



Research article

Novel prognostic signature for lung adenocarcinoma based on immune-related mRNA pairs

Jiawei Yue^{a,1}, Hui Guo^{b,1}, Jinhong Ma^b, Weifeng Shi^{b,**}, Yumin Wu^{b,c,*}^a Department of Orthopaedics, The Third Affiliated Hospital of Soochow University, Changzhou, Jiangsu, 213003, China^b Department of Laboratory Medicine, The Third Affiliated Hospital of Soochow University, Changzhou, Jiangsu, 213003, China^c Jiangsu Key Laboratory for Carbon-Based Functional Materials and Devices Institute of Nano and Soft Materials (FUNSOM) College of Nano Science & Technology (CNST) Suzhou, Jiangsu, 215123, China

ARTICLE INFO

Keywords:

Immune-related mRNA
Lung adenocarcinoma
Survival prognosis
Risk model

ABSTRACT

Lung adenocarcinoma (LUAD) is a highly lethal malignant tumor. While the involvement of multiple mRNAs in the progression of LUAD is well established, the potential diagnostic value of immune-related mRNAs (irmRNAs) in LUAD remains largely unexplored. In this study, we utilized RNA-seq, clinical data, and immune-related gene information from LUAD patients to identify differentially expressed immune-related mRNAs (DEirmRNAs) and developed a predictive risk model based on specific DEirmRNA pairs closely linked with patient prognosis. We classified patients into high-risk and low-risk groups and analyzed factors such as survival rate, clinical characteristics, gene enrichment, immune cell infiltration, tumor mutation load, and drug susceptibility. We confirmed the expression levels of these DEirmRNAs in tumor tissues using qRT-PCR assay. Our results showed that the low-risk group had a longer survival time and lower tumor mutation burden (TMB) and microsatellite instability (MSI) compared to the high-risk group. The high-risk group also had a significant reduction in the number of certain immune cells and a lower half-maximum inhibitor concentration (IC50). We identified specific DEirmRNA pairs that were up-regulated or down-regulated in tumor tissues compared to adjacent tissues. Our prognostic risk model based on DEirmRNA pairs could be used to predict the prognosis of LUAD patients and provide reference for better treatment.

1. Introduction

Lung cancer is a devastating disease and remains one of the most common and deadly forms of cancer worldwide [1,2]. According to recent statistics, in 2021 alone, there were 2.35 million new cases of lung cancer and 1.8 million deaths attributed to this disease - accounting for nearly a quarter (22.5 %) of all cancer-related deaths [3,4]. Lung cancer is mainly related to smoking and usually be divided into small cell lung cancer (SCLC) and non-small cell lung cancer (NSCLC) [5]. NSCLC is the more common of the two, accounting for approximately 85 % of all cases, with lung adenocarcinoma (LUAD) being the most prevalent subtype [6,7]. Previous

* Corresponding author. Department of Laboratory Medicine, The Third Affiliated Hospital of Soochow University, Changzhou, Jiangsu, 213003, China.

** Corresponding author.

E-mail addresses: swf67113@163.com (W. Shi), wuyumin1116@yeah.net (Y. Wu).

¹ These authors have contributed equally to this work.

<https://doi.org/10.1016/j.heliyon.2024.e24397>

Received 11 May 2023; Received in revised form 16 December 2023; Accepted 8 January 2024

Available online 20 January 2024

2405-8440/© 2024 The Authors. Published by Elsevier Ltd. This is an open access article under the CC BY-NC-ND license (<http://creativecommons.org/licenses/by-nc-nd/4.0/>).

reports indicated that the incidence of lung cancer, especially LUAD, in non-smoking women is increasing year by year [8–10]. While radiotherapy, chemotherapy, and immunotherapy have significantly improved the prognosis and quality of life for LUAD patients, the survival rate for this type of cancer remains relatively low [11]. Moreover, since patients respond inconsistently to these treatments, as well as the varying side effects, it is necessary to provide individualized treatment for LUAD patients, which can contribute to prolonging survival and improving outcomes [12]. Therefore, it is of great significance to develop a reliable risk model for an accurate prognosis of the disease. Such a model would enable doctors to choose more personalized treatments for each patient based on their specific condition.

Typically, the primary tumor, regional lymph nodes, and distant metastasis are used to assess the stage of the tumor and provide individualized treatment to patients [13]. However, due to the molecular heterogeneity of the disease, similar clinical cases may have different outcomes. Hence, it is imperative to search for a suitable and new diagnostic marker that can effectively evaluate the outcome of LUAD patients [14,15]. Previous genetic techniques have demonstrated that certain differentially expressed mRNAs could serve as promising biomarkers for predicting the outcomes and survival of patients [16]. Additionally, the study found that low expression of four specific mRNAs (CCR2, CCR4, P2RY12, and P2RY13) directly affected immune cell infiltration in the tumor microenvironment, indicating that these mRNAs could be useful prognostic markers in patients with pancreatic cancer [17]. It is important to note that this study was conducted on patients with pancreatic cancer, and the findings may not be directly applicable to patients with other types of cancer or other diseases. Further research will be needed to validate these findings and determine the potential clinical usefulness of these mRNAs as prognostic markers. Immune-related mRNA (irmRNA) possess advantages over other mRNAs due to their crucial role in immune responses, specific expression in immune cells and tissues, differential expression patterns in response to immune stimuli or diseases, and potential as therapeutic targets. These features make immune-related mRNAs a valuable focus of research in immunology and provide opportunities for developing new diagnostic and therapeutic approaches for immune-related disorders [18–20]. The study conducted by Zhai et al. suggest that differentially expressed irmRNAs (DEirmRNAs) could be potentially useful biomarkers for predicting prognosis and monitoring disease progression in patients with LUAD [21]. However, further research is needed to validate these findings and to develop more accurate prognostic models for LUAD. Additionally, it will be important to understand the biological functions of these DEirmRNAs in order to identify potential targets for new therapeutic strategies.

Traditional diagnosis of cancer is often based on single gene diagnosis or quantification of single or multiple mRNAs. Therefore, our study aims to provide personalized treatment by developing a novel algorithm that retrieves, identifies, and screens DEirmRNA pairs. These pairs can be used to evaluate survival rates, gene enrichment, clinical characteristics, anti-tumor drug resistance, and immune cell infiltration in LUAD patients. By doing so, we hope to improve patient outcomes and prognosis.

2. Methods

2.1. Transcriptome database retrieval and differential expression analysis

We obtained RNA-seq data and patient information on LUAD from the UCSC-XENA database (<https://xenabrowser.net/datapages/>). The GTF files and mRNA expression profiles were extracted from RNA-seq data using the Ensembl database (<http://asia.ensembl.org>). Immune-related genes were collected from the Imm Port database (<http://www.immport.org>). We identified irmRNA by examining the co-expression correlation between immune-related genes and all mRNAs. IrmRNAs that met certain criteria (Pearson immune gene correlation coefficients of >0.4 with p value < 0.001) were considered for further analysis. The LIMMA package was used to identify differentially expressed irmRNAs based on the following criteria: $|\logFC| > 1$ and false discovery rate (FDR) < 0.05 . The DEirmRNAs that met the above-mentioned criteria were classified as differentially expressed immune-related mRNA.

2.2. Screening DEirmRNA pairs

0 or 1 matrix was constructed through single cyclic pairing, and an effective DEirmRNA pair was obtained by the iterative loop. An X was constructed to replace pairs of mRNA A and mRNA B. When mRNA A expression value was greater than that of mRNA B, X was defined as 1; otherwise, X was defined as 0. The vertical axis of the matrix represents each sample, and the horizontal axis represents each DEirmRNA pairs. For any mRNA pairs with the expression quantity of 0 or 1, the relationship between prognosis and mRNA pairs was not considered, since the mRNA pairs without a certain rank failed to accurately predict the outcome of LUAD patients. When the mRNA pairs with 0 or 1 expression accounted for between 20 % and 80 % of total pairs, this was considered an effective match.

2.3. Building the prognostic risk model

We used modified lasso regression and univariate analyses to filtrate DEirmRNA pairs associated with prognosis. Then, we performed a 10-fold cross-validation with 1000 repetitions and 1000 random stimuli per cycle by lasso regression analysis. When the frequency of each DEirmRNA pair in 1000 cycles was more than 100, the DEirmRNA pair was selected to create the prognostic risk model. As described in the literature [22] the area under the curve was used to find the critical point, after which LUAD patients were reclassified as high- or low-risk groups. Univariate and multivariate regression analyses were used to assess the independent prognostic efficiency of the prognostic risk model.

