A survival model for prognostic prediction based on ferroptosis-associated genes and the association with immune infiltration in Lung squamous cell carcinoma

Yanyi Lu
The Second Affiliated Hospital of Zunyi Medical University

Yunliang Cao
The Second Affiliated Hospital of Zunyi Medical University

Yunan Wang
The Second Affiliated Hospital of Zunyi Medical University

Mengjia Wu
The Second Affiliated Hospital of Zunyi Medical University

Bo He
The Second Affiliated Hospital of Zunyi Medical University

Junzhu Xu
The Second Affiliated Hospital of Zunyi Medical University

Zixuan Su
The Second Affiliated Hospital of Zunyi Medical University

Wei Hu (✉ huweikeyan@163.com)
The Second Affiliated Hospital of Zunyi Medical University

Research Article

Keywords: Lung squamous cell carcinoma, Ferroptosis, Prognosis, Immune infiltration, TCGA, GEO (Public gene expression profile data set)

Posted Date: May 5th, 2022

DOI: https://doi.org/10.21203/rs.3.rs-1597903/v1

License: Creative Commons Attribution 4.0 International License. Read Full License
Abstract

Lung squamous cell carcinoma (LUSC) is the primary pathological type of lung cancer with a less favourable prognosis. This study attempts to construct a ferroptosis-associated signature associated with OS, which can predict the prognosis of LUSC and explore its relationship with immune infiltration. A 5 ferroptosis-associated genes model was constructed by the LASSO-penalized regression analysis to predict the prognosis of patients with LUSC in the TCGA database and validated in GEO and TCGA databases. Patients were stratified into high-risk and low-risk groups by median value of the risk scores, and the former prognosis was significantly worse (P<0.001). Additionally, we found a certain association between the two risk groups and immune infiltration through CIBERSORT. Meanwhile, the differentially expressed genes (DEGs) between normal and tumor tissue were used to perform functional analysis, which shows significant association with leukocyte transendothelial migration pathways in the TCGA cohort. Besides, immune cell infiltration analysis confirmed that M2 macrophages were significantly highly expressed in the high-risk group. Overall, the model successfully established by ferroptosis-associated genes suggests that ferroptosis may be related to immune infiltration in LUSC.

1. Introduction

With an estimated 2.2 million new cases and 1.8 million deaths, lung cancer is the second most commonly diagnosed cancer and the leading cause of cancer death in 2020 [1]. According to histological types, lung cancers are mainly classified into small cell lung cancer (SCLC) and non-small cell lung cancer (NSCLC). NSCLC represents 87% of all lung cancers, and approximately one-third of NSCLC are reported as LUSC [2]. LUSC is associated with poor clinical prognosis and lacks targeted agents or other treatments compared to lung adenocarcinoma, although immunotherapy comes into play [3, 4]. The reasons for the dismal prognosis are the absence of valid therapeutic targets and a shortage of effective prognostic biomarkers for guidance on cancer therapy. Therefore, it is essential to identify a prognostic model, and further research on this basis to find possible factors affecting the prognosis to provide treatment options and improve the prognosis in patients with LUSC. The rapid development of Bioinformatics analysis has enabled the access to gene expression and clinical information through different databases, such as TCGA, GEO and ICGC, etc. These resources provide a novel way to establish a prognostic prediction model to evaluate prognosis in LUSC, which can also be used to give more ideas and directions for in-depth internal mechanism research and then guide clinical practice.

Ferroptosis was recently identified form differs from apoptosis and other primary forms of regulated cell death (RCD) in many aspects as a new way of cell death by Brent R. Stockwell’s laboratory in 2012 [5]. Morphologically, ferroptosis occurs mainly in cells as reduced mitochondrial volume, increased bilayer membrane density and reduction or disappearance of mitochondrial cristae [6]; Biochemically, there is intracellular glutathione (GSH) depletion and decreased activity of glutathione peroxidase 4 (GPX4), resulting in a large amount of ROS, which promotes ferroptosis [7]; Genetically, ferroptosis is a biological process regulated by multiple genes.
Although the contribution of ferroptosis to physiology remains obscure, its role in a variety of human pathological states such as neurotoxicity, traumatic brain injury, and malignancies has been confirmed [8, 9]. Previous studies have analyzed the role of ferroptosis in malignancies from different perspectives. On the one hand, some studies believe that ferroptosis might suppress the immune system and allow part of tumor cells growth, because it affects antitumor immunity by cooperating with multiple different cell types present in the tumor microenvironment [10]. On the other hand, there are also research consider ferroptosis can be induced by tumor microenvironment-responsive multistaged liposome through amplifying oxidative stress for inhibited tumor growth and cancer cell proliferation [11].

