adhesion, plasma membrane, secretory granule membrane, collagen-containing extracellular matrix, MHC class II protein complex, cytokine binding, immunoglobulin binding, chemokine activity, chemokine receptor binding, cytokine activity, CCR chemokine receptor binding. Also, the above MFs are associated with surface receptor activity and protein binding. This study also showed that DEGs were most enriched in immune-related typical pathway: cytokine-cytokine receptor interaction, cell adhesion molecules (CAMs), intestinal immune network for IgA production, Phagosome, chemokine signaling pathway. Our results confirmed previous reports on genomic alterations and tumor microenvironment impact cancer proliferation and invasion and roles of immune cells and their interactions with cancer cells in LUSC [41–43].

In addition, we further clarified the interplay among DEGs by using the STRING tool to construct protein-protein interaction (PPI) networks. The results demonstrated that all the significant genes are involved in immune response, and the top six genes (C3AR1, CSF1R, CCL2, CCR1, TYROBP, CD14) were identified as hub genes. One research suggested that the exclusion of CD8 T cells from the tumor islets was associated with poorer clinical outcomes and lower lymphocyte activity in LUSC patients, but CSF1R blockade enhanced the migration and infiltration of CD8 T cells to the tumor islets [44]. CSF1R inhibitors represent a new immune-modulatory in tumor therapy [45]. A previous study reported that CCL2 overexpression was associated with total survival

and progression-free survival in patients with LUSC [46]. Wang et al showed that CCR1 knockdown suppresses NSCLC cell invasion[47]. Although reports about C3AR1, and CD14 are lacking, the between those genes and immune-associated disease and other tumors has been reported [48,49]. These data suggest that our results based on the TCGA database have predictive value. Notably, most of the genes we identified have not been previously reported in LUSC, suggesting that these genes should be further validated.

Increasing evidence suggested a significant correlation between immune cell infiltration and patient outcomes [50,51]. The current study revealed positive correlations between hub genes expression and immune cell infiltration (C3AR1, CSF1R, CCL2, CCR1, TYROBP, CD14). Therefore, these genes may provide more information about immune cell infiltration and clinical outcomes in LUSC patients.

Many studies have explored the correlation between gene expression and tumorigenesis and prognosis of NSCLC or LUAD in tumor/immune microenvironment. Recently, there are several treatment methods for targeting TME-related genes and have got good results in clinical or clinical trial. We firstly used the ESTIMATE algorithm to explore the association between immune/stromal scores and LUSC clinicopathologic and got TME-related genes as the novel potential biomarkers for LUSC. Although our study has achieved very valuable results, this study was conducted in small cohorts and has significant limitations and clinical trials still need.

Conclusions

In conclusion, targeted therapy and immunotherapy progress rapidly, however, the overall progression-free survival rate in LUSC patients remains unsatisfactory. To improve the clinical outcomes of LUSC patients, there is an urgent need to monitor biomarkers of targeted treatment response. Using bioinformatics methods, we focused on TME-related genes of LUSC based on TCGA database and screened five potential prognostic genes (C3AR1, CSF1R, CCL2, CCR1, CD14), two of which (C3AR1 and CD14) lack relevant research in LUSC. Although the five prognostic markers identified in our study may still need lots of clinical trials to validation, these genes may provide prospects for prognostic evaluation of LUSC patients.

Abbreviations

BP	Biological processes
CC	Cellular components
DEGs	Differentially expressed genes
ESTIMATE	Estimation of Stromal and Immune cells in Malignant Tumor tissue using Expression data
GO	Gene Ontology
GEPIA	Gene Expression Profiling Interactive Analysis
KEGG	Kyoto Encyclopedia of Genes and Genomes
LUAD	Lung adenocarcinoma
LUSC	Lung squamous cell carcinoma
MF	Molecular functions
NSCLC	Non-small cell lung cancer
OS	Overall survival
PPI	Protein-protein interaction
TCGA	The Cancer Genome Atlas
TIMER	Tumor Immune Estimation Resource
TME	Tumor microenvironment

Declarations

Authors' contributions

YQQ and XC designed the study. RL and HYY performed the data collection. XL and JXZ analyzed the data. XC drafted the manuscript. YQQ revised the manuscript. All authors read and approved the final manuscript.

