

# The pattern of alternative splicing in lung adenocarcinoma shows novel events correlated with tumorigenesis and immune microenvironment

Wensheng Qiu (✉ [wsqiuqd@163.com](mailto:wsqiuqd@163.com))

---

Research article

**Keywords:** LUAD, prognosis, alternative splicing, immune, tumor microenvironment

**Posted Date:** October 13th, 2021

**DOI:** <https://doi.org/10.21203/rs.3.rs-960327/v1>

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

[Read Full License](#)

---

# Abstract

Lung adenocarcinoma (LUAD) is the leading cause of cancer deaths worldwide due to the lack of early diagnostic markers and specific drugs. Previous studies have shown the association of LUAD growth with aberrant alternative splicing (AS). Herein, clinical data of 535 tumor tissues and 59 normal tissues were extracted from the TCGA database. Each sample was analyzed using the ESTIMATE algorithm; a comparison between higher and lower score groups (stromal or immune) was made to determine the overall- and progression-free survival-related differentially expressed AS (DEAS) events. We then performed unsupervised clustering of these DEASs, followed by determining their relationship with survival rate, immune cells, and the tumor microenvironment (TME). Next, two prognostic signatures were developed using bioinformatics tools to explore the prognosis of cases with LUAD. Five OS- and six PFS-associated DEAS events were implemented to establish a prognostic risk score model. When compared to the high-risk group (HRG), the PFS and OS of the low-risk group (LRG) were found to be considerable. Additionally, a better prognosis was found considerably associated with the ESTIMATE score of the patients as well as immune cells infiltration. Our analysis of AS events in LUAD not only helps to clarify the tumorigenesis mechanism of AS, but also provides ideas for revealing potential prognostic biomarkers and therapeutic targets.

## Introduction

lung carcinoma (LC) is known as one of the most prevalent forms of carcinoma and is the underlying reason for cancer mortality in China (1). In the US, it possesses a 5-year survival rate of 14% (2). Delayed diagnosis, particularly after metastasis, is one of the major causes for its poor prognosis for LC (3). Approximately 85% of all cases of LC have non-small cell LC (NSCLC), including varying histological types, such as LUAD, squamous cell carcinoma, and large cell LC, with LUAD being the most common type of NSCLC (4). Presently, there are no reliable biomarkers available for predicting prognosis in patients suffering from LUAD, which, in turn, requires the need for the identification of suitable prognostic markers.

Alternative splicing (AS) is a critical post-transcriptional process that involves the formation of protein isoforms with varied structural and functional characteristics. AS has a considerable contribution in the modification of > 95% of human genes and is thus, commonly used to explore proteome diversity and cellular complexity (5–7). Numerous investigations have revealed the association between AS abnormal expression and the pathogenesis of several cancers, including LUAD(8–11). The splice factor SRSF1 has been shown to modulate PTPMT1 alternative splicing to regulate LC cell radioresistance (12). Also, whole-genome analysis of AS events in LC has resulted in the identification of several candidate splicing factors, which might act as therapeutic targets of LUAD, and help in disease prognosis of patients via the construction of gene signatures (13, 14), demonstrating the role of AS in LC.

AS is directly linked with tumorigenesis and critically involved in the formation of TME, which includes tumor cells, tumor-associated fibroblasts, immune/inflammatory cells, microvessels, stromal tissues, and

many cytokines and chemokines(15–19). The prognosis of cancer patients is known to be directly related to the immune cell count in the TME, which can act as a useful prognostic marker (20–22). However, except for some preliminary studies on LC-related AS events (13, 14) and immunological microenvironment (23, 24), there is a lack of sufficient data on the immunological relevance of AS events.

Here, we integrated TME and AS events to conduct a first-of-its-kind analysis of prognostic variables in patients with LUAD. Firstly, the immune and stromal scores of patients were procured with LUAD through exploring public databanks and applying the ESTIMATE algorithm. Next, the Kaplan–Meier (K-M) plots were generated by using the obtained data to investigate prognostic variations between higher and lower score groups (stromal or immune). Subsequently, we searched for DEAS through comparative analysis of AS events in a higher and lower score group (stromal or immune). Then, we created two OS- and PFS-based prognostic signatures and determined that prognostic indicators were independent of other pathophysiological markers. Finally, we discerned distinct AS-based LUAD clusters. We studied the relationship between AS-based clusters and the pathophysiological factors and other immunological characteristics, allowing an improved understanding of the prognosis of patients with LUAD.

## Methods

### Collection of Data according to TCGA

We extracted clinical data regarding LUAD tissues and normal lung tissues from TCGA (<https://www.cancer.gov/>) database and regarding the AS events and the related values of percent-splice-in (PSI) from the TCGA spliceseq databank. The AS events were quantified and analyzed using these PSI values (0-1). To confirm the consistency of underlined AS events, we used samples (%) with PSI values larger than 75 %. Finally, the study was done on 490 patients with LUAD.

### Analyzing the link between stromal- or immune-scores and prognosis of LUAD patients

To assess the impact of the microenvironment on tumor cells regarding the evolution of genomic changes, the immune- and stromal-scores for the outcomes of the expression of mRNA were evaluated by using the ESTIMATE algorithm in the R program(25). The ESTIMATE algorithm provided these scores via ssGSEA. For each sample, the expression pattern was rank-ordered and normalized. After that, a statistic was calculated by integrating the difference between the empirical cumulative distribution function, which is similar to the one used in gene set-enrichment analysis but based on absolute expression rather than differential expression.

### Measurement and prognostic significance of immune- and stromal-scores

The ESTIMATE algorithm was employed for the evaluation of the immune- or stromal-scores, and the R-script was obtained from the website (<https://sourceforge.net/projects/estimateproject/>). These values were utilized to categorize the patients into higher and lower score groups (immune or stromal). Then, K-M survival plots were created to determine the prognostic significance of each group.

## Screening of DEASs on the basis of immune and stromal scores

To investigate the differences in prognosis between the higher and lower score groups (immune/stromal), we analyzed differential expression with PSI values of AS events. We set a restricted requirement of  $|\log_2 \text{FC}| > 0$  and  $\text{FDR/adjusted } P < 0.05$ , corresponding to the elevated and decreased expression of AS events, accordingly, due to the low PSI values of several AS events (26). The plots of heatmaps and volcano were created via the pheatmap program and ggplot2 in the R computer program, accordingly. Consequently, the intersecting AS events were chosen as DEASs and then further evaluated using a Venn diagram. Furthermore, to display the intersections between the seven types of DEASs in LUAD, the plot of UpSet was constructed using the package of UpsetR in R (27).

## Functional enrichment analysis

Following that, an enrichment analysis of the associated parent genes was conducted. Metascape was employed to evaluate KEGG pathways ([www.metascape.org](http://www.metascape.org)) and GO analyses. The data obtained from KEGG and GO analyses were used to indicate the first 20 key pathways, if possible.

## Evaluation of Consensus Clustering and Immune Microenvironment (IME)

The cohort of TCGA LUAD was categorized using hierarchical consensus clustering. Patients were grouped into multiple clusters using the ConsensusClusterPlus program, which was utilized to cluster in an unbiased and unsupervised behavior (28). The Elbow approach and the Gap statistic were used to confirm the ideal number of clusters, resulting in a reliable categorization. The Wilcoxon rank-sum test, often known as the K-W test, was performed to compare immune cells and the TME in the three groups.

## Development of DEASs based on prognostic model

To determine the predictive relevance of DEASs in LUAD patients, researchers used univariate Cox-regression analysis (CRA) to discover survival-correlated DEASs. Following that, the regression of least absolute shrinkage and selection operator (LASSO) was implemented to exclude potential predictors with nonzero coefficients, minimizing model overfitting and obtaining the simplest (smallest parameter) model possible (Gao et al. 2010). Based on negative log-likelihood statistics and Akaike Information Criterion, a multivariate-CRA was performed to thoroughly assess how each DEAS contributed to prognosis, verifying DEASs participated in the final prognostic signature. Using the findings of the multivariate-CRA and the PSI values, we estimated the risk scores of individuals with LUAD by the given formula:

$$\text{Risk Score} = \sum_{i=0}^n \beta_i * G_i$$

$\beta_i$  is the coefficient of gene  $i$  in the multivariate Cox analysis;  $G_i$  is the expression value of gene  $i$ ; and  $n$  is the number of genes in the signature.

On the basis of their score of median risk, cases with LUAD were separated into LRG and HRG, and K–M survival curves were formed to indicate the varied prognoses. We also created ROC curves for 1, 2, and 3 years to show the predicted signatures' percipience.

### **Development of nomograms integrating DEAS signatures and clinical variables**

Clinical variables, such as age, gender, TNM stage, and American Joint Committee on Cancer (AJCC) stage were obtained from the cBioPortal database(29) . For individuals with LUAD, we used a combination of signatures and clinical factors to perform univariate CA. For the multivariate CRA, factors with a  $P < 0.05$  were chosen as independent prognostic variables. The “rms” program in R was then used to create two prognostic nomograms based on the independent prognostic parameters to predict OS and PFS in LUAD patients. To distinguish between the two nomograms, the C-index was used, followed by the construction of calibration curves to appraise the concordance between actual and nomogram-predicted results.

### **Gene set variation analysis (GSVA) between HRG and LRG**

The “GSVA” R package was utilized to carry out the GSVA enrichment analysis to evaluate variations in the activation of the biological cascade between the HRG and LRG (30). An adjusted  $P$ -value  $< 0.05$  was regarded as statistically considerable.

### **Building of potential SF–AS regulatory network**

The SpliceAid2 databank (<https://www.introni.it/splicing.html>) was employed to obtain splicing factor (SF) data (31, 32). The analysis of Pearson correlation was carried out for evaluation of the link between the values of SFs and PSI in CAAS events. An association was found when the absolute value of the correlation coefficient was  $> 0.5$  and the  $P$ -value was  $< 0.001$ . Cytoscape (version3.8.2) was used to create the correlation plot.

### **Statistical analysis**

The R computer program (ver. 4.1.0) was employed for statistical evaluations.  $P < 0.05$  (two-sided) was regarded as statistically considerable.