2.4. Validation of the prognostic risk model

Kaplan-Meier analysis was used to assess the survival situations between groups and drew the survival curve to verify the critical point we established. We used the Chi-square test to assess the relationship between the risk model and pathological features of clinical cases, thus providing the clinical significance of this model. We then used Wilcoxon rank test to calculate the difference between clinicopathological features and risk score. Univariate and multivariate Cox regression analyses with risk scores and clinicopathological features were performed to determine whether this risk model could be used as an individual diagnostic factor to evaluate the status of LUAD patients. The tags are described as the following: $P < 0.001 = ***$, $P < 0.01 = **$, $P < 0.05 = *$, and, $P > 0.05 = ns$.

2.5. Gene set enrichment analysis (GSEA)

We utilized the R language package “clusterprofiler” for analysis, which was divided into GSEA-GO, GSEA-KEGG, and GSEA-Reactome gene pathway enrichment analysis.

2.6. Tumor mutation load

The calculation of tumor mutation load was performed by determining the tumor mutation burden (TMB) using TCGA mutation data (TCGA. LUAD. varscan). Additionally, MSI data were obtained from the study conducted by Bonneville et al. [23].

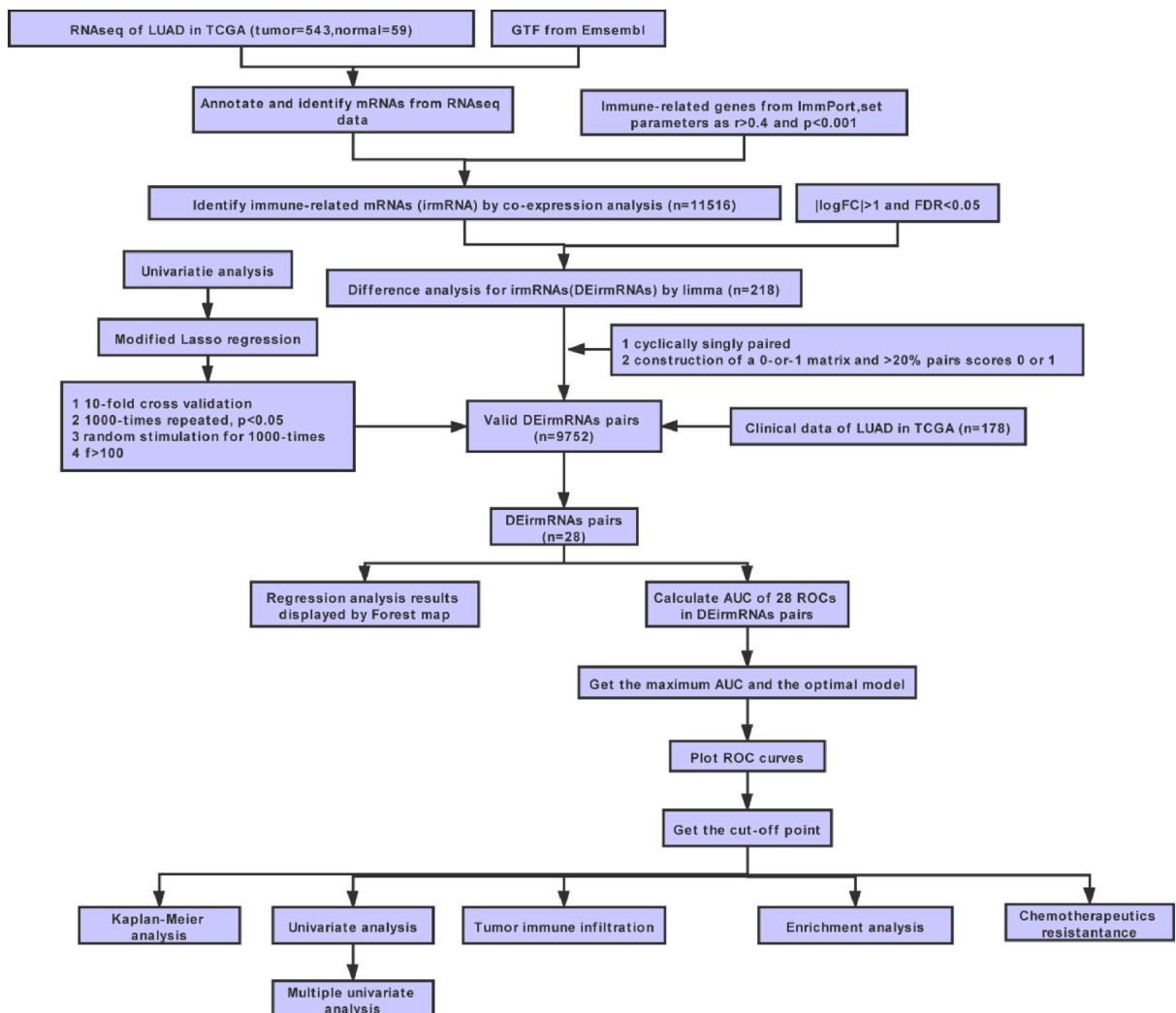


Fig. 1. Work flow of our study.

2.7. Analysis of anti-tumor immune cells

We downloaded immune cell infiltration data of LUAD patients from the ImmuCellAI database. To evaluate the differential expression of immune cells among the patients, we employed the Wilcoxon symbolic rank test.

2.8. Drug resistance to antineoplastic drugs

We utilized an R package called “prpoptc” to calculate the half-maximum inhibitor concentration (IC50) of lung cancer drugs. To evaluate the differences in IC50 between the patients, we employed the Wilcoxon symbolic rank test.

2.9. Tissue samples

All tissue samples were collected at the Third Affiliated Hospital of Soochow University (Changzhou, China) in accordance with the study protocol. Informed consent was obtained from all patients. The study was conducted in strict compliance with the ethical guidelines of the Third Affiliated Hospital of Soochow University. We collected both tumor and adjacent non-tumor tissues from surgically resected primary LUAD samples (n = 20 tumors, 20 normal tissues).

2.10. RNA extraction and PCR

Total RNA was isolated from adjacent and tumor tissues of LUAD patients by Trizol (Takara, Japan), followed by simultaneous

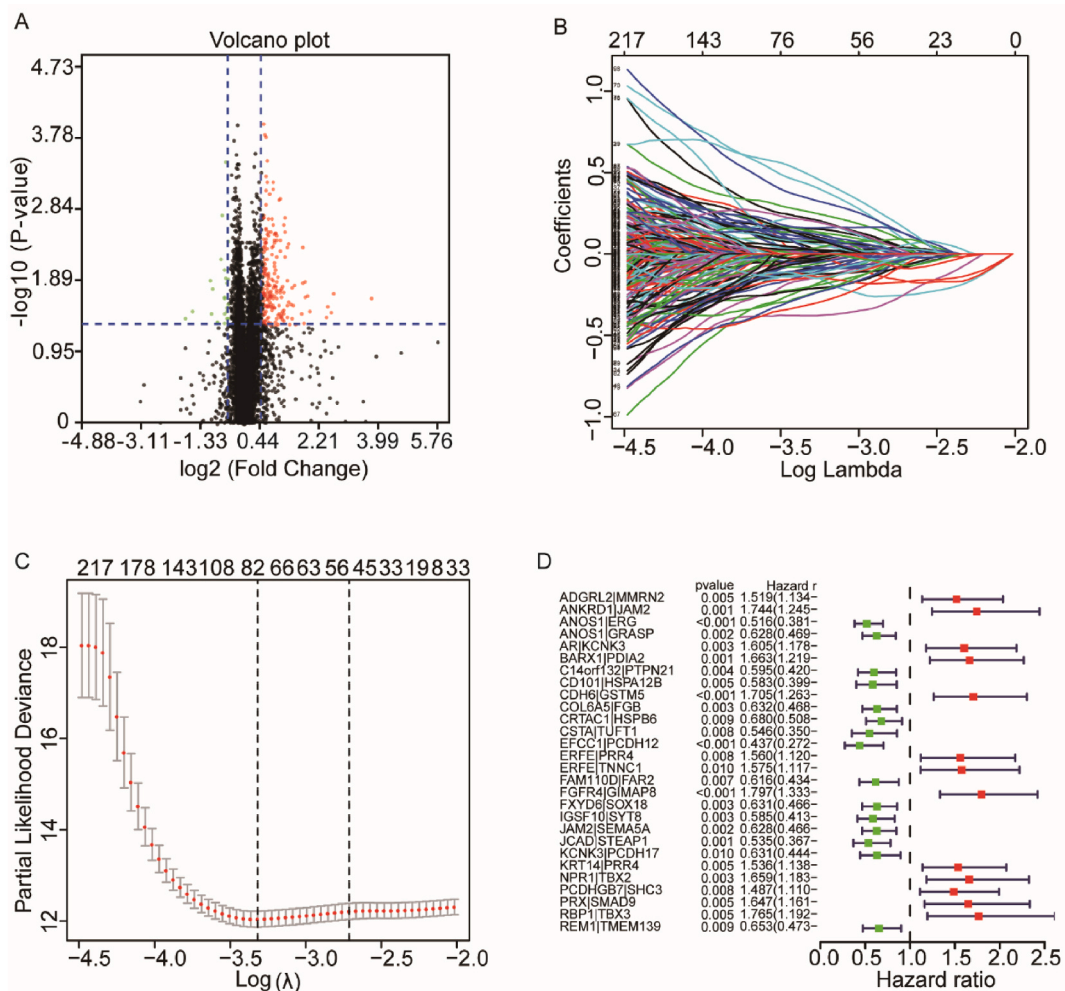


Fig. 2. Construction of a prognostic risk model by using DEirmRNA pairs. (A) The volcano plot of the differentially expressed immune-related mRNAs (DEirmRNAs). (B) The distribution of the lasso coefficients for 28 DEirmRNA pairs. (C) The partial likelihood deviation of the LASSO coefficient distribution. (D) The Forest plot shows the univariate regression analysis of 28 DEirmRNA pairs.

generation of cDNA using the single cell sequence specific amplification kit (Takara, Japan) according to the manufacturer’s instructions. qRT-PCR was performed using SYBR reagents (Bio-Rad, USA) with the CFX96 Touch Real-Time PCR Detection System (Bio-Rad, USA) following the manufacturer’s instructions. β -actin was used as the internal reference gene and relative gene expression levels were calculated by the $2^{-\Delta\Delta Ct}$ method. The sequence of primers used in this study are listed in [Supplementary Table 1](#).

