

# A Novel Ferroptosis-Related Genes Model for Prognosis Prediction of Lung Adenocarcinoma

Fei Li

Cancer Hospital of China Medical University: Liaoning Cancer Institute and Hospital

Dongcen Ge

Dalian Medical College: Dalian Medical University

Shu-lan Sun (✉ [sunshulan@cancerhosp-ln-cmu.com](mailto:sunshulan@cancerhosp-ln-cmu.com))

Cancer Hospital of China Medical University: Liaoning Cancer Institute and Hospital

<https://orcid.org/0000-0002-7624-1797>

---

## Primary research

**Keywords:** Ferroptosis, prognosis prediction model, Lung Adenocarcinoma

**Posted Date:** April 12th, 2021

**DOI:** <https://doi.org/10.21203/rs.3.rs-354347/v1>

**License:** © ⓘ This work is licensed under a Creative Commons Attribution 4.0 International License.

[Read Full License](#)

---

# Abstract

**Background.** Ferroptosis is a newly discovered form of cell death characterized by iron-dependent lipid peroxidation. The aim of this study is to investigate the relationship between ferroptosis and the prognosis of lung adenocarcinoma (LUAD).

**Methods.** RNA-seq data was collected from the LUAD dataset of The Cancer Genome Atlas (TCGA) database. We used ferroptosis-related genes as the basis, and identify the differential expression genes (DEGs) between cancer and paracancer. The univariate Cox regression analysis were used to screen the prognostic-related genes. We divided the patients into training and validation sets. Then, we screened out key genes and built a 5 genes prognostic prediction model by the applications of the least absolute shrinkage and selection operator (LASSO) 10-fold cross-validation and the multi-variate Cox regression analysis. We divided the cases by the median value of risk score and validated this model in the validation set. Meanwhile, we analyzed the somatic mutations, and estimated the score of immune infiltration in the high- and low-risk groups, as well as performed functional enrichment analysis of DEGs.

**Results.** The result revealed that the high-risk score triggered the worse prognosis. The maximum area under curve (AUC) of the training set and the validation set of in this study was 0.7 and 0.69. Moreover, we integrated the age, gender, and tumor stage to construct the composite nomogram. The charts indicated that the AUC of cases with survival time of 1, 3 and 5 years are 0.698, 0.71 and 0.73. In addition, the mutation frequency of patients in the high-risk group was higher than that in the low-risk group. Simultaneously, DEGs were mainly enriched in ferroptosis-related pathways by analyzing the functional results.

**Conclusion.** This study constructed a novel LUAD prognosis prediction model base on 5 ferroptosis-related genes, which can provide a prognostic evaluation tool for the clinical therapeutic decision.

## Background

The ferroptosis is an iron-dependent form of regulated cell death (RCD) that is discovered in recent years that may be varied from apoptosis, necrosis, and autophagy (1). The implementation of ferroptosis requires the activation of the following three ferroptosis features: (A) the oxidation of phospholipids containing polyunsaturated fatty acids (PUFA), (B) the presence of redox active iron, and (C) the loss of lipid peroxide repair abilities(2). With the deepening analyzing of ferroptosis, the induction of ferroptosis has emerged as a factor that is relevant with multiple human pathology including tumors. Furthermore, preliminary evidence suggests that ferroptosis plays a vital role in regulating carcinoma growth, such as renal cancer, pancreatic cancer, non-small-cell lung cancer (NSCLC) and diffuse large B-cell lymphoma(3). Multiple previous evidences suggest that ferroptosis suppresses tumor growth and progression and the induction of ferroptosis has emerged as a promising therapeutic alternative to trigger cancer cell death(4).

Lung cancer ranks first in terms of incidence among malignancies and is the chief leading cause of tumor-related death worldwide. 85% of lung cancer may be NSCLC, of which lung adenocarcinoma (LUAD) is one of the most frequent subtypes(5). With the wide-spreading application of targeted drugs and immune checkpoint inhibitors, the therapeutic methods of patients with LUAD have undergone the revolutionary change. Whereas, the prognosis of patients with metastasis or recurrence are still woefully poor(6). Besides, the overall survival of patients with lung cancer varies significantly across the world with a 5-year survival rate of 21.2% in the United States, which can be higher than that of China(7). In recent years, previous studies have reported that up-regulation of GSH synthesis pathway in NSCLC cells can trigger the suppression of ferroptosis(8),(9). Preliminary evaluations have explored that the ferroptosis-related genes such as NFS1 are highly expressed in LUAD cells(10). Previous studies have disclosed that ferroptosis are associated with the prognosis of renal carcinoma and hepatocellular carcinoma(11),(12). However, whether these ferroptosis-related genes are correlated with LUAD patient prognosis remains largely unknown.

The ferroptosis-related genes RNA sequencing expression data and corresponding clinical information of LUAD patients were downloaded from the public databases. The expression of ferroptosis-related genes was related with survival outcomes. In the present study, we divided the sample into a training set and a validation set based on random stratified sampling of tumor stages. Then, we established the multi-gene LUAD prognosis prediction model and calculated risk scores by combining least absolute shrinkage and selection operator (LASSO) ten-fold cross-validation and single-factor and multi-factor Cox regression analysis. Finally, we further verified the established model in the validation set and the overall sample set to test the fitting degree of the model. Moreover, in order to explore the inner molecular mechanism of the prognostic difference, we further performed immune and biological functional enrichment analysis to explore the underlying mechanisms of various prognosis.

## Methods

### Data acquisition and preliminary processing

The LUAD is a frequent subtype of NSCLC. This study involved a total of 592 Raw count RNA-Seq data samples, single-nucleotide variation (SNV) statistics and corresponding clinical information. Furthermore, the above data was derived from the LUAD data set of TCGA database (<https://portal.gdc.cancer.gov/>), in which the cancerous tissue sample was 533 cases, and the paracancerous tissue sample was 59 cases. Simultaneously, we used the DESeq2 + variance stabilizing transformation (VST) method to normalize the mRNA count expression data. Based on the previous research, we selected 60 ferroptosis-related genes as the candidate gene set (Table S1)(4, 13-15).

### Differential expression analysis and model establishment

The "DESeq2" R package(16)was utilized to analyze the differential expression genes of cancerous and paracancerous tissue samples in the LUAD statistics set. The selection of DEGs was based on the false discovery rate (FDR) <0.05. The Cox proportional hazards model was a type of semiparametric regression

model. The model took survival outcome and time as the dependent variables, and simultaneously analyzed the impact of multiple factors on survival time. The most crucial concept in the model was hazard ratio (HR). Specifically, in carcinoma explorations,  $HR > 1$  and  $HR < 1$  indicated a poor prognostic and protective factor respectively. The VST expression matrix was utilized to perform univariate Cox regression analysis for ferroptosis-related genes sets and overall survival time, and then to evaluate the impacts of each gene on prognosis. P values were adjusted with the Benjamini & Hochberg (BH) method. We selected genes with  $HR \neq 1$  and  $p < 0.05$  as prognostic-related genes. The prognostic-related ferroptosis-related genes and DEGs between cancerous and paracancerous tissue were taken the intersection. We performed downstream analysis on genes which had indispensable impacts on prognosis among the genes obtained above.

The "glmnet" R package was utilized to perform LASSO regression analysis on the target gene set (17) and to conduct multi-gene prognostic prediction model. In order to avoid over-fitting and obtain reliable results, we applied the ten-fold cross-validation to acquire optimal lambda value that came from the minimum partial likelihood deviance and screened out representative genes. We selected the few candidates to establish a multivariate Cox regression model(18), utilized the coxzph function to take PH test for each factor, have calculated the VIF and correlation coefficient of each factor in each regression model, and determined the possible collinearity of factors. We re-modeled the variables that conformed to the PH hypothesis and collinearity tests. The risk scores of the patients were calculated according to the modeling results. The formula was established as follows: score = sum (each gene's expression \* corresponding coefficient). The expression of each gene was the normalized mRNA expression, and the coefficient was the results of multivariate Cox regression analysis.

To exploring the survival predictive feasibility of risk score, which was obtained by multi-gene prognostic prediction model, we used the "survival" and "survminer" packages and the Kaplan-Meier method to estimate the survival curve. The Time-dependent ROC curve was plotted by the "survival ROC" R package, graphically displaying the predictive abilities of the model at different times.

### **Constructing and verifying the prognostic composite nomogram**

For the sake of establishing a more reliable prediction method that can be applied in clinical practice, we constructed the composite nomogram through the "rms" package. Similarly, the clinical information features and risk scores were combined to establish a multi-variate Cox regression model. Based on the influenced degrees of each impacting factor on the outcome variable, each influencing factor were assigned points to value its level, and then were summed up to acquire the total scores. Finally, through the function conversion relationships between the total scores and the probability of outcome events, we calculated the predicted probability of the individual's survival time.

