Research article

A novel signature of seven aging-related genes for risk stratification, prognosis prediction and benefit evaluation of chemotherapy, and immunotherapy in elderly patients with lung adenocarcinoma

Yi Tian [a], Wenya Zhao [a], Chenjing Lin [a], Yang Chen [b], Qiaoxin Lin [b], Yiru Liu [a], Dianna Gu [b,**], Ling Tian [a,*]

[a] Department of Central Laboratory, Shanghai Chest Hospital, Shanghai Jiao Tong University School of Medicine, Shanghai, 200030, China
[b] Department of Medical Oncology, The First Affiliated Hospital of Wenzhou Medical University, Wenzhou, 325000, China

A R T I C L E   I N F O

A B S T R A C T

Background: Aging, a multifaceted biological process, is thought to be associated with lung adenocarcinoma (LUAD) development and progression. However, it is unclear whether aging-related genes (ARGs) can predict tumor risk, chemotherapy and immunotherapy benefits, and prognosis in LUAD patients at different ages.
Methods: Gene expression datasets and clinical information of LUAD patients were downloaded from TCGA and GEO database. Univariate and multivariate Cox regression, and lasso algorithm were employed to identify the ARG signatures. Patients were stratified into high-risk and low-risk groups to evaluate the predictive accuracy using Kaplan-Meier curves, ROC curves, and time-dependent AUC. A nomogram was established to predict the survival probability. GSEA revealed potential pathways, and CIBERSORT indicated different immunologic status. TIDE score was used to predict the potential tumor response to immune checkpoint inhibitors, and GDSC was employed to evaluate the sensitivity of chemotherapeutic drugs. The correlation of TIDE score and patient age, as well as that of ARGs and patient age was investigated. And cell Culture and RT-qPCR for external validation for key gene.
Results: A novel gene signature based on seven ARGs was established, including BMP15, CD79A, CDKN3, CDX2, COL1A1, DKK1, and GRIK2. Our model demonstrated exceptional prediction accuracy for elderly LUAD patients of 71–90 years old. A nomogram model was constructed to predict the survival probability, and the C-index value was 0.737, indicating our prognostic nomogram model has high accuracy. Through external RT-qPCR validation, we found that CD79A expression in H1299 was higher than that of BEAS-2B. And novel immunotherapy and chemotherapy regimens were accordingly proposed for the elderly LUAD patients.
Conclusion: We identified a novel gene signature based on seven ARGs for risk stratification, prognosis prediction and benefit evaluation of immunotherapy and chemotherapy in elderly LUAD patients.

---

* Corresponding author.
** Corresponding author.
E-mail addresses: yinuo801@126.com (D. Gu), TL09168@hotmail.com (L. Tian).

1. Introduction

Lung cancer ranked the second most diagnosed cancer with an estimated 2.2 million new cases (11.4 %), and remained the leading cause of cancer death with an estimated 1.8 million deaths (18 %) worldwide by the GLOBOCAN estimates for 2020 [1]. Non-small cell lung cancer (NSCLC) comprises approximately 85 % of all lung cancer cases, with lung adenocarcinoma (LUAD) representing the most significant pathological subtype of primary lung cancer [2]. Despite of advances in cancer therapy in recent decades, LUAD patients have a dismal prognosis, with a 5-year survival rate of less than 21.7 % [3]. Fortunately, the rapid advances in biomedical technology, especially in genome sequencing, have led to a better understanding of tumor biology and more effective diagnosis and treatment of cancers.

Aging is a complex biological process that is characterized by the progressive loss of physiological integrity as individuals age. Recently, 12 major hallmarks of aging were proposed [4], and a synergistic, bidirectional correlation between aging and cancer was suggested with many shared characteristics, including genomic instability, telomere attrition, mitochondrial dysfunction, cellular senescence, and so on [5]. The latest statistics revealed that the incidence of cancer gradually increased with age, especially after the age of 50, and lung cancer was taken for the leading cause of death in people over 50 [6]. It is, therefore, crucial to identify aging-related genes (ARGs) as a prognostic signature for the elderly population. Needless to say, some ARGs have been identified and developed for risk stratification and prognosis prediction in LUAD [7–9]. However, the age stratification of LUAD patients did not been taken into sufficient consideration. Therefore, it is necessary to figure out the association between ARGs and the age of LUAD patients, and to further identify the associated ARGs for risk stratification, prognosis prediction and benefit evaluation of chemotherapy and immunotherapy.

Herein, we studied the relevant prognostic genes and corresponding clinical information collected from the TCGA (The Cancer Genome Atlas) LUAD database, and identified a novel gene signature based on seven ARGs, including BMP15, CD79A, CDKN3, CDX2, COL1A1, DKK1, and GRIK2, which showed exceptional prediction accuracy for elderly LUAD patients of 71–90 years old, particularly in terms of risk stratification, prognosis prediction and benefit evaluation of chemotherapy and immunotherapy.

2. Materials and methods

2.1. Data collection

The RNA sequence data of 539 LUAD samples and 59 normal ones, along with corresponding clinical information, were collected from TCGA (https://portal.gdc.cancer.gov/) database. In 539 patients with LUAD, we used 59 adjacent normal samples from TCGA to identify differentially expressed genes (DEGs) and also to compare the differences in prognostic gene expression between tumor and normal tissues. While these 59 normal tissue samples were excluded for the subsequent tumor patient survival analysis. Also, 66 tumor samples with incomplete clinical data were excluded for subsequent survival analysis of LUAD patients. The final sample size for further survival analysis was therefore 473 cases. The data were standardized and normalized using the Variance Stabilizing Transformation (VST) function from DESeq2 before further analysis [10,11]. Additionally, GSE30219 and GSE72094 datasets were downloaded from the Gene Expression Omnibus (GEO) database (https://www.ncbi.nlm.nih.gov/geo/) for validation purposes. Only patients with intact survival data were included, resulting in 293 and 398 patients for the two datasets, respectively. "Limma" R package was used to eliminate the batch effect on the data. Log2 transformation was performed for normalization. The data structures and clinical information are presented in Supplementary Table S1.

2.2. Identification of differentially expressed aging-related genes

A comprehensive list of aging-related genes (ARGs) (Supplementary Data) was generated from the GeneCards website (https://www.genecards.org/) [12], and the relevance scores of the selected ARGs exceed 5 points. Then, we employed the "DESeq2" R package to systematically screen the differentially expressed genes (DEGs) within the TCGA LUAD database [10]. Subsequently, we generated a volcano plot based on the up- and down-regulated genes, utilizing the "ggplot2" R package. To establish statistical significance, we set the adjusted p-value (adj. p) < 0.05 and $|$ log2 fold change (FC)$| > 2$ as the threshold for identifying DEGs. Finally, we utilized the "VennDiagram" R package to generate a Venn diagram of the overlapping genes between the DEGs and ARGs datasets. The genes that overlapped between these two datasets were then integrated into subsequent analysis.

To delve deeper into the function and pathway of the DEGs, we conducted a thorough analysis utilizing Gene Ontology (GO) and Kyoto Encyclopedia of Genes and Genomes (KEGG) methodologies. GO analysis was performed based on the "MSigDB C5: GO gene sets" database. For GO and KEGG enrichment analysis, we set the parameter p-value cutoff to 0.05, with a p-value less than 0.05 considered the enrichment results statistically significant. Additionally, we utilized advanced tools such as the "clusterProfiler" [13, 14], to craft a comprehensive and visually engaging circular pathway map that accurately reflects the results of our analysis.

Further, the Gene Set Enrichment Analysis (GSEA) was employed to investigate distinct potential pathways for high-risk and low-risk groups, utilizing the gene sets database "c2. all.v2022.1. Hs.symbols.gmt" [15]. Subsequent to conducting 1000 permutations, gene sets with p-values below 0.05 and a false discovery rate (FDR) below 0.25 were recognized as significantly enriched.

Finally, we performed single-sample gene set enrichment analysis (ssGSEA) to investigate differences in the distribution of ARGs between different age groups by "GSVA" R package.