## **Results**

### **The link between immune- and stromal-scores and the rate of Prognosis in LUAD Patients**

The transcriptome outcomes of 490 LUAD cases were taken from the TCGA databank, and the ESTIMATE algorithm was employed to identify their scores (immune/stromal). Then we separated the patients with LUAD into higher and lower score groups (stromal/immune) and examined the variations in prognosis between the two groups using the mean of stromal and immunological scores. Based on the data obtained from K-M curves, lower scores (immune and stromal) were substantially linked with the lower

OS in cases with LUAD (Fig. 1A–1B). Although no considerable link was detected between immunological and stromal scores with PFS in patients with LUAD (Supplement Fig. 1A, 1B). In addition, the differences in TNM staging between the high and low groups were also compared, and it was found that the T and M stages were statistically significant between the high and low immune groups (Fig. 1C,1D), and the M staging was statistically significant between the high and low stromal groups (Fig. 1E).

### **DEAS Events between Higher and Lower Score Groups (Immune/stromal)**

RNA-Seq data was utilized to create integrated AS event profiling, which led to 43948 AS events being identified. To guarantee high stringency, the AS events were evaluated using several filters (Standard Deviation  $\geq 0.01$ , average PSI value  $\geq 0.05$ ). As a result, 30569 AS events were discovered. The AS events in LUAD patients were then sorted, and it has been revealed that ES (Exon Skip) was the most common AS event, succeeded by AT (Alternate Terminator), and AP (alternate promoter). The intersections between AS events and the accompanying gene intersections were visualized using the UpSet plot (Fig. 2A). The DEAS events were then evaluated by the comparison of higher and lower score groups (immune or stromal), which were presented using heatmaps (Fig. 2B,2D) and volcano plots (Fig. 2C,2E). Thus, in the immune groups, 147 DEAS events were highly expressed while the expression of 131 DEAS events was decreased (Supplement Table 1). In the stromal groups, 199 DEAS events were highly expressed while the expression of 179 DEAS events was decreased in the underlined group (Supplement Table 2). The UpSet plots of DEAS events and the related gene intersections of the immune as well as stromal groups are depicted in Fig. 2F,2G, respectively. The Venn diagram program was employed to evaluate 96 DEAS events that were upregulated and 71 DEAS events that were downregulated (Fig. 3A,3B). The upset plot corresponding to the intersection DEAS events was also shown in the Supplement Fig. 1C.

The biological events and pathways were then evaluated by the GO and KEGG pathway analysis on the parent genes of DEAS events (Fig. 3C,3D). In the BP category, the most enriched pathways were immune response-activating cell surface receptor cascade of signaling and immune response-regulating cell surface receptor cascade of signaling; in the CC category, the most enriched GO terms were adherens junction and cell leading edge; in the MF category, cell adhesion molecule binding and Ras guanyl-nucleotide binding were the primary functions of these genes. For the KEGG pathway enrichment study, the major pathways were Axon guidance and T cell receptor signaling. The enrichment analysis results show that the parental genes of these AS events are closely related to the immune response.

### **AS-based clusters are considerably linked with prognosis, and also molecular, and immune characteristics**

The AS profiling revealed that the AS events among LUAD patients were incredibly heterogeneous. This finding makes us wonder whether we can predict the clinical outcome of each patient through the analysis of changes in the expression of AS events. Furthermore, unsupervised consensus analysis was carried out to explore whether AS had any discernible patterns (Fig. 4A). We found three distinct AS-based molecular clusters based on the consensus matrix heatmap: C1 (n = 128, 26.3%), C2 (n = 174, 35.7%),

and C3 (n = 185, 38.0%). To better understand the clinical effects of the evaluated AS clusters, we looked at the correlations between clusters and clinicopathological features, as depicted in Fig. 4B. The stage of TNM and survival status (OS and DFS) were not distributed in a random manner across clusters. Furthermore, the link of K–M analysis of cluster with prognosis yielded distinct survival patterns. Notably, C3 was linked to a positive OS, whereas C1 and C2 were related to a negative OS (Fig. 4C). DFS, on the other hand, did not differ considerably among the clusters (Supplement Fig. 1D).

In addition, differences in the immunological microenvironment amongst AS clusters were investigated. The scores (immune and stromal) were determined using the ESTIMATE method to measure the existence of immune and stromal cell infiltration in samples of the tumor. We found that the immune and stromal score were all highest in C3 (Fig. 4D,4E). Furthermore, using ssGSEA, a heatmap was created to examine the relative abundances of the 23 immune infiltrating cell subpopulations among different AS clusters (Fig. 4F). Higher scores (immune and stromal) were related to the C3 cluster, while lower scores (immune and stromal) were connected with the C1 and C2 clusters. The biological behaviors among these different AS patterns were then investigated using a GSVA enrichment analysis. B cell receptor signaling pathway, cytokine-cytokine receptor interaction, natural killer cell mediated cytotoxicity, leukocyte transendothelial migration, T cell receptor signaling pathway and Toll like receptor signaling pathways were among the enrichment cascades presented by the C2 and C3, as depicted in Fig. 4G,4H. The results show that patients with LUAD have a better prognosis when their immunological and stromal scores are greater.

### **Establishment and assessment of the prognostic signature for LUAD patients**

Biomarkers for early illness identification and potential therapeutic targets are still a hot topic in medicine. Previous research has found that abnormal AS events in the initial phases of cancer and are used as prognostic markers in a variety of cancers (33, 34). We generated signatures based on the DEASs to extract the underlying predictive value of individual DEASs, followed by a univariate CRA. The findings demonstrated that in LUAD patients, 39 and 12 intersecting AS events were strongly linked with OS and PFS, accordingly (fig 5A,6A). To avoid model overfitting, the LASSO regression was adopted to select the optimal OS-(fig 5B,5C) and PFS-(fig 6B,6C) related DEASs to construct the prediction models. After performing multivariate CRA, 11 DEASs were used to create two prognostic signatures, *i.e.*, 5 DEASs for the OS signature, 6 DEASs for the PFS signature, and one overlapping DEAS. We estimated each LUAD patient's risk score using the formula in the methods, and on the basis of the median of the risk score, we separated the cases into two groups: HRG and LRG. The AUCs of the OS signature to predict 1, 2, and 3 years OS were 0.709, 0.656, and 0.669, accordingly, according to time-dependent ROC curves (fig 5D). Patients in the HRG had a lower OS than those in the LRG, according to the K-M curves (fig 5E).

The PFS signature's time-dependent ROC curves were also developed in the same way. A median was used to determine the risk score cutoff, and 208 and 209 patients were categorized into two groups: HRG and LRG, respectively. The signature had AUCs of 0.723, 0.688, and 0.702 for predicting PFS after 1, 2, and 3 years, accordingly (fig 6D). The PFS was better in the patients with LR scores (fig 6E). The

underlined data revealed that both signatures were considered to predict the survival rate of LUAD patients. Additionally, heatmaps (fig 5F,6F), survival status plots (fig 5G,6G), and risk score plots (fig 5H,6H) were created to clearly illustrate variations in prognosis and AS patterns.

### **Construction of nomograms based on DEAS signature as well as clinical parameters.**

To enhance the clinical use of the underlined prognostic signatures, two complete nomograms integrating independent clinical characteristics were created. To investigate independent OS and PFS prognostic variables, we first used univariate and multivariate CRA. The obtained data revealed that Risk scores, T, and N were three independent variables associated with OS and PFS (Supplement Table 3,4). These findings showed that both DEASs-based signatures might be employed to predict the prognosis of LUAD cases independently.

Following that, on the basis of the independent prognostic factors, two novel nomograms were created for predicting OS (fig 7A) and PFS (fig 7E). The C-indices for the OS and PFS nomograms were 0.704 (95% CI 0.661–0.747) and 0.69 (95% CI 0.647–0.733), accordingly. The nomograms' significant calibration suggested that the outcomes predicted by the nomograms showed consistency with the observed data (fig 7B-D,7F-H).

### **Potential Regulatory Network Between SFs and AS Events**

SFs are protein factors involved in the splicing process of RNA precursors and are closely related to AS events. We retrieved 71 SFs from the SpliceAid2 databank to investigate the fundamental regulatory network between SFs and AS events in LUAD cases. We identified 42 SFs (blue) that were substantially connected to 39 AS events linked with survival, comprised of 9 favorable AS events (green) and 30 adverse AS events (red) with OS signature using Spearman correlation analyses (fig 8A). The PFS signature was screened out of 35 SFs (blue) that were substantially connected to 11 survival-associated AS events, with 6 adverse AS events (red) and 5 favorable AS events (green) (fig 8B). Many SFs were found to be associated with numerous AS events and to play opposing roles in regulating various AS events. Different SFs may also regulate a certain AS event. The underlined phenomenon elucidates that a single transcript can result in many splicing events.

### **Analyzes of SOD2-78301-AT Regulating Relationship**

In the multivariate CRA, SOD2-78301-AT was evaluated to be an independent OS- and PFS-associated AS and was added into both signatures as a DEAS. The expression of SOD2-78301-A T and the prognosis of LUAD patients were also investigated. A low level of SOD2-78301-AT was associated with considerably superior OS (fig 9A) and PFS (fig 9B), according to K-M survival curves. The link between SOD2|78301|AT and its only SF (ESRP2) was also investigated, and the findings verified their negative regulatory interaction (fig 9C). Finally, in LUAD, we investigated the link between SOD2|78301|AT's parental gene SOD2 and immune cells. There was a statistically considerable association between SOD2 expression in LUAD and the presence of immunological infiltrates (CD4 + T cells, B cells, macrophages, CD8 + T cells,

neutrophils, and dendritic cells) ( $P < 0.05$ ) (fig 9D). GISTIC 2.0 defines somatic copy number alterations (SCNAs) as arm-level deletion, deep deletion, arm-level gain, diploid/normal, and strong amplification. In LUAD, box plots were used to indicate the distributions of all immune subcategories at each copy number status with SOD2 (fig 9E).