We further verified the accuracy of the nomogram by applying the model used C statistics, the calibration plot and Time-dependent ROC curve. Simultaneously, in order to predict the proximity between the probability of survival and the actual results numerically, we calibrated the nomogram model. We utilized the Hosmer-Lemeshow (H-L) type of  $\chi^2$  statistics and divided the data into 3 groups (third quantiles) in

ascending order based on the predicted probability generated by the model to calculate the  $\chi^2$  statistic. Furthermore, the average predicted probability was compared with the actual event rate estimated using the Kaplan-Meier method.

### **The enrichment analysis of DEGs function and the enrichment score of immune gene set**

Differential expression analysis on high-risk groups and low-risk groups divided by risk score was performed to acquire DEGs. The "clusterprofiler" R package(19)was applied for these DEGs to perform enrichment analysis of gene ontology (GO) and Kyoto Encyclopedia of Genes and Genomes (KEGG) pathways.

We utilized the "gsva" R package to perform a single-sample gene set enrichment analysis (ssGSEA) on 29 immune gene sets(20)involving 16 immune cells and 13 immune-related activation pathways (21) (Table S2). Simultaneously, we scored the degrees of immune infiltration of each above sample.

### **Statistical Analysis**

The "sample" R package was used to implement in random stratified sampling. Meanwhile, the differentially expressed heatmap and SNV mutation landscape were respectively completed by the "pheatmap" and "maftools" R packages (22). Wilcoxon-test was used to test the difference between the groups of the boxplot. The p value of Kaplan-Meier survival curve was log-rank tests. R software (Version 3.6.0) was used for all statistical analyses. The flowchart of this study was shown in figure 1.

## **Results**

### **Identifying prognostic related DEGs which might be relevant with ferroptosis**

The LUAD project data set had a total of 533 cases of cancerous tissue samples and 59 cases of paracancerous tissue samples. Among them, there were 502 cases with clinical information, which had total survival time recorded (Table 1). Besides, the DESeq2 was used to process and analyze the mRNA expression data of the original count for obtaining DEGs (FDR < 0.05). The acquired DEGs were sorted according to the value of log2foldchange and were displayed them with the heatmap (Figure 2a). The results of univariate Cox regression analysis revealed that a total of 11 genes in the 60 ferroptosis-related genes were significantly related to the prognosis of patients with LUAD (*adj p* < 0.05). Taking the intersection of these 11 genes and DEGs, it can be found that these 11 genes were differentially expressed in cancerous and paracancerous tissue cancers. And the results of Cox regression analysis were indicated in the forest plot (Figure 2b).

### **Multi-gene prognostic model in the training set**

We randomly stratified samples of all patients according to tumor stage and divided them into training set (n=251) and validation set (n=251). Simultaneously, we used LASSO-Cox regression analysis to the above-mentioned set of prognostic-related genes related to ferroptosis that we have identified (Figure 3a

and 3b). The results revealed that a total of 5 genes passed the PH hypothesis (Figure S1) and collinearity test (correlation coefficient less than 0.5) (Figure 3c), namely ACSL4, GSS, ACSL3, PEBP1 and PGD. Plotting a forest plot for these five genes disclosed that ACSL4, GSS, ACSL3 and PGD were the risk factors, and the adjust  $p$  value of ACSL4 was less than 0.05. In this study, PEBP1 was a protective factor ( $adj P < 0.05$ ) (Figure 3d). We divided the cases into groups according to the level of expression of each gene to plot a survival curve and have displayed it in Figure 4. Besides, the threshold for the expressed level of each gene was calculated by the "surv\_cutpoint" function of the "survminer" R package (Figure S2). Furthermore, the risk score of each sample was calculated based on the modeling results, risk score = (0.2226 \* ACSL4 expression + 0.2373 \* GSS expression + 0.4369 \* ACSL3 expression - 0.4417 \* PEBP1 expression + 0.1431 \* PGD expression). And the expression level of each gene was normalized. We divided the cases into high-risk groups and low-risk groups based on the risk score value of each patient. The grouped threshold was determined as the median value of the risk score of all patients in the training set, and the median value was equal to 6.64. By utilizing the KM algorithm to analyze the overall survival time and plot the survival curve. It can be seen that the high-risk group had a worse survival status and survival time than the low-risk group ( $p = 0.0041$ ). We plotted the ROC curve to evaluate the predictive ability of the risk score model, and the highest AUC might be 0.7 (Figure 5a).

### **Verification of the prognostic model**

To determine the reliability of the established model and the accuracy of the prediction, we used the results of the model construction to the verification set, and internally verified the results. We used the same calculation methods to calculate the risk score for the validation set and took the same threshold (cut off = 6.64) to divide patients into high-risk and low-risk group. Moreover, we utilized the KM algorithm to estimate the survival status of the validation set and plotted the survival curve. Similarly, it can be seen that the prognosis of the high-risk group was significantly worse than that of the low-risk group ( $p = 0.00042$ ), and the maximum AUC was 0.69 (Figure 5b). Finally, we combined the training and validation set, calculated the risk score of the 502 patients according to the similar models, as well as used the same threshold to perform survival analysis and plot the survival curve (Figure 5c). The results were divided into high-risk group 257 cases and low-risk group 245 cases. Overall, the high-risk group in this study had a worse prognosis ( $p < 0.0001$ ), and the maximum AUC were 0.69.

### **Construction and verification of composite nomogram combining clinical information**

To better apply the model to the actual clinical situation, we incorporated clinical indicators into the evaluations of the model and constructed composite nomogram to predict the survival probabilities of patients. We performed a multivariate Cox regression analysis on age, gender, tumor stage (I/II vs III/IV) and risk score as variables and incorporated above factors into the model as variables to draw a nomogram (Figure 6a). In this figure, the line segment corresponding to each variable was marked with a scale, which can represent the value range of the variable, and the length of the line segment reflected the contribution of the factor to survival. The points in the figure represented the individual scores corresponding to each variable under different values and Total Points. The corresponding individual

scores after taking all the variables were added up to the total score. The last three lines were the survival rates at 1, 3 and 5 years in the corresponding total scores. For the sake of verifying the prognostic value of the nomogram model, the predictive ability of the nomogram was evaluated by tROC, and the AUC corresponding to 1, 3 and 5 years of survival was 0.698, 0.718, 0.731, respectively (Figure 6b). The prediction result of C statistics on the nomogram model was 0.677 (95% CI, 0.633–0.721). The check chart indicated that there may be the good agreement between the predicted and actual results (Figure 6c).

### **SNV mutation landscape variation between high and low risk groups**

We downloaded the SNV data corresponding to 502 samples, grouped them based on the risk score, and plotted the mutation landscape for display (Figure 7). Each figure indicated the 20 genes with the highest mutation frequency, and different colors were corresponded to various mutation types. In the figure, the upper bar graph was the total number of mutations in each sample, and the right bar graph was the mutation frequency of each gene in all samples. It can be deemed that the genes with higher mutation frequency in the high-risk group and the low-risk group had higher consistency, whereas the mutation frequency can be higher in the high-risk group, TP53 (53%), TTN (50%), MUC16 (42%), CSMD3 (40%), RYR2 (39%).

### **The GO and KEGG enrichment analysis**

To further explore the differences in biological functions and pathways between the high-risk and low-risk groups, we used DESeq2 to perform differential expression analysis on the expression levels of each sample in the high-risk and low-risk groups ( $\log_{2}FC > 1$ ,  $FDR < 0.05$ ). The DEGs of the two groups were screened out as the gene set for GO and KEGG enrichment analysis. The results of pathway enrichment and GO enrichment were shown as Figure 8, Supplementary Table 3-4. From these charts, it can be seen that multiple pathways were related to metabolism from these charts and figures. The main enriched pathways included the neuroactive ligand-receptor interaction, metabolism of xenobiotics by cytochrome P450, steroid hormone biosynthesis, staphylococcus aureus infection pathway, IL-17 signaling pathway and retinol metabolic pathway etc. The results of GO enrichment mainly may focus on the biological functions and processes such as keratinization, axoneme assembly, and microtubule bundle formation. Furthermore, the enrichment results availed to explain the reasons for the variations in prognosis from a functional level. It can also be seen from the results that some pathways and functions were closely relevant with ferroptosis or iron metabolism. The above specified a basis for further research.