Ferroptosis has gradually shown attractive antitumor effects. The activation of malignant tumour cells' death or growth restriction caused by different ferroptosis inducers or induction methods confirms its inhibitory influence on malignancies. For example, inhibition of sigma 1 receptor (S1R), which is heavily expressed in hepatocytes, can promote ferroptosis in hepatocellular carcinoma (HCC) cells [12]. In triple-negative breast cancer (TNBC) cells, inhibition of activation of the MUC1-C/xCT signaling pathway can induce ferroptosis [7].

In previous cases, constructed model based on ferroptosis-associated genes can be used for prognostic prediction in malignancies and provide ideas for us to take a closer look [13–16]. This study tried to establish and verify a prognostic prediction model based on ferroptosis-associated DEGs from the TCGA database. Additionally, we investigated the correlation between tumor immune microenvironment, including the immune cell infiltration and the enrichment of immune-related pathways with the two risk groups that were stratified based on the risk score of this model.

2. Results

2.1. Identification of prognostic ferroptosis-associated DEGs in the TCGA cohort

A total of 502 LUSC patient's genes expression and clinical information from the TCGA cohort and 247 LUSC patient's genes expression from the GEO cohort were enrolled. The detailed clinical characteristics of these patients are summarized in Table 1. We have obtained 60 ferroptosis-associated genes through previous research, 52 DEGs of them were differentially expressed between tumor and normal tissues and 5 of them were correlated with OS in the univariate Cox regression analysis in TCGA cohort (Fig. 1a). Take the intersection of the 52 DEGs and 5 prognostic related genes from the previous step, a total of 5 prognostic and ferroptosis-associated DEGs were screened out (Fig. 1b). The heat map shows the differential expression of the 5 genes between tumor and normal tissues, ALOX5 and DPP4 were significantly higher in normal tissues, while FADS2, NOX1, and PHKG2 were significantly higher in LUSC tissues (Fig. 1c). We used the Wilcoxon-rank sum test to analyze the relationship between the 5 ferroptosis-associated genes expression in different tissues respectively, and the results showed that the 5 ferroptosis-associated genes expression was significantly differences in LUSC tissues and normal tissues in the TCGA cohort, the result shows that ALOX5 and DPP4 were significantly higher in normal
tissues, while FADS2, NOX1, and PHKG2 were significantly higher in LUSC tissues (all adjusted P < 0.05) (Fig. S1). Subsequently, we used the Wilcoxon signed-rank test to determine the 5 ferroptosis-associated genes expression in 46 LUSC tissues and matched adjacent normal tissues in the TCGA cohort. The expression of ALOX5 and DPP4 were significantly higher in 46 matched adjacent normal tissues than in LUSC tissues, adversely, FADS2, NOX1, and PHKG2 were significantly higher in matched adjacent normal tissues, which is consistent with the previous unpaired analysis results (all adjusted P < 0.05) (Fig. S2). Because these 5 genes are significantly related to OS performed by univariate Cox regression analysis (Fig. 1a), Kaplan–Meier survival analysis was performed to examine the influence of different expression levels of these 5 genes on prognosis of LUSC respectively. The results showed that patients with low PHKG2 expression experienced a shorter OS duration than those with high PHKG2 expression (P = 0.026) and the patients with high DPP4 expression experienced a shorter OS duration than those with low DPP4 expression (P = 0.043) (Fig. S3). Meanwhile, we also explored the relationship between the 5 genes and different clinical characteristics respectively. Among those 5 genes, the expression of ALOX5 was obviously correlated with part of T stage and N stage and gender; the expression of DPP4 was obviously correlated with part of N stage; the expression of NOX1 was obviously correlated with part of T stage (all adjusted P < 0.05). (Fig. S4). The interaction network among these 5 genes shows that there is a certain correlation between them, and the correlation coefficient is 0.4 (Fig. 1d). PPI network analysis was performed using the online database STRING with interaction score of 0.4 as the threshold. There are 4 connects and 4 nodes, further proves the correlation between the 5 genes. (Fig. 1e).
Table 1
Clinical characteristics of the LUSC patients used in this study.

