


The Lung Adenocarcinoma Microenvironment Mining and Its Prognostic Merit

Technology in Cancer Research & Treatment
 Volume 19: 1-16
 © The Author(s) 2020
 Article reuse guidelines:
sagepub.com/journals-permissions
 DOI: 10.1177/1533033820977547
journals.sagepub.com/home/tct


Rongchang Zhao, MM¹ , Dan Ding, BM¹, Wenyan Yu, MM², Chunrong Zhu, MD², and Yan Ding, BM¹

Abstract

Background: As a common pathological type of lung cancer, lung adenocarcinoma (LUAD) is mainly treated by surgery, chemotherapy, targeted therapy and radiotherapy. Although a relatively mature treatment system has been established, there are few studies on the microenvironment of LUAD. **Material and Methods:** The immune and stromal scores of patients from the LUAD cohort in the TCGA database were obtained by using ESTIMATE. The relationship of immune and stromal scores with the clinicopathological characteristics and overall survival of LUAD patients was assessed by R. GO, KEGG and Cox regression analyses were employed to analyze intersecting genes and to identify reliable prognostic markers. The identified genes were also analyzed in the GEPIA database to assess their correlations with survival, and these relationships were verified with the Kaplan-Meier Plotter database. **Results:** The immune score was related to the survival time and tumor topography of LUAD patients. There was a significant correlation between stromal score and tumor metastasis. Through multivariate analysis, stage (HR = 1.640, 95% CI = 1.019-2.642, $P = 0.042$) and risk score (HR = 1.036, 95% CI = 1.026-1.046, $P < 0.001$). The genes (ARHGAP15, BTLA, CASS4, CLECL1, FAM129C, STAP1, TESPA1, and S100P) showed credible prognostic value in LUAD patients in TCGA through GEPIA database online analysis and verification in the Kaplan-Meier plotter database. **Conclusions:** In the microenvironment of lung adenocarcinoma, the differentially expressed genes screened by immune score and stromal score have certain value in evaluating the survival/prognosis of patients, as well as the invasion and progression of tumors.

Keywords

lung adenocarcinoma, TCGA, tumor microenvironment, ESTIMATE algorithm, overall survival

Received: June 08, 2020; Revised: October 21, 2020; Accepted: November 09, 2020.

Introduction

Lung carcinoma has become the main cause of cancer-related death, with an incidence rate of 11.6% and a mortality rate of 18.4%, in 185 countries worldwide.¹ Therefore, as the most common form of lung malignancy, LUAD deserves further study. To better understand the correlation between gene expression differences and clinical prognosis in tumor patients, a comprehensive gene data bank such as The Cancer Genome Atlas (TCGA) were utilized to classify and find genomic abnormalities in large sample populations around the world.² Considering the insufficient understanding of LUAD, it is essential to explore the evolution and progression of tumor cells in the human body by using these gene expression data.

The survival of humans is inseparable from the environment of the earth. Similarly, the survival of tumor cells cannot be considered as separate from the surrounding microenvironment. The tumor microenvironment (TME) usually consists of

non-tumor cells, including immune cells, fibroblasts, endothelial cells, and adipocytes.³ Solid malignancies are characterized by the uncontrolled proliferation and survival of normal cells caused by severe hypoxia and acidic stress, resulting in structural and functional abnormalities of the microenvironment. Therefore, a more detailed understanding of how adaptive responses to the microenvironment drive tumor progression is critical for the development of more rational cancer treatment

¹ Department of Oncology, Taixing people's Hospital Affiliated to Bengbu Medical College, Taixing, China

² Department of Oncology, The First Affiliated Hospital of Soochow University, Suzhou, China

Corresponding Author:

Yan Ding, Department of Oncology, Taixing people's Hospital, Chang Zheng road, Taixing, Jiangsu 225400, China.
 Email: zhao8296@163.com



strategies.^{4,5} In the microenvironment of tumors, the dynamic activity of molecules and cells reflects the nature of tumor evolution and is the basis of tumor immune escape, growth and metastasis.⁶ In 2013, an algorithm called ESTIMATE (Estimation of STromal and Immune cells in MAlignant tumor tissues using Expression data) was designed to assess tumor purity based on gene expression data.⁷ The algorithm is based on the analysis of specific gene expression profiles of stromal and

immune cells to predict the infiltration of non-tumor cells. It has been applied to prostate cancer,⁸ breast cancer,⁹ colon cancer,¹⁰ cutaneous melanoma,¹¹ glioblastoma,¹² and clear cell renal cell carcinoma¹³ so far, showing the effectiveness of this algorithm based on big data. However, to the best of our knowledge, the practicability of this algorithm has not been verified in LUAD. Therefore, we investigated the utility of LUAD on stromal and/or immune scores in detail.

Table 1. ESTIMATE Scores of Tumor Tissue.

| ID | StromalScore | ImmuneScore | ESTIMATEScore |
|------------------------------|--------------|-------------|---------------|
| TCGA-78-8648-01A-11R-2403-07 | 1991.679 | 2899.817 | 4891.496 |
| TCGA-50-5049-01A-01R-1628-07 | 1723.087 | 3096.331 | 4819.418 |
| TCGA-86-8671-01A-11R-2403-07 | 1458.873 | 3230.176 | 4689.049 |
| TCGA-44-6777-01A-11R-1858-07 | 1992.313 | 2654.481 | 4646.793 |
| TCGA-99-8028-01A-11R-2241-07 | 1399.158 | 3063.459 | 4462.616 |
| TCGA-50-5055-01A-01R-1628-07 | 1376.85 | 2950.409 | 4327.259 |
| TCGA-38-7271-01A-11R-2039-07 | 1281.863 | 3014.707 | 4296.57 |
| TCGA-75-6205-01A-11R-1755-07 | 1353.09 | 2929.736 | 4282.826 |
| TCGA-MP-A4TI-01A-21R-A24X-07 | 1259.481 | 2992.082 | 4251.563 |
| TCGA-50-8459-01A-11R-2326-07 | 1818.454 | 2415.926 | 4234.381 |

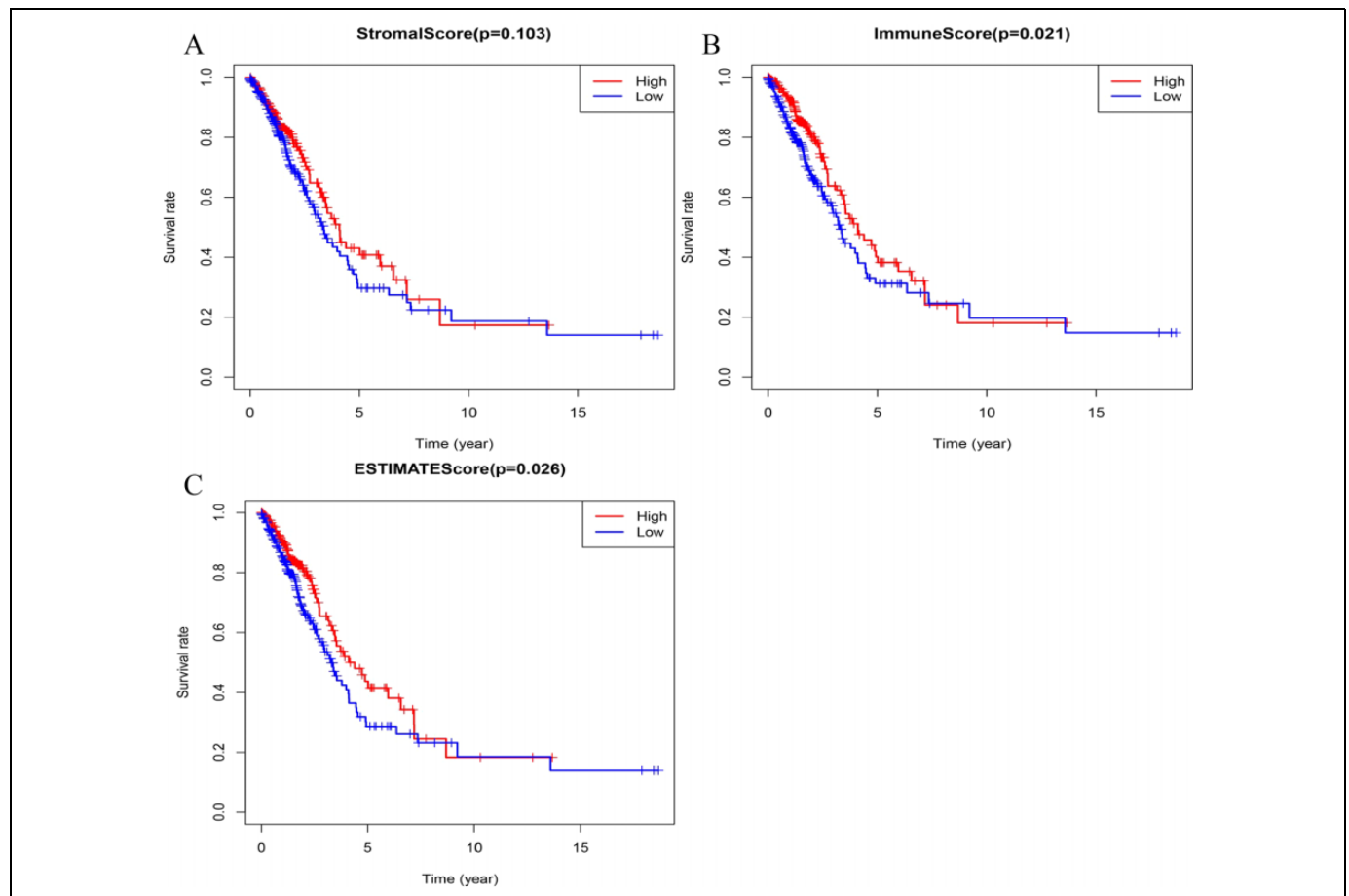


Figure 1. (A) Stromal score, (B) immune score, (C) ESTIMATE score. There was significant difference in survival time between low immune score and high immune score groups. Similar results were found between the low ESTIMATE score and high ESTIMATE score groups.