### **The correlation between risk score and immune status**

According to the threshold of risk score, we wanted to further explore the correlation between risk score and immune status. Thus, we acquired a total of 29 immune-related gene sets, which may involve multiple immune cells and immune-related functions or pathways. To quantify the enrichment degree of transcriptome data in the 29 immune gene sets, we used the ssGSEA method to estimate each sample and draw a boxplot to compare the difference between the high-risk and low-risk group (Figure 9). In

these immune cell genes concentration of aDCs, iDCs, B-cells, Mast Cells, Neutrophils, NK cells, T helper cells, Th2 cells and TIL, the enrichment scores of high and low groups were significantly different ( $p < 0.05$ ). Simultaneously, the enrichment scores of these immune function gene sets in HLA, inflammation promoting, MHC class I, T cell co-stimulation and Type II IFN Response can be also significantly different between the two groups ( $p < 0.05$ ).

## Discussion

The current study systematically explained the correlation between 60 ferroptosis-related genes and the prognosis of LUAD. We linked these genes to the overall survival time of patients with LUAD. In this study, the patient population was divided into training set and validation set by random stratified sampling method. The prognostic prediction model of 5 genes was established for the training set samples by the methods of combining LASSO ten-fold cross-validation and univariate and multivariate Cox regression analysis and was applied in actual clinical practice.

The previous evidences have confirmed that these ferroptosis-related genes play various roles in the entire ferroptosis process. Nevertheless, the true impact of each gene on the prognosis of specific cancer types was still unclear. We combined the mRNA expression levels of each gene with actual clinical characteristics to analyze the variations in expression between cancerous and paracancerous tissue. Simultaneously, we integrated the quantitative data of these gene expressions with the survival status to predict the survival status. Besides, we can divide the cases into groups according to the degree of risk, to explore the differential genes between groups and corresponding biological functions and pathways which produced variations. In conclusion, these above results will reveal the role which the target genes play in the occurrence and development of LUAD.

In the construction of the model, we involved a total of 5 genes (ACSL3, ACSL4, GSS, PEBP1, PGD). These genes were closely associated with the process of ferroptosis. ACSL3 was required for exogenous monounsaturated fatty acids to protect cells against ferroptosis. Negatively it was correlated with ferroptosis sensitivity(23). ACSL4 can be an essential proferroptotic gene. It knocked down inhibited erastin-induced ferroptosis, whereas overexpression restored ferroptosis sensitization. Re-expression of Flag-tagged human wild-type (WT) ACSL4 (ACSL4-Flag) in *Acs14* KO Pfa1 cells restored full sensitivity to ferroptosis induction. Inhibition showed significantly prolonged survival compared to vehicle-treated mice. And knockout of ACSL4 in ferroptosis-sensitive cells conferred protection from erastin- and RSL3-induced cell death(24) (25). GSS, which was glutathione synthetase. The glutathione-dependent lipid hydroperoxidase GPX4 prevented ferroptosis by converting lipid hydroperoxides into non-toxic lipid alcohols. Next, Elevated levels of PEBP1 resulted in increased sensitivity of HK2 cells to RSL3 whereas lowered contents of PEBP1 in HAEC and HT22 cells were associated with decreased sensitivity to ferroptosis(26). And the PGD was required in erastin-induced ferroptosis(1). We modeled and calculated the corresponding risk score based on the expression data of 5 candidate genes, and divided cases into high-risk and low-risk groups. This model can predict the survival of patients well in the training set, the validation set and the overall sample. Meanwhile, we took gender, age, and tumor stage into

consideration, plotting the composite nomogram that was more in line with clinical reality. Among them, the gender and age had relatively small effects on the prognosis, whereas the tumor staging and risk score had a greater impact on the prognosis. The above results were in accord with our investigation expectations.

To explore the factors for the variations in prognosis between the high and low risk groups, we compared the SNV levels and analyzed the differences in biological functions of the two groups. Simultaneously, we grouped the acquired SNV statistics according to high and low risk scores. We counted the mutation frequency of each gene and the number of mutations in each sample and displayed the mutations landscape of the two groups as the result. The results revealed that the high-risk group appeared to have a higher frequency of mutations in the same gene. Moreover, the five genes with the highest mutation frequency in the high-risk group, which included TP53 (53%), TTN (50%), MUC16 (42%), CSMD3 (40%) and RYR2 (39%) were all 5–13% higher than those in the low-risk group. It was the simple mutation frequency to find the source of the variations between the two groups. The academic information at the genome level needs to be further explored. We hoped that the underlying reasons for the difference in prognosis can be explained more clearly in the subsequent analysis.

Nevertheless, in the process of functional research, we performed functional enrichment analysis and immune infiltration score on DEGs between the two groups. It can be seen from the results of pathway enrichment that the main enriched pathways involved cytochrome P450, steroid hormones, IL-17 signaling pathways, and staphylococcus aureus infection pathways etc. The cytochrome P450 oxidoreductase (POR) was the necessary condition for triggering the ferroptosis in carcinoma cells(27). Moreover, POR boosted the execution of ferroptosis by engaging in the peroxidative modification of phospholipids in the cell membrane. PPAR $\alpha$ , a member of the steroid hormone receptor superfamily, can inhibit iron overload and lead to ferroptosis by combining with the GPX4 and facilitating its expression. The preliminary evaluations indicated that the steroid hormones in adrenal sebaceous cells were highly affected by ferroptosis induction(28). Iron played a crucial role in the continuous evolution process for staphylococcus aureus by establishing a variety of efficient iron transportation systems. On the one hand, staphylococcus aureus took up the hemoglobin in the red blood cells of the host and the serum through the transport systems encoded by the iron-regulated surface determinants located in the cell wall. On the other hand, the staphylococcus aureus acquired the iron by the siderophores with high affinity for iron. The previous studies showed that the reactive oxygen species can result in the drug resistance in staphylococcus aureus infection(29). Whereas, in GO enrichment, biological functions such as microtubule bundle formation and keratinization were potentially closely relevant with the iron(30, 31). Therefore, we concluded that the variations in these pathways might be the inherent factors that can trigger the differences between the high-risk and low-risk groups. Furthermore, it provided the clinical physicians and academic professionals with a biological function-based guide for understanding the mechanism of ferroptosis in patients with the lung cancers.

In spite of the prognostic model construction can play a vital role in the TCGA-LUAD cohort, we only used the steps of internal verification and overall verification in the verifying process of the model. Specifically,

the external validation set was not included to examine the final results. In functional analysis, we utilized the DEGs of the high-risk and low-risk groups for enrichment analysis. Whereas, the functions of these pathways were not be experimentally verified. In consequence, this study was clearly explaining how these pathways and functions triggered the prognostic variations.

In general, we had initially determined a prognostic prediction model based on 5 ferroptosis-related genes by the analysis of patients with LUAD. In the training and validation set, this model might be significantly related to overall survival time. Furthermore, taking clinical indicators into consideration can also provide greater possibilities for clinical practice. The results offered the reliable methods for predicting the prognosis of patients with LUAD and pointed out the new direction for explaining the prognostic variations among the patients with LUAD.

## **Conclusions**

This study constructed a novel LUAD prognosis prediction model base on 5 ferroptosis-related genes, which can provide a reliable prognostic evaluation tool for the clinic and assist the clinical therapeutic decision.

## **Declarations**

### **Ethics approval and consent to participate**

Not applicable

### **Consent for publication**

Not applicable

### **Availability of data and materials**

The datasets used and/or analysed during the current study are available from the corresponding author on reasonable request.

### **Competing interests**

The authors declare that they have no competing interests

### **Funding**

This work was supported by the Young Scientist and Technology Innovation Project of Shenyang City (RC190434) to SL. S.

### **Authors' contributions**

FL designed the study. DG and FL analyzed the data. SL S prepared the manuscript.