Acknowledgements

Not applicable.

Competing interests

The authors declare that they have no competing interests.

Availability of data and materials

Publicly available databases were analyzed in our study. The gene expression data from TCGA can be found in <https://portal.gdc.cancer.gov/>. Tumor-infiltration immune cells from TIMER is available in <https://cistrome.shinyapps.io/timer/>.

Consent for publication

Not applicable.

Ethics approval and consent to participate

Consent for participation for all patients was obtained through The Cancer Genome Atlas Project.

Funding

This work was supported by the start-up fund from the National Natural Science Foundation of China (81372333), the Science and Technology Foundation of Shandong Province (2014GSF118084), the Major Scientific and Technological Innovation Project of Shandong Province (2018CXGC1212), the CSCO-Qilu Cancer Research Fund (Y-Q201802-014) and the Medical and Health Technology Innovation Plan of Jinan City (201805002).

References

1. Bray F, Ferlay J, Soerjomataram I, Siegel RL, Torre LA, Jemal A. Global cancer statistics 2018: GLOBOCAN estimates of incidence and mortality worldwide for 36 cancers in 185 countries. *CA Cancer J Clin.* 2018;68(6):394-424.
2. Siegel RL, Miller KD, Jemal A. Cancer statistics, 2020. *CA Cancer J Clin.* 2020;70(1):7-30.
3. Relli V, Trerotola M, Guerra E, Alberti S. Abandoning the Notion of Non-Small Cell Lung Cancer. *Trends Mol Med.* 2019;25(7):585-94.
4. Herbst RS, Morgensztern D, Boshoff C. The biology and management of non-small cell lung cancer. *Nature.* 2018;553(7689):446-54.
5. Gao C, Zhuang J, Zhou C, Ma K, Zhao M, Liu C, et al. Prognostic value of aberrantly expressed methylation gene profiles in lung squamous cell carcinoma: A study based on The Cancer Genome Atlas. *J Cell Physiol.* 2019;234(5):6519-28.
6. Chiang AC, Herbst RS. Frontline immunotherapy for NSCLC - the tale of the tail. *Nat Rev Clin Oncol.* 2020;17(2):73-4.
7. Reck M, Rodriguez-Abreu D, Robinson AG, Hui R, Csoszi T, Fulop A, et al. Pembrolizumab versus Chemotherapy for PD-L1-Positive Non-Small-Cell Lung Cancer. *N Engl J Med.* 2016;375(19):1823-33.
8. Zhang P, Kang B, Xie G, Li S, Gu Y, Shen Y, et al. Genomic sequencing and editing revealed the GRM8 signaling pathway as potential therapeutic targets of squamous cell lung cancer. *Cancer Lett.* 2019;442:53-67.
9. Shames DS, Wistuba, II. The evolving genomic classification of lung cancer. *J Pathol.* 2014;232(2):121-33.
10. Dieterich LC, Bikfalvi A. The tumor organismal environment: Role in tumor development and cancer immunotherapy. *Semin Cancer Biol.* 2019.
11. Wu T, Dai Y. Tumor microenvironment and therapeutic response. *Cancer Lett.* 2017;387:61-8.
12. Zhang X, Yuan X, Shi H, Wu L, Qian H, Xu W. Exosomes in cancer: small particle, big player. *J Hematol Oncol.* 2015;8:83.
13. Gajewski TF, Schreiber H, Fu YX. Innate and adaptive immune cells in the tumor microenvironment. *Nat Immunol.* 2013;14(10):1014-22.
14. Galon J, Pagès F, Marincola FM, et al. The immune score as a new possible approach for the classification of cancer. *J Transl Med* 2012;10:1. DOI:10.1186/1479-5876-10-1.
15. Yue C, Ma H, Zhou Y. Identification of prognostic gene signature associated with microenvironment of lung adenocarcinoma. *PeerJ.* 2019;7.
16. Yoshihara K, Shahmoradgolji M, Martinez E, Vegesna R, Kim H, Torres-Garcia W, et al. Inferring tumour purity and stromal and immune cell admixture from expression data. *Nat Commun.* 2013;4:2612.
17. Zhao K, Yang H, Kang H, Wu A. Identification of Key Genes in Thyroid Cancer Microenvironment. *Med Sci Monit.* 2019;25:9602-8.
18. Li S, Xu W. Mining TCGA database for screening and identification of hub genes in kidney renal clear cell carcinoma microenvironment. *J Cell Biochem.* 2019.
19. Jia D, Li S, Li D, et al. Mining TCGA database for genes of prognostic value in glioblastoma microenvironment. *Aging (Albany NY).* 2018;10:592-605. DOI:10.18632/aging.101415.
20. Tomczak K, Czerwińska P, Wiznerowicz M. Review The Cancer Genome Atlas (TCGA): an immeasurable source of knowledge. *Współczesna Onkologia.* 2015;1A:68-77.
21. Ritchie ME, Phipson B, Wu D, Hu Y, Law CW, Shi W, et al. limma powers differential expression analyses for RNA-sequencing and microarray studies. *Nucleic Acids Res.* 2015;43(7):e47.
22. Szklarczyk D, Gable AL, Lyon D, Junge A, Wyder S, Huerta-Cepas J, et al. STRING v11: protein-protein association networks with increased coverage, supporting functional discovery in genome-wide experimental datasets. *Nucleic Acids Res.* 2019;47(D1):D607-D13.
23. 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;45(W1):W98-W102.