## Discussion

Immunotherapy has revolutionized the treatment of cancers over recent decades and possesses a considerable role in LUAD treatment. A randomized Phase III trial comparing atezolizumab with docetaxel in cases with earlier treated progressive NSCLC, which revealed that atezolizumab outperformed docetaxel in terms of OS. Across programmed death-ligand 1 and histological subcategories, atezolizumab was found to improve the survival rate. Patients who received atezolizumab had fewer treatment-associated adverse events of grade 3 or 4 than those who received docetaxel (35). Durvalumab indicated clinically substantial enhancements in OS and PFS than the standard of care in strongly pretreated patients with metastatic NSCLC. Durvalumab + tremelimumab indicated numerical enhancements in OS and PFS relative to standard of care (36). Immunotherapy has given people with LC novel hope, but it is not appropriate for all of them. To enhance the prognosis of LC patients, more research is required to identify immune-associated prognostic markers and to develop innovative therapeutic options. The immune-associated DEAS events were employed to build a model of risk score that accurately predicted the outcomes of LUAD patients in our investigation.

Herein, the ESTIMATE algorithm was employed to determine scores (immune as well as stromal) of LUAD generated from the TCGA database from the microenvironment's perspective. We then used K–M curves to predict the prognosis of LUAD and indicated that cases with a greater immune/stromal score exhibited a higher chance of survival. We also evaluated immune/stromal score-associated DEASs and then chosen the optimal DEASs linked with the survival using LASSO CRA through comparing the profiles of transcriptional expression within LUAD cases with higher against lower immune/stromal scores. In addition, the final prognostic signature was developed, demonstrating that it is capable of accurate prediction. The LRG, in particular, had a higher chance of surviving than the HRG. Furthermore, the risk score can be used for predicting LUAD patient survival as an independent factor. This characteristic, when taken as a whole, holds a lot of promise for predicting LUAD patients' survival.

The biological functions, such as adherens junction, cell-matrix adhesion, lysosome, and others, were found to be closely linked with the development, growth, and progression of tumors, based on GO and KEGG enrichment analysis; Adherens junctions are significantly linked with the invasive and migratory potential of cancerous cells(37), for instance, E-cadherin is one of the key constituents of adherens junction, and a tumor suppressor and its loss is linked with a bad prognosis in a wide spectrum of cancers, including prostate cancer (38) and neck and head cancer (39). Cell-matrix adhesion has also been linked to the progression of many cancers (40–42). Lysosomes considerably contribute to the cell's degradation, and defects in them can lead to unregulated cell proliferation (43). All of the above shows

that the biological effects related to AS events are inseparable from the occurrence and development of tumors.

We also performed an analysis of unsupervised cluster to divide the TCGA LUAD cohort into three subgroups. The patients with a good prognosis had the greatest immune, stromal, and ESTIMATE scores, as well as the greatest level of immune cell infiltration, according to our findings. Immune cells are key prognostic factors in patients with LUAD in earlier research (44). For instance, the increase in several targetable immune checkpoint molecules is linked to EMT and the microenvironment of inflammatory tumors (45). Öjlert et al. discovered that a higher immune score and great estimations of numerous adaptive immune cell types were linked to higher PFS in LUAD patients(23). Jones et al. discovered that the immunological signatures of cytotoxic lymphocytes and T-cell trafficking were linked to the prognosis of female malignancies like LUAD, showing that immune cell infiltration in the TME influenced treatment and survival results in LUAD cases(46).

In the current study, we also evaluated the prognostic value of DEASs. An OS-prognostic signature based on 5 DEASs (MKL1|62349|AP, ICAM3|47503|RI, SOD2|78301|AT, CLASP2|63869|AP, and EXOC6|12541|AP) and PFS-prognostic signature based on 6 DEASs (DNMT3A|52857|AT, NCK1|66941|AP, SOD2|78301|AT, IFRD1|81447|AP, SH3KBP1|88642|AP, and ABI1|11048|ES) was constructed. Interestingly, among the splicing events, one overlapping AS event (SOD2|78301|AT) was discovered with substantial variations based on OS and PFS concurrently, implying that SOD2|78301|AT is the most plausible independent prognostic factor. Its parental gene SOD2 is a crucial enzyme for scavenging ROS generated by mitochondria, which is necessary for cellular homeostasis(47). Reported studies have been revealed that SOD2 has been linked with the energy metabolism of colorectal cancer (48), oxidative stress in endocrine cancer (49), and stem cell reprogramming in breast cancer(50). Other parental genes in the two signatures, in addition to this one, have been identified to perform a considerable task in the existence and progression of many cancers to differing degrees, as well as the prognosis of cancer patients. An elevated expression of MKL1 indicates an unfavorable prognosis in papillary thyroid cancer patients and enhances nodal metastases (51). CLASP2 has been linked to the EMT and early progression of bladder cancer (52) and has also been shown to predict bladder cancer prognosis(53). In several malignancies, DNMT3A has been shown to influence cellular apoptosis(54), cisplatin resistance (55), tumor growth, and metastasis(56). NCK1-AS1 enhances the expression of NCK1 to aid carcinogenesis and chemo-resistance in ovarian cancer, according to Chang et al. (57). Elevated expression of IFRD1 indicates a low survival rate in patients suffering from human colon cancers, according to Lewis et al.(58). ABI1 has been linked to the EMT induction in prostate carcinogenesis (59), ovarian cancer metastasis(60), and neuroblastoma development, invasion, and metastasis in several investigations(61). These risk variables were nearly all linked to cancer patient's prognosis, even though there is little connected study on LUAD, which will require more research in the future.

There is no denying that our study still has significant flaws. First and foremost, an independent external validation cohort to confirm the performed DEAS-based predictive risk score signature would be preferable. Secondly, some experimental validation would have been beneficial to further confirm our

data. These are also our future research directions, and we will verify these results in advanced research settings.

In conclusion, our study developed a risk score model on the basis of 11 prognostic DEASs for predicting survival rate in LUAD patients. Notably, this study also shed light on the complexity as well as the diversity of the immune microenvironment in LUAD patients that might explore the lack of therapeutic success in these individuals.

## Conclusion

Our research showed for the first time that AS events were closely related to tumorigenesis and immune microenvironment in LUAD. We constructed two prediction models (OS and PFS) based on survival-related AS events. A total of 11 AS events are included in the models, 5 for OS and 6 for PFS, each of which is an important risk predictor. These findings provide new evidences for the interpretation of tumorigenesis and new ideas for the tumor treatment.

## References

1. Chen W, Zheng R, Baade P, Zhang S, Zeng H, Bray F, et al. Cancer statistics in China, 2015. 2016;66(2):115-32.
2. Torre L, Bray F, Siegel R, Ferlay J, Lortet-Tieulent J, Jemal AJCacjfc. Global cancer statistics, 2012. 2015;65(2):87-108.
3. Riihimäki M, Hemminki A, Fallah M, Thomsen H, Sundquist K, Sundquist J, et al. Metastatic sites survival in lung cancer. 2014;86(1):78–84.
4. Ettinger D, Wood D, Akerley W, Bazhenova L, Borghaei H, Camidge D, et al. Non-Small Cell Lung Cancer Version 6.2015. 2015;13(5):515–24.
5. Nilsen TW, Graveley BR. Expansion of the eukaryotic proteome by alternative splicing. *Nature*. 2010;463(7280):457–63.
6. Baralle FE, Giudice J. Alternative splicing as a regulator of development and tissue identity. *Nature reviews Molecular cell biology*. 2017;18(7):437–51.
7. Ge Y, Porse BT. The functional consequences of intron retention: alternative splicing coupled to NMD as a regulator of gene expression. *BioEssays: news reviews in molecular cellular developmental biology*. 2014;36(3):236–43.
8. Bonnal SC, López-Oreja I, Valcárcel J. Roles and mechanisms of alternative splicing in cancer - implications for care. *Nature reviews Clinical oncology*. 2020;17(8):457–74.
9. Climente-González H, Porta-Pardo E, Godzik A, Eyrales E. The Functional Impact of Alternative Splicing in Cancer. *Cell reports*. 2017;20(9):2215–26.
10. de Necochea-Campion R, Shouse GP, Zhou Q, Mirshahidi S, Chen CS. Aberrant splicing and drug resistance in AML. *J Hematol Oncol*. 2016;9(1):85.

11. Wan L, Yu W, Shen E, Sun W, Liu Y, Kong J, et al. SRSF6-regulated alternative splicing that promotes tumour progression offers a therapy target for colorectal cancer. *Gut*. 2019;68(1):118–29.
12. Sheng J, Zhao Q, Zhao J, Zhang W, Sun Y, Qin P, et al. SRSF1 modulates PTPMT1 alternative splicing to regulate lung cancer cell radioresistance. *EBioMedicine*. 2018;38:113–26.
13. Li Y, Sun N, Lu Z, Sun S, Huang J, Chen Z, et al. Prognostic alternative mRNA splicing signature in non-small cell lung cancer. *Cancer letters*. 2017;393:40–51.
14. Zhao D, Zhang C, Jiang M, Wang Y, Liang Y, Wang L, et al. Survival-associated alternative splicing signatures in non-small cell lung cancer. *Aging*. 2020;12(7):5878–93.
15. Frankiw L, Baltimore D, Li G. Alternative mRNA splicing in cancer immunotherapy. *Nature reviews Immunology*. 2019;19(11):675–87.
16. Li ZX, Zheng ZQ, Wei ZH, Zhang LL, Li F, Lin L, et al. Comprehensive characterization of the alternative splicing landscape in head and neck squamous cell carcinoma reveals novel events associated with tumorigenesis and the immune microenvironment. *Theranostics*. 2019;9(25):7648–65.
17. Yao J, Caballero OL, Huang Y, Lin C, Rimoldi D, Behren A, et al. Altered Expression and Splicing of ESRP1 in Malignant Melanoma Correlates with Epithelial-Mesenchymal Status and Tumor-Associated Immune Cytolytic Activity. *Cancer immunology research*. 2016;4(6):552–61.
18. Ye ZS, Zheng M, Liu QY, Zeng Y, Wei SH, Wang Y, et al. Survival-associated alternative splicing events interact with the immune microenvironment in stomach adenocarcinoma. *World journal of gastroenterology*. 2021;27(21):2871–94.
19. Hanahan D, Weinberg RA. Hallmarks of cancer: the next generation. *Cell*. 2011;144(5):646–74.
20. Bruni D, Angell HK, Galon J. The immune contexture and Immunoscore in cancer prognosis and therapeutic efficacy. *Nature reviews Cancer*. 2020;20(11):662–80.
21. Fridman WH, Zitvogel L, Sautès-Fridman C, Kroemer G. The immune contexture in cancer prognosis and treatment. *Nature reviews Clinical oncology*. 2017;14(12):717–34.
22. Varn FS, Wang Y, Mullins DW, Fiering S, Cheng C. Systematic Pan-Cancer Analysis Reveals Immune Cell Interactions in the Tumor Microenvironment. *Cancer research*. 2017;77(6):1271–82.
23. Öjlert ÅK, Halvorsen AR, Nebdal D, Lund-Iversen M, Solberg S, Brustugun OT, et al. The immune microenvironment in non-small cell lung cancer is predictive of prognosis after surgery. *Molecular oncology*. 2019;13(5):1166–79.
24. Shaul ME, Eyal O, Guglietta S, Aloni P, Zlotnik A, Forkosh E, et al. Circulating neutrophil subsets in advanced lung cancer patients exhibit unique immune signature and relate to prognosis. *FASEB journal: official publication of the Federation of American Societies for Experimental Biology*. 2020;34(3):4204–18.
25. Yoshihara K, Shahmoradgoli M, Martínez E, Vegesna R, Kim H, Torres-Garcia W, et al. Inferring tumour purity and stromal and immune cell admixture from expression data. *Nature communications*. 2013;4:2612.