2.3. Construction and validation of the prognostic ARG signature

We used the TCGA dataset for constructing the model, and the GEO database for model validation. Univariate Cox regression analysis was employed based on 261 ARGs to scrutinize potential genes that are linked to overall survival (OS) [16]. When using Cox regression models and the Kaplan-Meier method, we utilized the "cox.zph" function to compute the p-value for testing the proportional hazards assumption by using Schoenfeld residuals base on "survival" R package. Check whether the p-value of each covariate is less than the specified threshold of 0.05. If the p-value is less than the threshold, the covariable is designated as significant, and deemed eligible for the subsequent step analysis. Then, these genes underwent the least absolute shrinkage and selection operator (LASSO) regression analysis based on the "glmnet" R package [17]. In our analysis, we utilized a 10-fold cross-validation method to determine the optimal value for LASSO regression. This process was executed using the "cv.glmnet" function within the "glmnet" R package. After comprehensive evaluation, the optimal λ value was identified as 0.02608918. Upon selecting the optimal λ coefficient, the pertinent genes were further shortlisted for multivariate Cox regression analysis, and the potential prognostic genes were eventually identified. Based on the correlation coefficient of these genes by multivariate Cox regression, the risk score was calculated using the following equation: risk score = expression level of gene 1 × risk coefficients 1 + expression level of gene 2 × risk coefficients 2 + … + expression level of gene X × risk coefficients X. Subsequently, the patients were categorized into high-risk and low-risk groups based on the median of the risk score, and the survival status of both groups was analyzed using the Kaplan-Meier (KM) curve. The receiver operating characteristic (ROC) curve and corresponding area under the curve (AUC) were utilized to assess the accuracy and dependability of the survival curve based on the "timeROC" R package [18].

We further subjected the potential prognostic ARGs to rigorous evaluation within the GEO validation database. The relevant gene expression and survival data was meticulously acquired, and the identical formula and coefficients were subsequently utilized to partition the patients into high-risk and low-risk cohorts. To further reinforce the credibility of our model, we generated a KM survival curve and ROC curve using the same methodology [18].

To explore whether there is a correlation between age and prognostic ARGs, the Spearman correlation coefficient method was used to scrutinize the correlation between these seven ARGs and patient age [19]. We analyzed the differences in the expression of ARGs across age groups. Upon identifying the key genes, we plotted survival curves for these relevant genes. Additionally, to further explore their relationship with survival outcomes, we segmented and analyzed data by grouping patients according to age and drew the survival curves for each group.

2.4. Development of a nomogram for prognosis prediction

To enhance the evaluation of patients' 1-, 3-, and 5-year survival rates, we employed univariate and multivariate Cox analyses to identify independent prognostic factors from a pool of potential parameters including age, gender, cigarette usage, pathological clinical stage, T stage, N stage, M stage, and risk score. Subsequently, utilizing the "survival" and "rms" packages in R, we established a nomogram and generated calibration curves to verify the authenticity and precision of the nomogram [20].

2.5. CIBERSORT and TIDE analysis for the evaluation of immunotherapy

To elucidate the immune cell infiltration in our aging prognosis model, the cell-type identification by estimating relative subsets of RNA transcripts (CIBERSORT) analysis (https://cibersort.stanford.edu/) was employed to explore the differences in immune cell classification between high-risk and low-risk groups. Upon obtaining the corresponding LM22 file, we utilized the "cibersort" R package for 1000 replacements, calculated the proportion of 22 immune cell types [21], and employed the "pheatmap" R package to generate the corresponding heat map. Further, we utilized the "tidyverse" R package and "ggpubr" R package to produce the corresponding box map and histogram.

To assess the potential clinical effectiveness of immunotherapy in distinct risk groups and assess the potential ability of tumor immune evasion, we utilized Tumor Immune Dysfunction and Exclusion (TIDE) score for further evaluation. We employed an online website (https://tide.dfci.harvard.edu/) to calculate the TIDE score and utilized the "ggpubr" and "ggplot2" R packages to produce a violin plot. A lower TIDE score suggests that immune checkpoint inhibitor (ICI) therapy is more effective for patients [22].

To further explore the potential association between ICI therapy and patient age, the Spearman's rank coefficient of correlation method was used to analyze the correlation between patient age and TIDE [19]. The survival curve between the high- and low-TIDE groups was plotted according to their median TIDE scores. Finally, we assessed differences in the distribution of TIDE scores across age groups.

2.6. GDSC analysis for chemotherapy evaluation

In order to evaluate the sensitivity of chemotherapeutic drugs, the Genomics of Drug Sensitivity in Cancer (GDSC) database (https://www.cancerRxgene.org/) was employed to evaluate the response to chemotherapy in high-risk and low-risk groups. We used "oncoPredict" R package to analyze drug sensitivity [23]. For GDSC analysis, we set a p-value less than 0.05 to consider drug sensitivity as statistically different.

2.7. Cell culture and RT-qPCR for external validation

The Human lung adenocarcinoma cell line (H1299) and normal pulmonary epithelial cells (BEAS-2B) were obtained from the Shanghai Institutes for Biological Sciences (Shanghai, China) cell bank. The H1299 and BEAS-2B cell lines were authenticated by short tandem repeat (STR) analysis, which were supported by GTB (Genetic Testing Biotechnology) technical services (Suzhou, China). The STR Profile Report for Cell Line Authentication confirming the identity of H1299 and BEAS-2B cell lines was available in Supplementary Report Document. The cells were cultured in Dulbecco's modified Eagle's medium (DMEM; Gibco, USA) supplemented with 10 % fetal bovine serum (FBS; Gibco, USA) at 37 °C in a 5 % CO2 incubator. To prevent bacterial contamination, penicillin-streptomycin solutions (Ncmbio, China) was also added to the cell culture media. Besides, during cell culture, we also used Mycoplasma Elimination Reagent (Biosharp, China) to prevent mycoplasma infection. Furthermore, to ensure the absence of mycoplasma contamination, we regularly performed testing with the MycoBlue Kit (FuHeng, China). Total RNA (1 μg) was extracted from both cell lines using RNA-easy Isolation Reagent (Vazyme, Nanjing, China). After the mRNA was reverse transcribed to cDNA, we performed real-time quantitative polymerase chain reaction (RT-qPCR) using the Applied Biosystems ViiA 7 (Life Technologies, USA). We analyzed gene expression using the $2^{-\Delta\Delta CT}$ method. The primer sequences were obtained from Sangon (Shanghai, China). Primers were listed in Table S2.

2.8. Statistical analysis

We conducted all statistical analyses using R version 4.2.1 software. T-test was utilized for two-group comparison, and Kruskal-Wallis test for three-group comparison. Patients were stratified into high-risk and low-risk groups based on the median of their respective risk scores. We utilized the KM analysis method to assess OS, and the Spearman correlation coefficient method to analyze the correlation. Cox regression and Lasso regression analysis were employed to screen the genes that were significantly associated with the survival and prognosis of the patients. Significance was defined as a p-value less than 0.05, with *, **, and *** denoting p-values less than 0.05, 0.01, and 0.001, respectively.