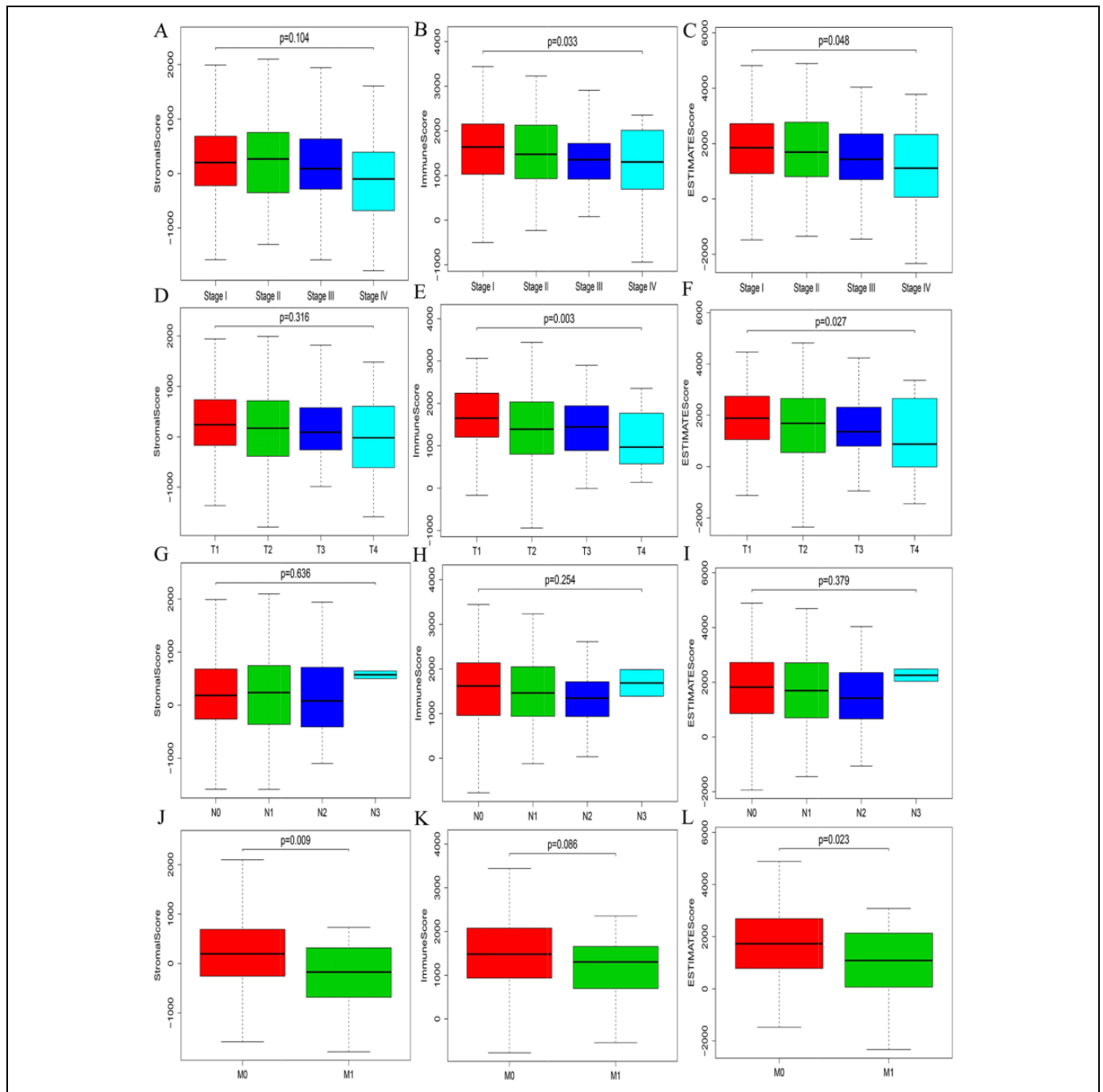


Figure 2. Relationship between tumor staging and (A) stromal, (B) immune, (C) ESTIMATE score. Relationship between T and (D) stromal, (E) immune, (F) ESTIMATE score. Relationship between N and (G) stromal, (H) immune, (I) ESTIMATE score. Relationship between M and (J) stromal, (K) immune, (L) ESTIMATE score.

Materials and Methods:

Data Download and ESTIMATE Scores

Gene expression quantification of transcriptome profiles from TCGA (<https://portal.gdc.cancer.gov/>) was downloaded. Perl (version 5.30.0) was employed to extract mRNA expression data from the downloaded data. The mRNA expression data of normal tissue samples and para-carcinoma tissue samples

were deleted by using the limma package, and the ESTIMATE scores were calculated by using the estimate package in R (version 3.6.1).

ESTIMATE Scores and Clinical Relationship

The clinical information of esophageal cancer was obtained from TCGA. Then, we used Perl to integrate ESTIMATE

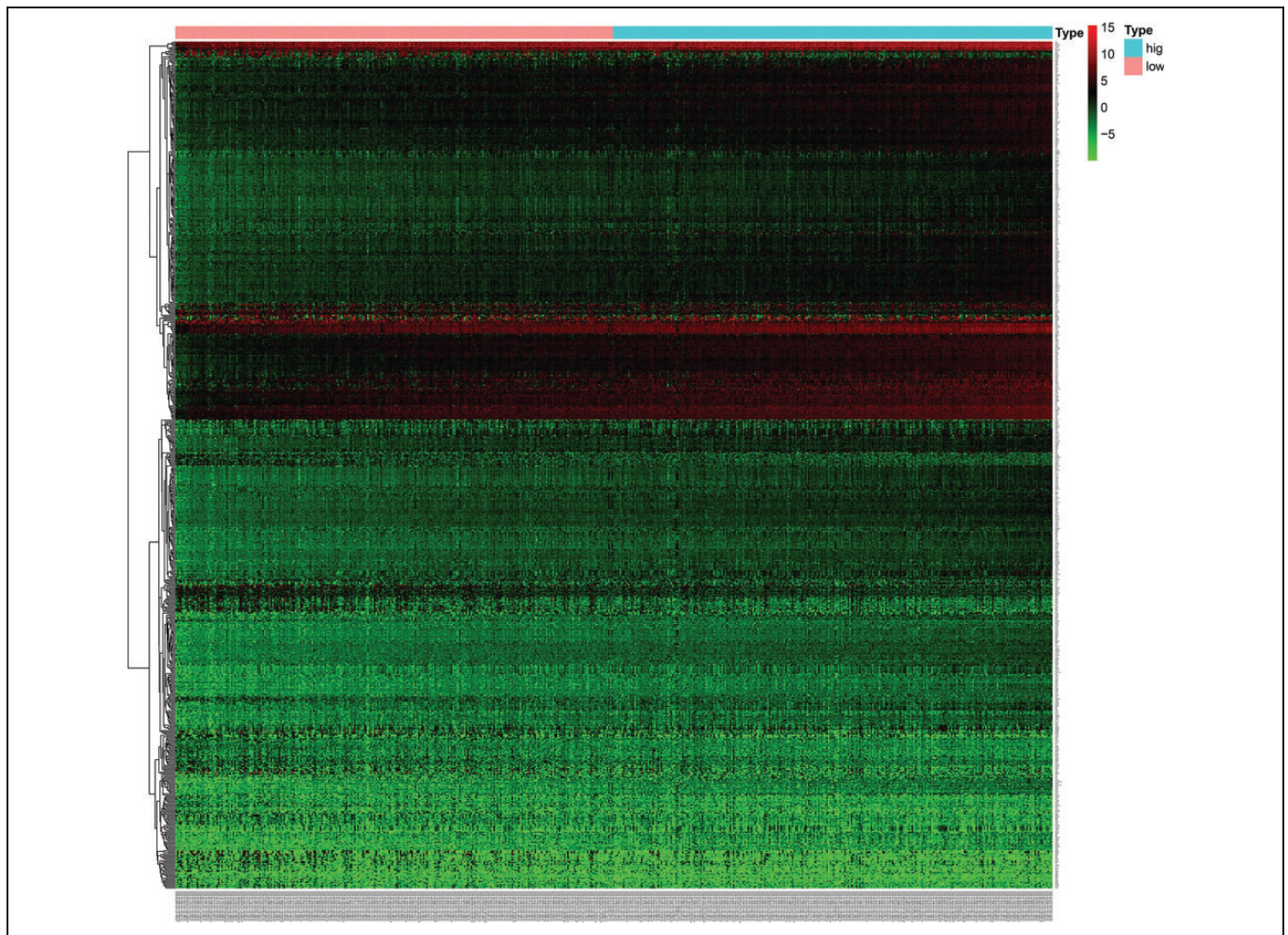


Figure 3. Heat map of DEGs related to stromal scores. Red represents genes with increased expression, green represents genes with decreased expression, and black represents genes with unchanged expression.

scores with data for survival, tumor stage, and tumor (topography) (T), lymph node (N), and metastasis (M) stage. Finally, the survival package in R was used for survival correlation analysis.

Differential Gene Analysis Was Performed in the Stromal Cell Group and the Immune Cell Group

The analysis of differentially expressed genes between the low- and high-score groups was executed based on the limma package, and the heat map was drawn in R. The cut-off criteria were set as follows: $|\log_2\text{fold change (logFC)}| > 1$; $P\text{-value} < 0.05$; FDR (false discovery rate) < 0.05 .

Functional Enrichment Analysis

R was used to draw the Venn diagram of upregulated differentially expressed genes and the Venn diagram of downregulated differentially expressed genes of the immune cell score group and stromal cell score group. Then, GO (Gene Ontology) and

KEGG (Kyoto Encyclopedia of Genes and Genomes) enrichment analyses were performed using the cluster Profiler R package. For both GO and KEGG, enrichment terms that met the cut-off criteria of $P\text{-value} < 0.05$ and Benjamin-Hochberg adjusted $P\text{-value} < 0.05$ were considered significant.

Survival Analysis

The intersecting genes were preliminarily assessed by univariate Cox regression analyses. Then, the multivariate Cox proportional hazards regression model was used to ensure the prognostic model, and risk scores were obtained. Subsequently, clinicopathological factors and risk scores undergo the same analytical procedures. Receiver operating characteristic (ROC) curve and the area under the ROC curve (AUC) for each clinicopathological element and risk score to measure the prognostic value of LUAD, which was employed by the “survivalROC,” with statistical significance defined as a $P\text{-value}$ less than 0.05. Genes related to the clinical characteristics of LUAD were obtained by integrating the intersection genes,

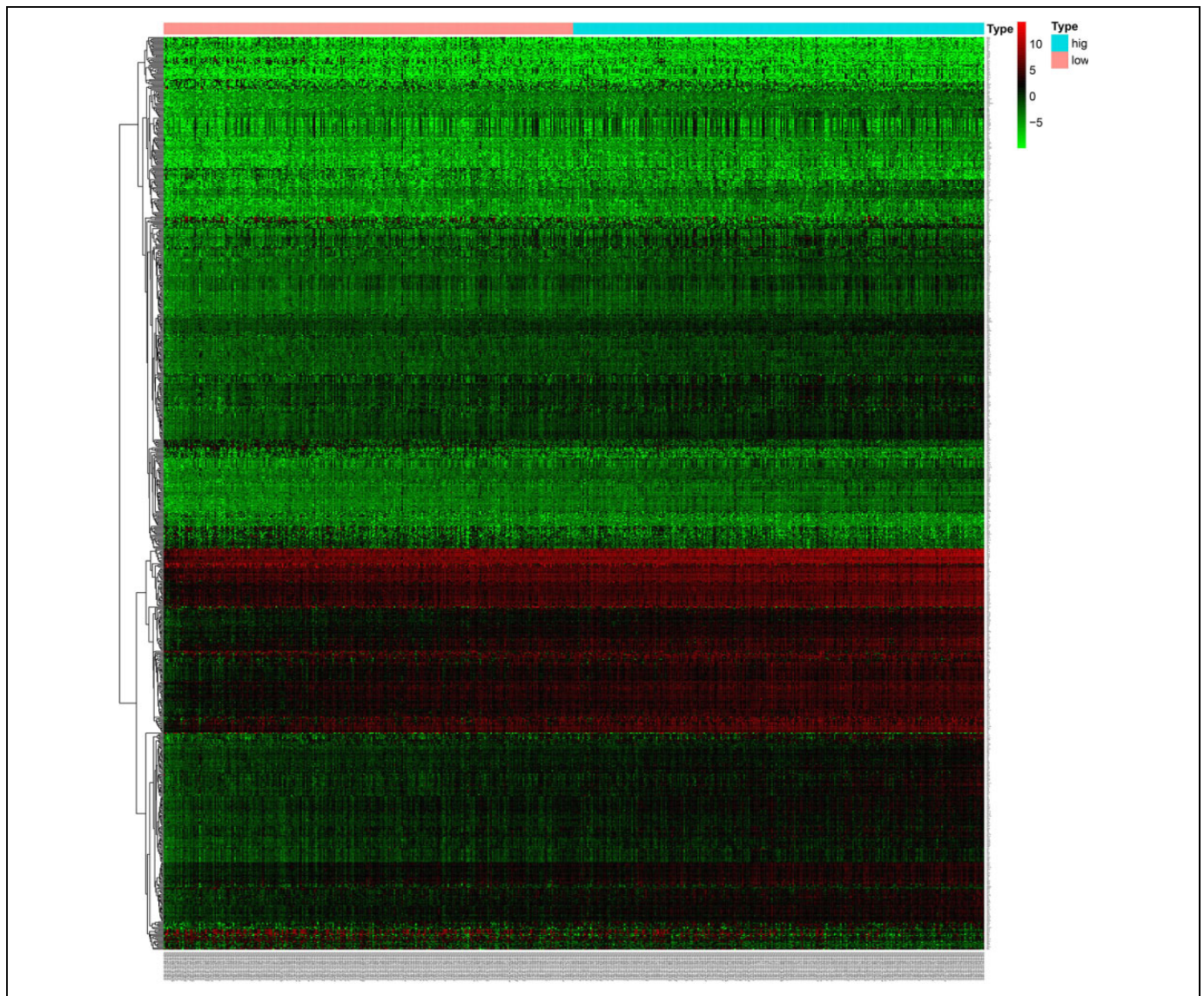


Figure 4. Heat map of DEGs related to immune scores. Red represents genes with increased expression, green represents genes with decreased expression, and black represents genes with unchanged expression.

clinical data, and risk score, with the P -value set to be less than 0.0001. All operations were implemented in R. The Gene Expression Profiling Interactive Analysis (GEPIA) database (<http://gepia.cancer-pku.cn/>) was used to determine the relationship between these genes and overall survival (OS). Finally, the Kaplan-Meier (KM) plotter database (<http://kmpplot.com/analysis/>) was used to verify the results of the survival analysis of these genes.

