Original Article

The role of immunogenic cell death in the prognosis and development of treatment strategies for non-small cell lung cancer: a multiomics and machine learning approach for predictive and personalized treatment

Yichen Sun[1#], Hao Chen[1#], Zhaoyang Wang[1], Rui Jiao[1,2], Franz Zehentmayr[3], Fabrizio Tabbò[4], Chengyang Wu[1], Tao Zhang[1], Hanyu Yan[1], Jian Wang[1*], Xiaolong Yan[1*]

[1]Department of Thoracic Surgery, Tangdu Hospital, Air Force Medical University, Xi'an, China; [2]School of Medicine, Northwest University, Xi'an, China; [3]Department of Radiation Oncology, University Hospital Salzburg, Paracelsus Medical University, Salzburg, Austria; [4]Oncology Unit, Department of Medicine, Michele and Pietro Ferrero Hospital, Verduno, Italy

Contributions: (I) Conception and design: Y Sun, H Chen, Z Wang; (II) Administrative support: X Yan, J Wang; (III) Provision of study materials or patients: X Yan, J Wang; (IV) Collection and assembly of data: Y Sun, H Chen; (V) Data analysis and interpretation: Y Sun, Z Wang; (VI) Manuscript writing: All authors; (VII) Final approval of manuscript: All authors.

[#]These authors contributed equally to this work as co-first authors.

[*]These authors contributed equally to this work.

Correspondence to: Xiaolong Yan, PhD; Jian Wang, PhD. Department of Thoracic Surgery, Tangdu Hospital, Air Force Medical University, 1 Xinsi Road, Xi'an 710038, China. Email: yanxiaolong@fmmu.edu.cn; wjtm1976@163.com.

Background: Non-small cell lung cancer (NSCLC) is the predominant histological subtype of lung cancer, whose diverse genomic landscape complicates prognosis and outcome prediction. In this context, immunogenic cell death (ICD), a distinct mechanism of cell death, may exert important antitumor effects. However, the specific role of ICD in NSCLC has not been clarified, and there is no suitable method for using ICD to achieve the treatment and prognosis assessment of NSCLC. The purpose of the current research is to develop a new approach to predict the survival prognosis and response to chemotherapy and targeted therapy in patients with NSCLC.

Methods: We used 101 combinations of 10 machine learning algorithms to construct an ICD-related signature (ICDRS). The predictive potential of this specific ICDRS for immune cell infiltration and therapeutic response was evaluated. We also examined the molecular mechanisms underlying the various responses of different ICDRS subpopulations, characterized the mutational landscape and tumor mutational burden (TMB), and assessed the applicability of the ICDRS in single-cell transcriptomic datasets. Gene expression patterns were subsequently validated with quantitative real-time polymerase chain reaction (qRT-PCR) and immunohistochemistry (IHC).

Results: We screened out five key ICDRS genes (*MMP14*, *ALDH2*, *FBP1*, *HLA-DRA*, *KCTD12*). Patients were classified into high-risk and low-risk groups according to the ICDRS score determined by the expression levels of these five genes. The high-risk group exhibited less favorable prognoses compared with the low-risk group, which demonstrated more positive outcomes. The low-risk group had more anticancer immune-cell infiltration and showed better response to chemotherapy and targeted therapy than the high-risk group ($P<0.05$).

Conclusions: The ICDRS exhibited excellent predictive performance and broad applicability, suggesting it as a powerful tool for prognosis and therapy response. The current study may contribute to more adequate patient selection in the context of tailored therapies.

Keywords: Immunotherapy; immunogenic cell death (ICD); machine learning; non-small cell lung cancer (NSCLC)

Submitted Jul 03, 2025. Accepted for publication Aug 12, 2025. Published online Aug 26, 2025.
doi: 10.21037/tlcr-2025-769
View this article at: https://dx.doi.org/10.21037/tlcr-2025-769

Introduction

The incidence and mortality rates of non-small cell lung cancer (NSCLC) are persistently high, with a significant proportion of patients diagnosed at advanced stages (1). Although numerous treatment methods have emerged for this disease, including surgery, chemotherapy, radiotherapy, and targeted therapy, drug resistance and various side effects during the treatment process have hindered their efficacy (2).

Both for NSCLC and small cell lung cancer (SCLC), which are the main entities of lung cancer, immune checkpoint inhibitors (ICIs) are becoming increasingly important in daily practice, especially for patients whose tumor does not have a targetable mutation (3-5). However, the heterogeneity of mutations within NSCLC tumors and significant interpatient variability can result in diminished responsiveness to ICI therapy and the development of drug resistance (6,7). Additionally, a lack of effective early assessment strategies, late symptom presentation and therefore delayed diagnosis have limited treatment options for NSCLC (8,9).

Immunogenic cell death (ICD) can be precipitated by a spectrum of therapeutic interventions, including specific chemotherapeutic agents, oncolytic virotherapy, physicochemical modalities, photodynamic therapy, radiation therapy, and immune-mediated processes (10). When the efflux of adenosine triphosphate (ATP), immunostimulatory cytokines and others damage-associated molecular patterns (DAMPs) are stimulated in an appropriate manner, they bind antigen-presenting cells, such as dendritic cells (DCs), at the site of imminent ICD. As a result, this triggers an immunogenic response in the tumor microenvironment (TME), promoting the establishment of tumor-specific immunity (11-14). ICD is sufficient to activate adaptive immune responses in the context of normal immune function. It has gained increasing attention as a potential strategy for addressing the challenging issues of low sensitivity to ICIs and immunologically cold tumors in cancer patients (15,16). While a growing number of studies have demonstrated that promoting ICD can enhance tumor responsiveness to ICIs, the significance of ICD-related genes for NSCLC patient survival, prognosis, and response to treatment has not been fully clarified thus far (17-20).

In this study, we intend to fill this knowledge gap by integrating 101 distinct machine learning algorithms within a novel computational framework to develop an ICD-related signature (ICDRS) in order to characterize the immune contexture of NSCLC. This methodology employs the complementary strengths of multiple algorithms to improve the predictive efficacy and precision of the ICDRS within the framework of NSCLC treatment. We evaluated the robustness of the ICDRS in predicting the survival outcomes of patients with NSCLC and clarified the associated molecular mechanisms and immune infiltration profiles. Additionally, we validated the ICRDS at the genomic and single-cell transcriptomic levels and assessed its ability to predict response to chemotherapy and targeted therapy. This approach can facilitate more rational and personalized prognostic evaluation, treatment, and care for patients with NSCLC, ultimately leading to better clinical outcomes. We present this article in accordance with the TRIPOD reporting checklist (available at https://tlcr.amegroups.com/article/view/10.21037/tlcr-2025-769/rc).

Highlight box

Key findings

- An immunogenic cell death-related signature (ICDRS) was constructed, and five key marker genes that can predict the prognosis and drug response of non-small cell lung cancer (NSCLC), including *MMP14*, were identified.

What is known and what is new?

- Immunogenic cell death is an important mode of cell death and is related to the malignant progression of cancer.
- This study clarifies the association of ICDRS and related genes with the prognosis and drug response of NSCLC.