3. Results

In this study, totals of 539 LUAD samples and 59 normal ones from TCGA data were utilized as the training group, and the GEO datasets GSE30219, GSE72094, and GSE68465 were employed as the validation group. The flowchart of this study is shown in Fig. 1.
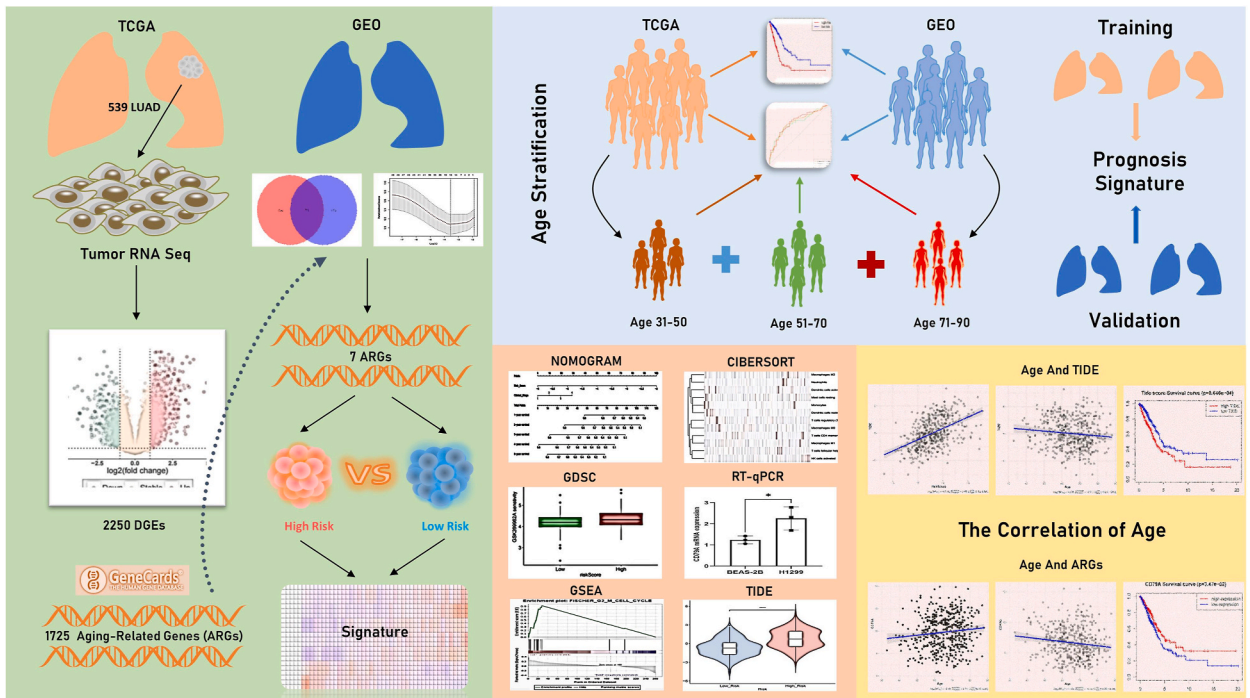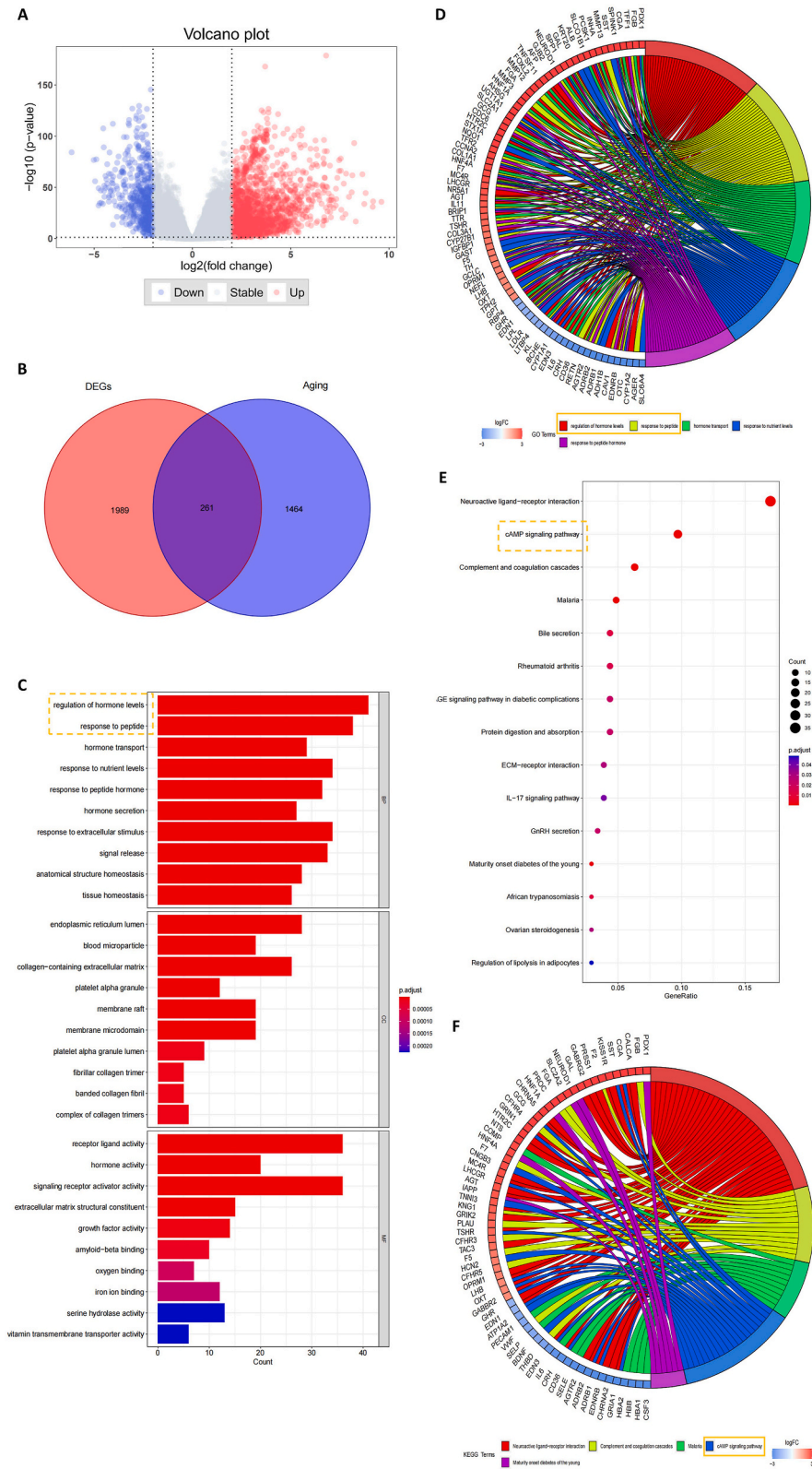

Fig. 1. The workflow diagram of this study.

Fig. 2. Identification and enrichment analysis of differentially expressed ARGs. (A) Volcano plot of DEGs in the TCGA LUAD database; (B) Venn graph illustrating the count of DEGs connected to ARGs; (C) The barplot of GO analysis including BP, CC and MF; (D) The circle map of the GO analysis; (E) KEGG analysis; (F) The circle map of the KEGG analysis.

3.1. Determination of differentially expressed ARGs and potential molecular pathway analysis of candidate genes

A total of 2250 DEGs were identified after analysis of the gene expression data of LUAD samples and corresponding normal tissues in TCGA cohort, comprising of 1681 up-regulated genes and 569 down-regulated genes, which were effectively displayed in the volcano plot (Fig. 2A). Subsequently, we intersected these DEGs with 1725 ARGs obtained from genecards (Supplementary Data), resulting in 261 candidate genes for further analysis, as depicted in the Venn graph (Fig. 2B).

Further, GO and KEGG analysis were performed on the 261 candidate DEGs. In the biological process (BP) analysis, these genes were found to enrich in the regulation of hormone levels, response to peptide (Fig. 2C). At the same time, through the circle diagram, we can see which genes are enriched in these pathways (Fig. 2D). Regarding to KEGG analysis, these genes were found to enrich in the cAMP signaling pathway (Fig. 2E). Meanwhile, through the circle diagram, we can see which genes are enriched in these pathways (Fig. 2F).

3.2. Construction and validation of a prognostic gene signature in TCGA training cohorts

To construct a prognostic gene signature, we first obtained 49 candidate genes through univariate Cox regression analysis. Subsequently, we applied LASSO regression analysis to these genes and selected the best λ coefficient, resulting in 15 differential genes (Fig. 3A & B). Finally, multivariate Cox analysis identified 7 prognostic genes, including BMP15, CD79A, CDKN3, CDX2, COL1A1, DKK1, and GRIK2, which were plotted on a forest map (Fig. 3C).

Further, the risk coefficients of these 7 genes were obtained through multivariate Cox regression analysis, they were -1.811558484, -0.133242782, 0.168764416, 0.177989831, 0.152898849, 0.158726098, and 0.139504727 for BMP15, CD79A, CDKN3, CDX2, COL1A1, DKK1, and GRIK2, respectively. Accordingly, we calculated the risk score using the following formula: risk score = (expression level of BMP15 $\times$ -1.811558484) + (expression level of CD79A $\times$ -0.133242782) + (expression level of CDKN3 $\times$ 0.168764416) + (expression level of CDX2 $\times$ 0.177989831) + (expression level of COL1A1 $\times$ 0.152898849) + (expression level of DKK1 $\times$ 0.158726098) + (expression level of GRIK2 $\times$ 0.139504727). The essential information about these 7 genes were summarized in Supplementary Table S3.
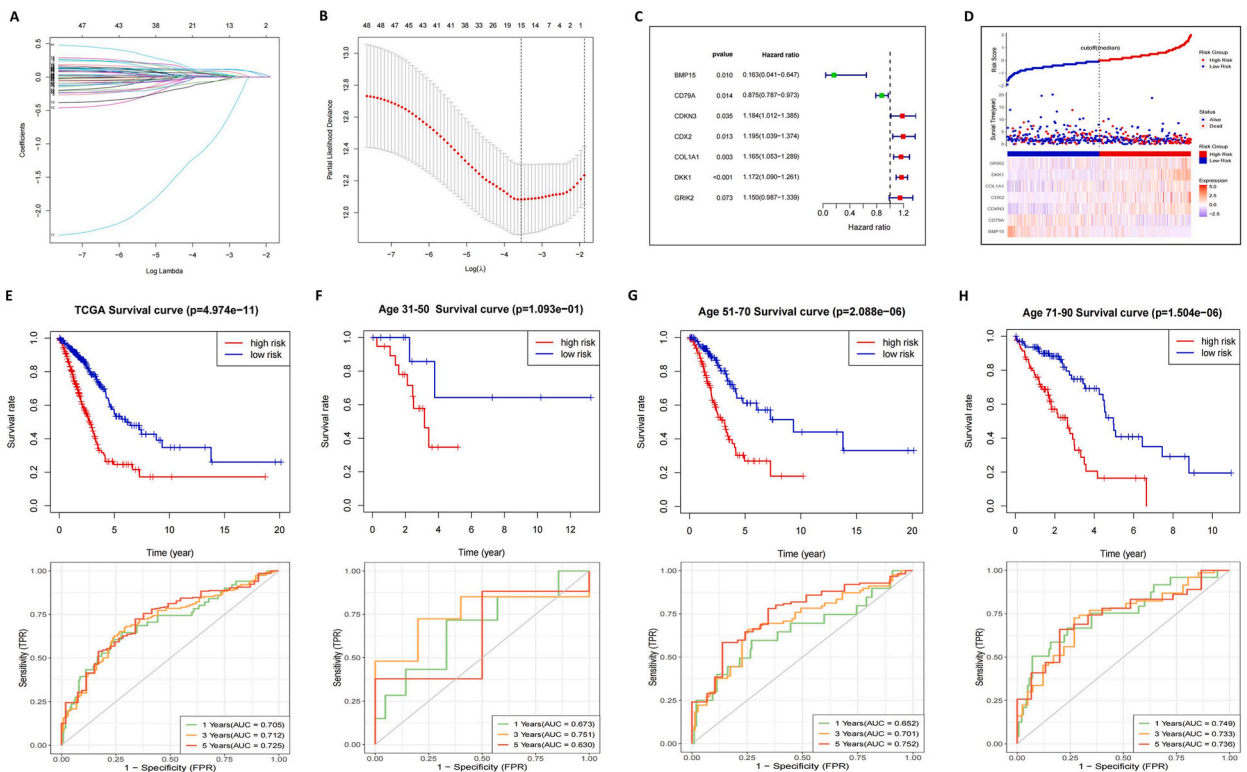