Results

Data Preprocessing and ESTIMATE Scores

We first extracted mRNA expression data from downloaded TCGA data, including 535 tumor tissue samples, 59 normal tissue samples and adjacent tissue samples. Then, we calculated the ESTIMATE scores of tumor tissue samples, and the

scores of 10 samples are shown in Table 1. Among 535 LUAD samples, the immune scores and stromal scores ranged from $-1,247.763$ to $3,377.796$ and $-2,350.481$ to $1,910.934$, respectively.

ESTIMATE Scores and Clinical Relationship

We used the above score data to integrate the survival data and found that there was no significant difference in survival time between the high and low stromal score groups (Figure 1A). However, the immune score (Figure 1B) was related to survival time. Here, our integrated survival data are the OS. We also analyzed the relationship between score and other clinical data. For example, tumor stage (T, N, M) was based on pathological results (Figure 2). This result suggested that the immune score was correlated with tumor stage

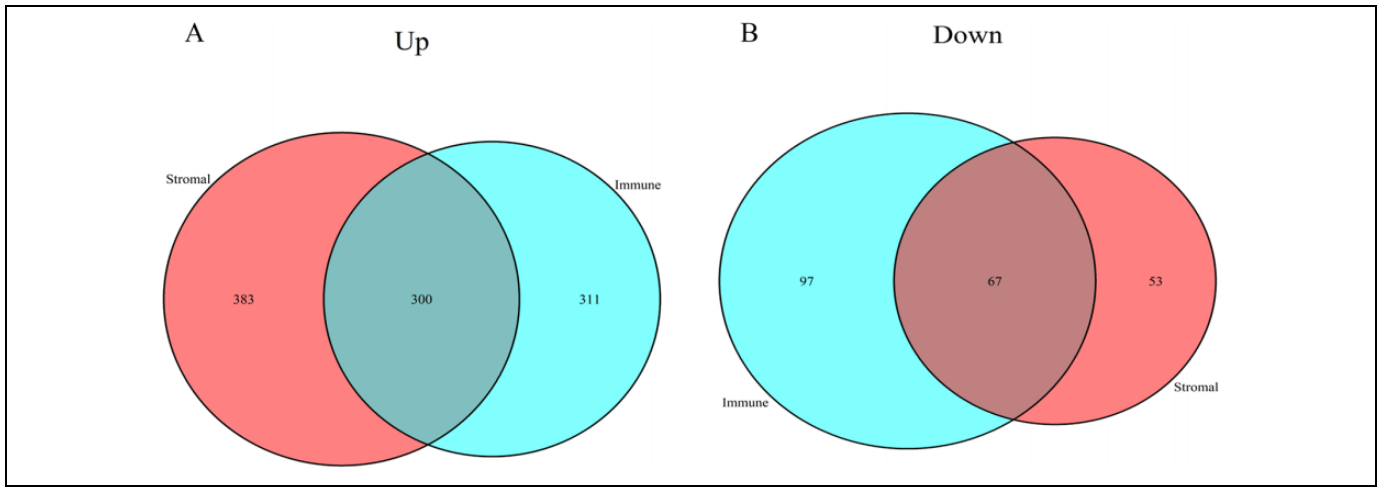


Figure 5. (A) The intersection of immune score and stromal score up-regulated genes. (B) The intersection of immune score and stromal score down-regulated genes.

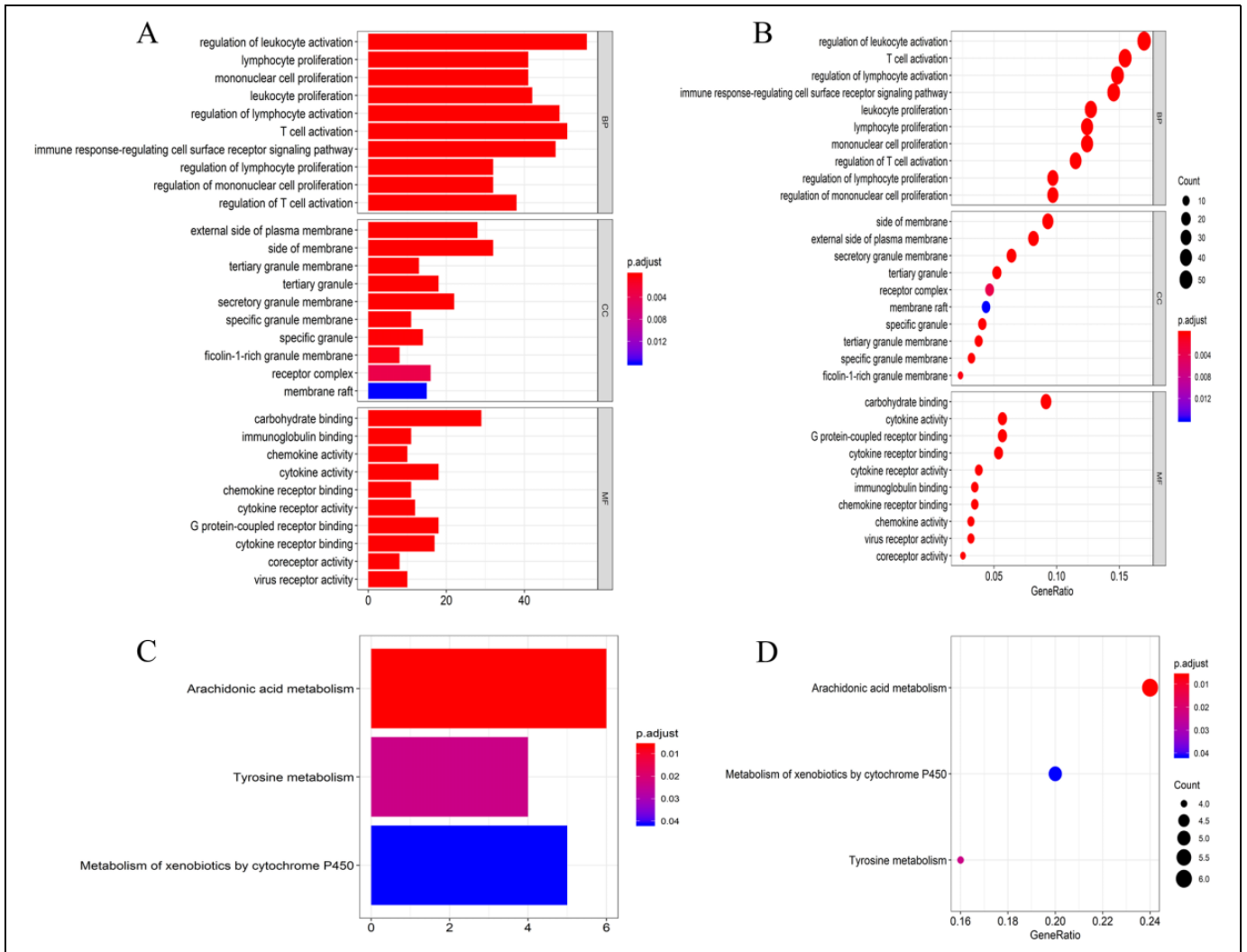


Figure 6. (A) Barplot of GO: BP, GO: CC, and GO: MF. (B) Dotplot of GO: BP, GO: CC, and GO: MF. (C) Barplot of KEGG pathways. (D) Dotplot of KEGG pathways. Functional enrichment analysis was performed in 367 commonly intersect genes.

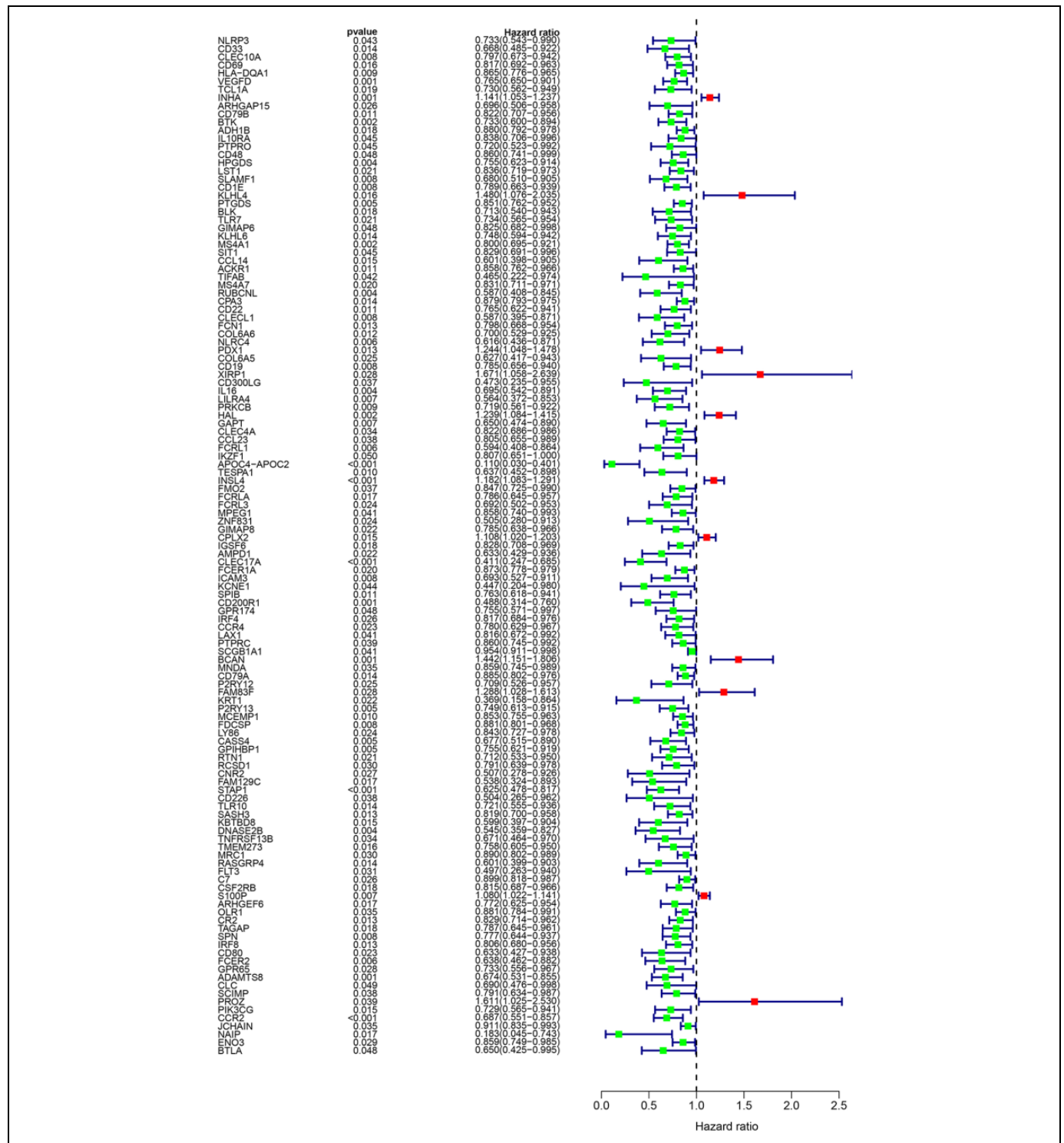


Figure 7. 125 genes associated with prognosis, after Univariate analysis.