What is the implication, and what should change now?

- This effective and robust ICDRS can help clinicians evaluate the prognosis of NSCLC patients and select treatment options based on the expression levels of relevant genes

Methods

Data collection and processing

The conceptual framework of this study is illustrated in a schematic diagram (Figure S1). NSCLC-related clinical

and transcriptomic data were obtained from The Cancer Genome Atlas (TCGA) database. Furthermore, single-cell omics sample data were sourced from the Gene Expression Omnibus (GEO) database, with these samples matched according to their identifiers, genomic data, and single-cell transcriptomic information.

We employed the R-based bioinformatics package "Seurat" (The R Foundation for Statistical Computing, Vienna, Austria) to process and interrogate single-cell RNA-sequencing data. Subsequently, we used the "Find Clusters" and "Find Neighbors" functions within the "Seurat" package to perform dimensionality reduction and cluster identification, respectively. The obtained cellular populations were subsequently visualized with the uniform manifold approximation and projection (UMAP) algorithm (21). Furthermore, we used the "inferCNV" R package to conduct copy number variation (CNV) analysis on different epithelial cells, selecting B cells from normal tissue as a reference. The GSE13213, GSE31210, and GSE72094 datasets, as well as internal and training sets, were selected to validate the predictive performance of the ICDRS (22).

Gene screening

Weighted gene coexpression network analysis (WGCNA) was used to elucidate inter-sample relationships and identify coexpressed gene modules. The "WGCNA" package was employed to develop gene coexpression networks within TCGA cohort (23,24). We calculated the appropriate soft threshold to fit a power-law distribution and conducted an area under the curve (AUC) analysis to screen for a set of genes that are strongly associated with ICD. Subsequently, we compared these identified genes with the differentially expressed gene set associated with lung cancer and the genes linked to lung cancer prognosis in TCGA database to identify a set of covariant genes.

Consensus clustering

The R package "ConsensusCluster" was used to group genes based on their expression patterns (25). To ensure that the results are robust comprehensive consensus scoring matrix was used to obtain an appropriate K value. Furthermore, the "principal component analysis" (PCA) package was used to verify the clustering results (26).

Immunological and therapeutic performance analysis

To account for the potentially confounding effects of algorithm selection, we employed a multifaceted approach using the following tools: tumor immune dysfunction and exclusion (TIDE), cell type identification by estimating relative subsets of RNA transcripts (CIBERSORT), XCell, estimation of stromal and immune cells in malignant tumor tissues using expression data (ESTIMATE) (27). Specifically, CIBERSORT was used to perform an immune infiltration analysis for both the high-risk and low-risk NSCLC patient groups, as well as to analyze the phenotypes and distributions of tumor-infiltrating immune cells. The "estimate" R package was used to determine the immunological purity of distinct subpopulations within the TME (27). Furthermore, gene set variation analysis (GSVA) was employed to assess gene set enrichment within particular samples. We calculated the immune, stromal and ESTIMATE scores, for quantitative assessment of immune and stromal components present in the tumor samples (27). Additionally, we analyzed the outcome of different subgroups to immunotherapy by means of the TIDE scores (27). The Wilcoxon test was used to detect differentially expressed immune checkpoint genes in various subsets of the ICDRS. A P value <0.05 was considered statistically significant.

Enrichment analysis

We applied the "clusterProfiler", "org.hs.egg.db", "enrichmentPlot", and "ggplot2" R packages to conduct the Gene Ontology (GO) enrichment analysis and the Kyoto Encyclopedia of Genes and Genomes (KEGG) pathway analysis. By identifying the enriched functions and molecular pathways that are significantly associated with the prognostic profiles of distinct ICDRS subsets, these methods can help elucidate the fundamental biological mechanisms that influence patient outcomes (27). The ICD scores across various samples were examined using gene set enrichment analysis (GSEA), with notable differences in 64 signaling pathways observed between the high- and low-risk patients. Subsequently, an in-depth enrichment analysis was conducted on a per-sample basis via GSVA (28,29). A P value <0.05, q-value [false discovery rate (FDR)] <0.05 and |log fold change (FC)| >0.1 was considered statistically significant (28).

ICDRS construction

To develop a robust and effective ICDRS, we employed a total of 10 distinct machine learning algorithms, with 101 combinations of these algorithms. These included random survival forest (RSF), elastic net (Enet), least absolute shrinkage and selection operator (Lasso), ridge, stepwise Cox, CoxBoost, partial least squares regression for Cox (plsRcox), supervised principal components (SuperPC), survival support vector machine (survival-SVM), and generalized boosted regression modeling (GBM) (30,31). The steps are outlined below:

(I) Univariate Cox regression analysis was used to identify the ICDR genes that are closely associated with prognosis in the TCGA-THCA cohort (32).
(II) To achieve 101 combinations of these 10 methods, 10-fold cross validation was used.
(III) All models were verified in the GSE13213, GSE31210, GSE72094 datasets (33).

Survival analysis and nomogram construction

In accordance with the algorithm, the data were categorized into a training set, test set, GSE13213 set, GSE31210 set, and GSE72094 set. Samples exceeding the optimal cutoff value were placed in the high-risk group, while those falling below the optimal cutoff value were placed in the low-risk group (32). To determine if the survival probability differed between the two groups, Kaplan-Meier (KM) analysis was performed via the "survminer" R package (34). To evaluate the prognostic independence of the ICDRS in NSCLC and its predictive significance together with clinical variables such as age, gender, and tumor-node-metastasis (TNM) stage for patient survival, we conducted a univariate Cox regression analysis (35). To enhance the predictive performance, we integrated the ICDRS with clinical characteristics to construct a nomogram, aiming to quantitatively predict NSCLC patient survival. Next, we used the concordance index (C-index) and calibration curve to assess the effectiveness of this nomogram (36).

Evaluation of therapeutic drug response

To better guide clinical treatment approaches, the "oncoPredict" R package was used to predict the sensitivity of patients with NSCLC from different ICDRS subgroups to treatments. The half-maximal inhibitory concentration (IC_{50}) values were calculated, with the Wilcoxon test subsequently used to examine the differences between the low- and high-risk groups (37).

ICDRS subgroup genomic variation

The "maftools" R package was used to characterize the mutational profiles of the different ICDRS subgroups (38). Waterfall plots were generated to visualize the mutational landscapes of both the high-risk and low-risk patients with NSCLC. Additionally, we calculated the tumor mutational burden (TMB), conducted survival analysis, and further evaluated the ability of the ICDRS to predict the TMB. A comprehensive CNV analysis was performed to delineate the prognostic significance of the key ICDRS genes with the most pronounced differential expression patterns.