26. Huang X, Liu J, Mo X, Liu H, Wei C, Huang L, et al. Systematic profiling of alternative splicing events and splicing factors in left- and right-sided colon cancer. *Aging*. 2019;11(19):8270–93.
27. Conway JR, Lex A, Gehlenborg N. UpSetR: an R package for the visualization of intersecting sets and their properties. *Bioinformatics*. 2017;33(18):2938–40.
28. Giulietti M, Piva F, D'Antonio M, D'Onorio De Meo P, Paoletti D, Castrignanò T, et al. SpliceAid-F: a database of human splicing factors and their RNA-binding sites. *Nucleic acids research*. 2013;41(Database issue):D125-31.
29. Gao J, Aksoy BA, Dogrusoz U, Dresdner G, Gross B, Sumer SO, et al. Integrative analysis of complex cancer genomics and clinical profiles using the cBioPortal. *Sci Signal*. 2013;6(269):pl1.
30. Hänzelmann S, Castelo R, Guinney J. GSVA: gene set variation analysis for microarray and RNA-seq data. *BMC Bioinform*. 2013;14:7.
31. Sveen A, Kilpinen S, Ruusulehto A, Lothe RA, Skotheim RI. Aberrant RNA splicing in cancer; expression changes and driver mutations of splicing factor genes. *Oncogene*. 2016;35(19):2413–27.
32. Dvinge H, Kim E, Abdel-Wahab O, Bradley RK. RNA splicing factors as oncoproteins and tumour suppressors. *Nature reviews Cancer*. 2016;16(7):413–30.
33. Bisognin A, Pizzini S, Perilli L, Esposito G, Mocellin S, Nitti D, et al. An integrative framework identifies alternative splicing events in colorectal cancer development. *Molecular oncology*. 2014;8(1):129–41.
34. Shen Q, Nam SW. SF3B4 as an early-stage diagnostic marker and driver of hepatocellular carcinoma. *BMB Rep*. 2018;51(2):57–8.
35. Fehrenbacher L, von Pawel J, Park K, Rittmeyer A, Gandara D, Ponce Aix S, et al. Updated Efficacy Analysis Including Secondary Population Results for OAK: A Randomized Phase III Study of Atezolizumab versus Docetaxel in Patients with Previously Treated Advanced Non-Small Cell Lung Cancer. 2018;13(8):1156–70.
36. Planchard D, Reinmuth N, Orlov S, Fischer J, Sugawara S, Mandziuk S, et al. ARCTIC: durvalumab with or without tremelimumab as third-line or later treatment of metastatic non-small-cell lung cancer. 2020;31(5):609–18.
37. Venhuizen J, Jacobs F, Span P, Zegers, MJSicb. P120 and E-cadherin: Double-edged swords in tumor metastasis. 2020;60:107–20.
38. Deep G, Jain A, Ramteke A, Ting H, Vijendra K, Gangar S, et al. SNAI1 is critical for the aggressiveness of prostate cancer cells with low. E-cadherin. 2014;13:37.
39. Yazdani J, Ghavimi M, Jabbari Hagh E. Ahmadpour FJMd, therapy. The Role of E-Cadherin as a Prognostic Biomarker in Head and Neck Squamous Carcinoma: A Systematic Review and Meta-Analysis. 2018;22(5):523–35.
40. Lechuga S, Amin P, Wolen A, Ivanov AJBebaMcr. Adducins inhibit lung cancer cell migration through mechanisms involving regulation of cell-matrix adhesion and cadherin-11 expression. 2019;1866(3):395–408.

41. Papoff G, Presutti D, Lalli C, Bolasco G, Santini S, Manelfi C, et al. CASP4 gene silencing in epithelial cancer cells leads to impairment of cell migration. cell-matrix adhesion tissue invasion. 2018;8(1):17705.
42. Singh C, Shyanti R, Singh V, Kale R, Mishra J, Singh RJB, et al. Integrin expression and glycosylation patterns regulate cell-matrix adhesion and alter with breast cancer progression. 2018;499(2):374–80.
43. Mutvei A, Nagiec M, Hamann J, Kim S, Vincent C, Blenis JJNc. Rap1-GTPases control mTORC1 activity by coordinating lysosome organization with amino acid availability. 2020;11(1):1416.
44. Zhang L, Chen J, Cheng T, Yang H, Li H, Pan C. Identification of the key genes and characterizations of Tumor Immune Microenvironment in Lung Adenocarcinoma (LUAD) and Lung Squamous Cell Carcinoma (LUSC). *J Cancer*. 2020;11(17):4965–79.
45. Lou Y, Diao L, Cuentas ER, Denning WL, Chen L, Fan YH, et al. Epithelial-Mesenchymal Transition Is Associated with a Distinct Tumor Microenvironment Including Elevation of Inflammatory Signals and Multiple Immune Checkpoints in Lung Adenocarcinoma. *Clinical cancer research: an official journal of the American Association for Cancer Research*. 2016;22(14):3630–42.
46. Jones WD, Michener CM, Biscotti C, Braicu I, Sehouli J, Ganapathi MK, et al. RNA Immune Signatures from Pan-Cancer Analysis Are Prognostic for High-Grade Serous Ovarian Cancer and Other Female Cancers. *Cancers*. 2020;12(3).
47. Cramer-Morales K, Heer C, Mapuskar K. Domann FJFrb, medicine. SOD2 targeted gene editing by CRISPR/Cas9 yields Human cells devoid of MnSOD. 2015;89:379–86.
48. Zhou C, Lyu LH, Miao HK, Bahr T, Zhang QY, Liang T, et al. Redox regulation by SOD2 modulates colorectal cancer tumorigenesis through AMPK-mediated energy metabolism. *Molecular carcinogenesis*. 2020;59(5):545–56.
49. Ashtekar A, Huk D, Magner A, La Perle KMD, Boucai L, Kirschner LS. Alterations in Sod2-Induced Oxidative Stress Affect Endocrine Cancer Progression. *J Clin Endocrinol Metab*. 2018;103(11):4135–45.
50. He C, Danes JM, Hart PC, Zhu Y, Huang Y, de Abreu AL, et al. SOD2 acetylation on lysine 68 promotes stem cell reprogramming in breast cancer. *Proc Natl Acad Sci USA*. 2019;116(47):23534–41.
51. Cheng X, Xu S, Pan J, Zheng J, Wang X, Yu H, et al. MKL1 overexpression predicts poor prognosis in patients with papillary thyroid cancer and promotes nodal metastasis. *Journal of cell science*. 2019;132(16).
52. Zhu B, Qi L, Liu S, Liu W, Ou Z, Chen M, et al. CLASP2 is involved in the EMT and early progression after transurethral resection of the bladder tumor. *BMC Cancer*. 2017;17(1):105.
53. Chen L, Xiong W, Guo W, Su S, Qi L, Zhu B, et al. Significance of CLASP2 expression in prognosis for muscle-invasive bladder cancer patients: A propensity score-based analysis. *Urol Oncol*. 2019;37(10):800–7.
54. Lu W, Lu T, Wei XJOr. Downregulation of DNMT3a expression increases miR-182-induced apoptosis of ovarian cancer through caspase-3 and caspase-9-mediated apoptosis and DNA damage response. 2016;36(6):3597–604.

55. Han X, Liu D, Zhou Y, Wang L, Hou H, Chen H, et al. The negative feedback between miR-143 and DNMT3A regulates cisplatin resistance in ovarian cancer. 2021;45(1):227–37.
56. Cao J, Huang Z, Ou S, Wen F, Yang G, Miao Q, et al. circ0093740 Promotes Tumor Growth and Metastasis by Sponging miR-136/145 and Upregulating DNMT3A in Wilms Tumor. 2021;11:647352.
57. Chang H, Li B, Zhang X, Meng X. NCK1-AS1 promotes NCK1 expression to facilitate tumorigenesis and chemo-resistance in ovarian cancer. Biochem Biophys Res Commun. 2020;522(2):292–9.
58. Lewis MA, Sharabash N, Miao ZF, Lyons LN, Piccirillo J, Kallogjeri D, et al. Increased IFRD1 Expression in Human Colon Cancers Predicts Reduced Patient Survival. Digestive diseases sciences. 2017;62(12):3460–7.
59. Nath D, Li X, Mondragon C, Post D, Chen M, White JR, et al. Abi1 loss drives prostate tumorigenesis through activation of EMT and non-canonical WNT signaling. Cell communication signaling: CCS. 2019;17(1):120.
60. Yu X, Liang C, Zhang Y, Zhang W, Chen H. Inhibitory short peptides targeting EPS8/ABI1/SOS1 tri-complex suppress invasion and metastasis of ovarian cancer cells. BMC Cancer. 2019;19(1):878.
61. Liu X, Peng H, Liao W, Luo A, Cai M, He J, et al. MiR-181a/b induce the growth, invasion, and metastasis of neuroblastoma cells through targeting ABI1. Molecular carcinogenesis. 2018;57(9):1237–50.