(Figure 2B), specifically with the characteristics of the T (Figure 2E). There was a significant correlation between stromal score and tumor metastasis (Figure 2J). As the comprehensive score, the ESTIMATE score is a combination of the immune score and stromal score.

Differentially Expressed Genes in the Matrix and Immune Score of Esophageal Cancer

The limma package was used to analyze the differences in gene expression data between the low- and high-score groups.

Table 2. Prognostic Model Parameters.

| ID | Coef | HR | HR.95L | HR.95H | P-value |
|-------------|--------|--------|--------|--------|----------|
| CLEC10A | 0.353 | 1.424 | 0.893 | 2.269 | 0.137 |
| HLA-DQA1 | -0.291 | 0.748 | 0.583 | 0.959 | 0.022 |
| VEGFD | -0.484 | 0.616 | 0.427 | 0.891 | 0.010 |
| TCL1A | 0.765 | 2.149 | 1.144 | 4.040 | 0.017 |
| INHA | 0.122 | 1.130 | 1.004 | 1.271 | 0.042 |
| ARHGAP15 | 1.246 | 3.477 | 1.790 | 6.753 | 2.33E-04 |
| ADH1B | 0.583 | 1.791 | 1.307 | 2.453 | 0.000 |
| IL10RA | 0.683 | 1.979 | 0.953 | 4.113 | 0.067 |
| PTPRO | 2.203 | 9.049 | 3.045 | 26.895 | 7.39E-05 |
| HPGDS | -0.831 | 0.435 | 0.255 | 0.745 | 0.002 |
| SLAMF1 | -2.007 | 0.134 | 0.036 | 0.498 | 0.003 |
| KLHL4 | 0.546 | 1.727 | 1.083 | 2.753 | 0.022 |
| KLHL6 | 0.643 | 1.903 | 0.972 | 3.727 | 0.061 |
| ACKR1 | -0.382 | 0.683 | 0.510 | 0.914 | 0.010 |
| MS4A7 | -0.810 | 0.445 | 0.254 | 0.778 | 0.005 |
| RUBCNL | -0.867 | 0.420 | 0.174 | 1.015 | 0.054 |
| CLECL1 | -1.452 | 0.234 | 0.070 | 0.778 | 0.018 |
| NLRC4 | -1.159 | 0.314 | 0.108 | 0.912 | 0.033 |
| COL6A5 | 0.696 | 2.006 | 1.040 | 3.867 | 0.038 |
| LILRA4 | -1.066 | 0.345 | 0.138 | 0.860 | 0.023 |
| GAPT | 1.359 | 3.892 | 1.518 | 9.978 | 0.005 |
| CLEC4A | 0.596 | 1.815 | 0.925 | 3.560 | 0.083 |
| CCL23 | 0.660 | 1.935 | 1.151 | 3.254 | 0.013 |
| IKZF1 | 1.171 | 3.225 | 1.073 | 9.692 | 0.037 |
| APOC4-APOC2 | -3.271 | 0.038 | 0.004 | 0.330 | 0.003 |
| TESPA1 | 1.263 | 3.536 | 1.178 | 10.615 | 0.024 |
| INSL4 | 0.216 | 1.241 | 1.078 | 1.430 | 0.003 |
| MPEG1 | -0.770 | 0.463 | 0.267 | 0.803 | 0.006 |
| CPLX2 | 0.113 | 1.120 | 0.983 | 1.276 | 0.089 |
| CLEC17A | -1.215 | 0.297 | 0.069 | 1.282 | 0.104 |
| ICAM3 | -1.167 | 0.311 | 0.109 | 0.892 | 0.030 |
| CD200R1 | -2.366 | 0.094 | 0.030 | 0.293 | 0.000 |
| LAX1 | 0.808 | 2.243 | 1.002 | 5.018 | 0.049 |
| MNDA | 0.961 | 2.613 | 1.341 | 5.094 | 0.005 |
| P2RY13 | -0.771 | 0.463 | 0.234 | 0.913 | 0.026 |
| MCEMP1 | -0.494 | 0.610 | 0.436 | 0.854 | 0.004 |
| LY86 | 0.476 | 1.610 | 0.930 | 2.787 | 0.089 |
| CASS4 | -0.653 | 0.521 | 0.251 | 1.080 | 0.079 |
| RCSD1 | -0.950 | 0.387 | 0.162 | 0.925 | 0.033 |
| FAM129C | -1.745 | 0.175 | 0.042 | 0.729 | 0.017 |
| STAP1 | -1.571 | 0.208 | 0.098 | 0.440 | 0.000 |
| CD226 | -2.529 | 0.080 | 0.011 | 0.599 | 0.014 |
| DNASE2B | -0.626 | 0.535 | 0.265 | 1.076 | 0.080 |
| TNFRSF13B | 2.396 | 10.982 | 3.522 | 34.238 | 3.63E-05 |
| TMEM273 | 1.344 | 3.834 | 1.866 | 7.878 | 0.000 |
| S100P | 0.070 | 1.073 | 0.983 | 1.170 | 0.117 |
| ARHGEF6 | 1.268 | 3.555 | 1.669 | 7.573 | 0.001 |
| CR2 | -0.480 | 0.619 | 0.471 | 0.812 | 0.001 |
| TAGAP | 0.560 | 1.750 | 0.918 | 3.335 | 0.089 |
| CLC | -0.546 | 0.579 | 0.366 | 0.917 | 0.020 |
| PROZ | 1.245 | 3.472 | 1.753 | 6.877 | 3.58E-04 |
| CCR2 | -1.298 | 0.273 | 0.119 | 0.625 | 0.002 |
| ENO3 | -0.220 | 0.803 | 0.652 | 0.988 | 0.038 |
| BTLA | 1.921 | 6.828 | 1.672 | 27.894 | 0.007 |

Compared to the low score group, the high score group had 683 upregulated genes and 120 downregulated genes; a heat map was drawn (Figure 3) based on the comparison of the stromal

Table 3. Multivariate Cox Regression Analysis of These 8 Genes.

| ID | β | exp(coef) | se(coef) | z | p |
|----------|---------|-----------|----------|--------|-------|
| ARHGAP15 | 0.488 | 1.628 | 0.502 | 0.971 | 0.332 |
| CLECL1 | -0.202 | 0.817 | 0.421 | -0.479 | 0.632 |
| TESPA1 | 0.412 | 1.510 | 0.486 | 0.849 | 0.396 |
| CASS4 | -0.588 | 0.556 | 0.277 | -2.126 | 0.034 |
| FAM129C | -0.411 | 0.413 | 0.413 | -0.994 | 0.320 |
| STAP1 | 0.844 | 0.430 | 0.334 | -2.528 | 0.012 |
| S100P | 0.189 | 1.207 | 0.111 | 1.699 | 0.089 |
| BTLA | 0.367 | 1.443 | 0.566 | 0.648 | 0.517 |

Note. β , regression coefficients; exp(coef), hazard ratio; se(coef), standard error; Z, $\beta / \text{se}(\text{coef})$.

score results. Similarly, 611 upregulated genes and 164 downregulated genes were found in the low score group compared to the high score group, and a heat map was created (Figure 4).

GO and KEGG Analyses

The Venn algorithm was used to overlap the results, which revealed a total of 300 common upregulated DEGs (Figure 5A) and 67 common downregulated DEGs (Figure 5B). Then, functional enrichment analysis was performed in 367 intersecting genes. The top 10 GO terms of the BP, CC, and MF categories revealed that these intersecting genes were relevant to the regulation of leukocyte activation, side of membrane and carbohydrate binding (Figure 6A and B). KEGG analysis demonstrated that DEGs were particularly enriched in arachidonic acid metabolism, tyrosine metabolism and xenobiotics metabolism by cytochrome P450 (Figure 6C and D).

Survival Analysis of Intersecting Genes

Univariate analysis of each intersecting gene revealed that 125 genes may be associated with prognosis ($P < 0.05$) (Figure 7). Subsequently, these genes were subjected to a multivariate comprehensive analysis, and 54 genes were selected as prognostic models (Table 2). Through the joint analysis of risk score and survival data, we obtained a trend map of patients with low risk score and patients with high risk (Figure 8A) and found that the number of patients who died gradually increased as the risk score increased (Figure 8B); we also determined the relationship between the expression of 54 genes and risk value (Figure 8C). To identify the prognostic models in predicting the clinical outcome of LUAD patients, we analyzed the difference in survival time between the low- and high-risk groups by plotting the K-M plots, and we found that the low-risk group lived longer than the high-risk group ($P = 0$, Figure 9A). As an important feature of tumors, clinicopathological factors were further analyzed, and T stage (HR = 1.623, 95% CI = 1.310-2.011, $P < 0.001$), N stage (HR = 1.793, 95% CI = 1.465-2.194, $P < 0.001$), stage (HR = 1.645, 95% CI = 1.397-1.937, $P < 0.001$) and risk score (HR = 1.041, 95% CI = 1.029-1.054, $P < 0.001$) were identified as significant in the univariate analysis (Figure 9B). However, in the subsequent multivariate

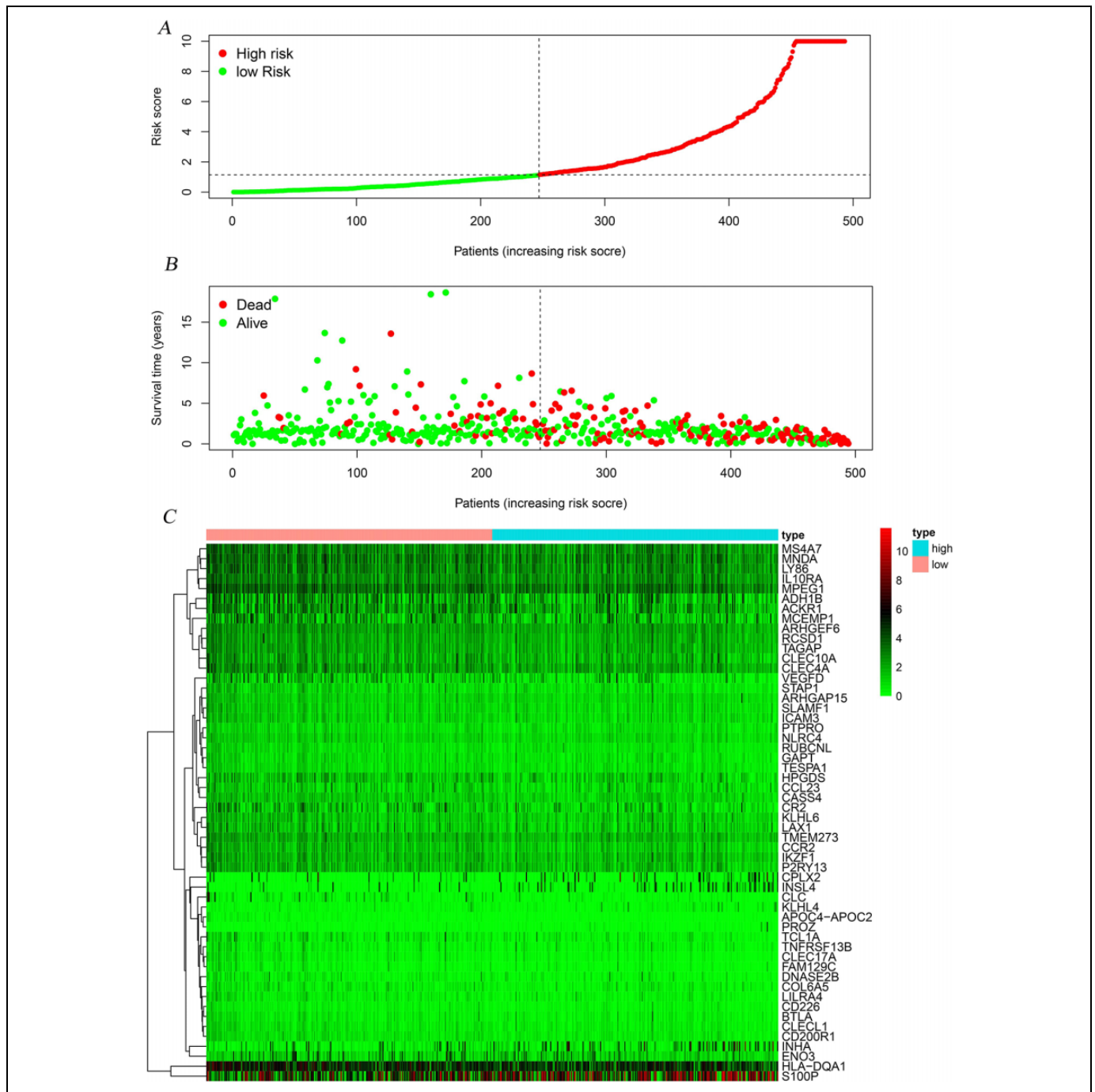


Figure 8. (A) Patient risk score distribution diagram. (B)The relationship between patient risk score and survival status. (C) The heat maps of risk scores and gene expression levels for 54 genes.