Quantitative real-time polymerase chain reaction (qRT-PCR)

Total RNA was extracted from the H1975, A549, and SK-LU-1 NSCLC cell lines and the normal human bronchial epithelial (HBE) cell line with the Super-Fast Pure Cell RNA Isolation Kit (RC112-00; Vazyme Biotech, Nanjing, China) following the manufacturer's instructions. RNA was reversely transcribed into complementary DNA (cDNA) with Script RT All-in-One Mix (FS-P1002; Starlighter, Beijing, China). Following amplification with SYBR Green qPCR Mix (FS Q1001; Starlighter), qRT-PCR was conducted with the CFX96 Real-Time PCR System (Bio-Rad Laboratories, Hercules, CA, USA). The cycle threshold (Ct) value comparison method and corresponding glyceraldehyde-3-phosphate dehydrogenase (GAPDH) expression levels were used for data processing and normalization. Data were obtained from experiments performed in triplicate. The findings are presented as the mean ± standard deviation (SD). The primer sequences used for qRT-PCR are included in the Appendix 1.

Immunohistochemistry (IHC)

This study enrolled 10 patients with NSCLC who were diagnosed and treated at the Department of Thoracic Surgery, Tangdu Hospital of Air Force Medical University (Xi'an, China) between January 4, 2022 and March 10, 2022 (for patient data please refer to https://cdn.amegroups.cn/static/public/tlcr-2025-769-1.xlsx). Ten groups of formalin-fixed NSCLC tissues were cut into 4-mm serial sections; placed in an oven at 80 ℃ for 2 hours; and dewaxed using

xylene, gradient ethanol (100% ethanol, 95% ethanol, 90% ethanol, 85% ethanol, and 75% ethanol); and Milli-Q water. Antigen retrieval was performed in [Tris(hydroxymethyl) aminomethane-ethylenediaminetetraacetic acid (Tris-EDTA)] (P0084; Beyotime, Shanghai, China). Endogenous peroxidase activity (P0100A; Beyotime) was blocked, and then the samples were incubated with an anti-matrix metallopeptidase 14 (MMP14) primary antibody (1:100; ab51074; Abcam, Cambridge, UK) overnight at 4 ℃. Subsequently, the IHC Detect Kit for Rabbit/Mouse Primary Antibody (PK10006; Proteintech, Rosemont, IL, USA) was used for detection, which was followed by observation via a slicing scanner (Olympus VS200, Olympus, Tokyo, Japan). The study was conducted in accordance with the Declaration of Helsinki and its subsequent amendments. The study was approved by the Ethics Committee of Tangdu Hospital of Air Force Medical University (ethical approval No. TDLL-202411-07), and informed consent was taken from all the patients.

Statistical analysis

R software version 4.4.1 was used to perform statistical analyses. In this study, Pearson correlation coefficients, Spearman correlation analysis, Wilcoxon tests, Kruskal-Wallis tests, chi-squared tests, and Student's *t*-tests were used. P values <0.05 were considered statistically significant. For survival analysis, univariate and multivariate Cox regression and KM analyses were conducted to determine the AUC values, optimal cutoff values, and individual C-indices via the "timeROC", "survminer", and "CompareC" packages, respectively. Receiver operating characteristic (ROC) curve analysis was performed to assess the diagnostic accuracy of the results. Moreover, two sophisticated dimensionality reduction methods, namely t-distributed stochastic neighbor embedding (t-SNE) and UMAP, were used to generate a visual representation of the data within a lower-dimensional space.

Results

WGCNA screened the potential gene modules associated with ICD

We grouped patients with NSCLC in the TCGA database into 11 major clusters according disease stage: T06, T08, T09, T18, T19, T20, T25, T28, T30, T31, and T34 (*Figure 1A*). Immune annotation on the different cases resulted in three major clusters: epithelial, immune, and stromal (*Figure 1B*). Subsequently, distinct cell type-specific marker genes were used to annotate the cells into nine major clusters: B cells, DCs, endothelial cells, epithelial cells, macrophages, monocytes, natural killer cells, smooth muscle cells, and T cells (*Figure 1C*). This analysis indicated that each sample contained multiple cell types, with each cell type being present in more than one sample. The heatmap showed the top five marker genes expressed by each cell type (*Figure 1D*). Following AUC scoring of these nine cell types, monocytes and macrophages displayed the highest scores (*Figure 1E*). To develop a gene coexpression network with equivalent sensitivity and robustness, we used the "WGCNA" R package. A soft threshold power was designated at β=7, which is in line with the recommended criteria for ensuring adherence to scale-free topology within the network architecture (*Figure 1F*). In addition, the set of covariant genes associated with lung cancer in the single-cell data was analyzed. Eight gene modules were obtained in the TCGA-THCA cohort via high dimensional-WGCNA (hdWGCNA) (*Figure 1G*). We then analyzed the AUC, nCount RNA, and nFeature RNA of these modules. The genes with P<0.001 and r>0.5 (*Figure 1H*) in the red and brown modules were screened, indicating that these genes displayed the highest correlations with ICD.

Identification of NSCLC-related genes in the potential ICD-related gene modules

Through the intersection of the 398 genes associated with ICD identified by WGCNA, 668 genes linked to lung cancer prognosis and 3,656 genes that were differentially expressed between normal and NSCLC tissues in the TCGA dataset, 29 genes related to both NSCLC prognosis and ICD were found (*Figure 2A*). We stratified the TCGA patient cohort into k distinct clusters, which were predicted on the transcriptional profiles of the 29 key genes. Consensus cluster analysis indicated an optimal cluster number (k) of 2 (*Figure 2B*), with significant differences observed between the two clusters (*Figure 2C*), which revealed differentially expressed ESTIMATE, immune, and stromal scores. Cluster 1 (C1), which showed higher expression levels of these genes, had higher scores for all three metrics compared with cluster 2 (C2) (P<0.001; *Figure 2D-2F*). A comparative analysis was subsequently performed to determine the distribution of immune cell infiltrates between clusters C1 and C2, which revealed notable infiltration disparities in the nine specific immune cell