## Declarations

Ethics approval and consent to participate: Not applicable

Consent for publication: Not applicable

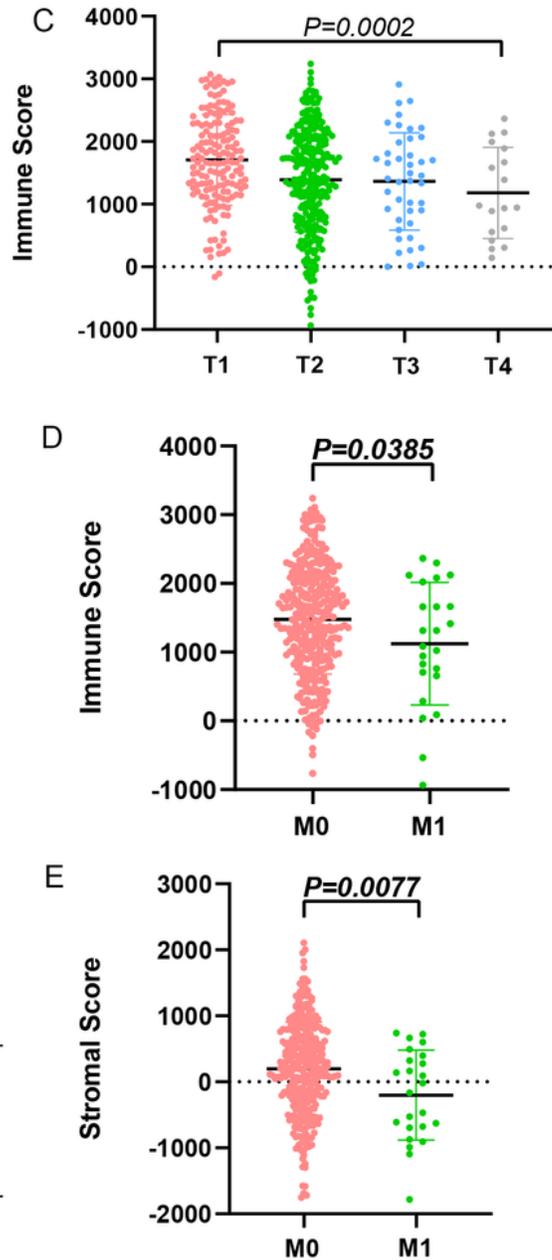
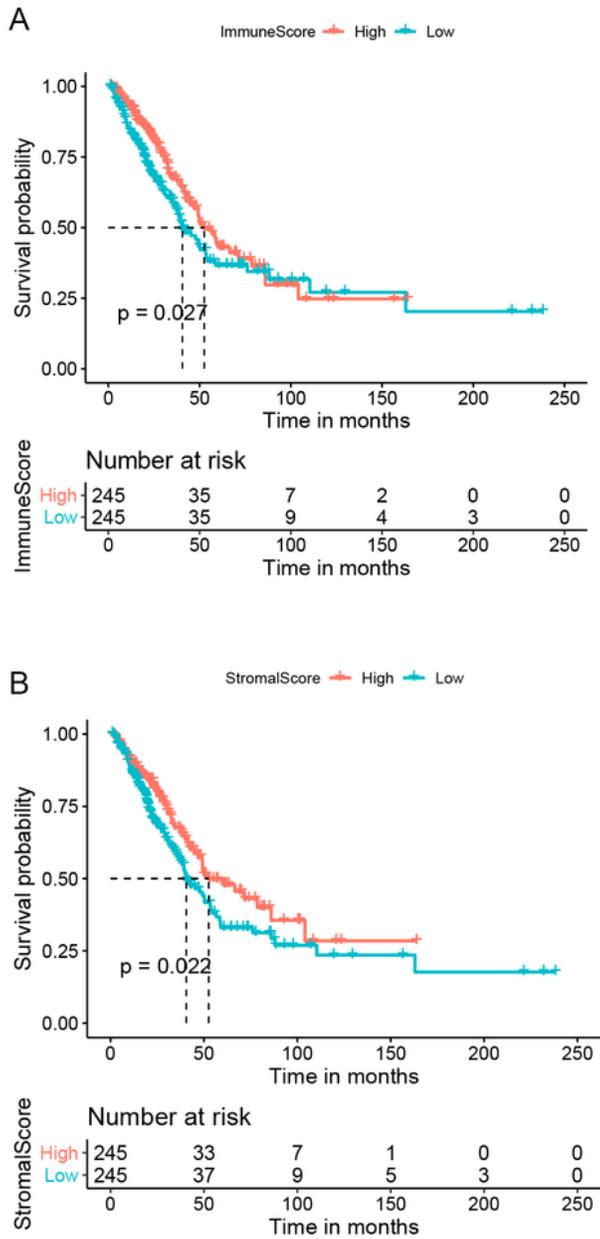
Competing interests: The datasets generated during and/or analysed during the current study are available in the TCGA repository. (<https://www.cancer.gov/>)

Funding: This study was supported by two grants: 1. Beijing Xisike Clinical Oncology Research Foundation (Grant No. Y-BMS2019-038) 2. WU JIEPING Medical Foundation (Grant No. 320.6750.2020-12-19)

Authors' contributions: Gongjun Wang analyzed the data and wrote the manuscript. Wensheng Qiu provided financial support. All authors contributed to the article and approved the submitted version.

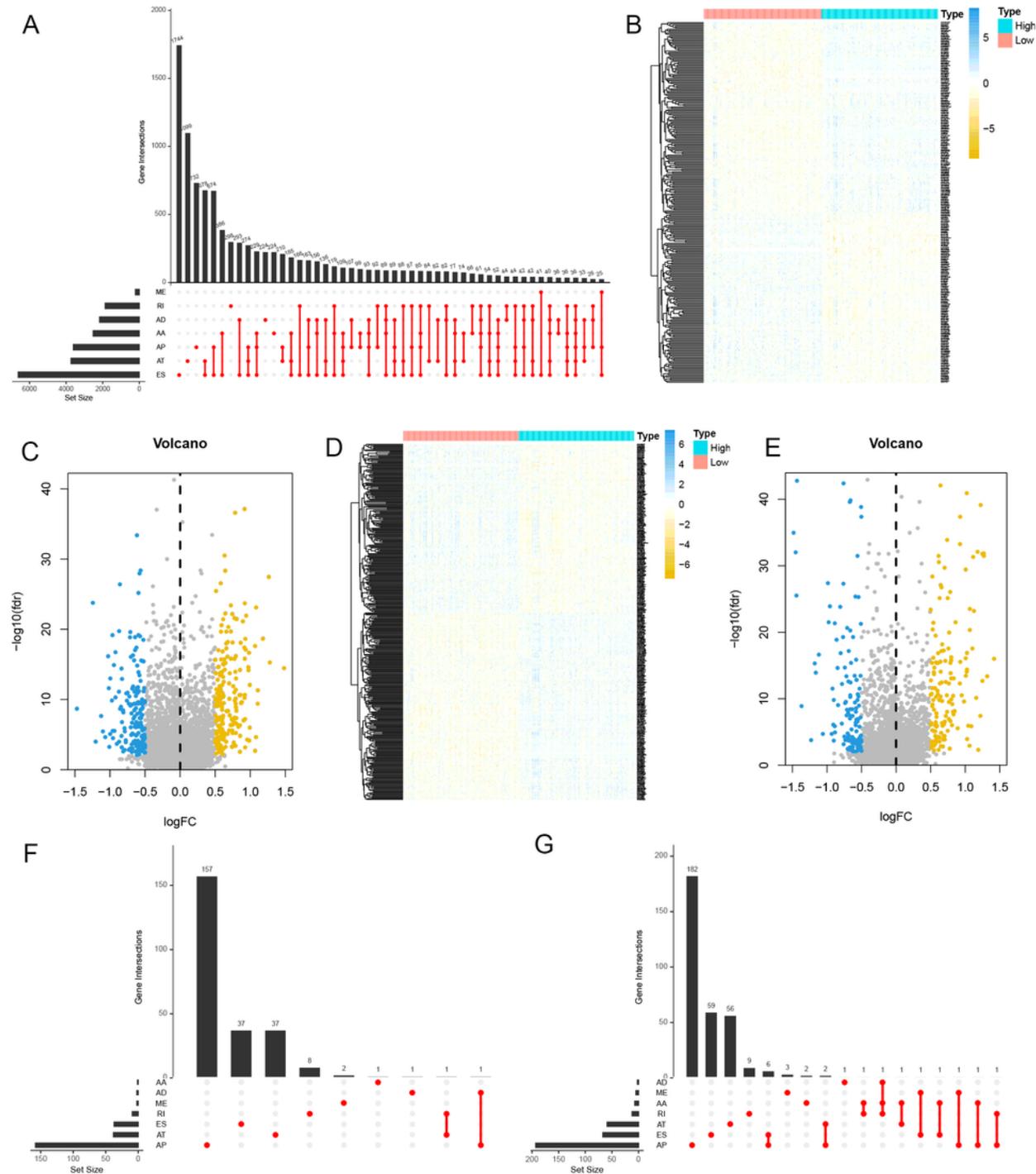
Acknowledgements: The authors gratefully acknowledge the data provided by patients and researchers participating in TCGA.

## Figures



**Figure 1**

The relationship between stromal/immune scores and prognosis of LUAD patients. (A) The K-M survival curves for OS of high and low immune scores groups. (B) The K-M survival curves for OS of high and low stromal scores groups. (C) Comparison of immune scores between T stages. (D) Comparison of immune scores between M stages. (E) Comparison of stromal scores between M stages.