24. Thul PJ, Lindskog C. The human protein atlas: A spatial map of the human proteome. *Protein Sci.* 2018;27(1):233-44.
 25. Li T, Fan J, Wang B, Traugh N, Chen Q, Liu JS, et al. TIMER: A Web Server for Comprehensive Analysis of Tumor-Infiltrating Immune Cells. *Cancer Res.* 2017;77(21):e108-e10.
 26. Nasim F, Sabath BF, Eapen GA. Lung Cancer. *Med Clin North Am.* 2019;103(3):463-73.
 27. Woodard GA, Jones KD, Jablons DM. Lung Cancer Staging and Prognosis. *Cancer Treat Res.* 2016;170:47-75.
 28. Chen M, Liu X, Du J, et al. Differentiated regulation of immune-response related genes between LUAD and LUSC subtypes of lung cancers. *Oncotarget.* 2017;8:133-144. DOI:10.18632/oncotarget.13346.
 29. Leonetti A, Facchinetti F, Rossi G, Minari R, Conti A, Friboulet L, et al. BRAF in non-small cell lung cancer (NSCLC): Pickaxing another brick in the wall. *Cancer Treat Rev.* 2018;66:82-94.
 30. Fernandez-Cuesta L, McKay JD. Genomic architecture of lung cancers. *Curr Opin Oncol.* 2016;28(1):52-7.
 31. Hisamatsu Y, Morinaga R, Watanabe E, Ohtani S, Shirao K. Febrile Neutropenia in a Patient with Non-Small Cell Lung Cancer Treated with the Immune-Checkpoint Inhibitor Nivolumab. *Am J Case Rep.* 2020;21:e920809.
 32. Casaluze F, Sgambato A, Maione P, Spagnuolo A, Gridelli C. Lung cancer, elderly and immune checkpoint inhibitors. *J Thorac Dis.* 2018;10(Suppl 13):S1474-S81.
 33. Youn B, Trikalinos NA, Mor V, Wilson IB, Dahabreh IJ. Real-world use and survival outcomes of immune checkpoint inhibitors in older adults with non-small cell lung cancer. *Cancer.* 2020.
 34. Zeng Q, Zhang W, Li X, Lai J, Li Z. Bioinformatic identification of renal cell carcinoma microenvironment-associated biomarkers with therapeutic and prognostic value. *Life Sci.* 2020;243:117273.
 35. Zeng D, Li M, Zhou R, Zhang J, Sun H, Shi M, et al. Tumor Microenvironment Characterization in Gastric Cancer Identifies Prognostic and Immunotherapeutically Relevant Gene Signatures. *Cancer Immunol Res.* 2019;7(5):737-50.
 36. Curry JM, Sprandio J, Cognetti D, Luginbuhl A, Bar-ad V, Pribitkin E, et al. Tumor microenvironment in head and neck squamous cell carcinoma. *Semin Oncol.* 2014;41(2):217-34.
 37. Givuchian KB, Garner C, Benz S, et al. An immunogenic NSCLC microenvironment is associated with favorable survival in lung adenocarcinoma. *Oncotarget.* 2019;10:1840-1849. DOI:10.18632/oncotarget.
- 26748.
38. Kadota K, Nitadori JI, Adusumilli PS. Prognostic value of the immune microenvironment in lung adenocarcinoma. *Oncoimmunology.* 2013;2(5):e24036.