Fig. 3. Construction of a prognostic gene signature in TCGA training cohorts. (A) LASSO coefficient curves for 15 genes in TCGA LUAD; (B) The optimal parameters (λ) were used for screening of prognostic gene signatures in TCGA LUAD by LASSO regression analysis; (C) Gene signatures of 7 ARGs were screened out by multivariate Cox analysis, and depicted in the forest map; (D) Distribution of risk scores, survival time, survival status, as well as gene expression heat map of the 7 prognostic genes between high-risk and low-risk patients in TCGA LUAD database; (E–H) The KM survival curve (Up) and AUC area (Down) of the low-risk and high-risk group in TCGA LUAD in total (E), and in patients aged 31–50 (F), 51–70 (G), and 71–90 years old (H), respectively.

Accordingly, based on the median risk score, patients were divided into high-risk and low-risk groups for TCGA LUAD, and in turn plotted their survival time and survival status, and also created a gene expression heat map of the 7 prognostic genes (Fig. 3D). Furthermore, the KM survival curve of the low-risk and high-risk groups was made and found that the survival probability of the low-risk group was significantly higher than that of the high-risk one. Meanwhile, the AUC areas of the 1-, 3- and 5-year ROC curve were also plotted, and results in 0.705, 0.712, and 0.725, respectively, indicating high accuracy to the prognostic model (Fig. 3E).

Then, we employed the GSEA software to scrutinize the enrichment pathways in both high-risk groups and low-risk ones. Our findings indicated that the high-risk group manifested significant enrichment in G2/M cell cycle, RNA Polymerase II Transcription, RhoA-induced transformation, and cycling genes (Supplementary Figs. S1A–S1D), respectively. Conversely, the low-risk group exhibited substantial enrichment in dominant-negative KRAS, metabolism of lipids, PI3K/Akt signaling pathway, and RB1 and TP53 targets (Supplementary Figs. S1E–S1H), respectively.

To investigate the effect of age on survival, the KM survival curves and AUC areas were plotted for 31–50 years, 51–70 years, and 71–90 years, respectively. We found that there was no statistically significant difference in survival time of patients aged 31–50 years old ($p > 0.05$) (Fig. 3F), while the 1-, 3-, and 5-year AUC areas of patients aged 51–70 were 0.605, 0.701, and 0.752, respectively (Fig. 3G), and those of patients aged 71–90 years old were 0.749, 0.73, and 0.736, respectively (Fig. 3H). These results suggested that the prognostic model performed better for individuals over 51 years old, particularly for those aged 71–90 years old.

To validate the accuracy of the prognostic model, we applied the same formula to stratify patients from GSE30219 and GSE72094 into high-risk and low-risk groups. We plotted scatter plots of the survival time, survival status, and corresponding heat maps of the two validation sets (Supplementary Figs. S2A and S3A). We also generated KM survival curves for the validation sets and found that the
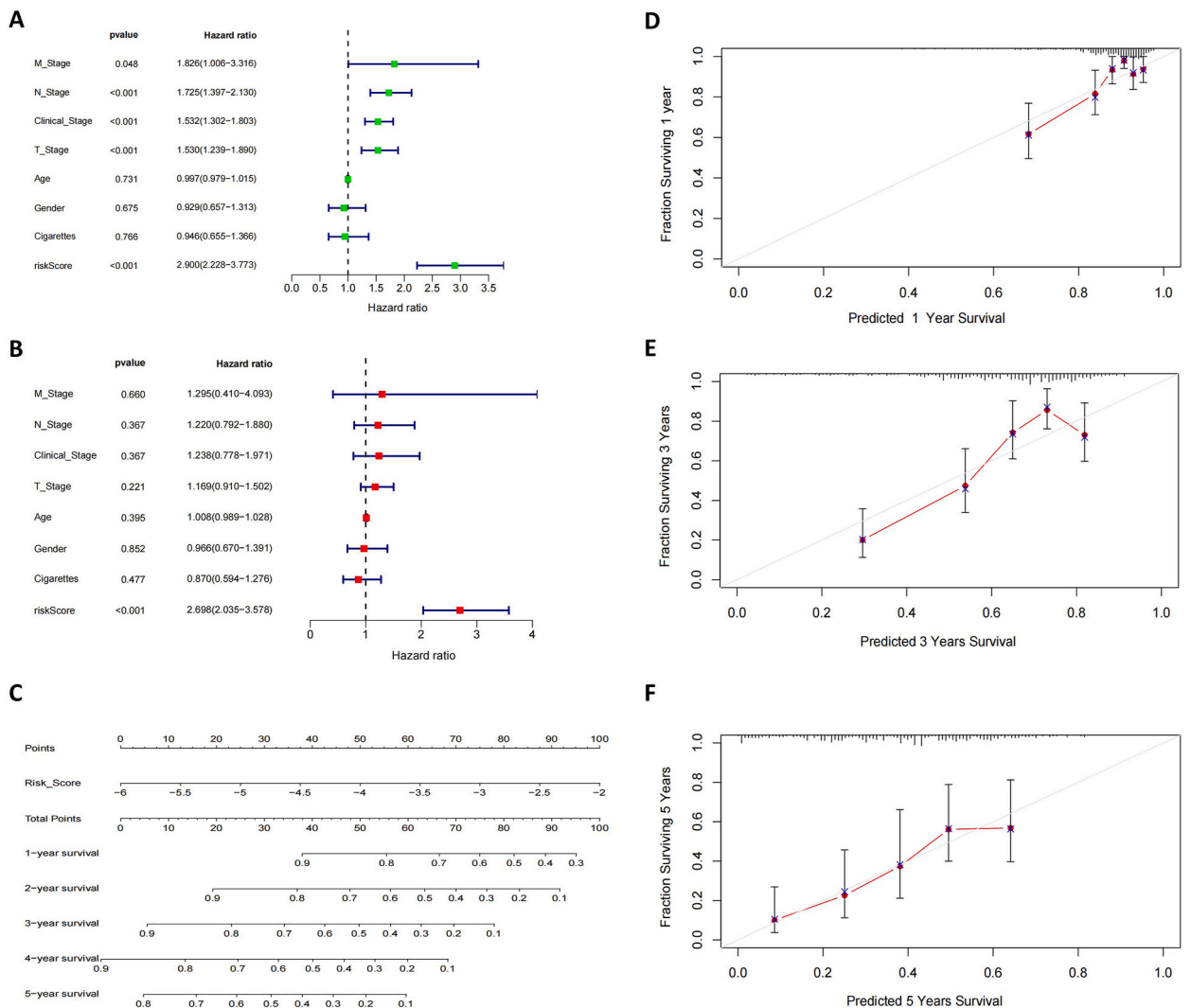

Fig. 4. Establishment of prognostic nomogram. (A–B) The results of univariate (A) and multivariate (B) COX regression analysis of clinical parameters; (C) The nomogram to predict the five-year survival probability of patients; (D–F) The calibration curves of 1-year (D), 3-year (E) and 5-year (F) survival, respectively.

survival probability of the low-risk group was higher than that of the high-risk group. In the GSE30219 validation set, the 1-, 3-, and 5-year AUC areas were 0.662, 0.698, and 0.682, respectively (Supplementary Fig. S2B). Similarly, we found that it also performed well in the GSE72094 validation set (Supplementary Fig. S3B). These findings indicated that the prognostic model was of certain predictive value in the validation sets.