**Figure 2**

DEAS events between high and low stromal/immune scores groups. (A) The UpSet plot of intersections between AS events and the corresponding gene intersections. The heatmaps (B, D) and volcano plots (C, E) of DEAS events between the high and low stromal/immune scores groups. (F, G) The UpSet plots of DEAS events and the corresponding gene intersections of the immune group and the stromal group.

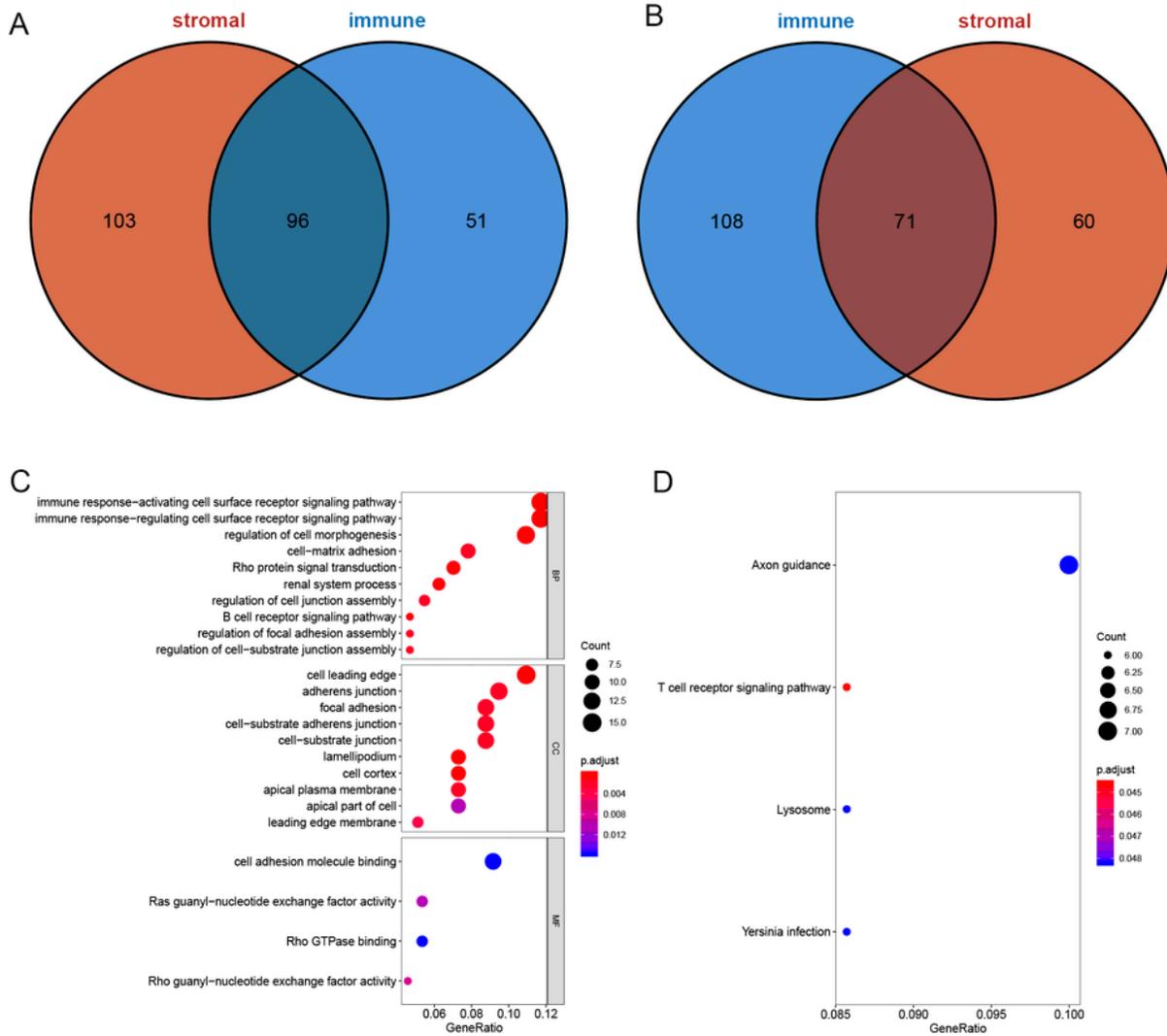


Figure 3

Intersection DEAS events and GO and KEGG pathway enrichment analysis. The up-regulated (A) or down-regulated (B) DEAS events in both stromal score and immune scores groups by Venn diagram. (C) GO analysis. (D) KEGG pathway enrichment analysis.

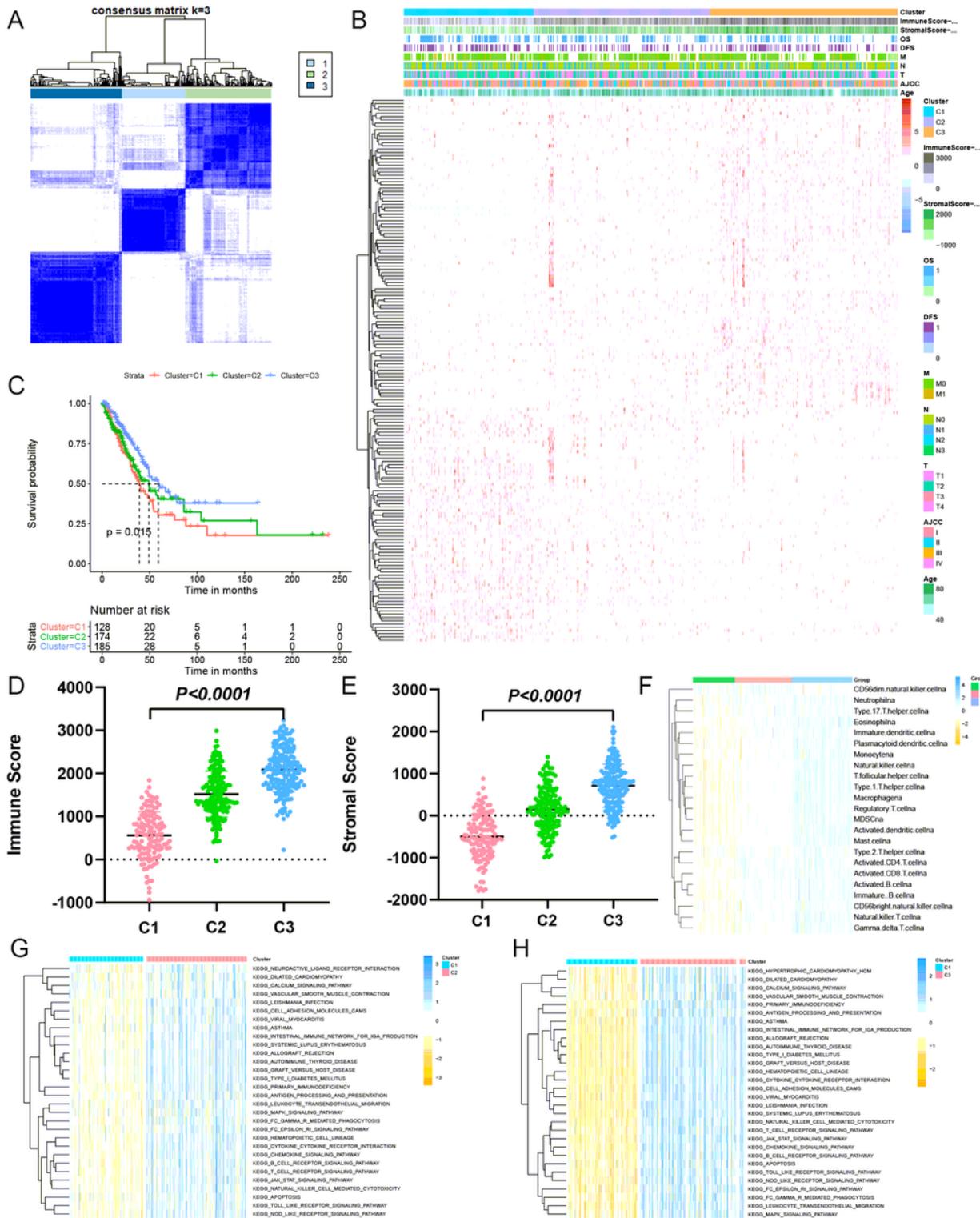


Figure 4

The immune microenvironment was closely related to the prognosis of LUAD patients. (A) TCGA LUAD cohort was clustered into three subgroups by unsupervised cluster analysis. (B) The heatmap of clusters

and clinical parameters. (C) K-M survival curves for OS of three clusters. (D, E) The comparison of immune scores and stromal scores between three clusters. (F) The comparison of immune cell infiltration between three clusters. (G, H) GSVA enrichment analysis showing the activation states of biological pathways in distinct clusters (C1-C2, C1-C3).