 39. Song Q, Shang J, Yang Z, Zhang L, Zhang C, Chen J, et al. Identification of an immune signature predicting prognosis risk of patients in lung adenocarcinoma. *J Transl Med.* 2019;17(1):70.
 40. Cancer Genome Atlas Research N. Comprehensive genomic characterization of squamous cell lung cancers. *Nature.* 2012;489(7417):519-25.
 41. Desrichard A, Kuo F, Chowell D, Lee KW, Riaz N, Wong RJ, et al. Tobacco Smoking-Associated Alterations in the Immune Microenvironment of Squamous Cell Carcinomas. *J Natl Cancer Inst.* 2018;110(12):1386-92.
 42. Seo JS, Lee JW, Kim A, Shin JY, Jung YJ, Lee SB, et al. Whole Exome and Transcriptome Analyses Integrated with Microenvironmental Immune Signatures of Lung Squamous Cell Carcinoma. *Cancer Immunol Res.* 2018;6(7):848-59.
 43. Jiang T, Shi J, Dong Z, Hou L, Zhao C, Li X, et al. Genomic landscape and its correlations with tumor mutational burden, PD-L1 expression, and immune cells infiltration in Chinese lung squamous cell carcinoma. *Journal of Hematology & Oncology.* 2019;12(1).
 44. Peranzoni E, Lemoine J, Vimeux L, Feuillet V, Barrin S, Kantari-Mimoun C, et al. Macrophages impede CD8 T cells from reaching tumor cells and limit the efficacy of anti-PD-1 treatment. *Proc Natl Acad Sci U S A.* 2018;115(17):E4041-E50.
 45. Cannarile MA, Weisser M, Jacob W, Jegg AM, Ries CH, Ruttinger D. Colony-stimulating factor 1 receptor (CSF1R) inhibitors in cancer therapy. *J Immunother Cancer.* 2017;5(1):53.
 46. Li L, Liu YD, Zhan YT, Zhu YH, Li Y, Xie D, et al. High levels of CCL2 or CCL4 in the tumor microenvironment predict unfavorable survival in lung adenocarcinoma. *Thorac Cancer.* 2018;9(7):775-84.
 47. Wang CL, Sun BS, Tang Y, Zhuang HQ, Cao WZ. CCR1 knockdown suppresses human non-small cell lung cancer cell invasion. *J Cancer Res Clin Oncol.* 2009;135(5):695-701.
 48. Su J, Cui J, Xue HT, et al. Study on the correlation between CD14 gene polymorphism and susceptibility to laryngeal cancer. *Eur Rev Med Pharmacol Sci.* 2017;21:4292-4297.
 49. Wu SY, Fan J, Hong D, et al. C3aR1 gene overexpressed at initial stage of acute myeloid leukemia-M2 predicting short-term survival. *Leuk Lymphoma.* 2015;56:2200-2202. DOI:10.3109/10428194.2014.986481.
 50. Li J, Liu C, Chen Y, Gao C, Wang M, Ma X, et al. Tumor Characterization in Breast Cancer Identifies Immune-Relevant Gene Signatures Associated With Prognosis. *Front Genet.* 2019;10:1119.
 51. Kikuchi T, Mimura K, Ashizawa M, Okayama H, Endo E, Saito K, et al. Characterization of tumor-infiltrating immune cells in relation to microbiota in colorectal cancers. *Cancer Immunol Immunother.* 2020;69(1):23-32.