2.11. Statistical analysis

The results obtained from clinical samples were analyzed using Prism software version 7.0 (GraphPad Software Inc, USA). The results are presented as means with standard errors. A two-group comparison was done by the two-tailed student’s t-test. A p value of less than 0.05 were considered statistically significant.

3. Results

3.1. Screening of DEirRNAs

As shown in [Fig. 1](#), the transcriptome data of LUAD from the UCSC-XENA database was explored, including 543 patients and 59

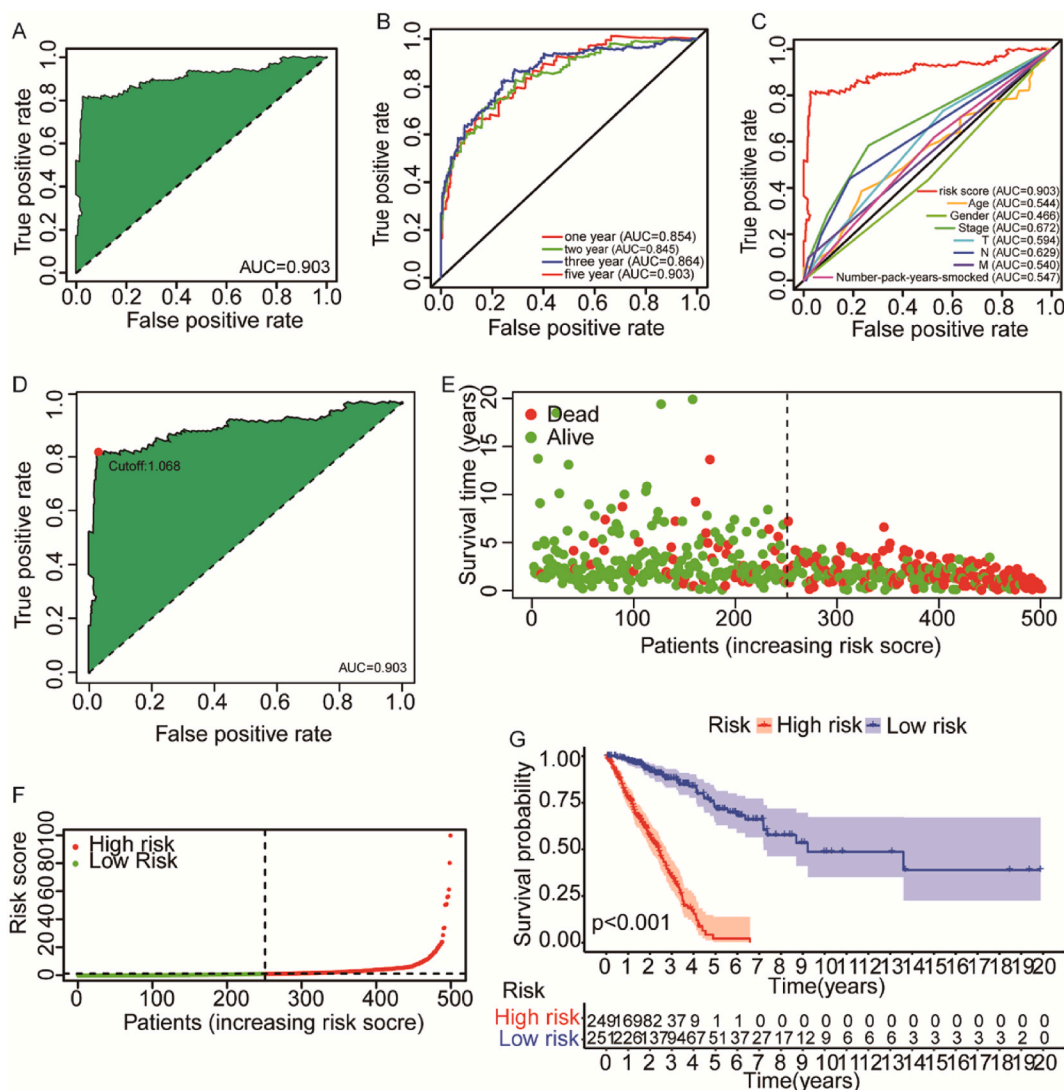


Fig. 3. Evaluate the prognostic risk model. (A) The ROC of the prognostic risk model with the best cut-off point. (B) The 1-, 2-, 3-, and 5-year ROC curves of the prognostic risk model showed that all AUCs were over 0.845. (C) Chi-square test analysis of the relationship between 5-year ROC curves and clinical characteristics. (D) The Akaike information criterion (AIC) value was used to verify the critical point. The distribution of the survival status (E) and risk scores (F) in the high-risk and low-risk group. (G) Kaplan-Meier analysis of LUAD patients.

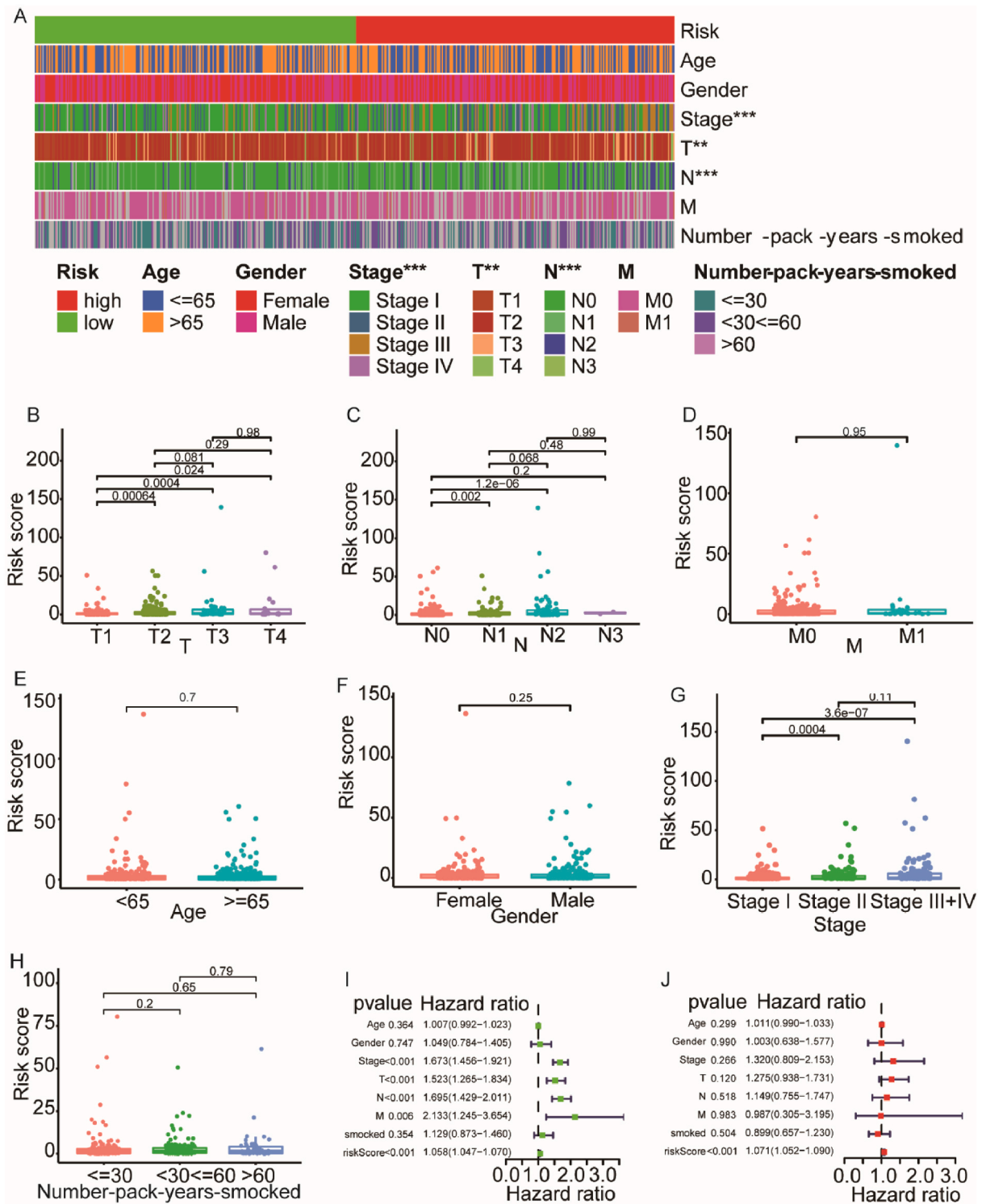


Fig. 4. Patients' evaluation by the prognostic risk model. (A) The strip chart showed the relationship between clinical stage (B), tumor stage (C-E), smoke status (F), age (G), gender (H) and risk scores. (I) Univariate Cox regression analysis showed that risk score and tumor stage were significantly correlated. (J) Multivariate regression analysis showed that only the risk score was regarded as an individual diagnostic factor.