analysis, only stage (HR = 1.640, 95% CI = 1.019-2.642, $P = 0.042$) and risk score (HR = 1.036, 95% CI = 1.026-1.046, $P < 0.001$) had a P value less than 0.05 (Figure 9C). To evaluate how well the prognostic effects of clinicopathological parameters on LUAD patients, we performed a time-dependent ROC curve analysis (Figure 9D). The AUC for the risk score was 0.845 and stage was 0.714. To clarify the relationship between the above 54

genes and clinicopathological characteristics, we performed a T test and set the P value to 0.0001. This result suggested that there were significant differences in the expression levels of 16 genes in different groups of T, N, stage and gender (Figures 10-12). Finally, the GEPIA website was used to analyze the survival prognosis of these 16 genes, among which ARHGAP15 ($P = 0.00059$, Figure 13A), BTLA ($P = 0.0076$,

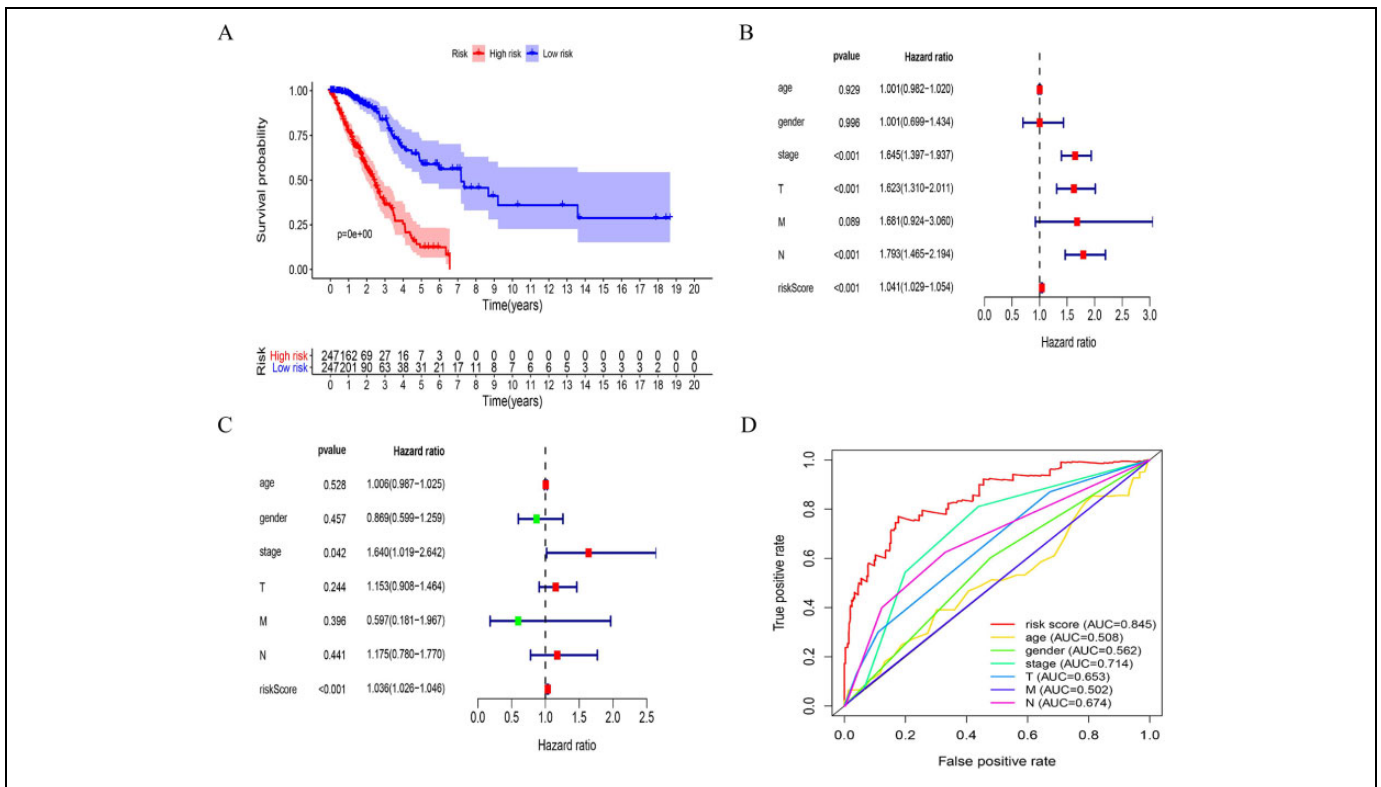


Figure 9. (A) Kaplan-Meier plot represents that patients in the low-risk group had significantly longer overall survival time than those in the high-risk group. (B) Univariate analysis of clinicopathological parameters and risk score. (C) Multivariate analysis of clinicopathological and risk score. (D) Time-dependent ROC curve analysis for survival prediction by the clinicopathological parameters and risk score.

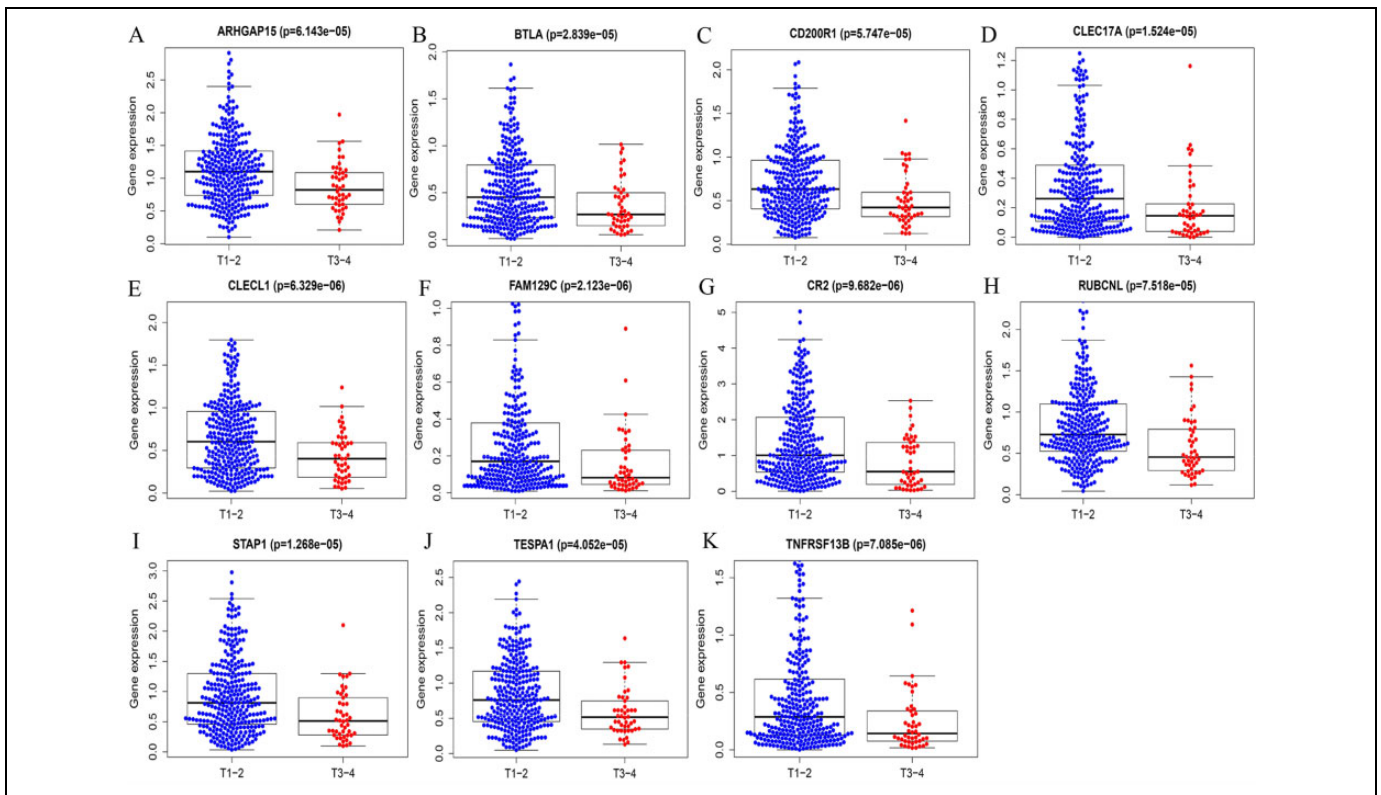


Figure 10. Differences in gene expression between different T groups. (A) ARHGAP15, (B) BTLA, (C) CD200R1, (D) CLEC17A, (E) CLECL1, (F) FAM129C, (G) CR2, (H) RUBCNL, (I) STAP1, (J) TESPA1, (K) TNFRSF13B.