<table>
<thead>
<tr>
<th></th>
<th>TCGA cohort</th>
<th>Merge cohort</th>
</tr>
</thead>
<tbody>
<tr>
<td>No. of patients</td>
<td>493</td>
<td>740</td>
</tr>
<tr>
<td>Age (median, range)</td>
<td>68(39–85)</td>
<td>66 (37–75)</td>
</tr>
<tr>
<td>Gender (%)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Female</td>
<td>128(26%)</td>
<td>180 (24%)</td>
</tr>
<tr>
<td>Male</td>
<td>365(74%)</td>
<td>560 (76%)</td>
</tr>
<tr>
<td>Stage (%)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>I</td>
<td>241</td>
<td>351</td>
</tr>
<tr>
<td>II</td>
<td>158</td>
<td>237</td>
</tr>
<tr>
<td>III</td>
<td>83</td>
<td>124</td>
</tr>
<tr>
<td>IV</td>
<td>7</td>
<td>17</td>
</tr>
<tr>
<td>unknown</td>
<td>4</td>
<td>11</td>
</tr>
<tr>
<td>Survival status</td>
<td></td>
<td></td>
</tr>
<tr>
<td>OS days(median)</td>
<td>660</td>
<td>620</td>
</tr>
</tbody>
</table>

2.2. Construction of a prognostic model in the TCGA cohort

We further used the LASSO Cox regression analysis to minimize the risk of overfitting. A risk model by these 5 ferroptosis-associated genes expression profile was selected in the LASSO, thus the predictive accuracy could be improved significantly. In detail, A 5-gene signature was identified based on the optimal value of $\lambda$ (Fig. S5). The optimal cut-off expression value of each gene indicated that high expression of ALOX5, DPP4, and FADS2 are correlated with a poor prognosis while NOX1 and PHKG2 are the opposite. The risk score was calculated as follows: \[ r = 0.116 \times \text{expression level of ALOX5} + 0.037 \times \text{expression level of DPP4} + (-0.306) \times \text{expression level of PHKG2} + 0.122 \times \text{expression level of FADS2} + (-0.526) \times \text{expression level of NOX1}. \] The patients were stratified into a high-risk group ($n = 246$) or a low-risk group ($n = 247$) according to the median value of the risk score (Fig. 2a). The high-risk group was found to be significantly associated with higher TNM stage, higher age in TCGA cohort (Table 2). In order to verify whether the model can effectively distinguish patients in different groups, we realize visualization through dimensionality reduction processing, PCA and t-SNE analysis indicated the patients in different risk groups were distributed in two directions (Fig. 2b-c) and patients with high-risk had a higher probability of death earlier than those with low-risk (Fig. 2d). Consistently, the Kaplan-Meier curve indicated that patients in the high-risk group had a significantly worse OS than their low-risk counterparts.
(Fig. 2e, P < 0.001). The predictive performance of the risk score for OS was evaluated by time-dependent ROC curves. The area under the curve (AUC) reached 0.67 at 1 year, 0.74 at 2 years, and 0.80 at 3 years proved the performance of this prognostic model (Fig. 2f).

Table 2
Baseline characteristics of the patients in different risk groups.

| Characteristics | TCGA cohort | | | |
|-----------------|-------------|----------|--------|
|                 | High risk   | Low risk | P value|
| Gender (%)      |             |          | 0.169  |
| Female          | 65 (26.4)   | 63 (25.5)|        |
| Male            | 181 (73.6)  | 184 (74.5)|       |
| Age (%)         |             |          | 0.042  |
| <60y            | 48 (19.5)   | 42 (17.0)|        |
| ≥ 60y           | 196 (79.7)  | 202 (81.8)|       |
| unknown         | 2 (0.8)     | 3 (1.2)  |        |
| TNM stage (%)   |             |          | 0.012  |
| +               | 199 (80.9)  | 200 (81.0)|       |
| +/−             | 44 (17.9)   | 46 (18.6)|        |
| unknown         | 3 (1.2)     | 1 (0.4)  |        |

2.3. Validation of the 5-gene signature in the GEO and TCGA cohort

We test the robustness of the model constructed by a cohort merged from GEO dataset and TCGA database, the patients from the merged cohort were also categorized into two risk groups by the median value calculated with the same formula as that from the TCGA cohort (Fig. 3a). Similar to the results obtained from the TCGA cohort, PCA and t-SNE analysis confirmed that patients in two subgroups were distributed in discrete directions (Fig. 3b-c). Likewise, patients in the high-risk group were encounter death earlier (Fig. 3d) and had a reduced survival time compared with those in the low-risk group (Fig. 3e, P < 0.05). Besides, the AUC of the model was 0.66 at 1 year, 0.69 at 2 years, and 0.73 at 3 years (Fig. 3f) proves that this prediction model can effectively predict the prognosis of patients further.