controls. Then, these data were annotated by using a GTF file. As a result, up to 11,516 irmRNAs were identified by co-expression analysis (Supplementary Table 2), of which 218 were identified as DEirmRNAs, among which 49 were decreased, and 169 were increased, respectively (Fig. 2A–C).

3.2. Creating and validating the prognostic risk model

A total number of 9752 valid DEirmRNA pairs were identified by 0 or 1 matrix screening among 218 DEirmRNAs using iterative cycles, and 28 DEirmRNA pairs were considered for the risk assessment model by univariate regression analysis and improved Lasso regression (Fig. 2D). Then, the AUC under the receiver operating characteristic (ROC) curve of the DEirmRNA pair was calculated, and the curve was drawn. The optimal DEirmRNA pair was determined by the highest point of the curve (0.903) (Fig. 3A). Next, the 1-, 2-, 3-, and 5-year ROC curves were created to verify the optimality. We found that the AUC values of these ROC curves were greater than 0.845 (Fig. 3B). Meanwhile, other clinical attributes, such as age, gender, and tumor stage, were compared with the 5-year ROC curves (Fig. 3C). Additionally, the Akaike information criterion (AIC) value was used to verify the critical point of 5-year ROC curve (Fig. 3D). Moreover, we extracted data of 491 patients with LUAD from the UCSC-XENA database, after which their risk scores were created and evaluated. According to the critical point determined by our algorithm model, these patients were reclassified into two groups: the high-risk group included 247 cases, where as the low-risk group included 244 cases. Fig. 3E and F shows the survival situation and risk scores of patients, and the final outcome of those with low-risk scores was better than that of those with high-risk scores. Moreover, they had longer survival (Fig. 3G).

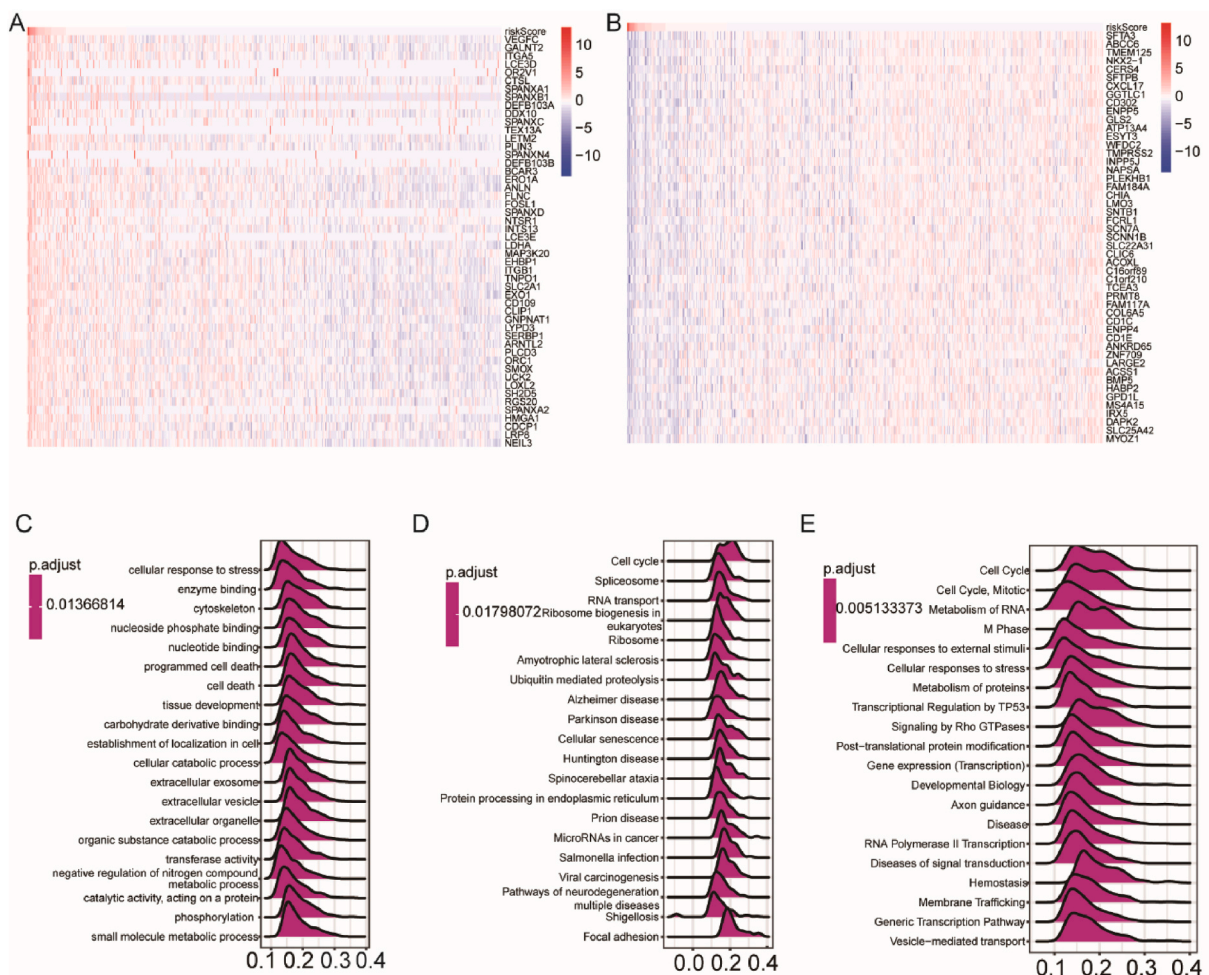


Fig. 5. Gene function enrichment analysis. (A) In risk score and gene correlation analysis, the heat map shows the expression of the top 50 genes positively correlated with the risk groups. (B) In risk score and gene correlation analysis, the heat map shows the expression of the top 50 genes negatively correlated with the risk groups. (C) GSEA-GO functional enrichment analysis. (D) GSEA-KEGG functional enrichment analysis. (E) GSEA-Reactome functional enrichment analysis.

3.3. Relationship between clinical characteristics and risk score

Multiple Chi-square tests were performed to assess the relationship between risk scores and clinical characteristics (Fig. 4A). As shown in Fig. 4B–H, the relationship of risk score with tumor stage, age, gender, clinical stage, and smoke status were obtained by rank-sum test. Moreover, univariate Cox regression analysis showed that the clinical stage ($P < 0.001$, HR = 1.673, 95 % CI [1.456–1.921]), T ($P < 0.001$, HR = 1.523, 95 % CI [1.265–1.834]), N ($P < 0.001$, HR = 1.695, 95 % CI [1.429–2.011]), M ($P < 0.01$, HR = 2.133, 95 % CI [1.245–3.654]), and risk scores ($P < 0.001$, HR = 1.058, 95 % CI [1.047–1.070]) were correlated with survival time (Fig. 4I). However, only the risk score ($P < 0.001$, HR = 1.071, 95 % CI [1.052–1.090]) demonstrated independent correlation with survival time, as evidenced by multivariate Cox regression analysis (Fig. 4J). Collectively, based on the outcomes of our comprehensive