## Acknowledgements

Not applicable

## References

1. Dixon SJ, Lemberg KM, Lamprecht MR, Skouta R, Zaitsev EM, Gleason CE, et al. Ferroptosis: an iron-dependent form of nonapoptotic cell death. *Cell*. 2012;149(5):1060–72.
2. Yang WS, Stockwell BR. Ferroptosis: Death by Lipid Peroxidation. *Trends Cell Biol*. 2016;26(3):165–76.
3. Xia X, Fan X, Zhao M, Zhu P. The Relationship between Ferroptosis and Tumors: A Novel Landscape for Therapeutic Approach. *Curr Gene Ther*. 2019;19(2):117–24.
4. Hassannia B, Vandenabeele P, Vanden Berghe T. Targeting Ferroptosis to Iron Out Cancer. *Cancer Cell*. 2019;35(6):830–49.
5. Herbst RS, Morgensztern D, Boshoff C. The biology and management of non-small cell lung cancer. *Nature*. 2018;553(7689):446–54.
6. Santarpia M, Aguilar A, Chaib I, Cardona AF, Fancelli S, Laguia F, et al. Non-Small-Cell Lung Cancer Signaling Pathways, Metabolism, and PD-1/PD-L1 Antibodies. *Cancers (Basel)*. 2020;12(6).
7. Allemani C, Matsuda T, Di Carlo V, Harewood R, Matz M, Niksic M, et al. Global surveillance of trends in cancer survival 2000-14 (CONCORD-3): analysis of individual records for 37 513 025 patients diagnosed with one of 18 cancers from 322 population-based registries in 71 countries. *Lancet*. 2018;391(10125):1023–75.
8. Ji X, Qian J, Rahman SMJ, Siska PJ, Zou Y, Harris BK, et al. xCT (SLC7A11)-mediated metabolic reprogramming promotes non-small cell lung cancer progression. *Oncogene*. 2018;37(36):5007–19.
9. Lai Y, Zhang Z, Li J, Li W, Huang Z, Zhang C, et al. STYK1/NOK correlates with ferroptosis in non-small cell lung carcinoma. *Biochem Biophys Res Commun*. 2019;519(4):659–66.
10. Alvarez SW, Sviderskiy VO, Terzi EM, Papagiannakopoulos T, Moreira AL, Adams S, et al. NFS1 undergoes positive selection in lung tumours and protects cells from ferroptosis. *Nature*. 2017;551(7682):639–43.
11. Wu G, Wang Q, Xu Y, Li Q, Cheng L. A new survival model based on ferroptosis-related genes for prognostic prediction in clear cell renal cell carcinoma. *Aging*. 2020;12(14):14933–48.
12. Liang JY, Wang DS, Lin HC, Chen XX, Yang H, Zheng Y, et al. A Novel Ferroptosis-related Gene Signature for Overall Survival Prediction in Patients with Hepatocellular Carcinoma. *Int J Biol Sci*. 2020;16(13):2430–41.
13. Stockwell BR, Friedmann Angeli JP, Bayir H, Bush AI, Conrad M, Dixon SJ, et al. Ferroptosis: A Regulated Cell Death Nexus Linking Metabolism, Redox Biology, and Disease. *Cell*. 2017;171(2):273–85.

14. Bersuker K, Hendricks JM, Li Z, Magtanong L, Ford B, Tang PH, et al. The CoQ oxidoreductase FSP1 acts parallel to GPX4 to inhibit ferroptosis. *Nature*. 2019;575(7784):688–92.
15. Doll S, Freitas FP, Shah R, Aldrovandi M, da Silva MC, Ingold I, et al. FSP1 is a glutathione-independent ferroptosis suppressor. *Nature*. 2019;575(7784):693–8.
16. Love MI, Huber W, Anders S. Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. *Genome Biol*. 2014;15(12):550.
17. Simon N, Friedman J, Hastie T, Tibshirani R. Regularization Paths for Cox's Proportional Hazards Model via Coordinate Descent. *J Stat Softw*. 2011;39(5):1–13.
18. Tibshirani R. The lasso method for variable selection in the Cox model. *Stat Med*. 1997;16(4):385–95.
19. Yu G, Wang LG, Han Y, He QY. clusterProfiler: an R package for comparing biological themes among gene clusters. *OMICS*. 2012;16(5):284–7.
20. Rooney MS, Shukla SA, Wu CJ, Getz G, Hacohen N. Molecular and genetic properties of tumors associated with local immune cytolytic activity. *Cell*. 2015;160(1–2):48–61.
21. Hanzelmann S, Castelo R, Guinney J. GSEA: gene set variation analysis for microarray and RNA-seq data. *BMC Bioinformatics*. 2013;14:7.
22. Mayakonda A, Lin DC, Assenov Y, Plass C, Koeffler HP. Maftools: efficient and comprehensive analysis of somatic variants in cancer. *Genome Res*. 2018;28(11):1747–56.
23. Magtanong L, Ko PJ, To M, Cao JY, Forcina GC, Tarangelo A, et al. Exogenous Monounsaturated Fatty Acids Promote a Ferroptosis-Resistant Cell State. *Cell Chem Biol*. 2019;26(3):420–32. e9.
24. Muller T, Dewitz C, Schmitz J, Schroder AS, Brasen JH, Stockwell BR, et al. Necroptosis and ferroptosis are alternative cell death pathways that operate in acute kidney failure. *Cell Mol Life Sci*. 2017;74(19):3631–45.
25. Doll S, Proneth B, Tyurina YY, Panzilius E, Kobayashi S, Ingold I, et al. ACSL4 dictates ferroptosis sensitivity by shaping cellular lipid composition. *Nat Chem Biol*. 2017;13(1):91–8.
26. Wenzel SE, Tyurina YY, Zhao J, St Croix CM, Dar HH, Mao G, et al. PEBP1 Wardens Ferroptosis by Enabling Lipoxygenase Generation of Lipid Death Signals. *Cell*. 2017;171(3):628–41. e26.
27. Zou Y, Li H, Graham ET, Deik AA, Eaton JK, Wang W, et al. Cytochrome P450 oxidoreductase contributes to phospholipid peroxidation in ferroptosis. *Nat Chem Biol*. 2020;16(3):302–9.
28. Weigand I, Schreiner J, Rohrig F, Sun N, Landwehr LS, Urlaub H, et al. Active steroid hormone synthesis renders adrenocortical cells highly susceptible to type II ferroptosis induction. *Cell Death Dis*. 2020;11(3):192.
29. Rowe SE, Wagner NJ, Li L, Beam JE, Wilkinson AD, Radlinski LC, et al. Reactive oxygen species induce antibiotic tolerance during systemic *Staphylococcus aureus* infection. *Nat Microbiol*. 2020;5(2):282–90.
30. Becci PJ, McDowell EM, Trump BF. The respiratory epithelium. VI. Histogenesis of lung tumors induced by benzo[a]pyrene-ferric oxide in the hamster. *J Natl Cancer Inst*. 1978;61(2):607–18.

31. Infante AA, Infante D, Chan MC, How PC, Kutschera W, Linhartova I, et al. Ferritin associates with marginal band microtubules. *Exp Cell Res.* 2007;313(8):1602–14.