2.4. Independent prognostic value of the 5-gene signature
Univariate and multivariate Cox regression analyses were carried out among the available variables to determine whether the risk score was an independent prognostic predictor for OS. In univariate Cox regression analyses, the risk score was significantly associated with OS in the TCGA and the merged cohort (HR = 2.844, 95% CI = 1.794–4.508, P < 0.001; HR = 1.595, 95% CI = 1.299–1.959, P < 0.001) (Fig. 4a, c). After correction for other confounding factors, the risk score still proved to be an independent predictor for OS in the multivariate Cox regression analysis (HR = 2.846, 95% CI = 1.788–4.529, P < 0.001; HR = 1.466, 95% CI = 1.187–1.810, P < 0.001) (Fig. 4b, d).

2.5. GO enrichment and KEGG pathway analyses

To elucidate the biological functions and pathways that were associated with the risk score, the DEGs were used to perform GO enrichment and KEGG pathway analyses in TCGA cohort. The result shows that various pathways including phagosome, pyrimidine metabolism, viral protein interaction with cytokine and cytokine receptor, and cytokine–cytokine receptor interaction, especially some immune-related pathways such as leukocyte transendothelial migration pathway were significantly enriched in cohort by KEGG pathway analyses (all adjusted P < 0.05, Fig. 5d, e, f). Interestingly, the DEGs were also obviously enriched in many immune-related biological processes, including leukocyte migration, humoral immune response, myeloid leukocyte migration, and leukocyte chemotaxis through GO enrichment (all adjusted P < 0.05, Fig. 5a, b, c). Taking into account the correlation between ferroptosis and immunity, enrichment analysis gives evidence that the model may also be related to immunity. Thus, in order to further confirm the above speculation, GSEA was followed performed to reveal differences between the high and low 5 ferroptosis-associated genes expression cohorts respectively. The results shows that multiple immune related pathways were significantly enriched and the most significantly enriched signaling pathways was leukocyte transendothelial migration pathway in the high ALOX5 and DPP4 and low NOX1 and PHKG2 expression phenotype that was consistent with the results of the previous KEGG pathway analyses (Fig s6, s7).

2.6. Single gene immune analysis

Because of the significantly associated with immune in our model, we wonder whether the 5 genes that involved in this model are also related to immune infiltration. The result shows that the relationship between the 5 ferroptosis-associated genes expression and immune cell infiltration (Fig. s8 a-e). At the same time, we extracted the expression levels of these 5 genes in tumors symbol respectively and divided the samples into high- and low-risk groups based on the median value, and then compared the differences in the content of each immune cell in the two groups (Fig. s9 a-e). We got the most relevant immune cells for the 5 ferroptosis-associated gene from the intersection of the two set and visualized using Venn plots (Fig. s10 a-e). Finally, the most significantly immune cells that related to each of the 5 ferroptosis-associated genes we filter out, which are ALOX5: memory B cells, resting CD4 memory T cells, follicular helper T cells, regulatory T (Tregs) cells, activated dendritic cells, M2 macrophages, activated mast cells, and neutrophils; DPP4: resting CD4 memory T cells, follicular helper T cells, M0 macrophages, M2 macrophages and, neutrophils; FADS2: M2 macrophages, activated dendritic cells, and neutrophils; NOX1: CD8 T cells, resting CD4 memory T cells, follicular helper T cells, M0 macrophages and, M2
macrophages; PHKG2: CD8 T cells, resting CD4 memory T cells, follicular helper T cells, resting natural killer (NK) cells, activated NK cells, M2 macrophages, resting dendritic cells, and eosinophils.