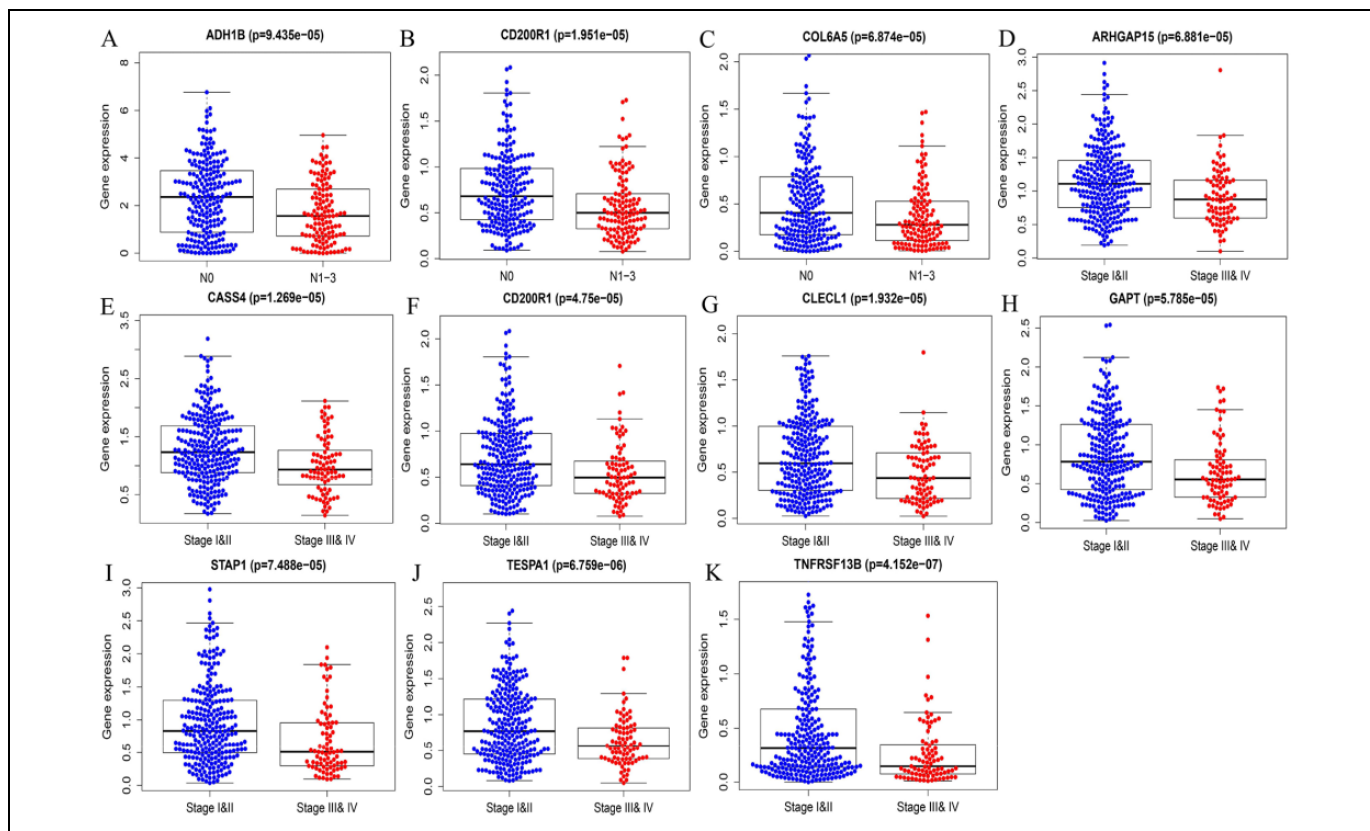


Figure 11. Differences in gene expression between different N groups. (A) ADH1B, (B) CD200R1, (C) COL6A5. Differences in gene expression between different stage groups. (D) ARHGAP15, (E) CASS4, (F) CD200R1, (G) CLECL1, (H) GAPT, (I) STAP1, (J) TESPA1, (K)TNFRSF13B.

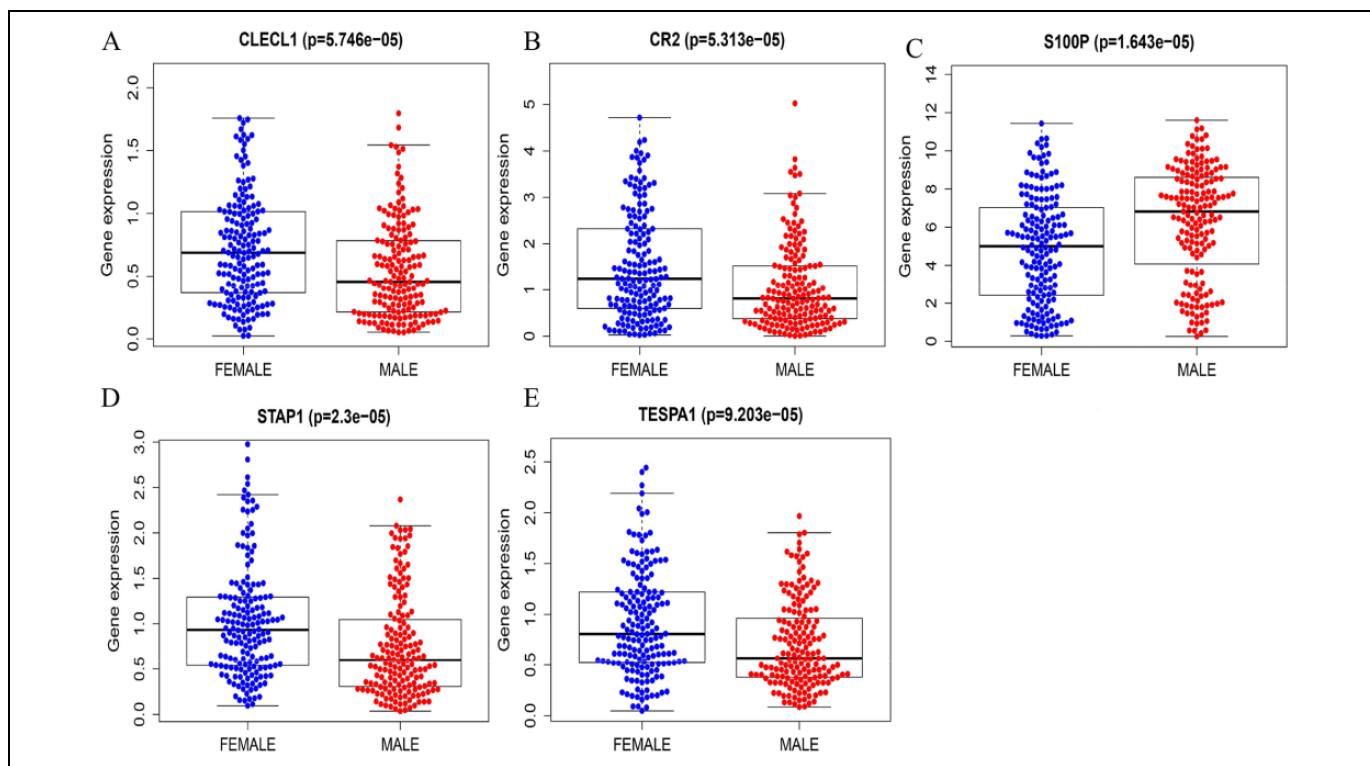


Figure 12. Differences in gene expression between different gender groups. (A) CLECL1, (B) CR2, (C) S100P, (D) STAP1, (E) TESPA1.

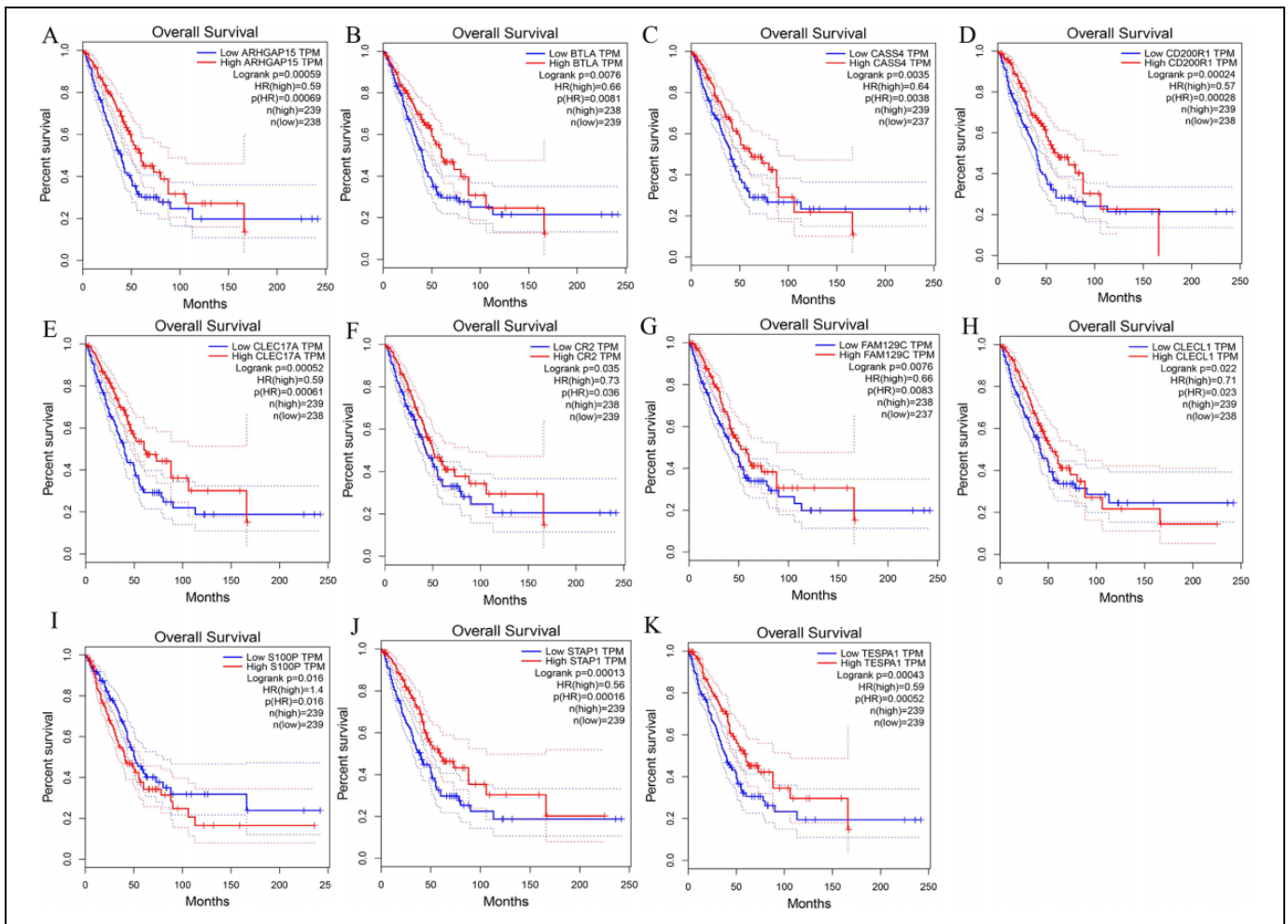


Figure 13. The overall survival of patients with low expressed genes was significantly shorter than those with high expression. (A) ARHGAP15, (B) BTLA, (C) CASS4, (D) CD200R1, (E) CLEC17A, (F) CR2, (G) FAM129C, (H) CLECL1, (J) STAP1, (K) TESPA1. The overall survival of patients with high expressed hub genes was significantly shorter than those with low expression. (I) S100P.

Figure 13B), CASS4 ($P = 0.0035$, Figure 13C), CD200R1 ($P = 0.00024$, Figure 13D), CLEC17A ($P = 0.00052$, Figure 13E), CR2 ($P = 0.035$, Figure 13F), FAM129C ($P = 0.0076$, Figure 13G), CLECL1 ($P = 0.022$, Figure 13H), S100P ($P = 0.016$, Figure 13I), STAP1 ($P = 0.00013$, Figure 13J), TESPA1 ($P = 0.00043$, Figure 13K), COL6A5 ($P = 0.098$), GAP1 ($P = 0.58$), KIAA0226 L ($P = 0.22$), TNFRSF13B ($P = 0.073$), ADH1B ($P = 0.17$).

Validation of the Selected Genes

After verification through the Kaplan-Meier plotter database, it was suggested that ARHGAP15 ($P = 1E-08$, Figure 14A), BTLA ($P = 0.00047$, Figure 14B), CASS4 ($P = 0.0021$, Figure 14C), CLECL1 ($P = 0.0015$, Figure 14D), FAM129C ($P = 0.0072$, Figure 14E), S100P ($P = 0.0064$, Figure 14F), STAP1 ($P = 0.00024$, Figure 14G) and TESPA1 ($P = 0.00097$, Figure 14H) were significantly correlated with prognosis. CD200R1 ($P = 0.21$) and CR2 ($P = 0.055$) showed no association with survival. CLEC17A was not found in the database.

Construction and Analysis of the Prognosis Risk Assessment Model of the 8 Genes

We performed multivariate Cox regression analysis of these 8 genes to obtain β (Table 3). The prognostic index = (0.488 * expression level of ARHGAP15) + (-0.202 * expression level of CLECL1) + (0.412 * expression level of TESPA1) + (-0.588 * expression level of CASS4) + (-0.411 * expression level of FAM129C) + (-0.844 * expression level of STAP1) + (0.189 * expression level of S100P) + (0.367 * expression level of BTLA). The LUAD patients were divided into low- and high-risk groups according to the median PI value (value = 1.036) of the prognostic risk score. There was a difference in survival time between the high-risk group and the low-risk group, and there was a certain accuracy in predicting the 3- and 5-year survival rates (Figure 15A-C).