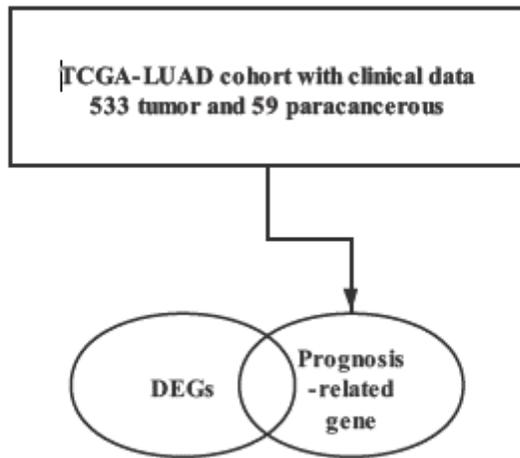
## Table

**Table 1. Clinical characteristics of LUAD patients**

**Clinical characteristics of LUAD patients**

|                                  | <b>Training Set</b> | <b>Validate Set</b> |
|----------------------------------|---------------------|---------------------|
| <b>No.of patient</b>             | 251                 | 251                 |
| <b>Age (mean (SD))</b>           | 64.57 (10.05)       | 65.86 (10.04)       |
| <b>Gender (%)</b>                |                     |                     |
| male                             | 123 (49.0)          | 109 (43.4)          |
| female                           | 128 (51.0)          | 142 (56.6)          |
| <b>Stage (%)</b>                 |                     |                     |
| Stage I                          | 1 (0.4)             | 4 (1.6)             |
| Stage IA                         | 73 (29.4)           | 57 (23.2)           |
| Stage IB                         | 63 (25.4)           | 72 (29.3)           |
| Stage II                         | 1 (0.4)             | 0 (0.0)             |
| Stage IIA                        | 22 ( 8.9)           | 28 (11.4)           |
| Stage IIB                        | 37 (14.9)           | 31 (12.6)           |
| Stage IIIA                       | 35 (14.1)           | 35 (14.2)           |
| Stage IIIB                       | 3 ( 1.2)            | 7 (2.8)             |
| Stage IV                         | 13 ( 5.2)           | 12 (4.9)            |
| <b>Pathologic T (%)</b>          |                     |                     |
| T1                               | 38 (15.1)           | 27 (10.8)           |
| T1a                              | 29 (11.6)           | 19 (7.6)            |
| T1b                              | 24 (9.6)            | 31 (12.4)           |
| T2                               | 78 (31.1)           | 82 (32.7)           |
| T2a                              | 40 (15.9)           | 41 (16.3)           |
| T2b                              | 11 (4.4)            | 16 (6.4)            |
| T3                               | 23 (9.2)            | 22 (8.8)            |
| T4                               | 7 (2.8)             | 11 (4.4)            |
| TX                               | 1 (0.4)             | 2 (0.8)             |
| <b>Pathologic N (%)</b>          |                     |                     |
| N0                               | 164 (65.3)          | 162 (64.8)          |
| N1                               | 47 (18.7)           | 47 (18.8)           |
| N2                               | 32 (12.7)           | 37 (14.8)           |
| N3                               | 1 (0.4)             | 1 (0.4)             |
| NX                               | 7 (2.8)             | 3 (1.2)             |
| <b>Pathologic M (%)</b>          |                     |                     |
| M0                               | 160 (64.3)          | 174 (69.9)          |
| M1                               | 7 (2.8)             | 10 (4.0)            |
| M1a                              | 1 (0.4)             | 1 (0.4)             |
| M1b                              | 4 (1.6)             | 1 (0.4)             |
| MX                               | 77 (30.9)           | 63 (25.3)           |
| <b>Vital status = Dead (%)</b>   | 92 (36.7)           | 90 (35.9)           |
| <b>Race (%)</b>                  |                     |                     |
| american indian or alaska native | 1 (0.4)             | 0 (0.0)             |
| asian                            | 4 (1.6)             | 3 (1.2)             |
| black or african american        | 26 (10.4)           | 26 (10.4)           |
| not reported                     | 26 (10.4)           | 28 (11.2)           |
| white                            | 194 (77.3)          | 194 (77.3)          |
| <b>Smoke_history = Yes (%)</b>   | 88 (35.1)           | 99 (39.4)           |
| <b>OS time (mean)</b>            | 890.17              | 926.26              |

**Figures**



**Figure 1**

Flowchart of this study

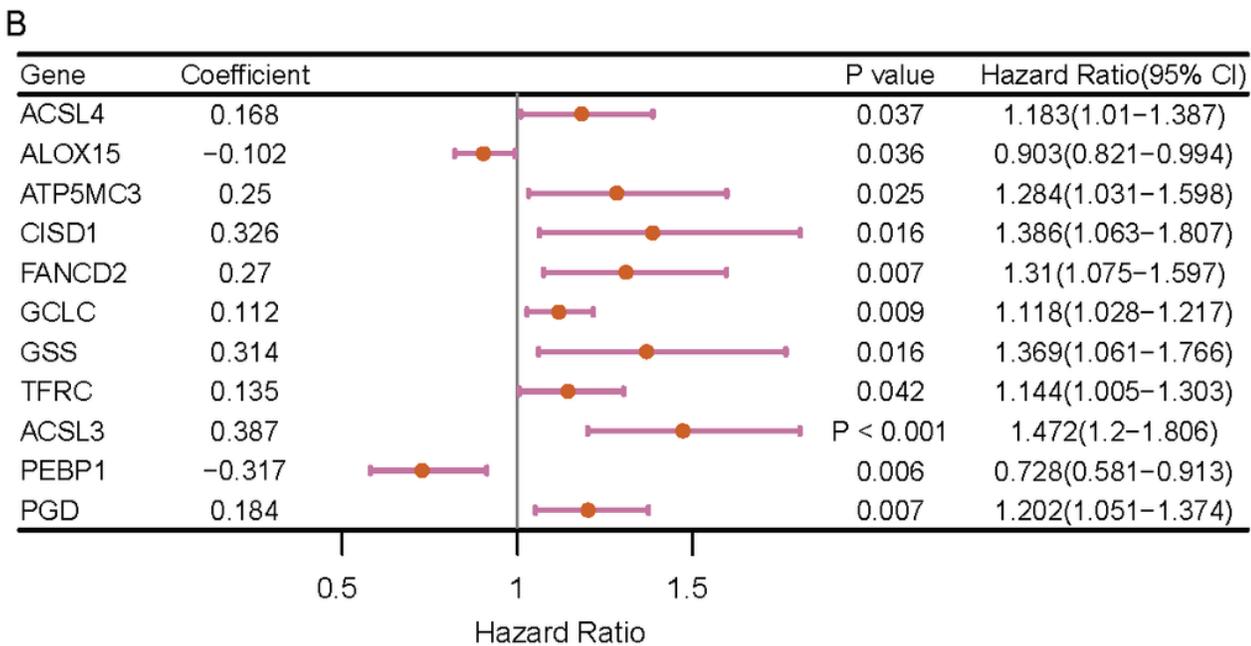
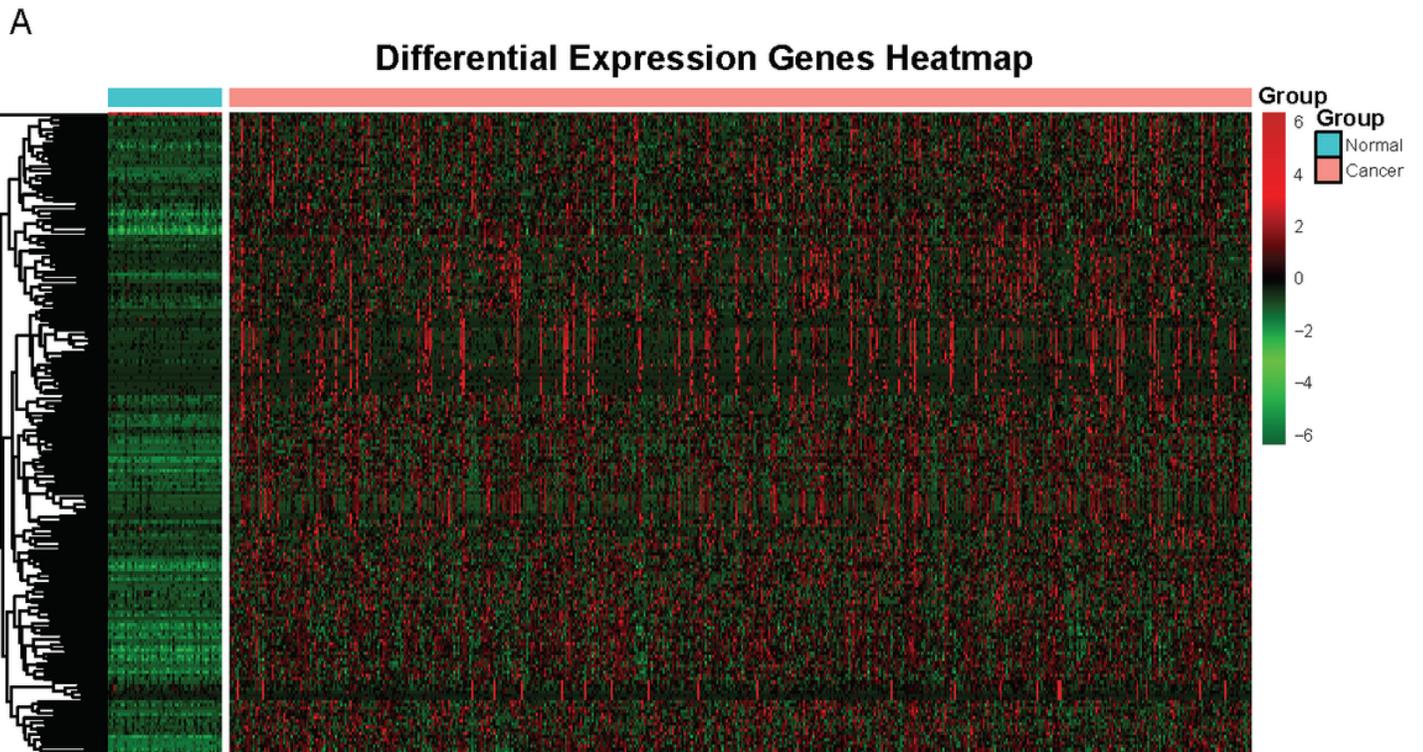


Figure 2 Data processing, screening of the DEGs.