### 2.7. Immune cell infiltration in different risk groups

Previous studies have revealed that tumor-infiltrating immune cells are related to prognosis [17]. In order to find more evidence to support the correlation between the high and low-risk groups and immune infiltration, we explored the infiltration of specific tumor immune cell subsets by CIBERSORT to get the 22 immune cells' abundance in different groups respectively. The radar charts depict a comparative summary of various immune cells in two risk groups (Fig. 6). We found that M2 macrophages and resting CD4 memory T cells were significantly highly expressed in the high-risk group, while follicular helper T cells, activated NK cells, activated dendritic cells were significantly higher in the low-risk group (Fig. 7). Therefore, we analyzed differences in immune infiltration between the tumor and normal tissues for 22 immune cells using CIBERSORT further. First, we presented the proportion of each immune cell in all samples using a bar plot (Figure s11b) and then used a heat map to compare the levels of immune cell infiltration between normal and LUSC tissues (Figure s11c). And low to moderate correlation was observed in various immunocyte subpopulations (figure s11a). To further explore the correlation between the risk score and immune status, we quantified the enrichment scores of diverse immune cell subpopulations, related functions or pathways with ssGSEA. To our surprise, all of them including immune cell subpopulations such as dendritic cells, B cells, CD8 + T cells, macrophages, mast cells, neutrophils, NK cells, helper T cells, tumor infiltrating lymphocyte, and Tregs, and related functions or pathways such as antigen-presenting cells (APC) co inhibition, APC co stimulation, chemokine receptor, check point, cytolytic activity, human leukocyte antigen, inflammation-promoting, major histocompatibility complex (MHC) class I, parainflammation, T cell co-inhibition, T cell co-stimulation, type I and type II interferons (IFN) response have higher scores in high-risk groups (Fig. 8, adjusted P < 0.001). The result reminds us that in LUSC, the ferroptosis-associated gene signature we construct is significantly correlated with immune infiltration. The results shown above all remind us that there are significant differences in the immune status of the high- and low-risk groups with this model in LUSC.

### 3. Discussion

LUSC, a malignancy associated with high mortality, is the leading cause of cancer death in 93 countries, and most countries are still observing a rising incidence of lung cancer [1]. To this end, there is an urgent need to develop a prognosis prediction model as a useful tool for prognosis predict and then provide guidance for clinical practice, may even dig out important information and conduct more in-depth mechanism discussions, then develop some new and effective treatment methods or targets.

In this study, we get 5 ferroptosis and prognosis associated DEGs from the 52 ferroptosis-associated DEGs and 5 prognosis related genes performed by univariate Cox analysis in LUSC from TCGA database, and then established a prognostic model using the expression profile of the 5 genes through LASSO Cox regression analysis and validated in an external cohort in order to ensure the authenticity and validity. Predictive power of the model was evaluated by the ROC curve analyses at last. The successful
establishment of this model suggests that there is a certain correlation between ferroptosis and the prognosis of LUSC patients. When we reviewed the 5 genes in the model further, we found that these 5 genes can be divided into two categories, including oxidant metabolism (ALOX5, NOX1, PHKG2) and lipid metabolism (DPP4, FADS2). These two types of genes play an important role in the occurrence and development of malignant tumors also confirms the scientific nature of our model in a certain way.

Ferroptosis as a new form of programmed cell death was mainly characterized by accumulation of reactive oxygen species (ROS) resulting from iron accumulation and lipid peroxidation [5]. Past research has shown that tumor cells can be induced to ferroptosis in many ways and then inhibit tumor growth. Furthermore, ferroptosis can trigger inflammation-associated immunosuppression in the tumor microenvironment, thus favouring tumor growth. In addition, previous studies have also shown that ferroptosis is also closely related to immunity and plays different roles in tumor cells through immunomodulatory. For one thing, ferroptosis exerts a cancer-promoting role in the immune system by suppressing antitumor immunity or causing its deficiency, such as ferroptotic cells can release some signals affect antigen-presenting cells (APCs) and other immune cells [18, 19]. For another thing, ferroptosis more often plays an anti-cancer role in the immune system, studies have revealed that ferroptosis can inhibit cancer cell proliferation and tumor growth by regulating tumor microenvironment (TME) and mediating the tumor suppressive activity [20].