Moreover, we divided patients by age to further investigate the prognostic impact of different age groups in the validation sets. The survival probability of the low-risk group aged 51–70 and 71–90 years old was significantly higher than that of the high-risk one. In the GSE30219 validation set, the 1-, 3-, and 5-year AUC areas for patients aged 51–70 years old were 0.582, 0.643, and 0.632, respectively (Supplementary Fig. S2C). For patients aged 71–90 years old, the 1-, 3-, and 5-year AUC areas were 0.622, 0.640, and 0.639, respectively (Supplementary Fig. S2D). Similarly, in the GSE72094 validation set, the 1-, 3-, and 5-year AUC areas for patients aged 51–70 years old were 0.530, 0.625, and 0.496, respectively (Supplementary Fig. S3C). For patients aged 71–90 years old, the 1-, 3-, and 5-year AUC areas were 0.721, 0.608, and 0.663, respectively (Supplementary Fig. S3D). However, the survival curves and ROC areas for patients aged 31–50 years old were not significant ($p > 0.05$) (Supplementary Figs. S2E and S3E). Given the small data size of these two 31–50 samples, we conducted another analysis using data from LUAD database (GSE68465), in which we identified 42 patients who met the study criteria. Our survival analysis revealed a statistically significant improvement in the probability of survival in the low-risk group compared to the high-risk group. Additionally, our ROC curve analysis demonstrated higher AUC values, with 1-, 3-, and 5-year AUC areas of 0.914, 0.756, and 0.598, respectively, in patients aged 31–50 years old (Supplementary Fig. S4A). Together, these results suggested that the prognostic model performed better for patients aged 71–90 years old in both validation sets.

Additionally, we also found that the 71–90 age group had the highest ssGSEA scores for ARGs, although there was no statistically significant difference. However, this suggests that ARGs may have potential relevance in the elderly population (Supplementary Fig. S4B). When comparing the distribution of risk scores across different age groups, we also discovered that the 71–90 age group exhibited the lowest median overall risk score (Supplementary Fig. S4C).
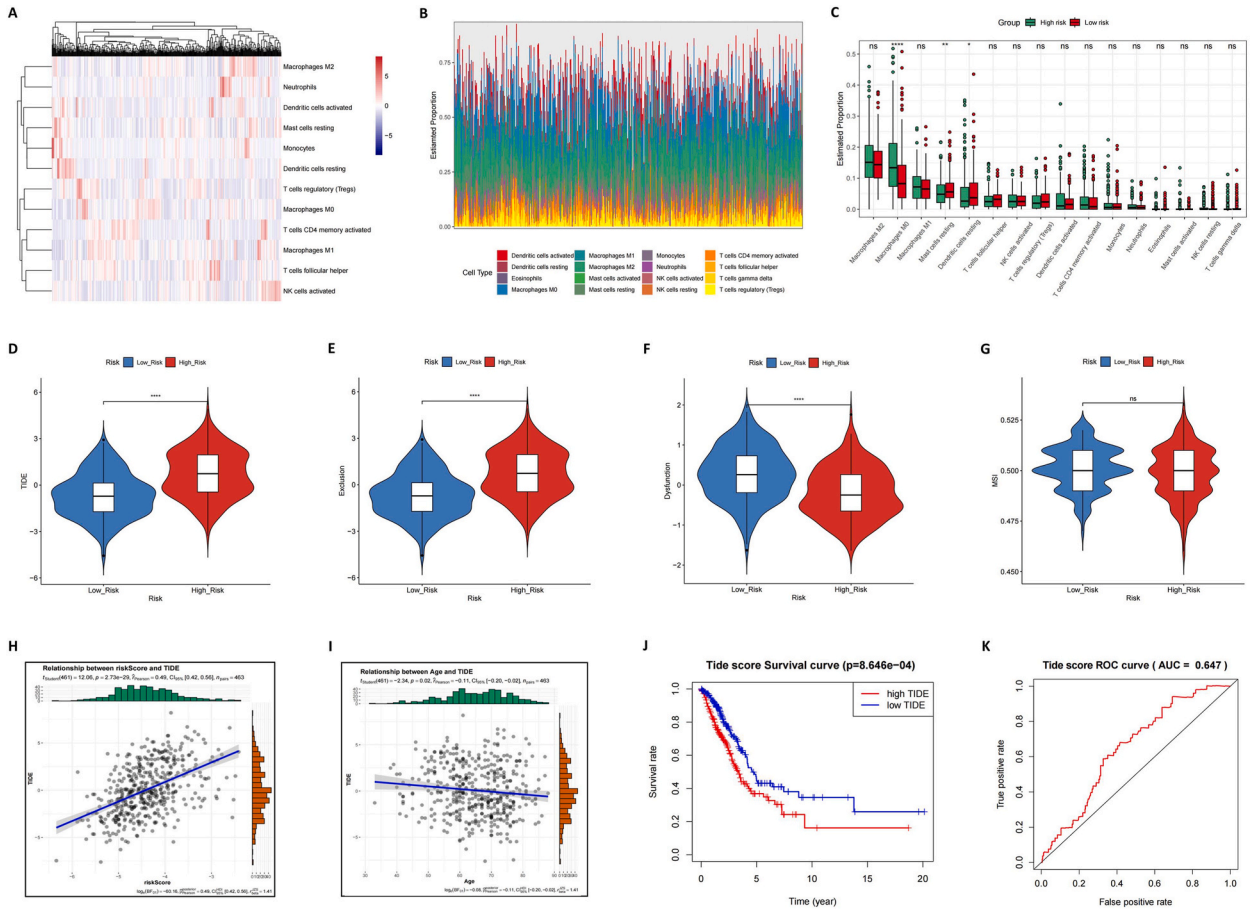
Fig. 5. CIBERSORT and TIDE analysis. (A–B) CIBERSORT analysis resulted in the heatmap (A) and histogram (B) of immune cell estimated proportion in all patients; (C) The box plot of immune cell estimated proportion in the low-risk group and the high-risk one; (D–G) TIDE score (D), Exclusion (E), Dysfunction (F), and MSI (G) between the high-risk groups and the low-risk ones were shown by Wilcoxon test; (H) Correlation risk score with TIDE score; (I) Correlation patient age with TIDE; (J) The survival curve of patients between low and high TIDE score in total; (K) The corresponding ROC curves to patients in total.

We also noticed that when we performed a survival analysis of seven genes in different age groups, there were no statistically significant differences for most of the single genes, such that CDKN3, DKK1, CDX2, COL1A1 and GRIK2 did not show a statistically significant association with patient prognosis in the 31–50 age group, while BMP15 did not show a significant association in all groups (Supplementary Figs. S4D–I). However, when these seven genes were analyzed in combination, the prognostic model demonstrated a statistically significant difference in patient survival (Fig. 3E–H). This observation might indicate that prognostic models incorporating multiple genes could have greater predictive power than those based on a single gene. Collectively, these seven genes play an important role in the survival of patients.