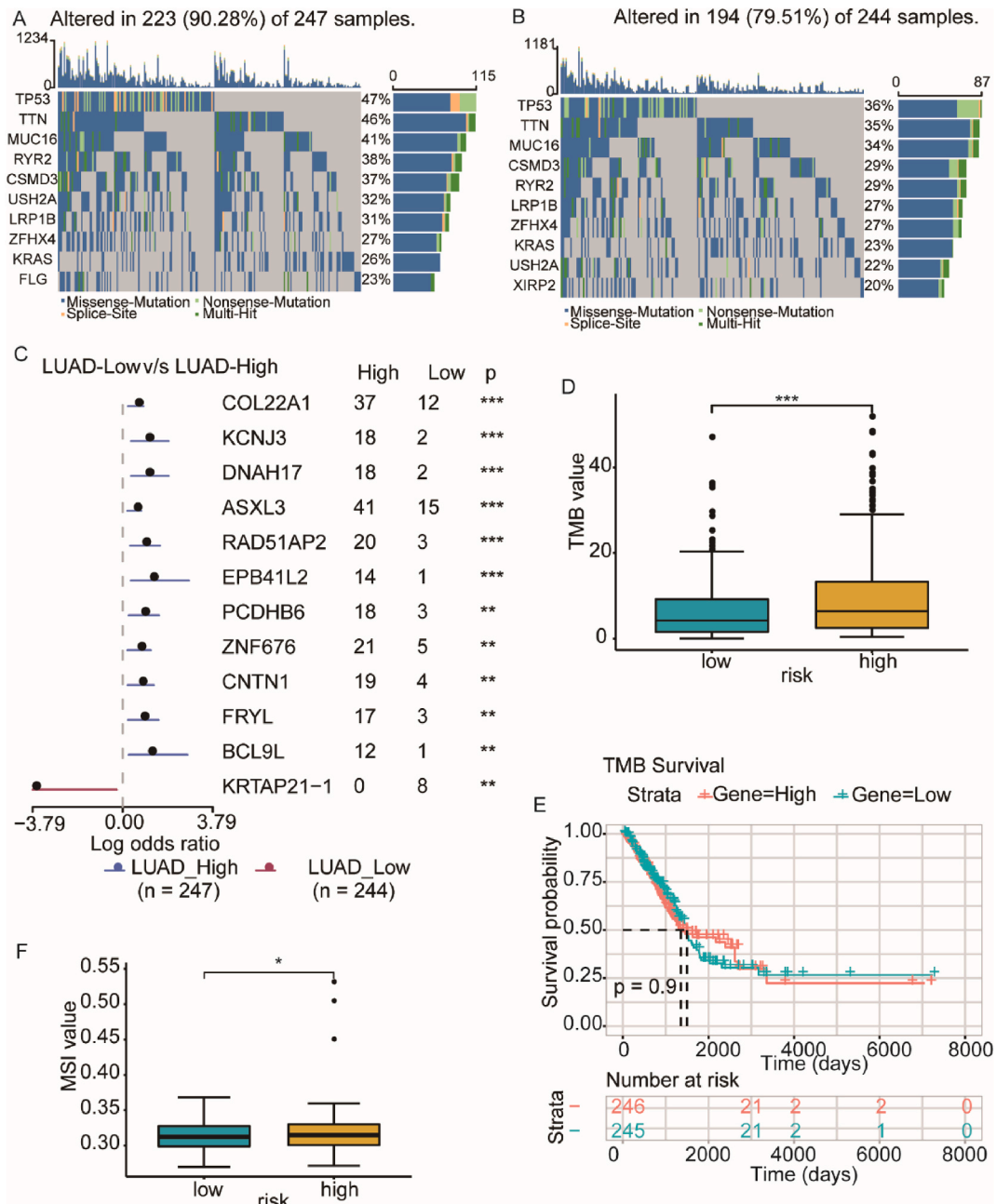


Fig. 6. Tumor mutation load. (A–B) Top 10 mutated genes in the two groups. (C) Comparison of mutated genes in two groups. (D) Different TMB values in two groups. (E) Multivariate analysis between TMB and survival prognosis of LUAD. (F) Differences in microsatellite instability in two groups. * $p < 0.05$, *** $p < 0.001$.

statistical analyses, it has been ascertained that the risk score emerges as the exclusive diagnostic factor that exerts a substantial impact on the survival time of patients. In contrast, although other variables display a certain degree of association, they fail to demonstrate an independent correlation with survival time.

3.4. Gene function enrichment analysis

A GSEA gene enrichment analysis as carried out to determine which pathway genes related to the high-risk and low-risk groups were mainly enriched. The heat maps in Fig. 5A,B showed the expression of the top 50 genes, which were positively and negatively correlated with the high- and low-risk groups. The enrichment analysis of GSEA-GO, GSEA-KEGG, and GSEA-Reactome gene pathways showed that these genes were primarily involved in 20 pathways, which were significantly enriched in chromatin separation, mitosis, cell adhesion, and molecular binding, as suggested by GSEA-GO (Fig. 5C) enrichment analysis. Nonetheless, GSEA-KEGG results (Fig. 5D) showed that genes were significantly enriched in cell cycle, RNA transport, adhesion plaques, virus infection, and calcium signaling pathways; GSEA-Reactome results (Fig. 5E) showed that genes were mainly enriched in cell cycle, RNA transport, gene transcription, and other pathways.

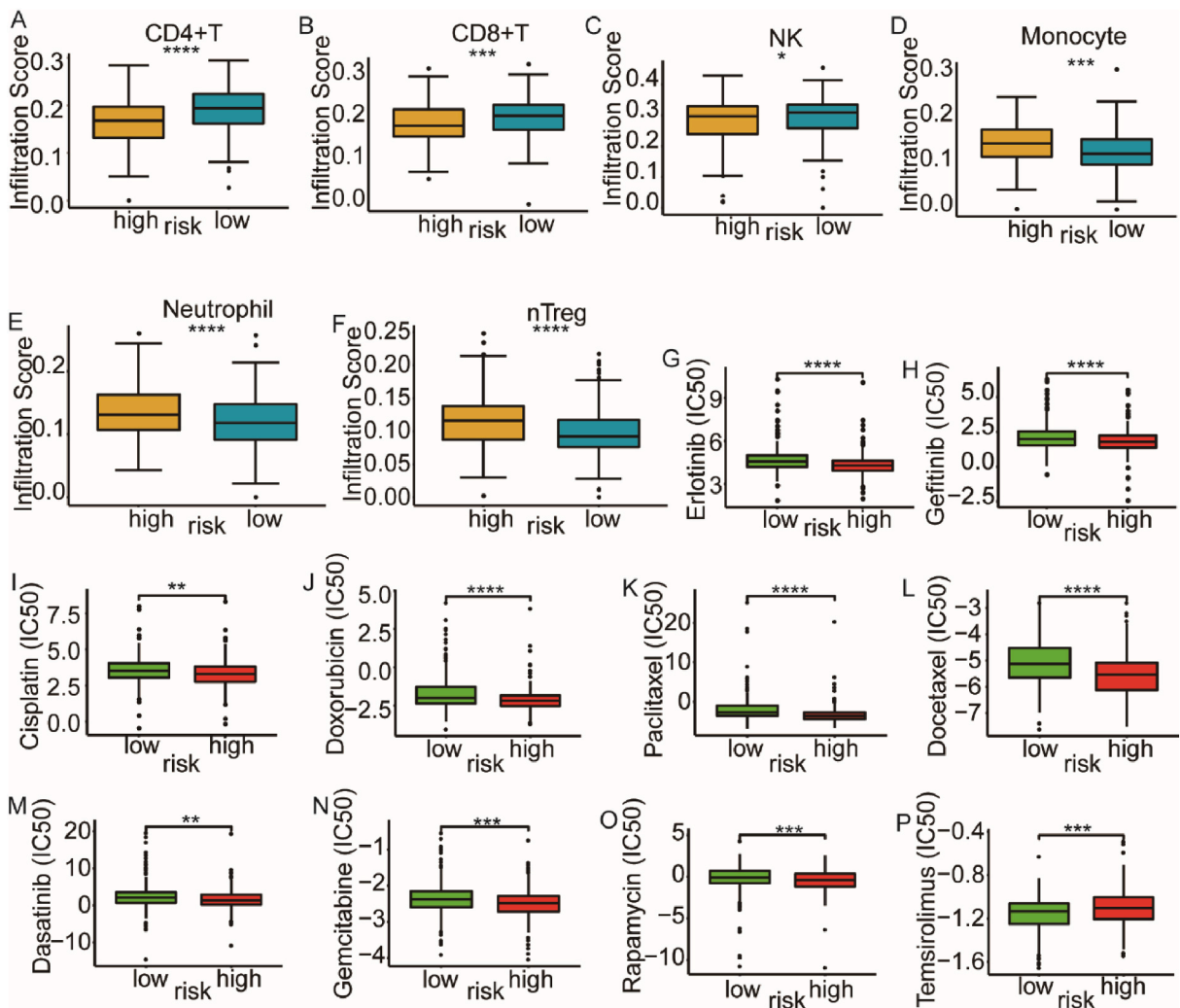


Fig. 7. Evaluation of immune cell infiltration by the prognostic risk model and drug susceptibility (IC50). (A) The abundance of CD4+T cells in two groups. (B) The abundance of CD8+T cells in the two groups. (C) The abundance of NK cells in the two groups. (D) The abundance of monocytes in the two groups. (E) The abundance of neutrophils in the two groups. (F) The abundance levels of nTregs in the two groups. (G–P) The sensitivity of targeted LUAD drugs, such as Erlotinib (G), Gefitinib (H), Cisplatin (I), Doxorubicin (J), Paclitaxel (K), Docetaxel (L), Dasatinib (M), Gemcitabine (N), Rapamycin (O), and Temsirolimus (P) in two groups. **p* < 0.05, ***p* < 0.01, ****p* < 0.001, *****p* < 0.0001.