Although studies have shown that ferroptosis has a certain significance with immune in tumorigenesis and development, it is still unclear whether it plays a same important role in LUSC. As we all know the occurrence and development of LUSC is significantly related to immunity [21]. In our study, DEGs were enriched in several immune-related biological process, such as leukocyte migration, myeloid leukocyte migration and leukocyte chemotaxis performed by GO enrichment. Besides, interestingly, we also found those biological process all act on leukocyte transendothelial migration pathway [22] which was indicated by KEGG pathway analyses and was consistent with the results of single gene enrichment analysis we did before. Trafficking of leukocytes is a key process for immune cell development and host defense. Leukocyte transendothelial migration (TEM) is a vital physiological process that occurs during both the adaptive and innate immune response and during routine immune surveillance can cause the occurrence of cancer further [23]. As the largest cells of the leukocyte family, particular macrophages, their infiltration often correlates with the aggravation of several diseases including cancers. Migration of macrophages across tissues can ensure efficient tissue infiltration [24]. Immune correlation analysis showed that M2 macrophages was significantly highly expressed in the high-risk group of this model, coincidentally, single gene immune cell infiltration analysis shows that almost all of the 5 genes are significantly related to immune cells especially M2 macrophages. M2 macrophages, an immune suppressor cells, can release suppressive factors such as ROS, to suppress T and NK cell functions and promote tumor growth and metastasis [25, 26]. Based on the above analysis, we have reason to speculate that M2 macrophages are regulated by the leukocyte transendothelial migration pathway to ensure their effective infiltration, thereby playing an immunosuppressive role in the immune microenvironment, and ultimately promoting tumor development in LUSC.
Moreover, many related functions or pathways such as type I and type II IFN response performed and parainflammation also have higher score in the high-risk group, they can play different roles in tumorigenesis and development such as promote tumor growth and proliferation and their migration to metastatic sites[27], evade immunosurveillance [28] and induce tumor cell cycle [29]. Notably, parainflammation, a low-grade form of inflammation, is widely prevalent in human cancer and relies mainly on macrophages in tissue [30, 31]. Although some high expression immune promotion related cells, functions and pathways was found in the high-risk group, considering that our model is used as a whole to evaluate the relationship between the two groups and immune infiltration, the difference in the prognosis of the two groups should be the result of their internal interaction, rather than attributed to a single factor alone.

In view of the above studies, we defined a prognostic model of 5 ferroptosis-associated genes and be independently associated with OS in both the derivation and validation cohorts although this study has some limitations such as small sample size. Functional analysis and immune cell infiltration analysis confirmed that there is a certain correlation between our model and immune infiltration. The underlying mechanisms between ferroptosis-associated genes and tumor immune infiltration in LUSC still poorly understood, nevertheless, this study providing us new ideas to explore the mechanism of initiation and progression of LUSC from the perspective of ferroptosis and immune cell infiltration.

4. Conclusions

In summary, our 5 ferroptosis-associated gene model can provide an evaluation reference for the survival outcome with LUSC. And this study proves that ferroptosis has a certain relationship with immune infiltration in LUSC, the infiltrating immune cells distinguished by this model in the high- and low-risk groups like M2 macrophages and immune-related pathways significantly related to the model such as leukocyte transendothelial migration. Based on this, we speculate that this model may alter the expression of M2 macrophages by regulating leukocyte transendothelial migration, thereby regulating the immune microenvironment of LUSC, and then affecting the prognosis of patients. This study can provide a new understanding of ferroptosis in LUSC’s development and progression. Given that our results are based on RNA seq technology, further research is needed to explore the prognostic value of this model.

5. Materials And Methods

5.1. Data Preparation

The RNA sequencing (RNA-seq) data and corresponding clinical information of 502 LUSC patients were downloaded from the TCGA database up to December 21, 2020 (https://portal.gdc.cancer.gov/repository). RNA-seq data and clinical information of another 249 tumor samples from the GEO database (GSE157009) were downloaded for analyses (https://www.ncbi.nlm.nih.gov/gds). The data from TCGA and GEO are both publicly available. So, this study was exempted from the approval of local ethics committees. The current research follows the
TCGA and GEO data access policies and publication guidelines. We use the scale method provided in the "limma" R package to normalize the gene expression profiles of each LUSC sample. Normalized read count values were used. 60 ferroptosis-associated genes were retrieved from the previous literature [32–35] and are provided in Supplementary Table S1.