## Figure 2

Data processing, screening of the DEGs

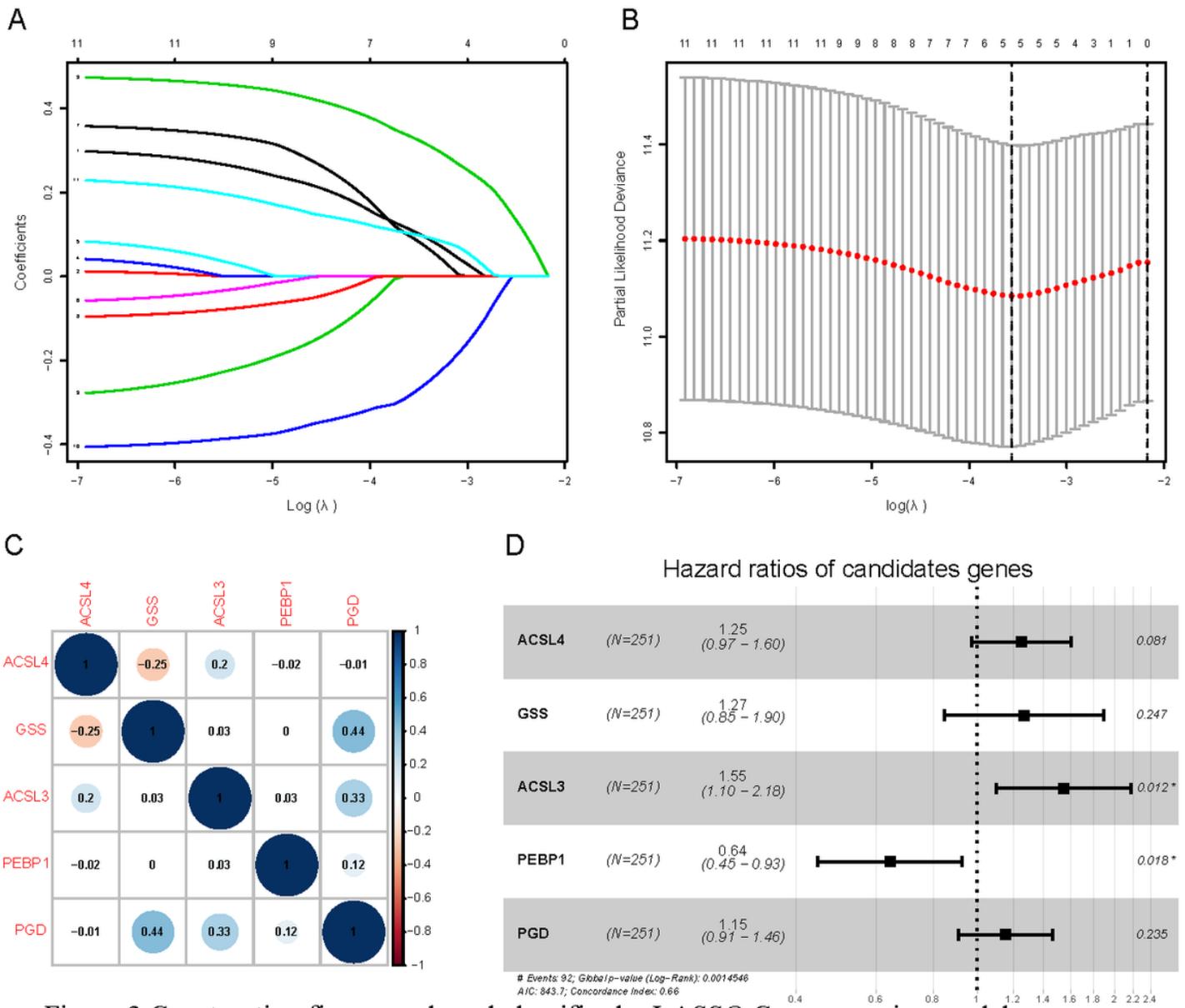


Figure 3 Constructing five-gene-based classifier by LASSO Cox regression model.

Figure 3

Constructing five-gene-based classifier by LASSO Cox regression model.

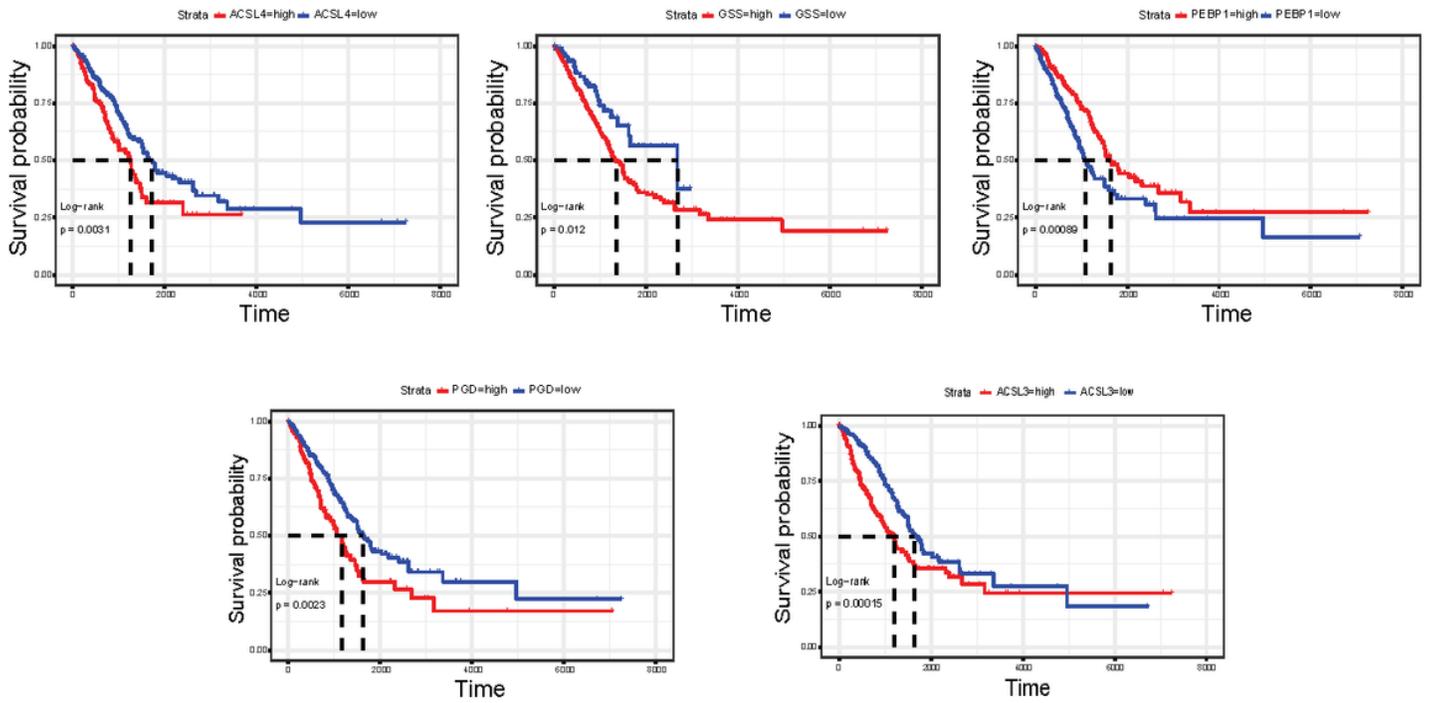


Figure 4 Survival analysis of the five candidate genes.

Figure 4

Survival analysis of the five candidate genes.