3.3. Establishment of a nomogram to predict the prognosis of patients with 5-year survival probability

We performed both univariate and multivariate Cox regression analyses on various clinical parameters, revealing risk score as a notable independent clinical prognostic parameter (Fig. 4A and 4B). Additionally, we constructed a nomogram to predict the five-year survival probability of patients (Fig. 4C), which was further validated by calibration curves for 1-, 3-, and 5-year survival (Fig. 4D, E & 4F). These curves demonstrated a close match between the predicted survival probabilities and the actual ones. Furthermore, our C-index calculation yielded a value of 0.737, indicating high accuracy of our prognostic nomogram model.
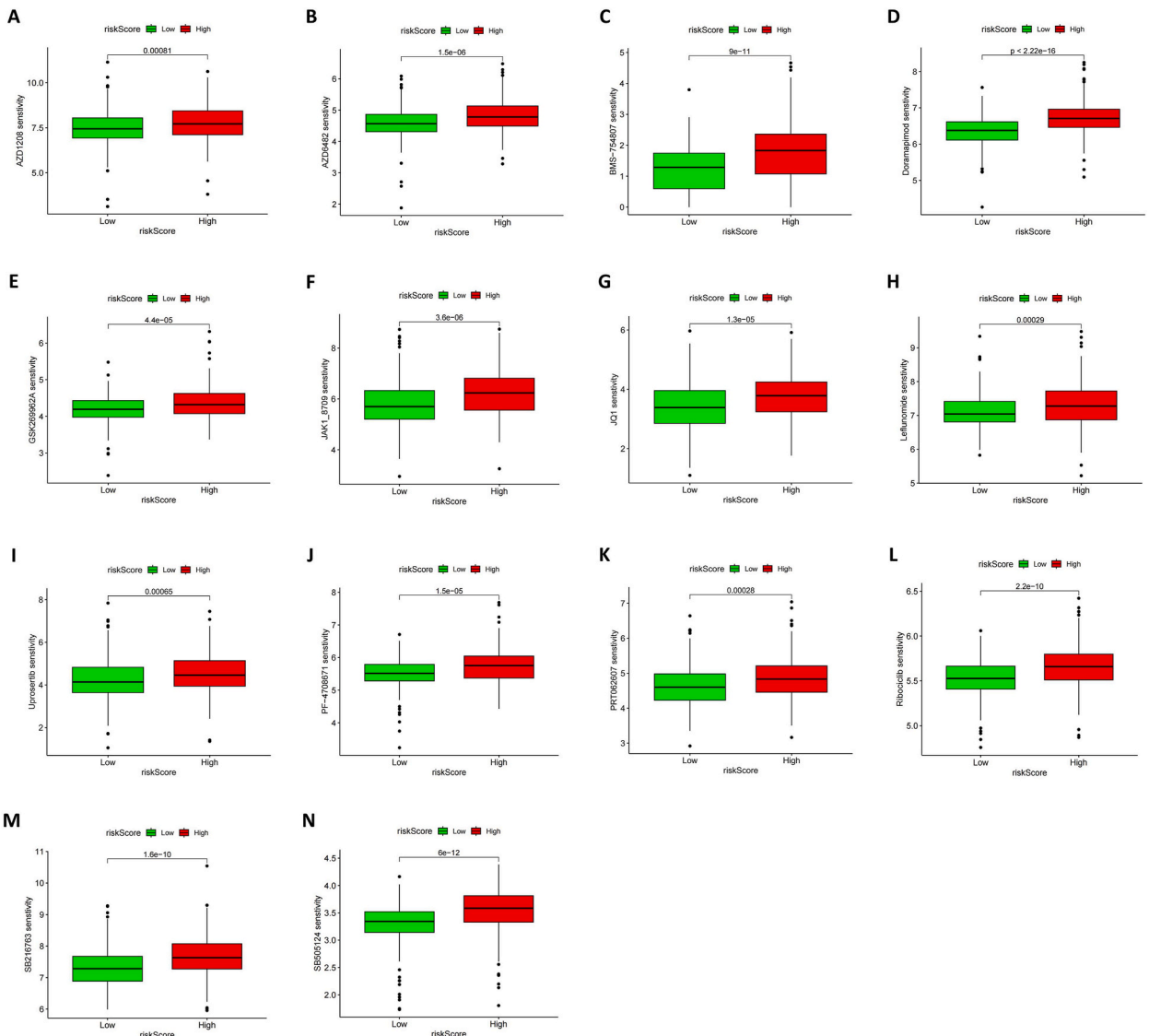

Fig. 6. Potential highly sensitive chemotherapeutic drugs to patients in high-risk group. (A–N) 14 highly sensitive chemotherapeutic drugs of AZD1208 (A), AZD6482 (B), BMS-754807 (C), Doramapimod (D), GSK269962A (E), JAK1_8709 (F), JQ1 (G), Leflunomide (H), Uprosertib (I), PF-4708671 (J), PRT062607 (K), Ribociclib (L), SB216763 (M), and SB505124 (N) were identified.

3.4. Potential benefits of immune checkpoint inhibitor (ICI) therapy in patients from different risk groups

Furthermore, we performed CIBERSORT analysis to explore the differences in immune cell classification between high-risk groups and low-risk ones. The heatmap and histogram of immune cell estimated proportion for all patients were shown (Fig. 5A and 5B). The box plot showed a significant reduction of macrophages M0, and upregulation of resting mast cells and dendritic cells in the low-risk group compared to that in the high-risk one (Fig. 5C).

We further evaluated the potential clinical effectiveness of immunotherapy and chemotherapy in patients from different risk groups. As known, TIDE score is considered to be a comprehensive indicator of the potential efficacy of immunotherapy, and the lower the score, the more suitable for immunotherapy. Our analysis revealed that both the TIDE Score and Exclusion Score were lower in the low-risk group compared to the high-risk group, suggesting a potential for enhanced immunotherapeutic effectiveness (Fig. 5D and 5E). Despite observing a higher Dysfunction Score in the low-risk group and no discernible difference in microsatellite instability (MSI) scores (Fig. 5F and 5G). Collectively, these findings suggested an increased sensitivity to immune checkpoint inhibitor (ICI) therapy in the low-risk group of LUAD patients.

We further correlated patient age with TIDE, and revealed that the risk score is positively associated with TIDE, indicating that a lower risk score corresponds to a lower TIDE score (Fig. 5H). Further analysis showed a negative correlation between patient age and TIDE, with TIDE scores declining gradually as patient age increased (Fig. 5I). To investigate the difference in survival time between high and low TIDE scores, we plotted corresponding survival curves and discovered that patients with low TIDE scores had longer survival times compared to those with high TIDE scores (Fig. 5J). The AUC value of patients in total was 0.647 (Fig. 5K). We further assessed the distribution of TIDE scores across different age groups, and found a trend toward lower median TIDE scores in the 51–70 group and the 71–90 group (Supplementary Fig. S4J). This implies that there may be a potential for immunotherapy responsiveness in these elderly populations based on low-risk score group.

3.5. Potential highly sensitive chemotherapeutic drugs to patients in different risk groups

After conducting "oncoPredict" R package analysis of our prognostic model, we obtained a list of chemotherapeutic drugs with $P < 0.05$. Among the high-risk group, we identified 14 highly sensitive chemotherapeutic drugs, including AZD1208, AZD6482, BMS-754807, Doramapimod, GSK269962A, JAK1_8709, JQ1, Leflunomide, Uprosertib, PRT062607, Ribociclib, SB216763, SB505124, and PF-4708671 (Fig. 6). Meanwhile, among the low-risk group, we identified 21 highly sensitive chemotherapeutic drugs, including 5-Fluorouracil, Alisertib, AZD6738, AZD7762, BDP-00009066, BI-2536, BMS-536924, BPD-00009900, Cediranib, ERK_6604, Foretinib, GDC0810, PD0325901, Savolitinib, SCH7 72984, Trametinib, Ulixertinib, VE821, VX-11e, Dasatinib, and YK-4-279 (Supplementary Fig. S5).
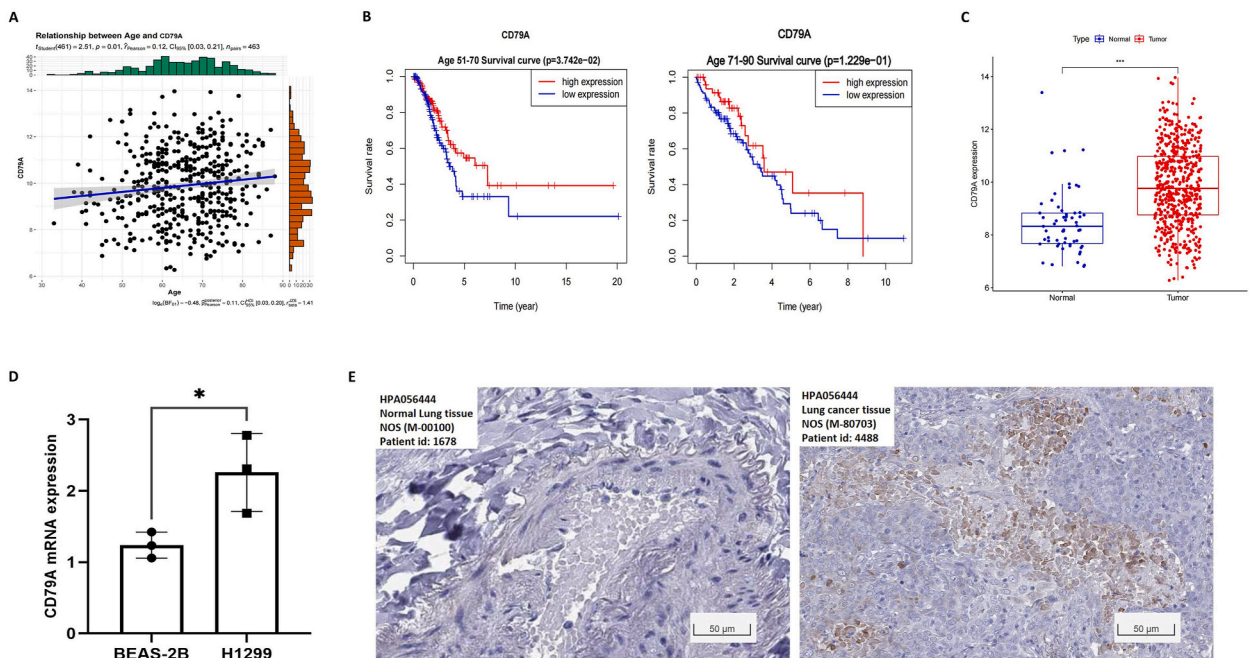
Fig. 7. External experimental validation results for key gene CD79A. (A) The Correlation between CD79A and age; (B) The KM survival curve of patient aged 51–70 (Left), and 71–90 years old (Right) to different expression levels of CD79A; (C) The differential expression of CD79A in the TCGA LUAD database; (D) The CD79A expression in H1299 and BEAS-2B by RT-qPCR results. (E) The CD79A expression in lung normal tissue (Left) and lung cancer tissue (Right) in The Human Protein Atlas.