5.2. Construction and validation of a prognostic ferroptosis-associated gene signature

We use the "limma" R package to identify the differentially expressed genes (DEGs) between tumor tissues and adjacent nontumorous tissues with a false discovery rate (FDR) < 0.05 in the TCGA cohort. The univariate Cox regression analysis was applied to investigate the association between ferroptosis-associated gene and overall survival (OS) of patients in the TCGA cohort. P values were adjusted by Benjamini& Hochberg (BH) correction. To minimize the risk of overfitting, the LASSO-penalized Cox regression analysis [36, 37] was used to construct a prognostic signature. The LASSO algorithm was used for variable selection and shrinkage with the "glmnet" R package. The independent variable in the regression was the normalized expression matrix of candidate prognostic DEGs, and the response variables were overall survival and status of patients in the TCGA cohort. Penalty parameter (λ) for the model was determined by tenfold cross-validation following the minimum criteria (i.e. the value of λ corresponding to the lowest partial likelihood deviance). The risk scores of the patients were calculated according to the normalized expression level of each gene and its corresponding regression coefficients. The formula was established as follows: \[ \text{score} = e^{\sum (\text{each gene's expression} \times \text{corresponding coefficient})} \]. The median value of the risk score stratified the patients into high-risk and low-risk groups. Based on the expression of genes in this model, PCA was performed with the "prcomp" function of the "stats" R package. Meanwhile, t-SNE was performed to explore the distribution of the two groups using the "Rtsne" R package. For the survival analysis, the optimal cut-off expression value of each ferroptosis-associated gene was determined by the "surv cutpoint" function of the "survminer" R package. The "survivalROC" R package was used to practice time-dependent ROC curve analyses to assess the predictive power of the gene signature.

5.3. Protein–protein interaction (PPI) network analysis

An interaction network for the overlapping prognostic DEGs was generated by the STRING database (version 11.0). STRING (https://string-db.org/) is a database that provides a function for predicted protein interactions, in which each PPI has one or more 'scores' that indicate the confidence in the interaction based on the available evidence. This score ranges from 0 to 1, with 1 being the highest possible confidence. The interaction relationships between the intersecting genes were acquired by PPI network analysis in STRING. Core genes were identified using the Cytoscape plug-in with the highest confidence (0.40) as a threshold.

5.4. Functional enrichment analysis

GO is a major bioinformatics tool for annotating genes and analyzing their biological processes including three independent branches: molecular function (MF), biological process (BP), and cellular component
KEGG is a database resource for understanding high-level functions and biological systems from large-scale molecular datasets generated by high-throughput experimental technologies. Using the R (version 4.0.2) software package ‘clusterProfiler’, ‘enrichplot’, and ‘ggplot2’ to perform GO and KEGG enrichment analyses based on the DEGs (\(|\log_{2}FC| \geq 1, FDR < 0.05\)) between the high-risk and low-risk groups. P values were adjusted with the BH method.

5.5. Gene set enrichment analysis (GSEA)

In this study, the reference gene set “c2.cp.kegg.v6.2.symbols.gmt” was downloaded from the Molecular Signatures Database (MSigDB) (http://software.broadinstitute.org/gsea/msigdb). We quarry significant pathways enriched in each 5 ferroptosis-associated genes performed by GSEA. Gene set arrangements were repeated 1,000 times for each analysis, and the expression level of 5 ferroptosis-associated gene were served as a phenotype label. The nominal P-value and normalized enrichment score (NES) were worked to analysis pathway enrichment. The NES, enrichment score (ES), false discovery rate (FDR) and P-value were considered four key statistics in the GSEA. A gene set was considered significantly enriched when the P-value was less than 0.05 and the FDR was less than 0.25.

5.6. single-sample gene set enrichment analysis (ssGSEA)

We performed ssGSEA to evaluate infiltrating score of 16 immune cells and the activity of 13 immune-related pathways through the "gsva" R package [38]. A log 2-foldchange was made between Gene expression profiles in the high- and low-risk groups. The difference immune status between the high- and low-risk groups was compared. Differences with an FDR-adjusted P < 0.05 were defined as significant. The annotated gene set file is provided in Supplementary Table S2.

5.7. Statistical analysis

We used student’s t-test compare the gene expression between tumor tissues and normal tissues. Differences in proportions were compared by the Chi-squared test. Mann-Whitney test with P values adjusted by the BH method was used to compare the ssGSEA scores of immune cells or pathways between the high- and low-risk group. Kaplan-Meier analysis was performed using the log-rank test to compare difference between the two groups. Univariate and multivariate Cox regression analyses were implemented to identify independent predictors of OS. All statistical analyses were performed with R software (Version 3.6.3) or R software (Version 4.0.3). If not specified above, a P value less than 0.05 was considered statistically significant, and all P values were two-tailed.