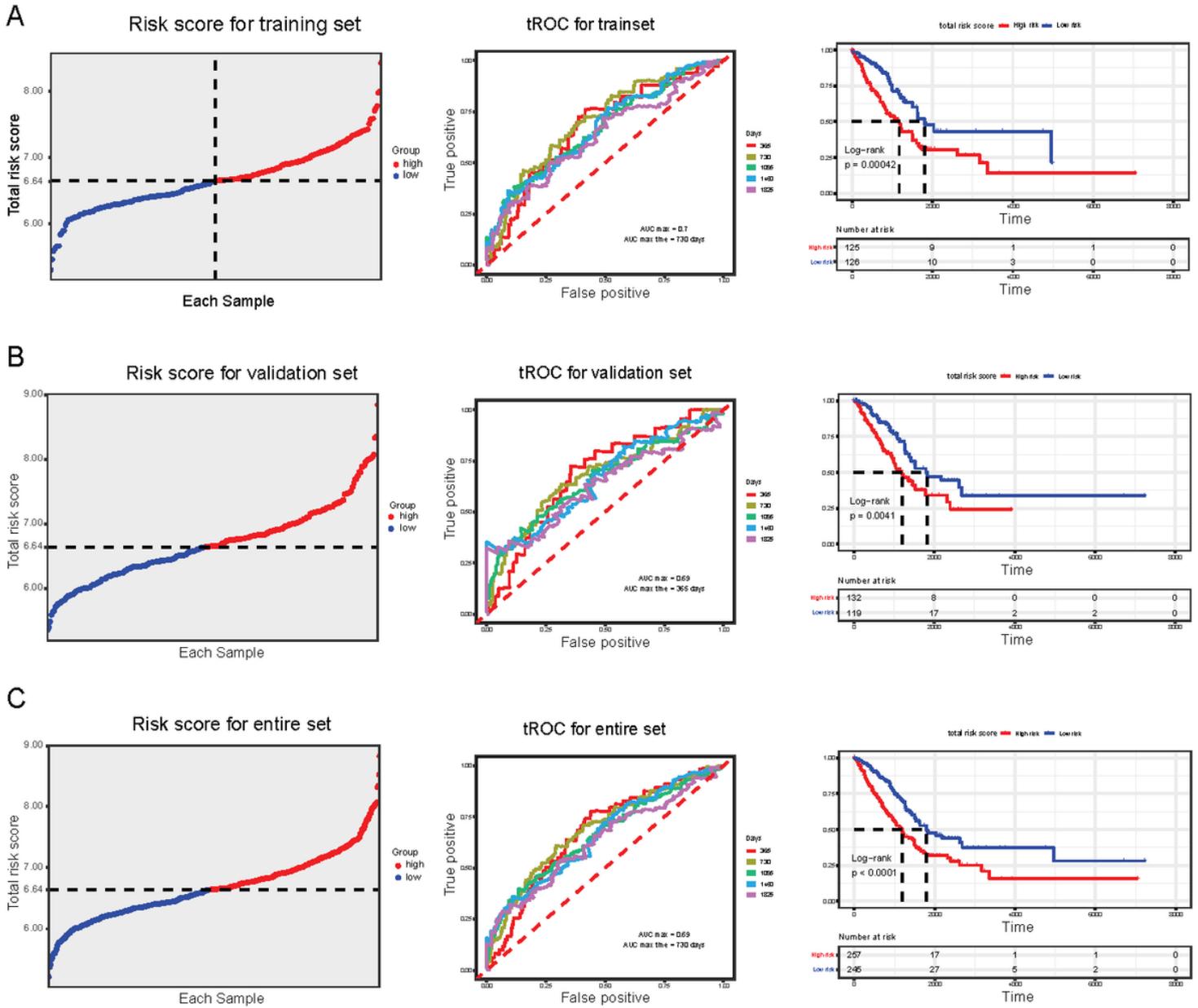


Figure 5 Prognostic analysis of five-gene signature.

Figure 5

Prognostic analysis of five-gene signature.

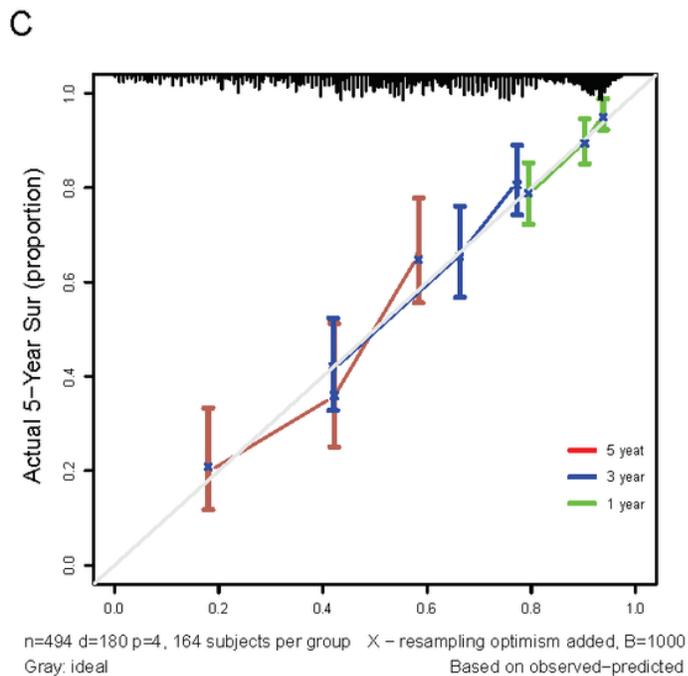
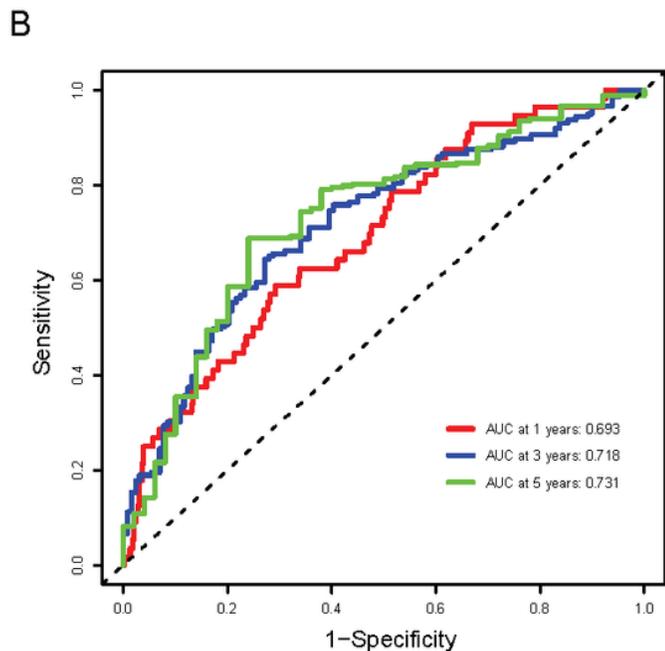
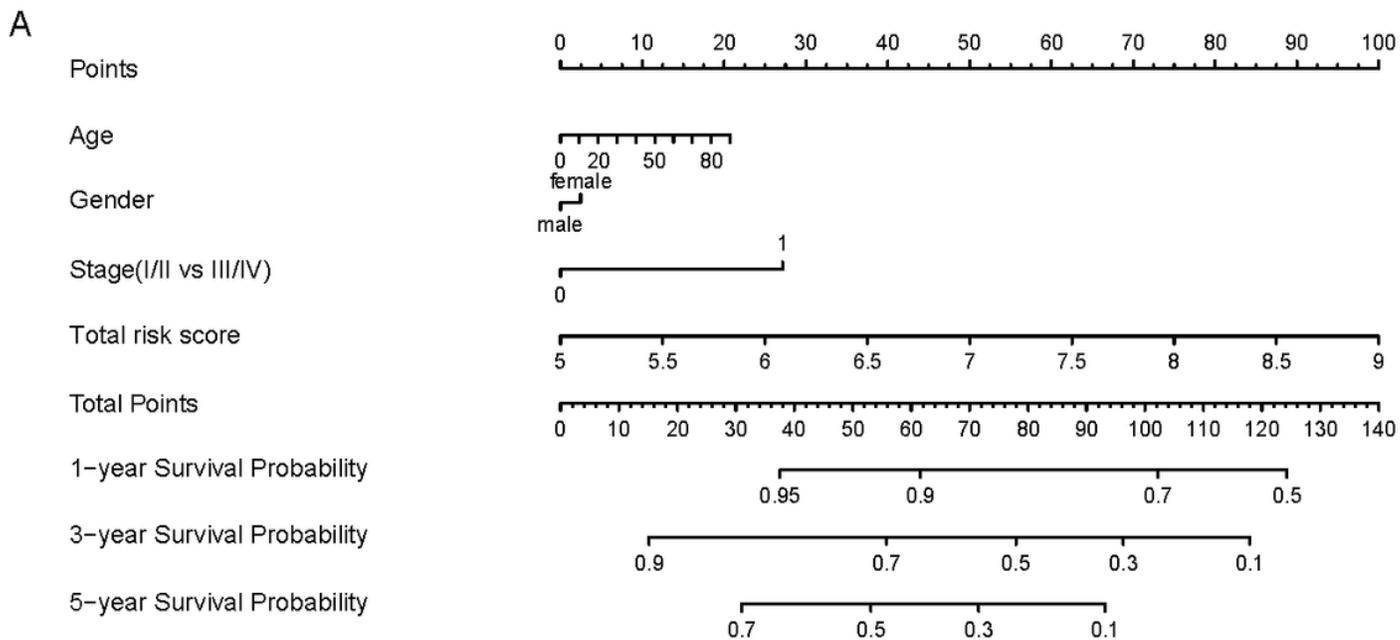


Figure 6 Construction of nomogram prognostic model.

**Figure 6**

Construction of nomogram prognostic model.