3.6. External experimental validation results for key gene CD79A

We would further figure out the relationship between these seven ARGs and the age of the patients, and found that there was statistical significance (P < 0.05) between patient age and aging-related gene CD79A (Fig. 7A), especially the high expression of CD79A in elderly patients (the 51–70 and 71–90 age group) (Supplementary Fig. S4K). In addition, the survival curve analyses also showed that a longer probability of survival was associated with the higher expression of CD79A in the 51–70 age group (Fig. 7B). As for the 71–90 age group, there is a trend towards improved survival for those with high CD79A expression, although there is no statistical significance (Fig. 7B). However, when CD79A were combined with other six genes, it was shown a statistically significant difference in patient survival in the 71–90 age group (Fig. 3H). Moreover, we analyzed the TCGA LUAD database and found that the expression of CD79A in cancer tissues was higher than that in normal tissues (Fig. 7C), and the external RT-qPCR validation experiment also showed that the expression of CD79A in H1299 was higher than that in BEAS-2B (Fig. 7D). Finally, the published immunohistochemical data from The Human Protein Atlas also confirmed that CD79A expression was also increased in lung cancer tissues compared to lung normal ones (Fig. 7E). These results indicated that CD79A might have significant implications in our prognostic model.

4. Discussion

Currently, the incidence rate of LUAD has surged to an alarming level, and the conventional treatment methods such as radiotherapy, chemotherapy, and surgery have demonstrated insufficient efficacy, with poor overall outcomes. This inadequacy can be attributed to the belated diagnosis of LUAD, which often results in a missed window for optimal treatment via surgical resection [24]. As we mentioned earlier, the incidence of lung cancer accounts for one of the leading causes of death, especially among the death over age of 50. Therefore, LUAD in the elderly demands comprehensive attention. It is imperative to identify novel aging genetic biomarkers that can accurately predict the prognosis of patients with LUAD in the geriatric population.

In this study, we employed a rigorous approach to identify potential ARGs by selecting differentially expressed genes from TCGA LUAD database. Our analysis yielded a total of 261 genes that exhibited differential expression patterns in ARGs. These ARGs were supposed to be involved in the regulation of hormone levels and response to peptide by GO analysis. As previous studies have demonstrated that surface polypeptide hormones play a critical role in the regulation of DNA damage repair and the DNA damage response (DDR), which in turn can impact tumor development [25]. Interestingly, DNA damage, is also one of the hallmarks of aging [4]. Moreover, KEGG analysis revealed that these ARGs were primarily enriched in the cAMP signaling pathways. Specifically, cAMP signaling was shown to enhance cisplatin-induced apoptosis in lung cancer cells by increasing HDAC8 expression through the protein kinase A (PKA) pathway [26]. Furthermore, GSEA analysis found that G2/M cell cycle and RNA polymerase II transcription were notably enriched in high-risk groups. Actually, an increasing number of drugs were used to inhibit tumors by inducing G2/M cell cycle arrest. For instance, ovatodiolide could ultimately lead to cell cycle arrest in the G2/M phase and resulted in apoptosis in lung cancer cells [27]. Additionally, some studies suggested that human RNA polymerase II-related factor-1 complex promoted lung cancer cell proliferation by regulating c-MYC transcription [28]. In the low-risk group, on the other hand, PI3K/Akt signaling pathway was observed to be increased and lung cancer KRAS decreased. Abnormal activation of PI3K/Akt pathway is widely believed to be related to tumorigenesis and development, and is involved in immune checkpoint regulation and immune checkpoint inhibitor sensitivity [29]. KRAS mutations are present in 35 % of LUAD, and several KRASG12C inhibitors, including sotorasib and adagrasib, are currently in development and show great therapeutic potentials [30].

Of note, some ARGs have been identified in LUAD by several research groups. Xu Q et al. reported that the ARG signature based on 7 genes could be used for risk stratification and prognosis prediction in LUAD [7]. Feng C et al. recently identified 17 genes as aging-related molecular subtype and prognostic signature in LUAD [8]. Zhang Q et al. have also recently identified 21 genes for prognosis and prediction of immunotherapy response [9]. However, these studies did not take into account the potential relationship between ARGs and patient age. Thereupon, we performed univariate, LASSO, and multivariate Cox regression analysis of aforementioned 261 genes, and finally identified 7 seven prognostic genes, including BMP15, CD79A, CDKN3, CDX2, COL1A1, DKK1, and GRIK2. These 7 genes obviously did not overlap with the aforesaid ARGs, indicating the independence of these prognostic genes. Additionally, we validated the key genes through in-vitro experiments.

According to the median of risk score, we divided the patients into high-risk and low-risk groups and plotted the survival curves, and found that the survival probability of the low-risk group was significantly higher than that of the high-risk one. Further, we explored the effect of patient age on survival probability, and noticed that this prognostic model showed the most significant effect on the survival probability of patients aged 71–90 years old. These findings were further validated in two GEO datasets of GSE30219 and GSE72094, suggesting some potential internal relationship between ARGs and the elderly patients of LUAD.

We further analyzed immune cell infiltration and TIDE scores of immune checkpoint inhibitors in both the high-risk and low-risk groups to explore their clinical treatment sensitivity. In terms of immune cell infiltration distribution, CIBERSORT analysis demonstrated that the number of M0 macrophages was higher in the high-risk group than in the low-risk one. Several studies revealed that an increase in M0 macrophages was found to be associated with poor early clinical prognosis in LUAD [31]. In contrast, resting mast cells and dendritic cells were more prevalent in the low-risk group, which is consistent with some previous studies that have suggested that resting mast cells and resting dendritic cells exhibited strong prognostic ability in LUAD [32]. Besides, TIDE analysis suggested that the low-risk group was more likely to benefit from ICI therapy. Furthermore, we also revealed a positive relationship between risk scores and TIDE, with lower TIDE scores recorded in low-risk group. Previous studies proved that TIDE score was a good predictor of ICB treatment, and a higher tumor TIDE predictive score was not only associated with poorer ICB response, but also with poorer survival

rate in patients receiving anti-PD1 and anti-CTLA4 therapy [33]. Meanwhile, we scrutinized the link between patient age and the seven pivotal prognostic ARGs, and found that CD79A were positively correlated with age. We also confirmed the high expression of CD79A in tumor tissues by in-vitro experiments. Interestingly, a distinct investigation identified that individuals expressing CD79A in classical Hodgkin's lymphoma tended to exhibit an older age of disease onset, with the age range clustering between 52 and 81 years old [34].

Moreover, we also identified distinct sets of responsive chemotherapeutic drugs via GDSC analysis. Among these, 14 potent agents were detected in the high-risk cohort whilst 21 efficacious drugs were identified within the low-risk group. These identified agents provided invaluable insights towards guiding clinical chemotherapy and treatment decisions.

Although we believe our prognostic model for predicting the survival rate of elderly patients is more accurate, it does have certain limitations. Firstly, while we did establish some association between ARGs and patient age in our prognostic model, the underlying mechanism remains unknown. Secondly, we did not analyze other type of cancers other than LUAD. These explorations are highly needed in the future. Thirdly, our study revealed variation in the effectiveness of prognostic models across different age groups, suggesting that models developed regardless of age may not adequately address the needs of all populations. Other prognostic models should be implemented in future clinical trials. Lastly, due to the limitations of the data available, our study did not include population studies of individuals under 30 years of age, suggesting the need for further research in this population.

5. Conclusion

In summary, we developed a robust prognostic model for aging genes based on the TCGA database. Notably, our model demonstrated exceptional prediction accuracy for patients aged 71–90 years old. Moreover, novel therapeutic strategies for immunotherapy and chemotherapy were proposed for the elderly patients with LUAD.

Ethics approval

Not applicable.

Data availability statement

All data collected and analyzed in this study can be downloaded from public databases including the Cancer Genome Atlas (https://portal.gdc.cancer.gov/), the Gene Expression Omnibus (https://www.ncbi.nlm.nih.gov/geo/). The data supporting the findings of this study are given in the main manuscript and supplementary files. All data generated during this study are available from the corresponding author on reasonable request.

CRediT authorship contribution statement

Yi Tian: Writing – original draft, Conceptualization. **Wenya Zhao:** Writing – review & editing. **Chenjing Lin:** Visualization, Investigation. **Yang Chen:** Software, Resources. **Qiaoxin Lin:** Formal analysis. **Yiru Liu:** Validation. **Dianna Gu:** Project administration. **Ling Tian:** Supervision, Methodology, Funding acquisition.

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.