Abbreviations

LUSC: Lung squamous cell carcinoma; DEGs: differentially expressed genes; SCLC: small cell lung cancer; NSCLC: non-small cell lung cancer; RCD: regulated cell death; GPX4: glutathione peroxidase 4; GSH: glutathione; HCC: hepatocellular carcinoma; TNBC: triple-negative breast cancer; ROS: reactive oxygen species; APCs: antigen-presenting cells; TME: tumor microenvironment; FDR: false discovery rate; OS:
overall survival; BH: Benjamini & Hochberg; MF: molecular function; BP: biological process; CC: cellular component; ES: enrichment score; NES: normalized enrichment score.

**Declarations**

**Acknowledgments**

We would like to thank TCGA and GEO database for their support and the LUSC patients and healthy subjects for their contribution.

**Author Contributions**

Conception and design of study: Wei Hu. Acquisition of data: Yanyi Lu and Yunliang Cao. Analysis and/or interpretation of data: Yanyi Lu, Yunliang Cao and Mengjia Wu. Drafting the manuscript: Yanyi Lu. Revising the manuscript for important intellectual content: Wei Hu, Yunan Wang and Bo He. Approval of the version of the manuscript to be published: Wei Hu, Yanyi Lu, Yun-liang Cao, Mengjia Wu, Yunan Wang, Bo He, Lei Zhou and Wei Hu.

**Funding**

This research was funded by Guizhou Provincial Health Commission, grant number: gzwjkj2020-1-034, Science and Technology Department of Guizhou Province, grant number: ZK [2021]-452, Dr Zunyi Medical University, grant number: [2017] No.19, Dr Zunyi Medical University, grant number: [2015] No.51 and Zunyi Medical University School-level Education Reform, grant number: XJJG2021-45.

**Ethics approval and consent to participate**

Not applicable.

**Consent for publication**

The authors declare that they consent for publication.

**Competing interests**

The authors declare that they have no conflicts of interest.

**References**


Figures
Figure 1

Identification of the candidate ferroptosis-associated genes in the TCGA cohort. a. Forest plots showing the results of the univariate Cox regression analysis between gene expression and OS. b. Venn diagram to identify differentially expressed genes between tumor and adjacent normal tissue intersect with OS-correlated genes. c. The 5 overlapping genes were all difference expression between normal and tumor tissue. d. The correlation network of candidate genes. The correlation coefficients are represented by different colors (positive correlation is red, negative correlation is blue). e. The PPI network downloaded from the STRING database indicated the interactions among the candidate genes.
Figure 2

Prognostic analysis of the 5 gene signature model in the TCGA cohort. a. The distribution by median value of the risk scores with this model. b. PCA plot. c. t-SNE analysis. d. The distributions of OS status, OS and risk score. e. Kaplan-Meier curves for the OS of patients in the high-risk group and low-risk group. f. AUC of time-dependent ROC curves verified the prognostic performance of the risk score.
Figure 3

Prognostic analysis of the 5-gene signature model in the merged cohort. a. The distribution by median value of the risk scores with this model. b. PCA plot. c. t-SNE analysis. d. The distributions of OS status, OS and risk score. e. Kaplan-Meier curves for the OS of patients in the high-risk group and low-risk group. f. AUC of time-dependent ROC curves verified the prognostic performance of the risk score.
Figure 4

Results of the univariate and multivariate Cox regression analyses regarding OS in the TCGA derivation cohort (a, b) and the merged validation cohort (c, d).
Figure 5

Results of GO (a, b) and KEGG analyses (d, e). The circle plots show the enrichment relationship between genes and the main enriched terms in GO (c) and KEGG analyses (f).
Figure 6

The radar charts depict a comparative summary of various immune cells in these two risk groups (*P< 0.05, **P< 0.01, ***P< 0.001).
Figure 7

The abundance distribution of specific immune cells in different risk groups.

Figure 8
Comparison of the ssGSEA scores between different risk groups in the TCGA cohort. The scores of 16 immune cells (a) and 13 immune-related functions (b) are displayed in boxplots. CCR, cytokine-cytokine receptor. Adjusted P values were showed as: ns, not significant; *, (P< 0.05; **, P< 0.01; ***, P< 0.001).

Supplementary Files

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

- SupplementaryMaterials.zip