3.5. Tumor mutation burden (TMB)

Proto-oncogene mutation is an important cause of tumorigenesis. Therefore, we analyzed mutation profiling in the high- and low-risk groups. We discovered that 90.28 % of the patients had gene mutations in the high-risk group, whereas 79.51 % of the patients had gene mutations in the low-risk group. Moreover, the frequency of TP53 gene mutation was the highest, and missense mutation was the dominant type (Fig. 6A and B). Fig. 6C shows different gene mutations in the high-risk and low-risk groups. Next, we calculated the TMB values, which were considered a predictive factor to evaluate the effectiveness of immune checkpoint therapy. Fig. 6D shows that the high-risk group had a higher TMB value; however, multivariate analysis found no association between TMB and survival of LUAD (Fig. 6F). Fig. 6E shows higher microsatellite instability (MSI) of LUAD patients in the high-risk group when comparing to the low-risk group. The tumor mutation burden has been recognized as a crucial determinant for the efficacy of targeted therapy and immunotherapy. This observation emphasizes the necessity for conducting further analysis in order to guide the implementation of

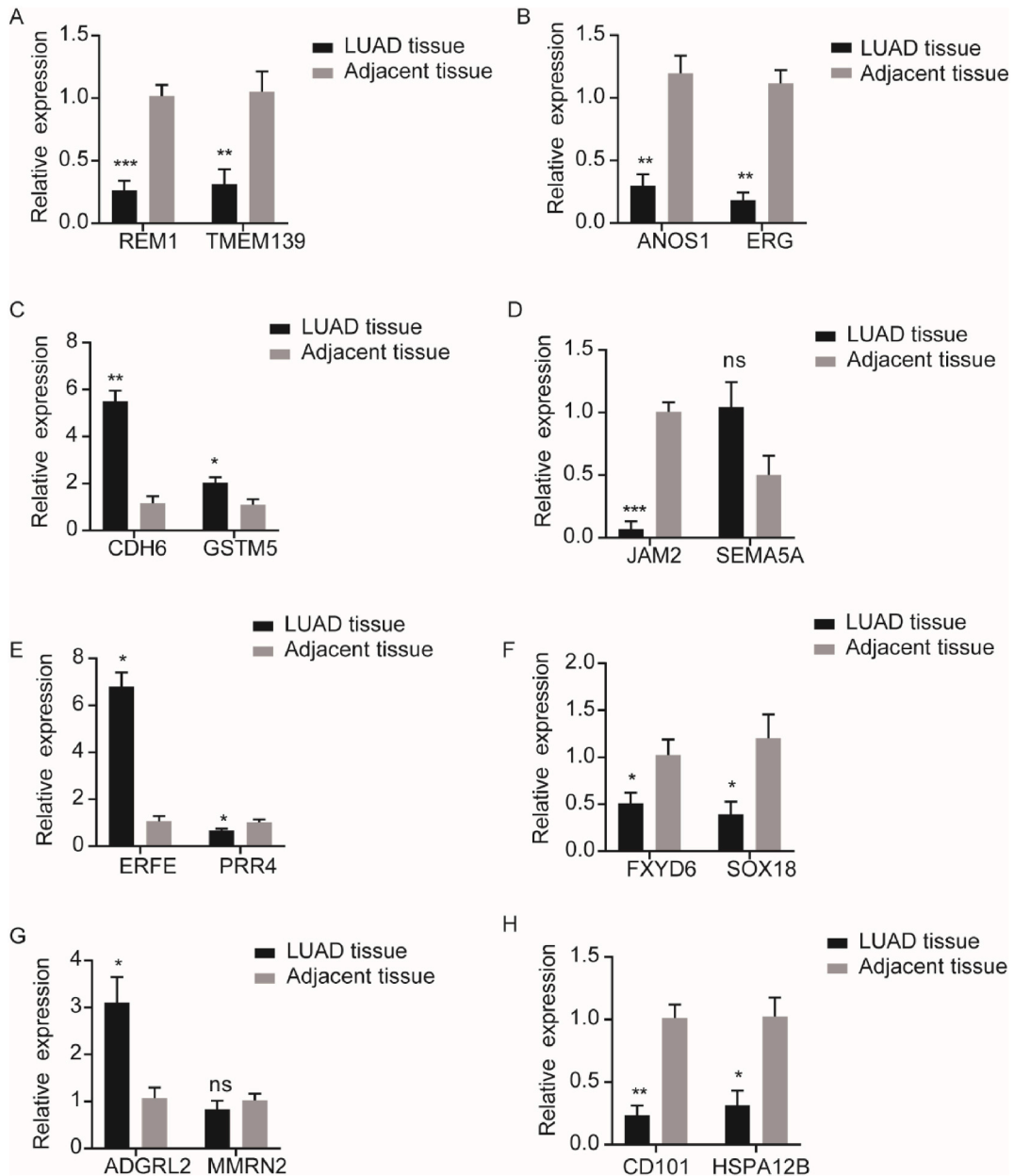


Fig. 8. Expression of DEirmRNA pairs in paracancerous and cancer tissues. (A–D) Expression of REM1/TMEM139, ANOS1/ERG, CDH6/GSTM5, and JAM2/SEMA5A from the risk model in adjacent and cancer tissue. (E–H) Expression of ERFE/PRR4, FXYD6/SOX18, ADGRL2/MMRN2, and CD101/HSPA12B from the risk model in adjacent and cancer tissue. **p* < 0.05, ***p* < 0.01, ****p* < 0.001, ns, no significance.

immunotherapy and targeted therapies.

3.6. Using the prognostic risk model to infer immune cell infiltration and drug resistance

Since tumor metastasis, diffusion, and deterioration are inseparable from the suppressive TME, we wanted to clarify the correlation between risk scores and immune cell infiltration. Our findings reveal a significant decrease in the abundance of infiltrating immune cells in the high-risk group compared to the low-risk group, indicating a state of immune cell deficiency in the high-risk group (Fig. S1A). Further analysis revealed that compared to the low-risk group, individuals in the high-risk group exhibited a significant decrease in the abundance of immune cells associated with anti-tumor immune responses, including CD4⁺ T cells, CD8⁺ T cells, NK cells, B cells, NKT cells, $\gamma\delta$ T cells, and Tfh cells. In contrast, the abundance of nTreg cells, which are associated with promoting tumor immune responses, was increased in the high-risk group relative to the low-risk group. Additionally, the abundance of exhausted CD8⁺T cells (PD-1⁺CD8⁺T cells) was upregulated in the high-risk group compared to the low-risk group. No statistically significant differences were observed in the abundance of other immune cells such as macrophages, dendritic cells, and iTreg cells. (Fig. 7A–F, Figs. S1B–I). We also predicted the treatment response of clinically used chemotherapy drugs and targeted drugs to guide clinical medication. As shown in Fig. 7G–P, cisplatin, paclitaxel, docetaxel, gemcitabine, as well as EGFR-related targeted drugs Erlotinib and Gefitinib, exhibited lower half-inhibitory concentrations in the high-risk group, indicating a higher sensitivity to these chemotherapy and targeted treatments in the high-risk group. Therefore, our model can provide valuable insights for clinical drug selection and treatment.

3.7. Expression of DEirmRNA pairs in adjacent and tumor tissues

In addition to database analysis, we also detected mRNA expression of relative DEirmRNA pairs in adjacent and tumor tissues by a quantitative reverse-transcription PCR (qRT-PCR) assay to verify the accuracy of the model. Results showed that compared to adjacent tissues, ERFE/PRR4, ADGRL2/MMRN2, and CDH6/GSTM5 were highly expressed, and ANOS1/ERG, CD101/HSPA12B, FXVD6/SOX18, JAM2/SEMA5A, and REM1/TMEM139 were lowly expressed (Fig. 8A–H). These results suggest that our risk model based on DEirmRNA pairs can be used to assist in determining the prognosis of patients with LUAD.

4. Discussion

In this study, we creatively selected DEirmRNA through the algorithm and paired it to establish a prognostic risk model. Moreover, to make our model more precise regarding prediction, we constructed a modified lasso regression and 10-fold cross-validation and incorporated factors into the Cox regression model through the ranking of the frequency of occurrence in the repetition process rather than the intersection. The maximum AUC value of the optimal model was obtained by calculating the AUC value of each ROC curve, and we used the AIC value to verify the critical point of our model. After distinguishing patients by the new method, we re-evaluated the survival results of the patients and analyzed their clinical characteristics by univariate and multivariate regression analyses. Additionally, the relationship of risk score with survival, clinical characteristics, gene enrichment, tumor mutation load, immune cell infiltration, and drug susceptibility of LUAD patients was analyzed.