Appendix A. Supplementary data

Supplementary data to this article can be found online at https://doi.org/10.1016/j.heliyon.2024.e33268.

References

[1] H. Sung, J. Ferlay, R.L. Siegel, et al., Global cancer statistics 2020: GLOBOCAN estimates of incidence and mortality worldwide for 36 cancers in 185 countries, Ca - Cancer J. Clin. 71 (2021) 209–249, https://doi.org/10.3322/caac.21660.
[2] A.A. Thai, B.J. Solomon, L.V. Sequist, et al., Lung cancer, Lancet 398 (2021) 535–554, https://doi.org/10.1016/S0140-6736(21)00312-3.
[3] D.S. Ettinger, D.E. Wood, D.L. Aisner, et al., Non-small cell lung cancer, version 3.2022, NCCN clinical practice guidelines in oncology, J. Natl. Compr. Cancer Netw. 20 (2022) 497–530, https://doi.org/10.6004/jnccn.2022.0025.
[4] C. López-Otín, M.A. Blasco, L. Partridge, et al., Hallmarks of aging: an expanding universe, Cell 186 (2023) 243–278, https://doi.org/10.1016/j.cell.2022.11.001.

[5] S. Terracina, G. Ferraguti, C. Petrella, et al., Characteristic hallmarks of aging and the impact on carcinogenesis, Curr. Cancer Drug Targets 23 (2023) ·doi, https://doi.org/10.2174/1568009622666220816120353.

[6] R.L. Siegel, K.D. Miller, N.S. Wagle, et al., Cancer statistics, 2023, Ca - Cancer J. Clin. 73 (2023) 17–48, https://doi.org/10.3322/caac.21763.

[7] Q. Xu, Y. Chen, An aging-related gene signature-based model for risk stratification and prognosis prediction in lung adenocarcinoma, Front. Cell Dev. Biol. 9 (2021) 685379, https://doi.org/10.3389/fcell.2021.685379.

[8] C. Feng, W. Che, H. Liang, et al., Mining database to identify aging-related molecular subtype and prognostic signature in lung adenocarcinoma, JAMA Oncol. 2022 (2022) 9142903, https://doi.org/10.1155/2022/9142903.

[9] W. Zhang, Y. Li, J. Lyu, et al., An aging-related signature predicts favorable outcome and immunogenicity in lung adenocarcinoma, Cancer Sci. 113 (2022) 891–903, https://doi.org/10.1111/cas.15254.

[10] M.I. Love, W. Huber, S. Anders, Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2, Genome Biol. 15 (2014) 550.

[11] S.-M. Zhang, W.L. Cai, X. Liu, et al., KDM5B promotes immune evasion by recruiting SETDB1 to silence retroelements, Nature 598 (2021) 682–687, https://doi.org/10.1038/s41586-021-03994-2.

[12] G. Stelzer, N. Rosen, I. Plaschkes, et al., The GeneCards suite: from gene data mining to disease genome sequence analyses, Curr Protoc Bioinformatics 54 (2016) ·doi, https://doi.org/10.1002/cpbi.5.

[13] G. Yu, L.-G. Wang, Y. Han, et al., clusterProfiler: an R package for comparing biological themes among gene clusters, OMICS 16 (2012) 284–287, https://doi.org/10.1089/omi.2011.0118.

[14] T. Wu, E. Hu, S. Xu, et al., clusterProfiler 4.0: a universal enrichment tool for interpreting omics data, Innovation 2 (2021) 100141, https://doi.org/10.1016/j.xinn.2021.100141.

[15] A. Subramanian, P. Tamayo, V.K. Mootha, et al., Gene set enrichment analysis: a knowledge-based approach for interpreting genome-wide expression profiles, Proc. Natl. Acad. Sci. U. S. A. 102 (2005) 15545–15550.

[16] K. Nagashima, Y. Sato, Information criteria for Firth's penalized partial likelihood approach in Cox regression models, Stat. Med. 36 (2017) 3422–3436, https://doi.org/10.1002/sim.7368.

[17] Q. Zhao, X. Shi, Y. Xie, et al., Combining multidimensional genomic measurements for predicting cancer prognosis: observations from TCGA, Briefings Bioinf. 16 (2015) 291–303, https://doi.org/10.1093/bib/bbu003.

[18] P. Blanche, J.-F. Dartigues, H. Jacqmin-Gadda, Estimating and comparing time-dependent areas under receiver operating characteristic curves for censored event times with competing risks, Stat. Med. 32 (2013) 5381–5397, https://doi.org/10.1002/sim.5958.

[19] P. Schober, C. Boer, L.A. Schwarte, Correlation coefficients: appropriate use and interpretation, Anesth. Analg. 126 (2018) 1763–1768, https://doi.org/10.1213/ANE.0000000000002864.

[20] A. Iasonos, D. Schrag, G.V. Raj, et al., How to build and interpret a nomogram for cancer prognosis, J. Clin. Oncol. 26 (2008) 1364–1370, https://doi.org/10.1200/JCO.2007.12.9791.

[21] B. Chen, M.S. Khodadoust, C.L. Liu, et al., Profiling tumor infiltrating immune cells with CIBERSORT, Methods Mol. Biol. 1711 (2018) 243–259, https://doi.org/10.1007/978-1-4939-7493-1_12.

[22] J. Fu, K. Li, W. Zhang, et al., Large-scale public data reuse to model immunotherapy response and resistance, Genome Med. 12 (2020) 21, https://doi.org/10.1186/s13073-020-0721-z.

[23] W. Yang, J. Soares, P. Greninger, et al., Genomics of Drug Sensitivity in Cancer (GDSC): a resource for therapeutic biomarker discovery in cancer cells, Nucleic Acids Res. 41 (2013) D955–D961, https://doi.org/10.1093/nar/gks1111.

[24] G.-G. Akhtar-Danesh, C. Finley, H.Y. Seow, et al., Change in treatment modality and trends in survival among stage I non-small cell lung cancer patients: a population-based study, J. Thorac. Dis. 12 (2020) 4670–4679, https://doi.org/10.21037/jtd-20-1387.

[25] V. Chesnokova, S. Melmed, Peptide hormone regulation of DNA damage responses, Endocr. Rev. 41 (2020) ·doi, https://doi.org/10.1210/endrev/bnaa009.

[26] J.-Y. Park, Y.-S. Juhnn, cAMP signaling increases histone deacetylase 8 expression via the Epac2-Rap1A-Akt pathway in H1299 lung cancer cells, Exp. Mol. Med. 49 (2017) e297.·doi, https://doi.org/10.1038/emm.2016.152.

[27] C.-Y. Yu, C.-L. Jerry Teng, P.-S. Hung, et al., Ovatodiolide isolated from Anisomeles indica induces cell cycle G2/M arrest and apoptosis via a ROS-dependent ATM/ATR signaling pathways, Eur. J. Pharmacol. 819 (2018) 16–29, https://doi.org/10.1016/j.ejphar.2017.09.050.

[28] X. Zhi, E. Giroux-Leprieur, M. Wislez, et al., Human RNA polymerase II associated factor 1 complex promotes tumorigenesis by activating c-MYC transcription in non-small cell lung cancer, Biochem. Biophys. Res. Commun. 465 (2015) 685–690, https://doi.org/10.1016/j.bbrc.2015.08.017.

[29] Y. He, M.M. Sun, G.G. Zhang, et al., Targeting PI3K/Akt signal transduction for cancer therapy, Signal Transduct. Targeted Ther. 6 (2021) 425, https://doi.org/10.1038/s41392-021-00828-5.

[30] M. Reck, D.P. Carbone, M. Garassino, et al., Targeting KRAS in non-small-cell lung cancer: recent progress and new approaches, Ann. Oncol. 32 (2021) 1101–1110, https://doi.org/10.1016/j.annonc.2021.06.001.

[31] X. Liu, S. Wu, Y. Yang, et al., The prognostic landscape of tumor-infiltrating immune cell and immunomodulators in lung cancer, Biomed. Pharmacother. 95 (2017) 55–61, https://doi.org/10.1016/j.biopha.2017.08.003.

[32] A. Zhang, J. Yang, C. Ma, et al., Development and validation of a robust ferroptosis-related prognostic signature in lung adenocarcinoma, Front. Cell Dev. Biol. 9 (2021) 616271, https://doi.org/10.3389/fcell.2021.616271.

[33] P. Jiang, S. Gu, D. Pan, et al., Signatures of T cell dysfunction and exclusion predict cancer immunotherapy response, Nat. Med. 24 (2018) 1550–1558, https://doi.org/10.1038/s41591-018-0136-1.

[34] A. Sakatani, T. Igawa, T. Okatani, et al., Clinicopathological significance of CD79a expression in classic Hodgkin lymphoma, J. Clin. Exp. Hematop. 60 (2020) 78–86, https://doi.org/10.3960/jslrt.20010.