Figures

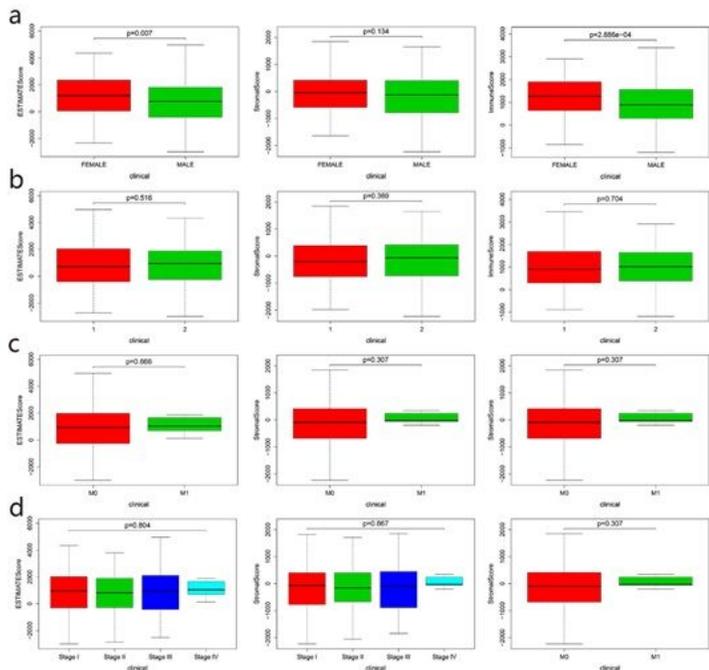


Figure 1
 Association between immune/stromal scores and clinical characteristics in LUSC. a A significant association was revealed between gender and the level of immune scores ($P=2.886e-04$) and ESTIMATE scores ($P=0.007$). b A significant association was revealed between age (1 means age ≤ 60 2 means age ≥ 60) and the level of immune scores ($P=0.009$), stromal scores ($P=0.017$), and ESTIMATE scores ($P=0.007$). c LUSC patients with distant metastases had higher immune scores ($P=0.88$, median scores 1034.73 VS 1016.71), stromal scores ($P=0.307$, median scores 56.57 VS -95.23), and ESTIMATE scores ($P=0.666$, median scores 1253.27 VS 948.14), although statistically not significant. d Stage IV cases have higher stromal scores (median scores -25.95 VS -65.06) and ESTIMATE scores (median scores 1038.32 VS 956.59) compared with stage I, although statistically not significant.