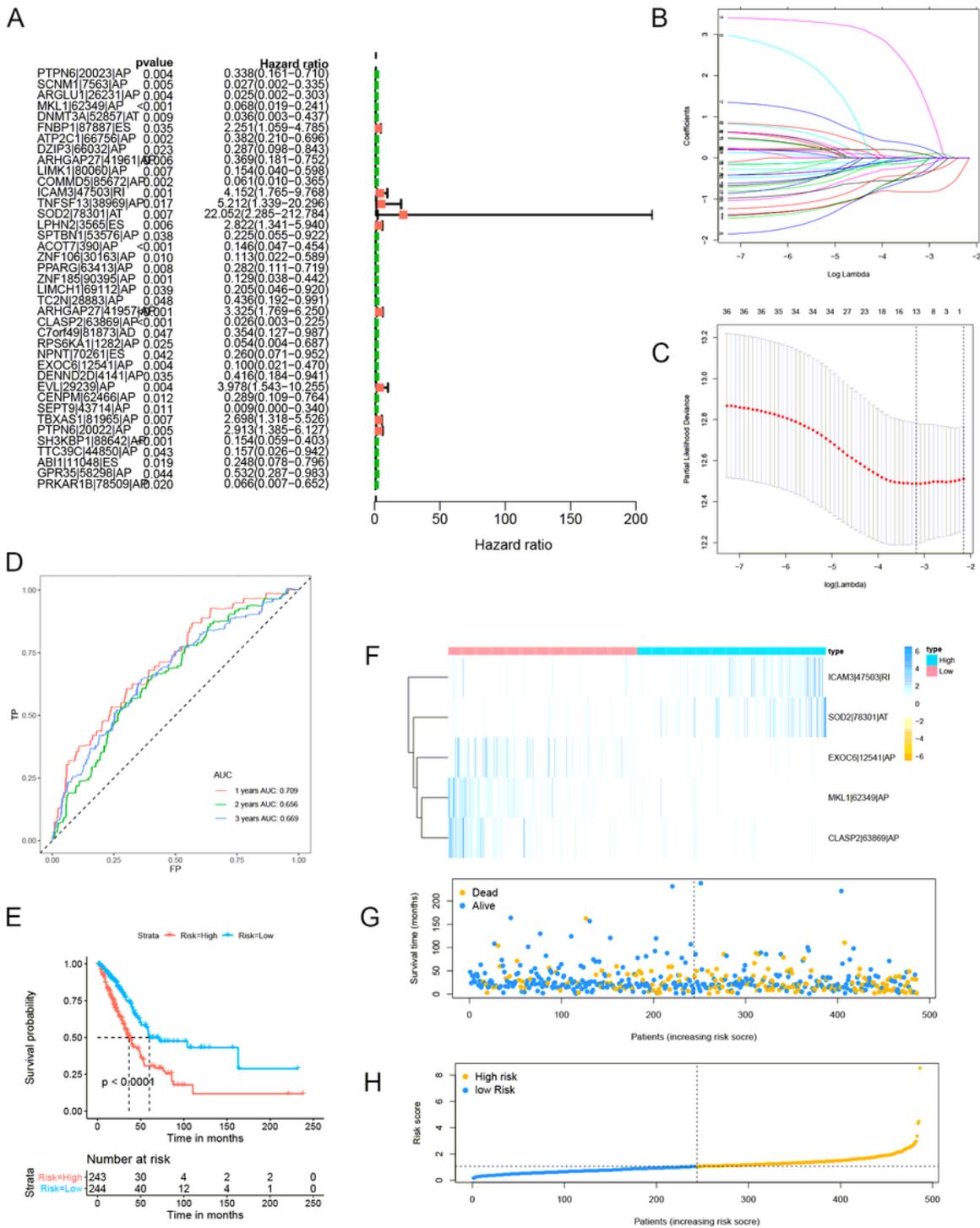
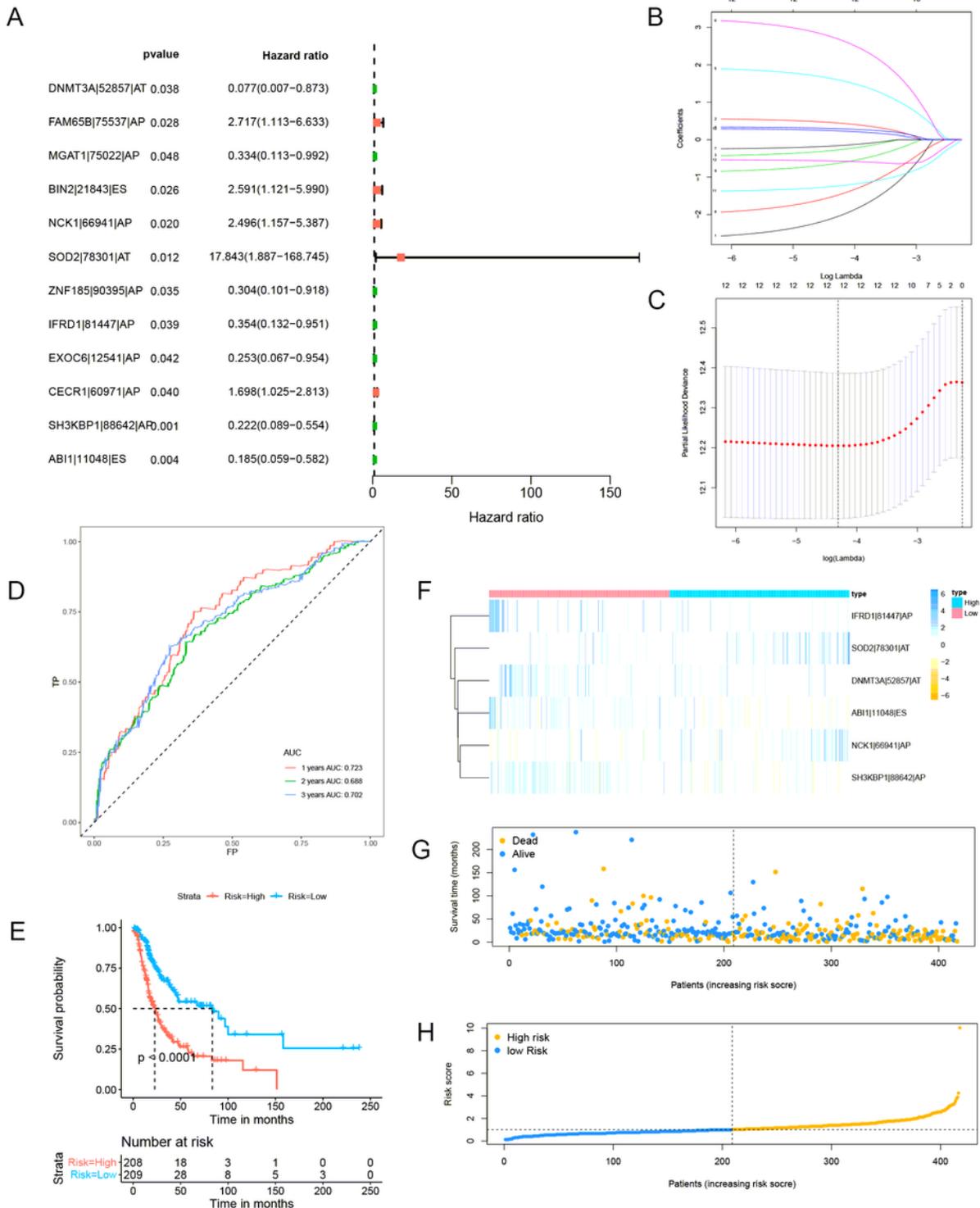


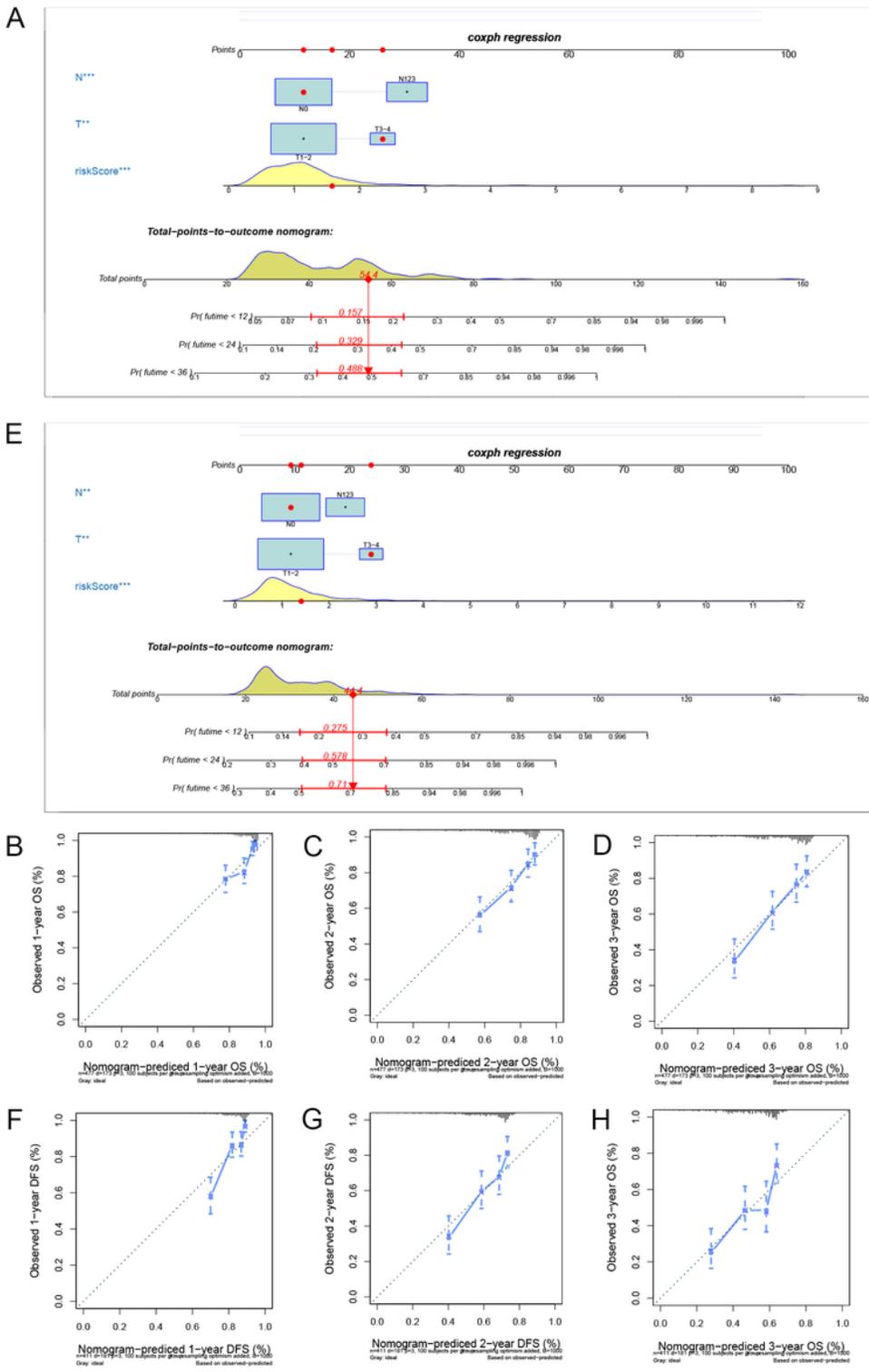
Figure 5

Establishment of a 5-DEAS-based OS signature. (A) The forest plot of the univariate Cox analyses. (B) LASSO coefficient profiles of the candidate survival-related DEASs. A coefficient profile plot was produced against the  $\log \lambda$  sequence related to OS. (C) Dotted vertical lines were drawn at the optimal values using the minimum criteria related to OS. (D) ROC curve analysis to evaluate the sensitivity and specificity of the gene signature. (E) K-M survival curve to test the predictive effect of the gene signature. (F) DEAS expression levels of the high-risk and low-risk groups with the OS signatures. (G) OS scatter plots for LUAD patients. (H) Risk score distribution of patients with the OS signatures.