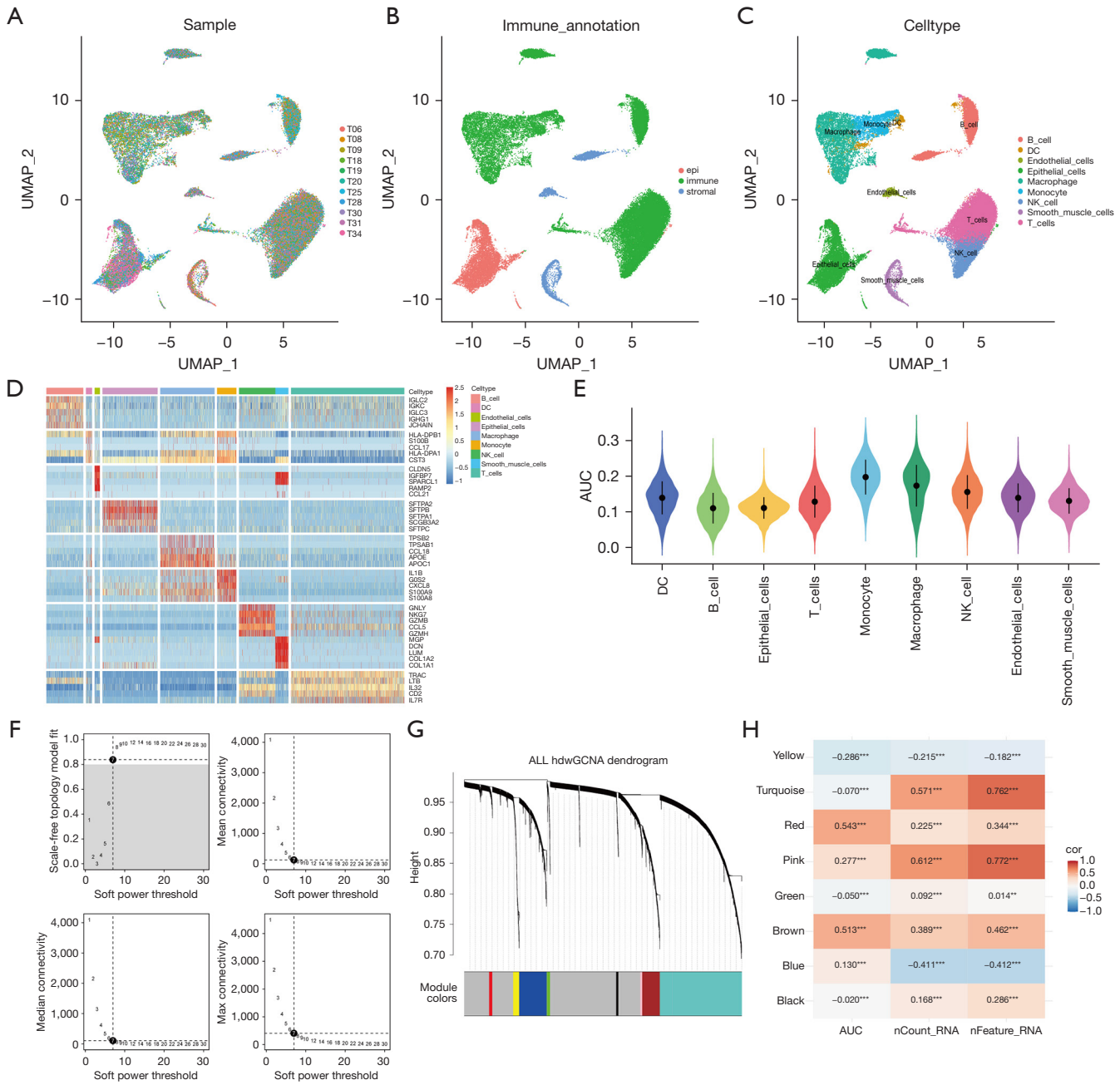

Figure 1 Screening of potential gene modules associated with ICD. (A-C) Patients divided into different clusters. (D) The top five marker genes expressed in each cell type. (E) The AUC scores of these 9 cell types. (F) Gene coexpression network with the optimal soft threshold (β=7). (G) Identification of covariant gene modules. (H) Gene modules associated with immunogenic cell death. **, P<0.01; ***, P<0.001. AUC, area under the curve; DC, dendritic cell; epi, epithelial; hdWGCNA, high dimensional weighted gene coexpression network analysis; ICD, immunogenic cell death; NK, natural killer; UMAP, uniform manifold approximation and projection.

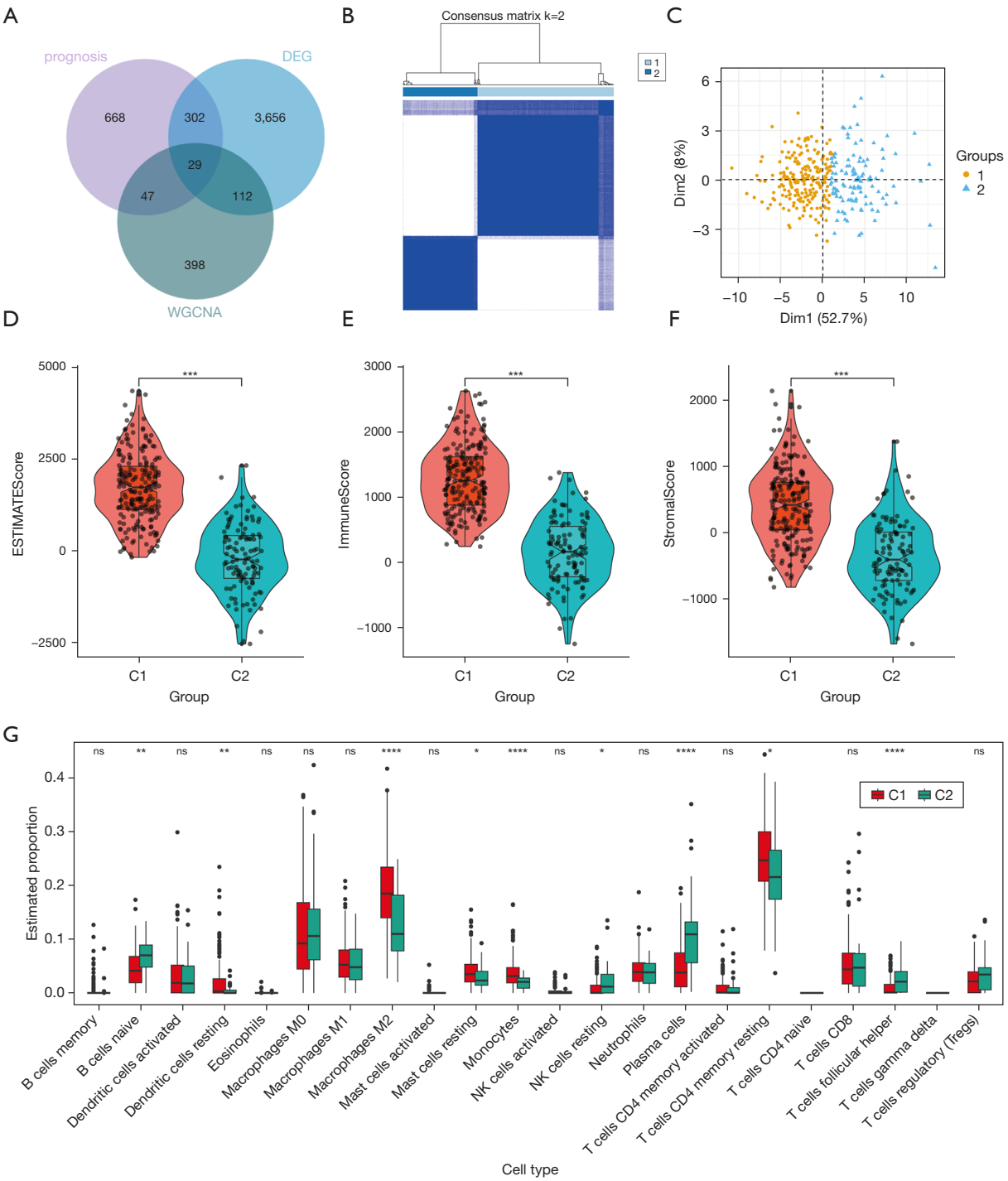

Figure 2 Identification of NSCLC-related genes in the ICD -related gene modules. (A) Genes associated with both immunogenicity death and lung cancer prognosis, along with the genes differentially expressed in normal tissues and NSCLC. (B) The consensus score matrix for all samples when k is equal to 2. (C) Yellow color represents C1 with higher levels of gene expression, and blue color represents C2 with lower levels of gene expression. (D-F) The ESTIMATE score, immune score, and stromal score of the two clusters. (G) Estimated proportion assessment for both clusters. *, P<0.05; **, P<0.01; ***, P<0.001; ****, P<0.0001; ns, not significant. C1, cluster 1; C2, cluster 2; DEG, differentially expressed gene; ICD, immunogenic cell death; NK, natural killer; NSCLC, non-small cell lung cancer; WGCNA, weighted gene coexpression network analysis.

populations, such as monocytes, between these two groups (*Figure 2G*).