Discussion

With the continuous innovation and development of gene detection technology, the use of gene sequencing technology

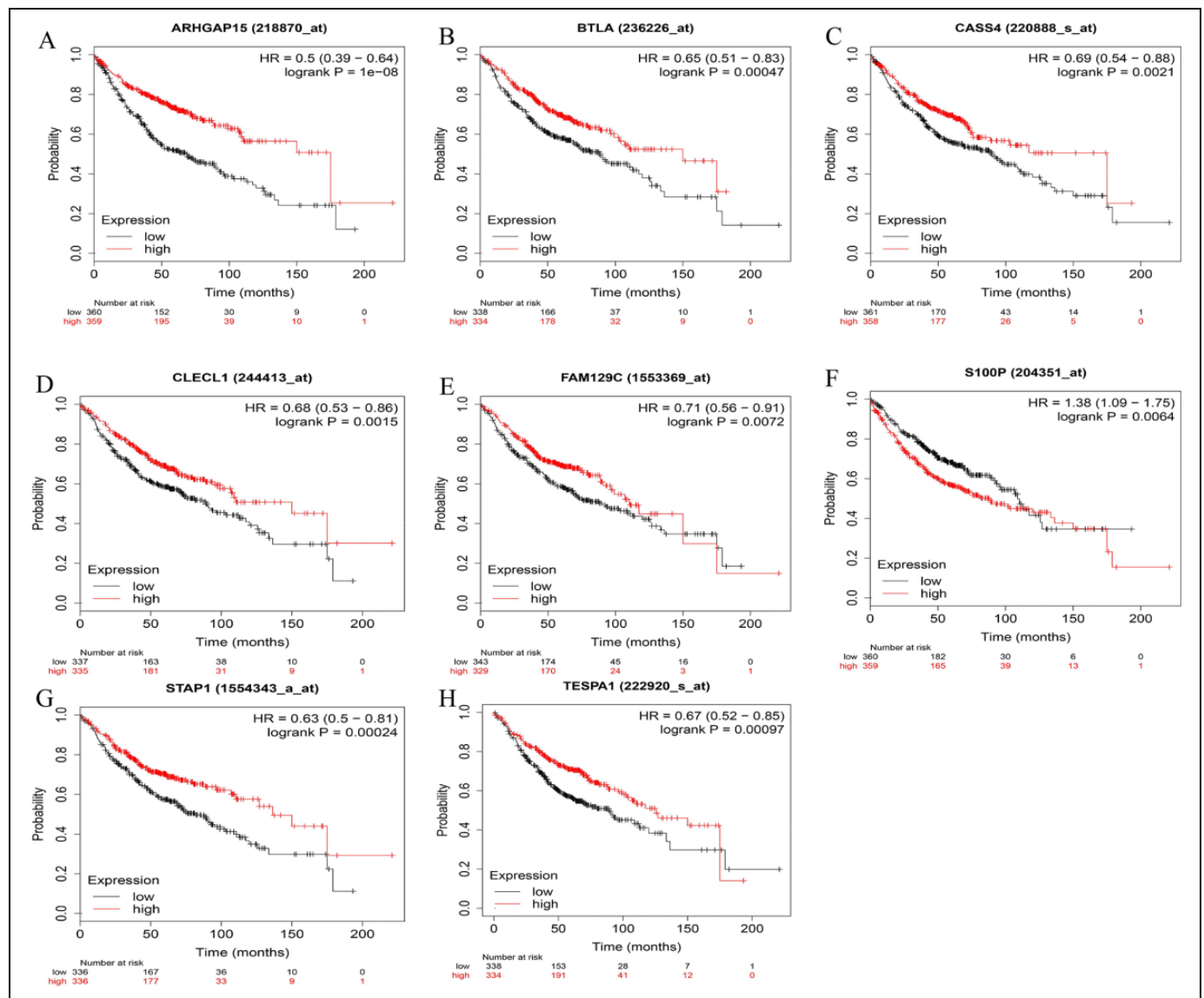


Figure 14. Validation of prognostic gene. The low expression group of (A) ARHGAP15, (B) BTLA, (C) CASS4, (D) CLECL1, (E) FAM129C, (G) STAP1 and (H) TESPA1 had a worse prognosis than the high expression group. The high expression group of (F) S100P had a worse prognosis than the low expression group.

to assess tumor tissue samples has progressed, and researchers are constantly exploring new targets and using statistical algorithms to perform external validation in LUAD. However, most of the current studies have failed to effectively classify and analyze the cellular components of tumor tissues, which may have a crucial impact on the characteristics of tumor treatment response, especially in precision immunotherapy.¹⁴

To gain insight into the invasion and progression of the tumor in patients with LUAD, we sought to explore the components of the TME and extract DEGs with important prognostic value. First, the relationship between stromal/immune scores and survival was analyzed, and it was found that there was a significant correlation. Then, we conducted further analysis of the relationship between pathological stage and score to determine the correlation between score and tumor

progression. Here, we found that the immune score and tumor stage had obvious relevance; specifically, the immune score was associated with the characteristics of the tumor size, location, and degree of invasion of the surrounding organs in esophageal cancer, which are reflected in the T part of the TNM stage, and stromal scores were associated with tumor metastasis. Next, we extracted 367 intersecting genes that participate in the immune response and extracellular matrix through differential expression analysis of transcripts in 535 LUAD samples with high versus low immune/stromal scores. We carried out functional analysis of these genes using the R package.¹⁵ Moreover, these 367 genes were subjected to univariate Cox regression analyses and multivariate Cox regression analyses to identify the prognostic model. Then, 54 genes were screened as a prognostic model, some of which, although their *P* value

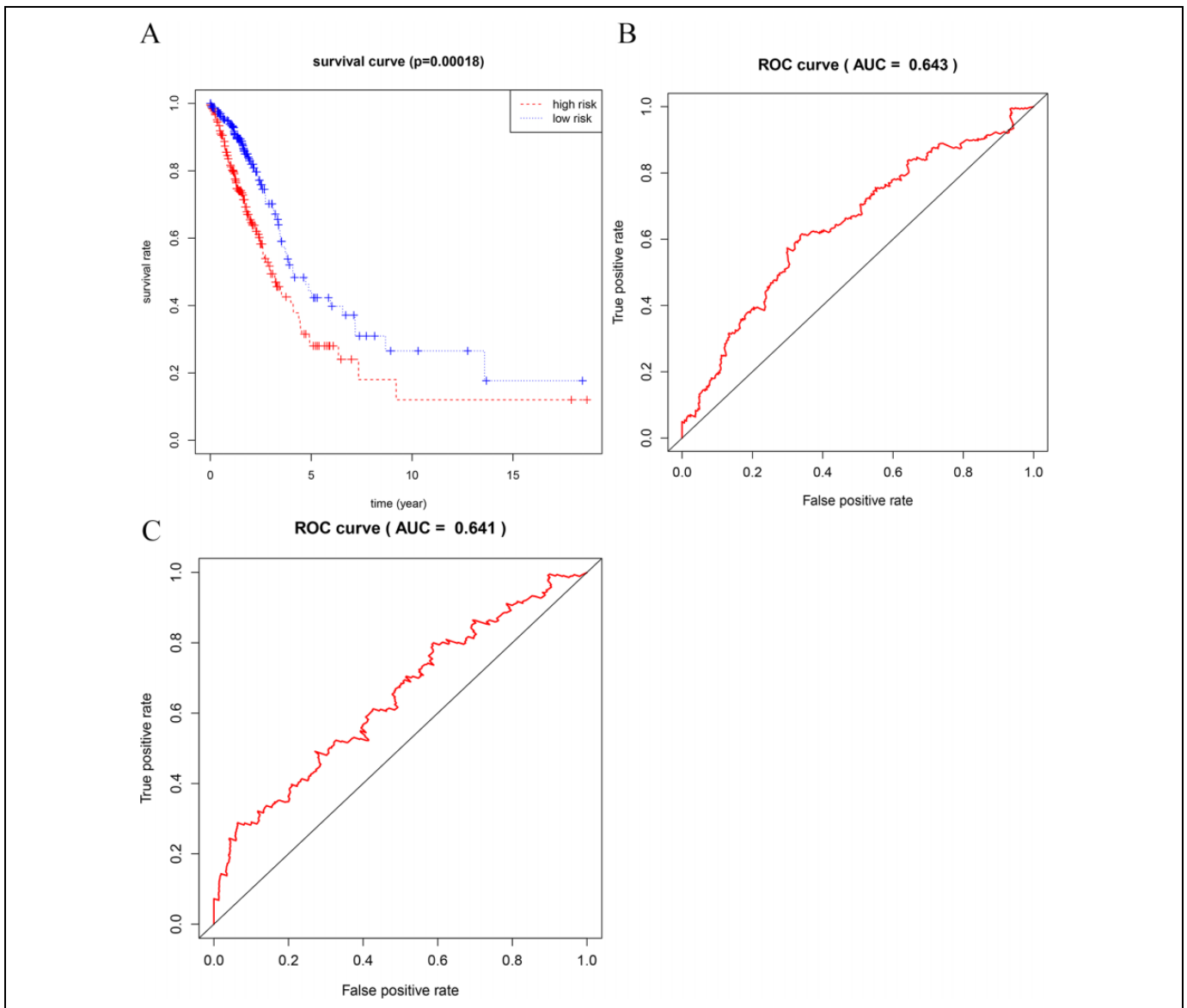


Figure 15. Kaplan–Meier and ROC curves for a linear risk model based on eight genes. (A) Log-rank test was used for the difference between high-risk group and low-risk group. (B) Time-dependent ROC curves analysis of eight genes for 3-year survival prediction. (C) Time-dependent ROC curves analysis for of eight genes 5-year survival prediction.

was greater than 0.05, were complementary to other genes. As the website for online analysis of TCGA data, GEPIA¹⁶ was used to analyze the prognosis of selected genes. The primary purpose of the Kaplan-Meier plotter database is a meta-analysis-based discovery and validation of survival biomarkers.¹⁷ Finally, 8 genes were identified as survival biomarkers after validation. This finding suggested that LUAD patients with downregulation of S100P had a better prognosis, while those with upregulation of ARHGAP15, BTLA, CASS4, CLECL1, FAM129C, STAP1 and TESPA1 had a better prognosis according to the Kaplan-Meier plotter database.¹⁸

As a member of the RhoGAP family, Rho GTPase activating protein 15 (ARHGAP15) plays a role in various biological processes of tumors, such as cell proliferation and migration.¹⁹⁻²²

It has been found in lung cancer cells that the upregulation of ARHGAP15 can inhibit cell proliferation, migration and invasion by reducing the expression of vascular endothelial growth factor (VEGF), matrix metalloproteinase-2 (MMP2), MMP9, and the phosphorylation of signal transducer and activator of transcription-3 (p-STAT3).²³ B and T lymphocyte associated (BTLA) plays a crucial role in the transmission of immunosuppressive signals.^{24,25} It has been found to increase the expression of CD4(+) and CD8(+) T cells in the pleural fluid of patients with lung cancer and can downregulate the expression of the T cell activation marker CD25 to inhibit the production of inflammatory cytokines.²⁶ In addition, a recent study found that high expression of BTLA was positively correlated with a high level of PD-L1, and it was also found that patients with negative

BTLA expression had longer relapse-free survival (RFS) than those with positive BTLA expression; in addition, patients who had negative expression of both had longer RFS than patients with positive expression of both.²⁷ Cas scaffold protein family member 4 (CASS4) has been found to promote non-small cell lung cancer (NSCLC) invasion by inhibiting the expression of E-cadherin, which involves the activation of the AKT signaling pathway.^{28,29} The expression of C-type lectin like 1 (CLECL1) can enhance the production of interleukin-4 and may be involved in the regulation of the immune response,^{30,31} but the relationship between CLECL1 and solid tumors has not been clarified. Because of its relationship with immune regulation, CLECL1 is a potential prognostic biomarker in LUAD. FAM129C is also known as niban apoptosis regulator 3, and it has only been found to be associated with the progression of ovarian cancer.³² Therefore, it can be used as a new target in the study of the mechanism of the occurrence and progression of LUAD, which is helpful for expanding the information available for this gene. The protein signal transducing adaptor family member 1 (STAP1) participates in a positive feedback loop by upregulating the activity of tyrosine protein kinase Tec, and variants of this gene are associated with elevated LDL cholesterol levels and an increased risk of coronary vascular disease.³³⁻³⁶ There are some studies that show that it is related to the B-cell receptor,³⁷ but the gene has not been explored in LUAD. At present, we lack an understanding of the mechanism of thymocyte expressed, positive selection associated 1 (TESPA1), in tumors, and only a few studies have suggested that it affects inositol 1,4,5-trisphosphate receptor-mediated Ca²⁺ signaling.^{38,39} Therefore, subsequent experiments are needed to further explore the role of TESP1 in the initiation and progression of LUAD. S100 calcium binding protein P (S100P) has been proven to be involved in the migration, invasion, and metastasis of lung cancer, and detection and targeting of this gene as a therapy is an attractive therapeutic strategy for lung cancer.^{40,41}

Conclusions

In general, the ESTIMATE algorithm has important significance for TME research. The immune score and stromal score are of great value in the prognostic evaluation of LUAD patients. In the microenvironment of LUAD, genes that affect the distribution of stromal cells and immune cells were screened by constructing a prognostic model and were further verified to obtain reliable prognostic monitoring indicators. New insight into the mechanism of LUAD could be provided by further study of these genes with prognostic value.