## Figure 6

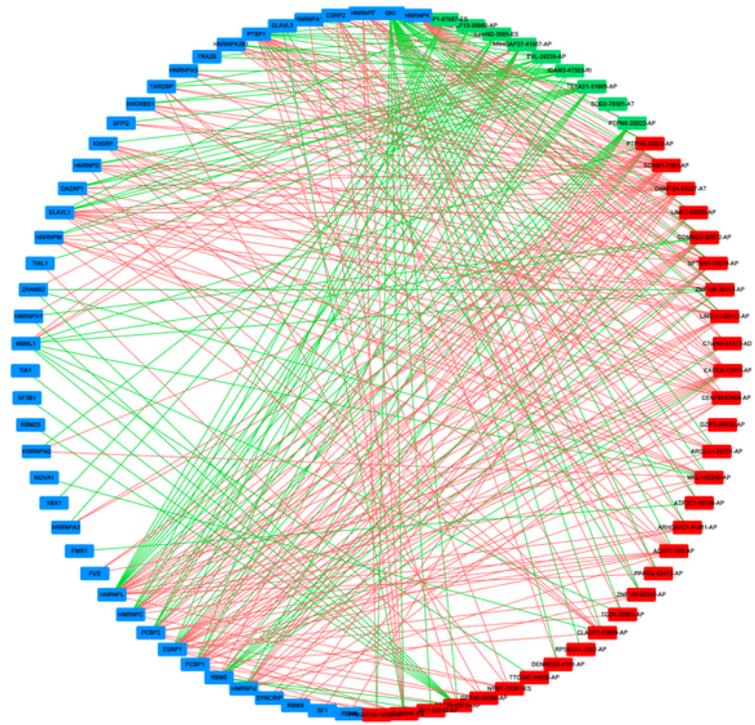
Establishment of a 6-DEAS-based PFS signature. (A) The forest plot of the univariate Cox analyses. (B) LASSO coefficient profiles of the candidate survival-related DEASs. A coefficient profile plot was produced against the  $\log \lambda$  sequence related to PFS. (C) Dotted vertical lines were drawn at the optimal values using the minimum criteria related to PFS. (D) ROC curve analysis to evaluate the sensitivity and specificity of the gene signature. (E) K-M survival curve to test the predictive effect of the gene signature. (F) DEAS expression levels of the high-risk and low-risk groups with the PFS signatures. (G) PFS scatter plots for LUAD patients. (H) Risk score distribution of patients with the PFS signatures.



**Figure 7**

Development of two nomograms combining DEAS-based signature and independent prognostic clinical variables to predict OS and PFS in LUAD patients. (A) Nomogram of OS combining the OS signature and three clinical variables of LUAD patients. (B-D) Calibration curves of the nomogram at 1, 2, and 3 years. (E) Nomogram of PFS combining the PFS signature and two clinical variables of LUAD patients. (F-H) Calibration curves of the nomogram at 1, 2, and 3 years.

A



B

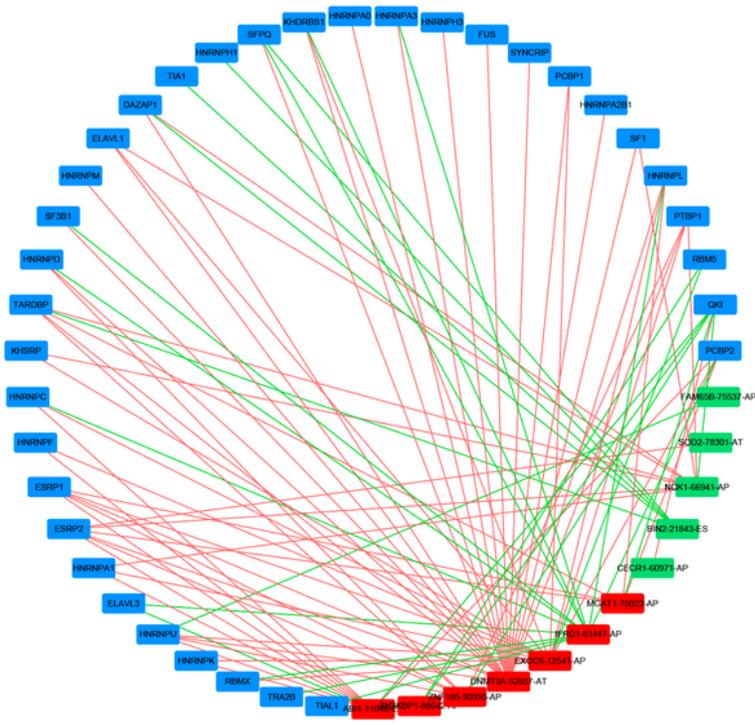
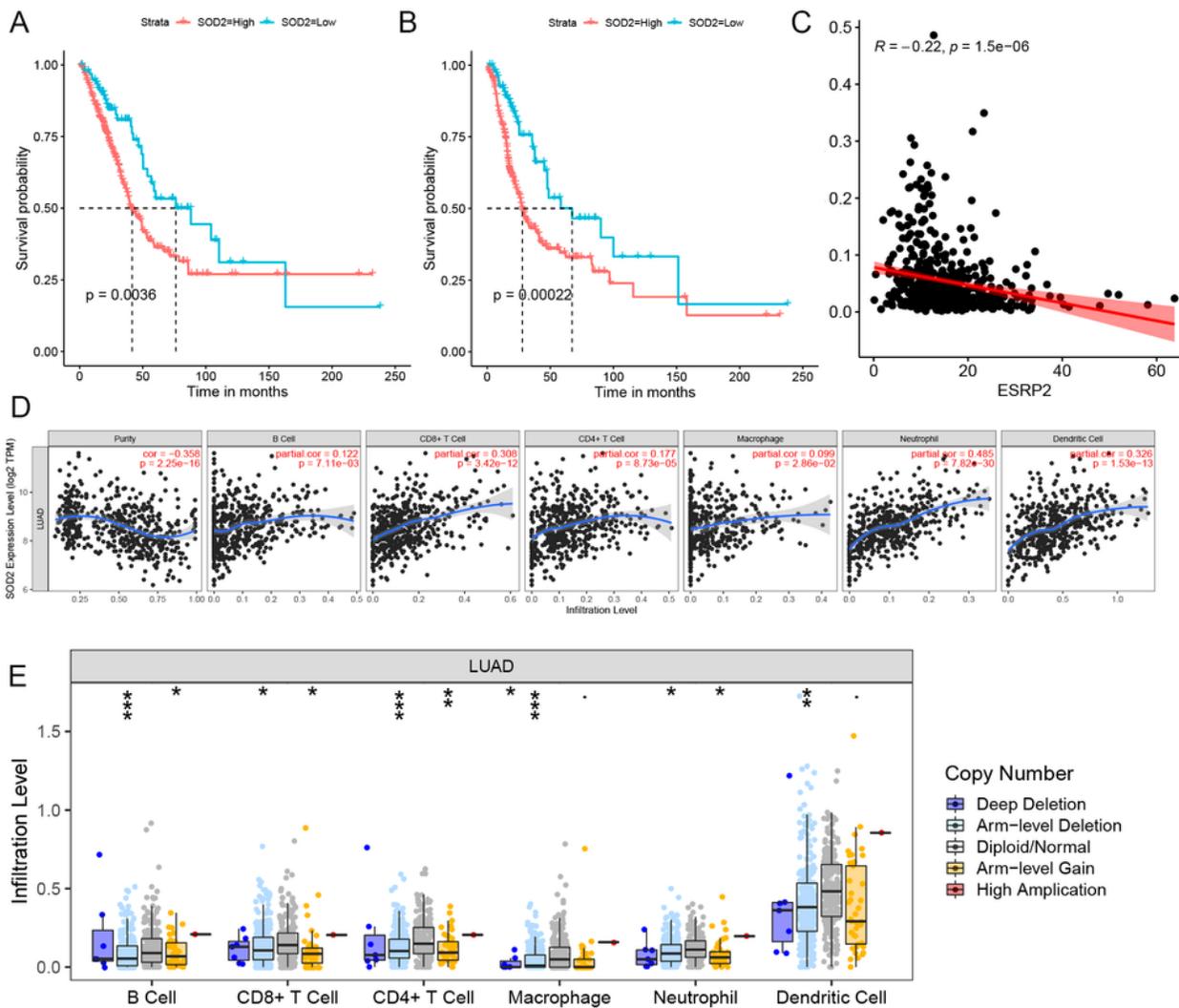


Figure 8

Regulatory network between SFs and AS events in LUAD. (A) OS (B) PFS



**Figure 9**

Analysis of the crucial SOD2-78301-AT DEAS event. The K-M survival curves for OS (A) and PFS (B) of high and low SOD2-78301-AT expression level groups. (C) The relationship between SOD2|78301|AT and ESRP2. (D) The correlation between SOD2 expression in LUAD and abundance of immune infiltrates. (E) The correlation between somatic copy number alterations (SCAN) and abundance of immune infiltrates of SOD2.

## Supplementary Files

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

- [SupplementFIG1.tif](#)
- [SupplementTable1.xlsx](#)
- [SupplementTable2.xlsx](#)
- [SupplementTable3.docx](#)
- [SupplementTable4.docx](#)