Development of an ICDRS via integrative machine learning

With the 29 ICD-associated NSCLC genes we constructed 101 models by combining 10 machine learning algorithms. We then calculated the C-index of various machine learning algorithms in multiple cohorts, including the GSE13213, GSE31210, GSE72094, internal, and training sets. This suggested that Lasso + StepCox (both) exhibited the highest average C-index value (0.631) and consistent performance across all validation sets (*Figure 3A*). ROC curve analysis indicated that the AUC values for the ICDRS in the training set were 0.735, 0.713, and 0.649 at 1, 3, and 5 years, respectively, demonstrating its excellent prediction accuracy (*Figure 3B*). According to the ICDRS, the five following key NSCLC prognosis-related genes were screened: *MMP14*; acetaldehyde dehydrogenase (*ALDH2*); fructose-bisphosphatase 1 (*FBP1*); major histocompatibility complex, class II, DR alpha (*HLA-DRA*); and potassium channel tetramerization domain containing 12 (*KCTD12*). To evaluate the prognostic relevance of these five genes in patients with NSCLC, we developed a risk score algorithm. This algorithm synthesized the expression profiles of these genes, assigning weights based on their expression levels, to categorize patients into high-risk and low-risk groups. KM survival analysis was used to assess the predictive accuracy of the ICDRS within our training cohort. This revealed a significantly superior survival probability for the low-risk group compared with that of the high-risk group (P<0.001; *Figure 3C*). Subsequently, we investigated the survival time and risk scores for both the high- and low-risk groups, which demonstrated a notable positive correlation between an elevated risk score and increased mortality rate (*Figure 3D*). We also examined the suitability and robustness of the ICDRS performance by repeating the above-mentioned steps for different datasets. Similar results were obtained with the test set (P=0.04), GSE13213 set (P=0.02), GSE31210 set (P=0.03), and GSE72094 set (P<0.001) (Figure S2A-S2H). Furthermore, qRT-PCR and IHC assays indicated that the mRNA and protein expression levels of these five key genes were in the expected range both in the normal tissue and NSCLC cohorts (Figure S3A-S3G).

Survival prediction by nomogram

The most important prognostic indicators for NSCLC are age, gender, and TNM staging (39-41). To evaluate the prognostic significance of ICDRS, we examined the correlations between the 5-gene-signature we found and these clinical parameters using univariate and multivariate Cox proportional hazards models. The univariate Cox regression analysis results indicated that the ICDRS, T stage, N stage, and M stage were significant risk factors for NSCLC prognosis [hazard ratio (HR) >1.5; P<0.001]. After multivariate Cox regression analysis, only the ICDRS was found to be an independent prognostic factor for NSCLC [HR =1.690; 95% confidence interval (CI): 1.397–2.044; P<0.001]. These data further substantiate the unique value of the ICDRS in the prognostic evaluation of NSCLC (*Figure 4A*,*4B*). To further evaluate the clinical applicability of the ICDRS, we constructed a nomogram incorporating the risk score, age, gender, T descriptors, N descriptors, M descriptors, and stage (*Figure 4C*). Additionally, the C-index results demonstrated an excellent predictive performance for both the ICDRS and the nomogram, particularly for predicting the 1- to 10-year OS of TCGA, which surpassed the performance of clinical features such as gender (*Figure 4D*). We then generated calibration curves, which indicated good agreement between the predicted value with the nomogram and the actually observed value (*Figure 4E*). These findings support the robustness and effectiveness of the ICDRS for the prognostic evaluation of NSCLC.

Pathway enrichment analysis

Genes that were differentially expressed between high- and low-risk groups underwent GO, Reactome, GSEA, and KEGG enrichment analyses, which suggested that they were strongly enriched in aspects such as nuclear division, pattern specification process, extracellular matrix organization, and the interaction between neuroactive ligands and receptors (*Figure 5A-5C*). Moreover, the low-risk group exhibited a significant enrichment in pathways associated with autoimmune thyroid disease, *Leishmania* infection, and cell cycle regulation, while the high-risk group was notably enriched for allograft rejection processes (*Figure 5D*). The molecular mechanisms underlying the differences in prognosis and treatment response between the high-risk and low-risk groups may be related to these characteristics.