We found that risk scores are reliable independent predictors for the prognosis of LUAD patients. We enriched and analyzed the differentially expressed genes in selected patients and found that these genes were remarkably enriched in chromatin separation, mitosis, cell cycle, adhesion plaque, transcription, and other pathways. The most basic feature of cancer is abnormal cell proliferation caused by uncontrolled cell division and imbalance of the cell cycle [24]. Additionally, adhesion plaques play an important role in cancer invasion and metastasis, resulting in cancer progression [25,26]. Our data suggested that high-risk LUAD patients are prone to have *TP53* gene mutations than low-risk patients. Research has demonstrated that the *TP53* gene is more prone to mutations than other genes, and mutations in the *TP53* tumor suppressor factor are a common cause of lung cancer, particularly in NSCLC [27]. MSI usually occurs during DNA replication because the function of base mismatch repair is damaged, resulting in base insertion and deletion, which change the length of MS sequences. Some studies have revealed that patients who had a higher MSI value could enhance the ability to inhibit tumor progression, which benefited the prognosis of cancer patients [28,29]. High-risk LUAD patients had a greater MSI value compared to the low-risk patients, indicated that they were more likely to benefit from immunotherapy. High TMB values were usually associated with an increased number of mutations in tumors and were regarded as a predictive factor in evaluating the effectiveness of immune checkpoint therapy. Previous studies have demonstrated TMB correlated with the effectiveness of immune checkpoint therapy, patients who have high TMB are more sensitive to immunotherapy and have a better outcome [30–32]. In our study, high-risk score LUAD patients had greater TMB value than those with low-risk scores, indicated that they had a better effect of immunotherapy and longer survival. However, Fig. 6F shows that there was no association between TMB and survival time of LUAD. Nonetheless, since TMB is associated with the ability of tumor cells to produce neoantigens, which correlates with antitumor immune responses, we also sought to explore the relationship between this model and immune cell infiltration.

According to our model, we observed a significant decrease in the abundance of infiltrating immune cells in the high-risk group compared to the low-risk group. CD4⁺ and CD8⁺ T cells are the primary cells that exert adaptive immunity, including immune surveillance and tumor cell clearance. CD4⁺ T cells exert their anti-tumor ability by secreting interferon- γ or modulating the activity of other immune cells to inhibit tumor progression, and CD8⁺T cells (CTL) are able to kill malignant cells upon recognition by T-cell receptor (TCR) of specific antigenic peptides presented on the surface of target cells. NK cells can release cytotoxic granules at sites of direct contact with Ag-bearing tumor cell. In the high-risk group, the decrease in immune cells (i.e. CD4⁺T, CD8⁺T and NK cells)

infiltrating the TME indicated that their anti-tumor response was inhibited or related to their poor prognosis. Furthermore, in addition to the significant decrease in the abundance of infiltrating immune cells in the high-risk group compared to the low-risk group, there was also an increase in exhaustion of CD8⁺ T cells, indicating a state of T cell dysfunction. In experimental studies, IC50 is often used as an index to measure drug efficacy. The stronger the drug efficacy, the lower the corresponding IC50 value [33]. EGFR is a highly mutated oncogene in NSCLC, and the FDA has approved the use of EGFR inhibitors to treat NSCLC [34], Gefitinib and Erlotinib both were EGFR inhibitors, in our study, we evaluated the relationship between IC50 of these antitumor drugs and risk scores. We observed a higher susceptibility of the high-risk groups to EGFR mutation-related targeted drugs, such as Erlotinib and Gefitinib, along with various chemotherapeutic agents, compared to the low-risk groups. Consequently, our findings suggest the potential of this model to offer personalized treatment for LUAD patients, particularly in the context of EGFR mutation-related targeted therapies. Furthermore, we propose that this model has the capability to mitigate the requirement for expensive sequencing expenses in a clinical setting. To better verify the results of our study, we also detected the expression level of relative DEirmRNA pairs in adjacent and cancer tissues. The results were in consistent with the previous data predicted by the prognostic risk model, suggesting that the prognostic risk model based on DEirmRNA pairs has great clinical potential in predicting the prognosis of LUAD.

5. Conclusions

In conclusion, we identified and validated 28 DEirmRNA pairs to create the prognostic risk model. The model was regarded as a good biomarker to predict the prognosis of LUAD patients and provided an important individualized clinical treatment for LUAD patients.

Data availability

The data used to support the findings of this study are included within the article.

Availability of data and materials

Publicly available datasets were analyzed in this study. This data can be found here: UCSC-XENA database, ImmPort database.

Ethics statement

The study was performed in strict compliance and approval by the ethical committee of the Third Affiliated Hospital of Soochow University.

CRedit authorship contribution statement

Jiawei Yue: Writing – original draft, Investigation, Formal analysis, Data curation, Wei Zhou, Methodology, Formal analysis. **Hui Guo:** Writing – original draft, Supervision, Software, Formal analysis, Data curation. **Jinhong Ma:** Validation, Software. **Weifeng Shi:** Writing – review & editing, Supervision. **Yumin Wu:** Writing – review & editing, Project administration, Funding acquisition, Formal analysis, Data curation, Conceptualization.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgments

This work was supported by grants from the National Natural Science Foundation of China (grant 81801568).

Appendix A. Supplementary data

Supplementary data to this article can be found online at <https://doi.org/10.1016/j.heliyon.2024.e24397>.

References

- [1] M. Cao, H. Li, D. Sun, W. Chen, Cancer burden of major cancers in China: a need for sustainable actions, *Cancer Commun.* 40 (5) (2020) 205–210.
- [2] M.B. Schabath, M.L. Cote, Cancer progress and priorities: lung cancer, *Cancer Epidemiol. Biomarkers Prev.* 28 (10) (2019) 1563–1579.
- [3] R.L. Siegel, K.D. Miller, H.E. Fuchs, A. Jemal, Cancer statistics, 2021, *CA Cancer J Clin* 71 (1) (2021) 7–33.