Declaration of Conflicting Interests

The author(s) declared no potential conflicts of interest with respect to the research, authorship, and/or publication of this article.


Ethical Statement

Our study did not require an ethical board approval because it did not contain human or animal trials.

Funding

The author(s) received no financial support for the research, authorship, and/or publication of this article.

ORCID iD

Rongchang Zhao  <https://orcid.org/0000-0002-8453-9446>

Supplemental Material

Supplementary material for this article is available online.

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. doi:10.3322/caac.21492
2. Cancer Genome Atlas Research Network. Comprehensive genomic characterization defines human glioblastoma genes and core pathways. *Nature.* 2008;455(7216):1061-1068. doi:10.1038/nature07385
3. Hanahan D, Coussens LM. Accessories to the crime: functions of cells recruited to the tumor microenvironment. *Cancer cell.* 2012; 21(3):309-322. doi:10.1016/j.ccr.2012.02.022
4. Brown JM, Wilson WR. Exploiting tumour hypoxia in cancer treatment. *Nat rev Cancer.* 2004;4(6):437-447. doi:10.1038/nrc1367
5. Neri D, Supuran CT. Interfering with pH regulation in tumours as a therapeutic strategy. *Nat rev Drug Discov.* 2011;10(10): 767-777. doi:10.1038/nrd3554
6. Jiang X, Wang J, Deng X, et al. Role of the tumor microenvironment in PD-L1/PD-1-mediated tumor immune escape. *Mol cancer.* 2019;18(1):10. doi:10.1186/s12943-018-0928-4
7. Yoshihara K, Shahmoradgoli M, Martinez E, et al. Inferring tumour purity and stromal and immune cell admixture from expression data. *Nat commun.* 2013;4:2612. doi:10.1038/ncomms3612
8. Shah N, Wang P, Wongvipat J, et al. Regulation of the glucocorticoid receptor via a BET-dependent enhancer drives antiandrogen resistance in prostate cancer. *ELife.* 2017;6:e27861. doi:10.7554/eLife.27861
9. Priedigkeit N, Watters RJ, Lucas PC, et al. Exome-capture RNA sequencing of decade-old breast cancers and matched decalcified bone metastases. *JCI insight.* 2017;2(17):e95703. doi:10.1172/jci.insight.95703
10. Alonso MH, Ausso S, Lopez-Doriga A, et al. Comprehensive analysis of copy number aberrations in microsatellite stable colon cancer in view of stromal component. *Br J Cancer.* 2017;117(3): 421-431. doi:10.1038/bjc.2017.208
11. Yang S, Liu T, Nan H, et al. Comprehensive analysis of prognostic immune-related genes in the tumor microenvironment of cutaneous melanoma. *J Cell Physiol.* 2020;235(2):1025-1035. doi:10.1002/jcp.29018
12. Jia D, Li S, Li D, Xue H, Yang D, Liu Y. Mining TCGA database for genes of prognostic value in glioblastoma microenvironment. *Aging.* 2018;10(4):592-605. doi:10.18632/aging.101415
13. Xu WH, Xu Y, Wang J, et al. Prognostic value and immune infiltration of novel signatures in clear cell renal cell carcinoma

- microenvironment. *Aging*. 2019;11(17):6999-7020. doi:10.18632/aging.102233
14. Li X, Wenes M, Romero P, Huang SC, Fendt SM, Ho PC. Navigating metabolic pathways to enhance antitumour immunity and immunotherapy. *Nat Rev Clin Oncol*. 2019;16(7):425-441. doi:10.1038/s41571-019-0203-7
 15. Yu G, Wang LG, Han Y, He QY. clusterProfiler: an R package for comparing biological themes among gene clusters. *OMICS*. 2012;16(5):284-287. doi:10.1089/omi.2011.0118
 16. 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. doi:10.1093/nar/gkx247
 17. Gyorffy B, Surowiak P, Budczies J, Lanczky A. Online survival analysis software to assess the prognostic value of biomarkers using transcriptomic data in non-small-cell lung cancer. *PLoS One*. 2013;8(12):e82241. doi:10.1371/journal.pone.0082241
 18. Hagenbaugh A, Sharma S, Dubinett SM, et al. Altered immune responses in interleukin 10 transgenic mice. *J Exp Med*. 1997;185(12):2101-2110. doi:10.1084/jem.185.12.2101
 19. Katoh Y, Katoh M. Identification and characterization of ARHGAP27 gene in silico. *Int J Mol Med*. 2004;14(5):943-947.
 20. Pan S, Deng Y, Fu J, et al. Decreased expression of ARHGAP15 promotes the development of colorectal cancer through PTEN/AKT/FOXO1 axis. *Cell Death Dis*. 2018;9(6):673. doi:10.1038/s41419-018-0707-6
 21. Takagi K, Miki Y, Onodera Y, et al. ARHGAP15 in human breast carcinoma: a potent tumor suppressor regulated by androgens. *Int J Mol Sci*. 2018;19(3):804. doi:10.3390/ijms19030804
 22. Sun Z, Zhang B, Wang C, et al. Forkhead box P3 regulates ARHGAP15 expression and affects migration of glioma cells through the rac1 signaling pathway. *Cancer Sci*. 2017;108(1):61-72. doi:10.1111/cas.13118
 23. Liu ZD, Mou ZX, Che XH, et al. ARHGAP15 regulates lung cancer cell proliferation and metastasis via the STAT3 pathway. *Eur Rev Med Pharmacol*. 2019;23(13):5840-5850. doi:10.26355/eurrev_201907_18326
 24. Watanabe N, Gavrieli M, Sedy JR, et al. BTLA is a lymphocyte inhibitory receptor with similarities to CTLA-4 and PD-1. *Nat Immunol*. 2003;4(7):670-679. doi:10.1038/ni944
 25. Quan L, Lan X, Meng Y, et al. BTLA marks a less cytotoxic T-cell subset in diffuse large B-cell lymphoma with high expression of checkpoints. *Exp Hematol*. 2018;60:47-56.e1. doi:10.1016/j.exphem.2018.01.003
 26. Wang XF, Chen YJ, Wang Q, et al. Distinct expression and inhibitory function of B and T lymphocyte attenuator on human T cells. *Tissue Antigens*. 2007;69(2):145-153. doi:10.1111/j.1399-0039.2006.00710.x
 27. Li X, Xu Z, Cui G, Yu L, Zhang X. BTLA expression in Stage I-III non-small-cell lung cancer and its correlation with PD-1/PD-L1 and clinical outcomes. *OncoTargets Ther*. 2020;13:215-224. doi:10.2147/ott.s232234
 28. Miao Y, Wang L, Liu Y, et al. Overexpression and cytoplasmic accumulation of Hsp1 is associated with clinicopathological parameters and poor prognosis in non-small cell lung cancer. *Tumour Biol*. 2013;34(1):107-114. doi:10.1007/s13277-012-0517-x
 29. Li A, Zhang W, Xia H, et al. Overexpression of CASS4 promotes invasion in non-small cell lung cancer by activating the AKT signaling pathway and inhibiting E-cadherin expression. *Tumour biology: the Journal of the International Society for Oncodevelopmental Biology and Medicine*. 2016;37(11):15157-15164. doi:10.1007/s13277-016-5411-5
 30. Zhao X, Singh S, Pardoux C, et al. Targeting C-type lectin-like molecule-1 for antibody-mediated immunotherapy in acute myeloid leukemia. *Haematologica*. 2010;95(1):71-78. doi:10.3324/haematol.2009.009811
 31. Sawcer S, Hellenthal G, Pirinen M, et al. Genetic risk and a primary role for cell-mediated immune mechanisms in multiple sclerosis. *Nature*. 2011;476(7359):214-219. doi:10.1038/nature10251
 32. Smebye ML, Agostini A, Johannessen B, et al. Involvement of DPP9 in gene fusions in serous ovarian carcinoma. *BMC Cancer*. 2017;17(1):642. doi:10.1186/s12885-017-3625-6
 33. Blanco-Vaca F, Martín-Campos JM, Pérez A, Fuentes-Prior P. A rare STAP1 mutation incompletely associated with familial hypercholesterolemia. *Clinica Chimica Acta*. 2018;487:270-274. doi:10.1016/j.cca.2018.10.014
 34. Amor-Salamanca A, Castillo S, Gonzalez-Vioque E, et al. Genetically confirmed familial hypercholesterolemia in patients with acute coronary syndrome. *J Am Coll Cardiol*. 2017;70(14):1732-1740. doi:10.1016/j.jacc.2017.08.009
 35. Fouchier SW, Dallinga-Thie GM, Meijers JC, et al. Mutations in STAP1 are associated with autosomal dominant hypercholesterolemia. *Circulation Research*. 2014;115(6):552-555. doi:10.1161/circresaha.115.304660
 36. Rose JE, Behm FM, Drgon T, Johnson C, Uhl GR. Personalized smoking cessation: interactions between nicotine dose, dependence and quit-success genotype score. *Mol Med (Cambridge, Mass)*. 2010;16(7-8):247-253. doi:10.2119/molmed.2009.00159
 37. Steeghs EMP, Bakker M, Hoogkamer AQ, et al. High STAP1 expression in DUX4-rearranged cases is not suitable as therapeutic target in pediatric B-cell precursor acute lymphoblastic leukemia. *Sci Rep*. 2018;8(1):693. doi:10.1038/s41598-017-17704-4
 38. Fujimoto T, Matsuzaki H, Tanaka M, Shirasawa S. Tespa1 protein is phosphorylated in response to store-operated calcium entry. *Biochem Biophys Res Commun*. 2013;434(1):162-165. doi:10.1016/j.bbrc.2013.02.128
 39. Parys JB, Vervliet T. New insights in the IP(3) receptor and its regulation. *Adv Exp Med Biol*. 2020;1131:243-270. doi:10.1007/978-3-030-12457-1_10
 40. Hsu YL, Hung JY, Liang YY, et al. S100P interacts with integrin $\alpha 7$ and increases cancer cell migration and invasion in lung cancer. *Oncotarget*. 2015;6(30):29585-29598. doi:10.18632/oncotarget.4987
 41. Tan BS, Yang MC, Singh S, et al. LncRNA NORAD is repressed by the YAP pathway and suppresses lung and breast cancer metastasis by sequestering S100P. *Oncogene*. 2019;38(28):5612-5626. doi:10.1038/s41388-019-0812-8