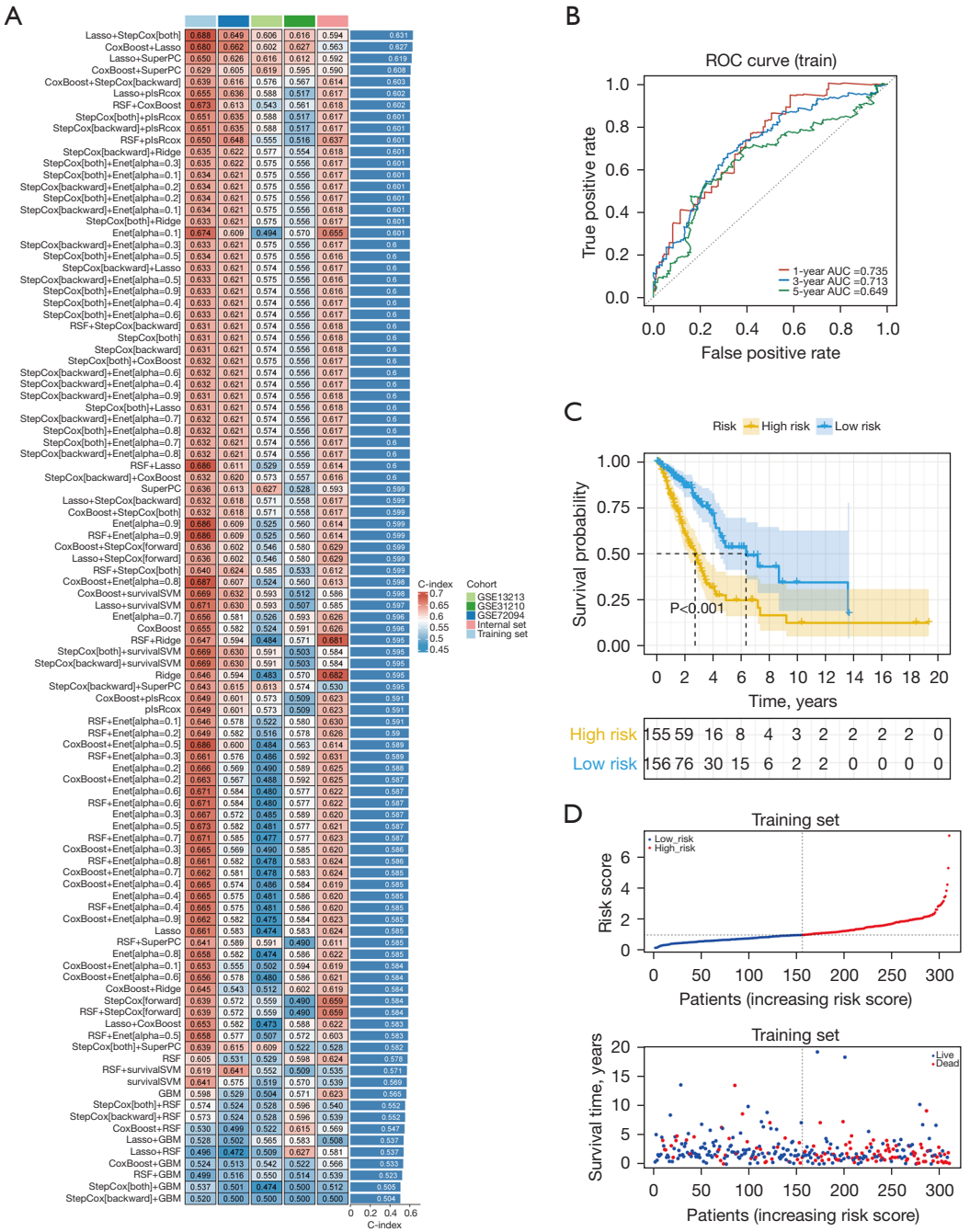

Figure 3 Construction of the ICDRS. (A) The total of 101 predictive models were generated by integrating 10 distinct algorithms. Subsequently, the C-index for each model was computed across the GSE13213, GSE31210, GSE72094, internal, and training sets. (B) ROC of patients with NSCLC at 1, 3, or 5 years as predicted by the ICDRS. (C,D) The predictive power of the ICDRS in the training set (P<0.001). AUC, area under the curve; ICDRS, immunogenic cell death-related signature; NSCLC, non-small cell lung cancer; ROC, receiver operating characteristic.

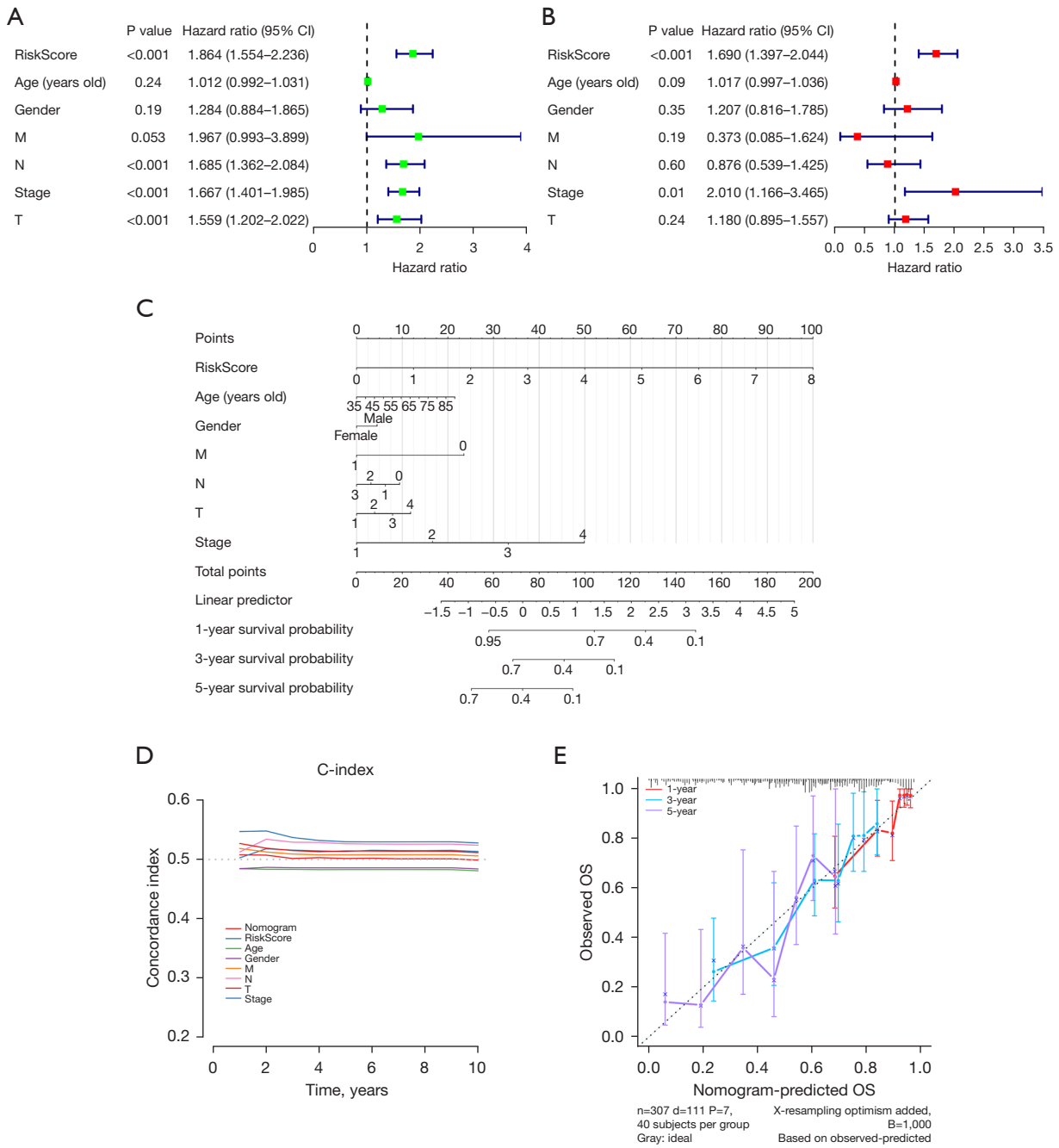

Figure 4 Construction and verification of the nomogram. (A,B) Univariate and multivariate regression analyses demonstrated the independent prognostic value of the ICDRS. (C) Nomogram predicting the prognosis of NSCLC. (D) The C-index for each predictor. (E) Calibration curves were used to evaluate the alignment between the predictions generated by nomograms and the actual observed outcomes. CI, confidence interval; ICDRS, immunogenic cell death-related signature; M, metastasis; N, node; NSCLC, non-small cell lung cancer; OS, overall survival; T, tumor.