- [4] L.P. Jiang, S.Q. Fan, Q.X. Xiong, Y.C. Zhou, Z.Z. Yang, G.F. Li, Y.C. Huang, M.G. Wu, Q.S. Shen, K. Liu, C.P. Yang, Y.B. Chen, GRK5 functions as an oncogenic factor in non-small-cell lung cancer, *Cell Death Dis.* 9 (3) (2018) 295.
- [5] A.A. Thai, B.J. Solomon, L.V. Sequist, J.F. Gainor, R.S. Heist, Lung cancer, *Lancet* 398 (10299) (2021) 535–554.
- [6] D.I. Suster, M. Mino-Kenudson, Molecular pathology of primary non-small cell lung cancer, *Arch. Med. Res.* 51 (8) (2020) 784–798.
- [7] W. Lin, Y. Chen, B. Wu, Y. Chen, Z. Li, Identification of the pyroptosis-related prognostic gene signature and the associated regulation axis in lung adenocarcinoma, *Cell Death Dis.* 7 (1) (2021) 161.
- [8] J.W. Ma, M. Li, Molecular typing of lung adenocarcinoma with computed tomography and CT image-based radiomics: a narrative review of research progress and prospects, *Transl. Cancer Res.* 10 (9) (2021) 4217–4231.
- [9] D.J.H. Shih, N. Nayyar, I. Bihun, I. Dagogo-Jack, C.M. Gill, E. Aquilanti, M. Bertalan, A. Kaplan, M.R. D'Andrea, U. Chukwueke, F.M. Ippen, C. Alvarez-Breckenridge, N.D. Camarda, M. Lastrapes, D. McCabe, B. Kuter, B. Kaufman, M.R. Strickland, J.C. Martinez-Gutierrez, D. Nagabhushan, M. De Sauvage, M. D. White, B.A. Castro, K. Hoang, A. Kaneb, E.D. Batchelor, S.H. Paek, S.H. Park, M. Martinez-Lage, A.S. Berghoff, P. Merrill, E.R. Gerstner, T.T. Batchelor, M. P. Frosch, R.P. Frazier, D.R. Borger, A.J. Iafate, B.E. Johnson, S. Santagata, M. Preusser, D.P. Cahill, S.L. Carter, P.K. Brastianos, Genomic characterization of human brain metastases identifies drivers of metastatic lung adenocarcinoma, *Nat. Genet.* 52 (4) (2020) 371–377.
- [10] S. Jia, L. Li, L. Xie, W. Zhang, T. Zhu, B. Qian, Transcriptome based estrogen related genes biomarkers for diagnosis and prognosis in non-small cell lung cancer, *Front. Genet.* 12 (2021) 666396.
- [11] B. Sun, E.D. Brooks, R. Komaki, Z. Liao, M. Jeter, M. McAleer, P.A. Balter, J.D. Welsh, M. O'Reilly, D. Gomez, S.M. Hahn, B. Sepesi, D.C. Rice, J.V. Heymach, J. Y. Chang, Long-term outcomes of salvage stereotactic ablative radiotherapy for isolated lung recurrence of non-small cell lung cancer: a phase II clinical trial, *J. Thorac. Oncol.* 12 (6) (2017) 983–992.
- [12] H. Wen, X. Lin, D. Sun, Gender-specific nomogram models to predict the prognosis of male and female lung adenocarcinoma patients: a population-based analysis, *Ann. Transl. Med.* 9 (22) (2021) 1654.
- [13] S.K. Kamarajah, W.R. Burns, T.L. Frankel, C.S. Cho, H. Nathan, Validation of the American joint commission on cancer (ajcc) 8th edition staging System for patients with pancreatic adenocarcinoma: a surveillance, epidemiology and end results (seer) analysis, *Ann. Surg. Oncol.* 24 (7) (2017) 2023–2030.
- [14] S. Wu, X. Dai, D. Xie, Identification and validation of an immune-related RNA signature to predict survival of patients with head and neck squamous cell carcinoma, *Front. Genet.* 10 (2019) 1252.
- [15] F.R. Hirsch, G.V. Scagliotti, J.L. Mulshine, R. Kwon, W.J. Curran Jr., Y.L. Wu, L. Paz-Ares, Lung cancer: current therapies and new targeted treatments, *Lancet* 389 (10066) (2017) 299–311.
- [16] M. Wu, X. Li, T. Zhang, Z. Liu, Y. Zhao, Identification of a nine-gene signature and establishment of a prognostic nomogram predicting overall survival of pancreatic cancer, *Front. Oncol.* 9 (2019) 996.
- [17] T. Fan, M. Zhu, L. Wang, Y. Liu, H. Tian, Y. Zheng, F. Tan, N. Sun, C. Li, J. He, Immune profile of the tumor microenvironment and the identification of a four-gene signature for lung adenocarcinoma, *Aging (Albany NY)* 13 (2) (2020) 2397–2417.
- [18] M. Majumder, P. Chakraborty, S. Mohan, S. Mehrotra, V. Palanisamy, HuR as a molecular target for cancer therapeutics and immune-related disorders, *Adv. Drug Deliv. Rev.* 188 (2022) 114442.
- [19] Q. Ren, Q. Li, C. Shao, P. Zhang, Z. Hu, J. Li, W. Wang, Y. Yu, Establishing a prognostic model based on immune-related genes and identification of BIRC5 as a potential biomarker for lung adenocarcinoma patients, *BMC Cancer* 23 (1) (2023) 897.
- [20] Y. Wang, J. Li, F. Wang, Y. Cui, L. Song, B. Ruan, Y. Yu, Analysis of ceRNA regulatory mechanism of rape pollen allergy based on whole-transcriptome sequencing of peripheral blood mononuclear cells, *J. Asthma Allergy* 16 (2023) 775–788.
- [21] Y. Zhai, B. Zhao, Y. Wang, L. Li, J. Li, X. Li, L. Chang, Q. Chen, Z. Liao, Construction of the optimization prognostic model based on differentially expressed immune genes of lung adenocarcinoma, *BMC Cancer* 21 (1) (2021) 213.
- [22] X. Qi, G. Chen, Z. Chen, J. Li, W. Chen, J. Lin, L. Lin, Construction of a novel lung adenocarcinoma immune-related lncRNA pair signature, *Int. J. Gen. Med.* 14 (2021) 4279–4289.
- [23] R. Bonneville, M.A. Krook, E.A. Kautto, J. Miya, M.R. Wing, H.Z. Chen, J.W. Reeser, L. Yu, S. Roychowdhury, Landscape of microsatellite instability across 39 cancer types, *JCO Precis Oncol* 2017 (2017).
- [24] A.W. Lau, H. Inuzuka, H. Fukushima, L. Wan, P. Liu, D. Gao, Y. Sun, W. Wei, Regulation of APC(Cdh1) E3 ligase activity by the Fbw7/cyclin E signaling axis contributes to the tumor suppressor function of Fbw7, *Cell Res.* 23 (7) (2013) 947–961.
- [25] L. Lewczuk, A. Pryczynicz, K. Guzinska-Ustymowicz, Cell adhesion molecules in endometrial cancer - a systematic review, *Adv. Med. Sci.* 64 (2) (2019) 423–429.
- [26] W. Wang, W. Lou, B. Ding, B. Yang, H. Lu, Q. Kong, W. Fan, A novel mRNA-miRNA-lncRNA competing endogenous RNA triple sub-network associated with prognosis of pancreatic cancer, *Aging (Albany NY)* 11 (9) (2019) 2610–2627.
- [27] X.D. Jiao, B.D. Qin, P. You, J. Cai, Y.S. Zang, The prognostic value of TP53 and its correlation with EGFR mutation in advanced non-small cell lung cancer, an analysis based on cBioPortal data base, *Lung Cancer* 123 (2018) 70–75.
- [28] H.M. Mohan, E. Ryan, I. Balasubramanian, R. Kennelly, R. Geraghty, F. Sclafani, D. Fennelly, R. McDermott, E.J. Ryan, D. O'Donoghue, J.M. Hyland, S. T. Martin, P.R. O'Connell, D. Gibbons, D. Winter, K. Sheahan, Microsatellite instability is associated with reduced disease specific survival in stage III colon cancer, *Eur. J. Surg. Oncol.* 42 (11) (2016) 1680–1686.
- [29] C.G. Kim, J.B. Ahn, M. Jung, S.H. Beom, C. Kim, J.H. Kim, S.J. Heo, H.S. Park, J.H. Kim, N.K. Kim, B.S. Min, H. Kim, W.S. Koom, S.J. Shin, Effects of microsatellite instability on recurrence patterns and outcomes in colorectal cancers, *Br. J. Cancer* 115 (1) (2016) 25–33.
- [30] R. Buttner, J.W. Longshore, F. Lopez-Rios, S. Merkelbach-Bruse, N. Normanno, E. Rouleau, F. Penault-Llorca, Implementing TMB measurement in clinical practice: considerations on assay requirements, *ESMO Open* 4 (1) (2019) e000442.
- [31] D.J. McGrail, P.G. Pilie, N.U. Rashid, L. Voorwerk, M. Slagter, M. Kok, E. Jonasch, M. Khasraw, A.B. Heimberger, B. Lim, N.T. Ueno, J.K. Litton, R. Ferrarotto, J. T. Chang, S.L. Moulder, S.Y. Lin, High tumor mutation burden fails to predict immune checkpoint blockade response across all cancer types, *Ann. Oncol.* 32 (5) (2021) 661–672.
- [32] R.M. Samstein, C.H. Lee, A.N. Shoushtari, M.D. Hellmann, R. Shen, Y.Y. Janjigian, D.A. Barron, A. Zehir, E.J. Jordan, A. Omuro, T.J. Kaley, S.M. Kendall, R. J. Motzer, A.A. Hakimi, M.H. Voss, P. Russo, J. Rosenberg, G. Iyer, B.H. Bochner, D.F. Bajorin, H.A. Al-Ahmadie, J.E. Chaff, C.M. Rudin, G.J. Riely, S. Baxi, A. L. Ho, R.J. Wong, D.G. Pfister, J.D. Wolchok, C.A. Barker, P.H. Gutin, C.W. Brennan, V. Tabar, I.K. Mellingerhoff, L.M. DeAngelis, C.E. Ariyan, N. Lee, W.D. Tap, M.M. Gounder, S.P. D'Angelo, L. Saltz, Z.K. Stadler, H.I. Scher, J. Baselga, P. Razavi, C.A. Klebanoff, R. Yaeger, N.H. Segal, G.Y. Ku, R.P. DeMatteo, M. Ladanyi, N.A. Rizvi, M.F. Berger, N. Riaz, D.B. Solit, T.A. Chan, L.G.T. Morris, Tumor mutational load predicts survival after immunotherapy across multiple cancer types, *Nat. Genet.* 51 (2) (2019) 202–206.
- [33] M. Joo, A. Park, K. Kim, W.J. Son, H.S. Lee, G. Lim, J. Lee, D.H. Lee, J. An, J.H. Kim, T. Ahn, S. Nam, A deep learning model for cell growth inhibition IC50 prediction and its application for gastric cancer patients, *Int. J. Mol. Sci.* 20 (24) (2019).
- [34] C. Cheng, Y. Zhao, E. Schaafsma, Y.L. Weng, C. Amos, An EGFR signature predicts cell line and patient sensitivity to multiple tyrosine kinase inhibitors, *Int. J. Cancer* 147 (9) (2020) 2621–2633.