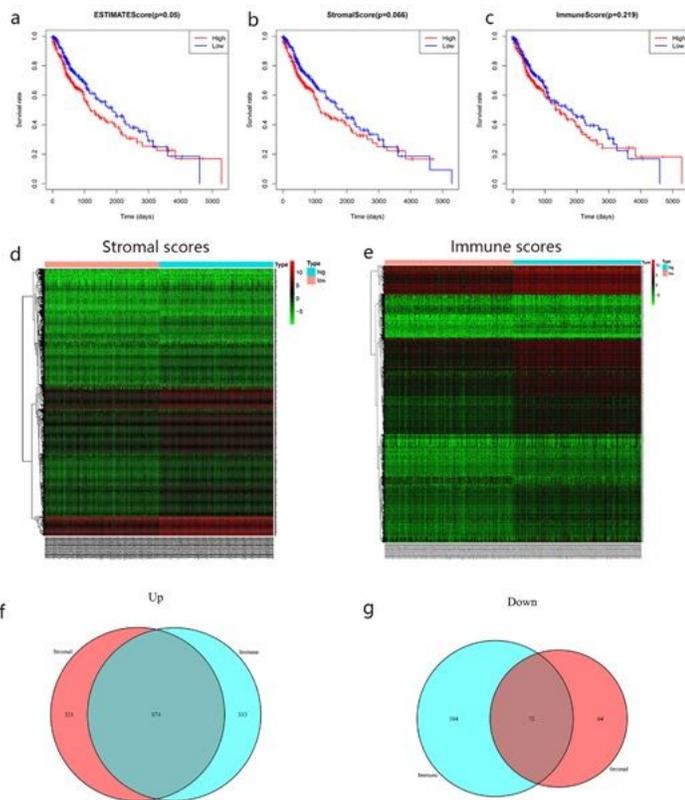


Figure 2

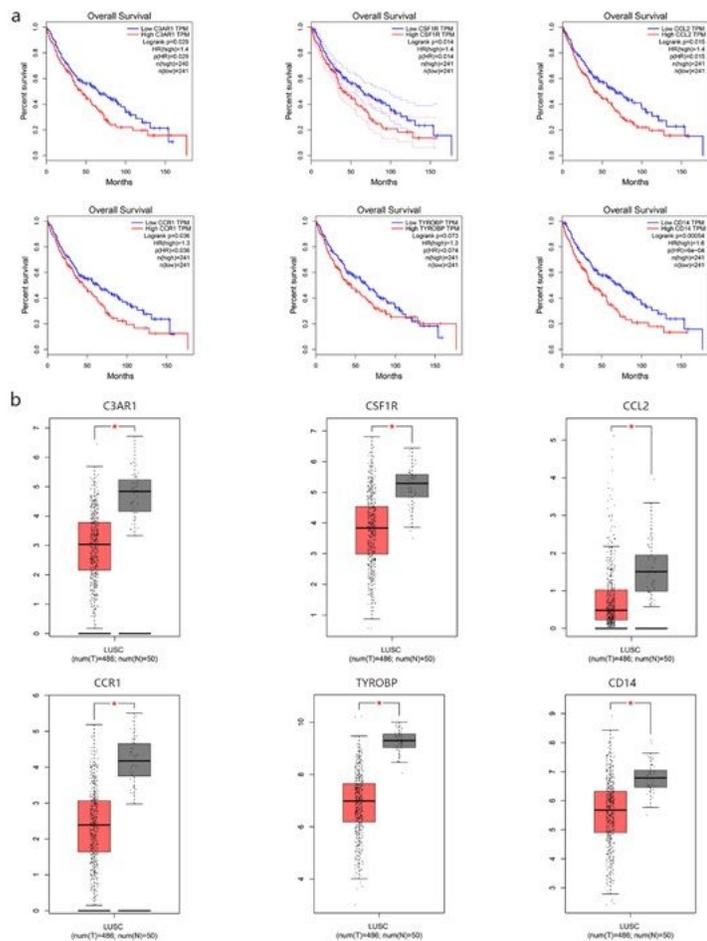


Figure 4

Correlation of expression of hub genes in overall survival in LUSC. a Kaplan-Meier survival curves were generated for hub genes extracted from the comparison of groups of high (red line) and low (blue line) gene expression. b shown Expression on Box Plots of tumor (red) and normal (grey).