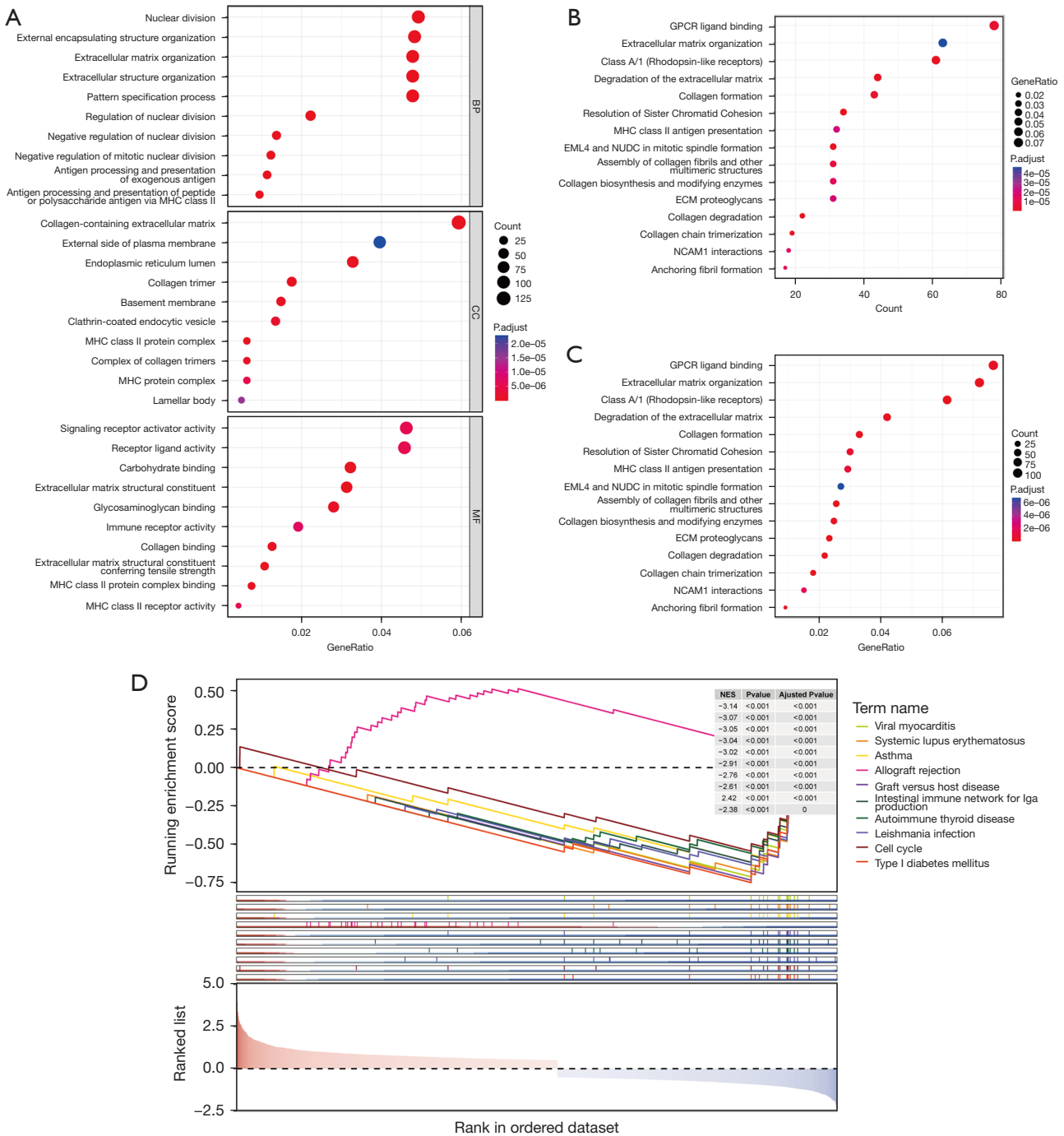

Figure 5 Enrichment analysis of the differential genes. (A-C) The differential gene expression between the high-risk and low-risk subgroups was examined through GO, Reactome, GSEA, and KEGG enrichment analyses. (D) Term name associated with the low- and low-risk subpopulation. BP, biological process; CC, cellular component; ECM, extracellular matrix; GO, Gene Ontology; GSEA, gene set enrichment analysis; KEGG, Kyoto Encyclopedia of Genes and Genomes; MF, molecular function; MHC, major histocompatibility complex; NES, normalized enrichment score.

The performance of the ICDRS in predicting immune cell infiltration and drug treatment responses

We subsequently conducted a comprehensive assessment of the prognostic value of the ICDRS in NSCLC by evaluating the TIDE, immune, stromal, and ESTIMATE scores within the high- and low-risk cohorts. Our analysis indicated that high-risk cohort had a significantly higher TIDE score compared with the low-risk cohort (P<0.001), suggesting a diminished response to ICI therapy and poorer prognosis for the high-risk patients. Conversely, the low-risk group demonstrated significantly better immune (P<0.001), ESTIMATE (P<0.001), and stromal (P<0.05) scores compared with the high-risk group, implying better immune cell infiltration and lower tumor purity in the low-risk group (*Figure 6A-6D*). These data highlight the potential of the ICDRS to predict immune cell infiltration and the response to immunotherapy in patients with NSCLC. Furthermore, the relationships between the five critically prognostic genes (*MMP14*, *ALDH2*, *FBP1*, *HLA-DRA*, *KCTD12*) and the extent of immune cell infiltration in NSCLC were investigated. Our analysis revealed significant correlations between these genes and 39 types of immune cells. Specifically, *MMP14* was significantly associated with cancer-associated fibroblast and 14 other immune cell types, *ALDH2* with hematopoietic stem cell and 5 others immune cell types, FBP1 with macrophage M2 and 16 others immune cell types, *HLA-DRA* with myeloid DC and 27 others immune cell types, and *KCTD12* with monocytes and 28 other immune cell types (*Figure 6E*). When comparing the immunological characteristics between the high-risk and low-risk groups, we identified a marked divergence in the levels of M0 macrophages, 9 types of immune cells, hepatitis a virus cellular receptor 2 (HAVCR2), and 9 key immune checkpoints. These findings may provide insights into the immunological mechanisms that underlie the prognostic variability between low- and high-risk patients with NSCLC (*Figure 6F,6G*) and suggest that these groups differ in their response to immunotherapy. Our results also indicated that nine drugs displayed significantly higher IC_{50} values in the high-risk group (Figure S4A-S4I).

Genetic variation in different ICDRS subgroups

In the mutational profiling analysis, the high-risk group primarily exhibited missense mutations, with single-nucleotide variant being the predominant variant type. The most prevalent mutation signature was cytosine > adenine, followed by cytosine > thymine. We examined the mutation information for each sample in the high-risk group and mutation type distribution, identifying the top 10 genes with the highest mutation frequencies in this group, which included Titin (*TTN*) (*Figure 7A*). Similarly, we determined the mutation landscape of the low-risk group. Both groups shared the same variant classification, variant type, and similar single-nucleotide variant class but differed in the mutation details for each sample, mutation patterns, and top 10 genes with the highest mutation frequencies (*Figure 7B*). To further evaluate the mutation profiles of the different risk groups and their impact on NSCLC prognosis, we generated a waterfall plot showing the mutation status of these highly mutated genes (*Figure 7C*). Subsequently, we stratified the patient samples into two distinct categories by TMB. However, this metric did not demonstrate a significant predictive value (P=0.65; *Figure 7D*). Notably, our analysis revealed that the high-risk group exhibited a significantly elevated TMB (P<0.001) compared with the low-risk group, indicating that the ICDRS has strong predictive ability regarding TMB (*Figure 7E*). Additionally, we investigated the CNV of the key ICDRS-associated genes and found that all five were primarily associated with amplification in copy number, with *HLA-DRA* having the highest frequency (*Figure 7F*).