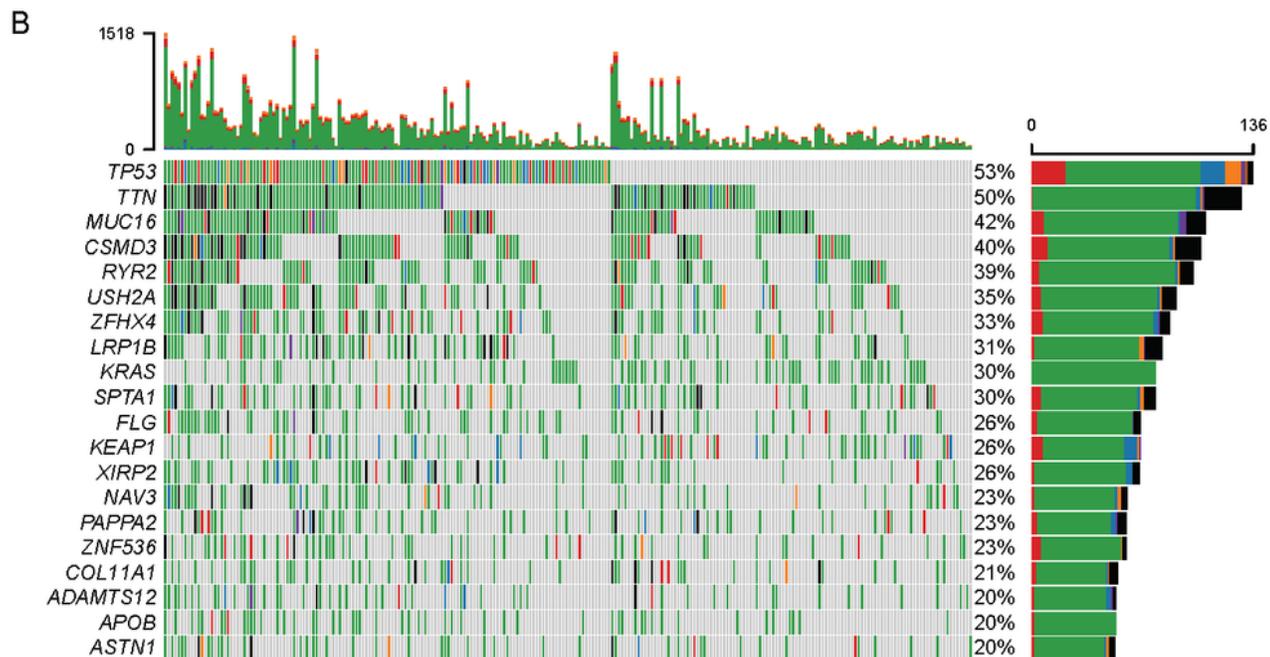
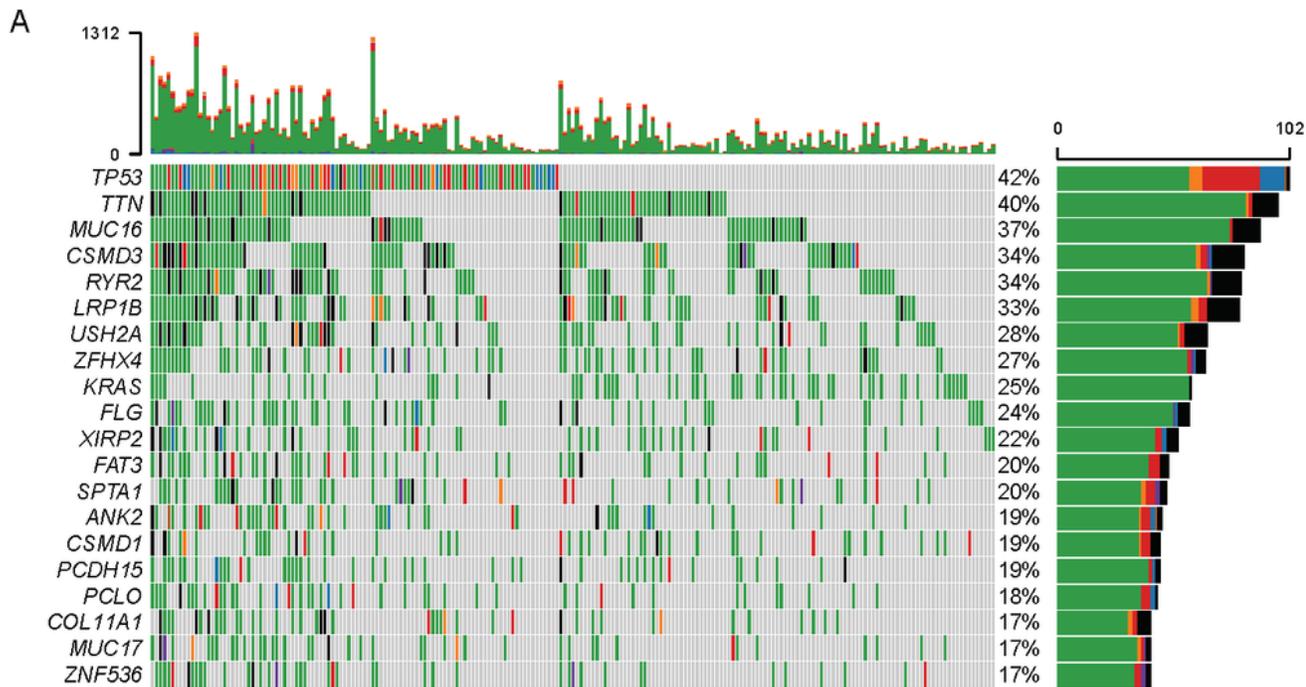


Figure 7 SNV mutation landscape variation between high and low risk groups.

**Figure 7**

SNV mutation landscape variation between high and low risk groups.



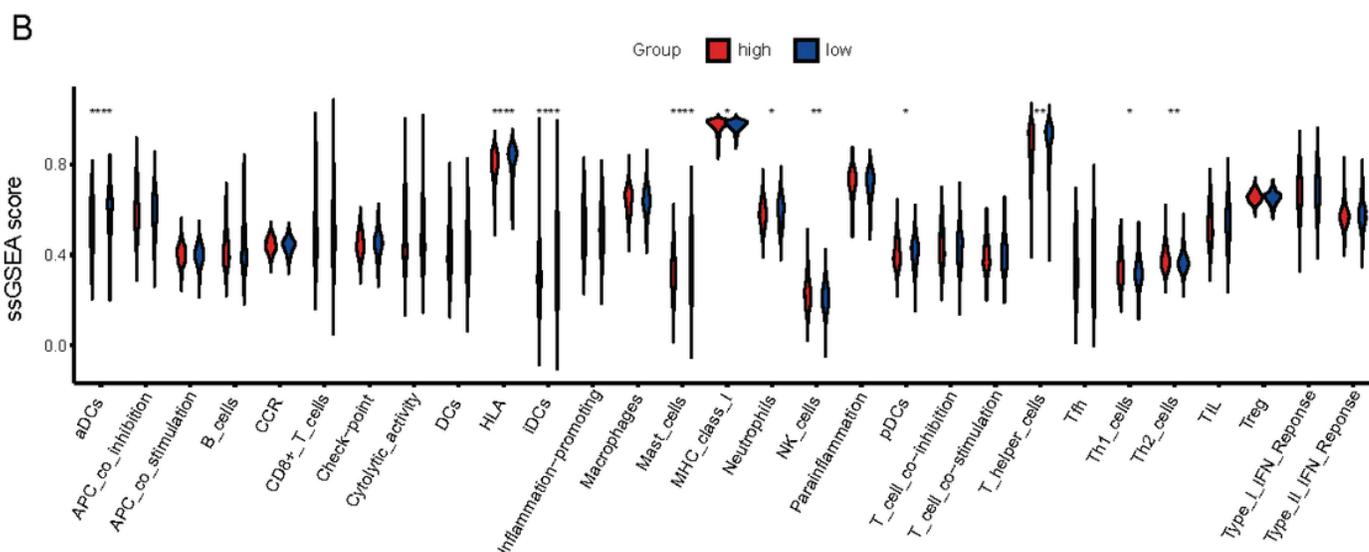


Figure 9 The correlation between risk score and immune status.

## Figure 9

The correlation between risk score and immune status.

## Supplementary Files

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

- [SupplementarytableS1.xlsx](#)
- [SupplementarytableS2.xlsx](#)
- [SupplementarytableS3.xlsx](#)
- [SupplementarytableS4.xlsx](#)
- [figS1.pdf](#)
- [figS2.pdf](#)