ICDRS at the single-cell transcriptomic level

We examined the expression profiles of *MMP14*, *ALDH2*, *FBP1*, *HLA-DRA*, and *KCTD12* across various cell types to assess the effectiveness of the ICDRS at the single-cell transcriptomic level. *ALDH2* was predominantly expressed in macrophages and epithelial cells, *FBP1* was abundant in macrophages, *HLA-DRA* showed high expression levels in B cells, *KCTD12* was highly expressed in endothelial cells and macrophages, and *MMP14* demonstrated high expression patterns in smooth muscle cells (*Figure 8A*). Pathologically, tumor cells are frequently classified as epithelial cells, but not all epithelial cells are classified as tumor cells (42). Using the distinct CNVs observed in tumor cells, we divided the epithelial cells in the GSE dataset in 11 subpopulations and evaluated their CNV status relative to B cells (*Figure 8B*). Kruskal-Wallis tests revealed a highly significant increase in CNVs in these epithelial tissues compared with the B cells ($P<2.2\times10^{-16}$), indicating that they could be considered tumor cells. Additionally, we identified cluster C3 as having the highest CNV score (*Figure 8C*). Subsequently, guided by the ICDRS, we grouped these epithelial cells into both

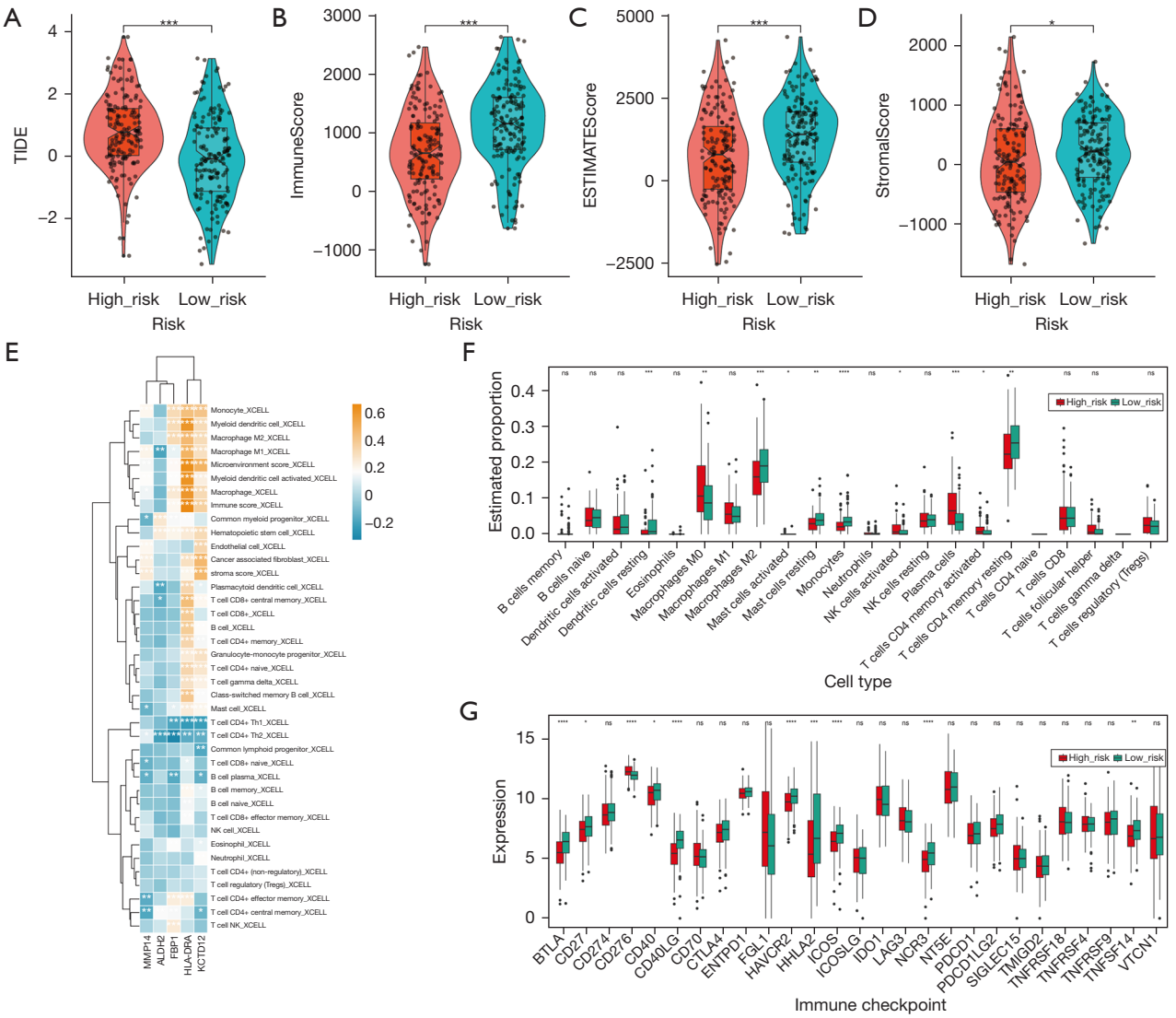

Figure 6 The predictive power of the ICDRS for immunization status. (A-D) TIDE (P<0.001), immune score (P<0.001), ESTIMATE score (P<0.001), and stromal score (P<0.05) were compared between the high- and low-risk subgroups. (E) Assessment of the abundance of immune cell infiltration corresponding to five ICDRS key genes. (F) Evaluation of the disparities in estimated proportion of immune cells between the high-risk and low-risk subgroups. (G) Assessment of variations in the expression of immune checkpoints between the high-risk and low-risk subgroups. *, P<0.05; **, P<0.01; ***, P<0.001; ****, P<0.0001; ns, not significant. ICDRS, immunogenic cell death-related signature; NK, natural killer; TIDE, tumor immune dysfunction and exclusion.

high-risk and low-risk groups and performed GSVA on their associated pathways. The results indicated that the high-risk group exhibited greater activity in the glycolysis, apical junction, and mitotic spindle pathways, while the low-risk group showed stronger activity in the bile acid metabolism and pancreas beta cell pathways (*Figure 8D*). Moreover, the findings of the enrichment analysis comparing the high-risk and low-risk groups indicated that coronavirus disease 2019 (COVID-19) and ribosomal activity are essential components contributing to the differences between high and low risk groups (*Figure 8E*).

Discussion

The incidence and mortality rates of NSCLC are persistently high, with a significant proportion of patients

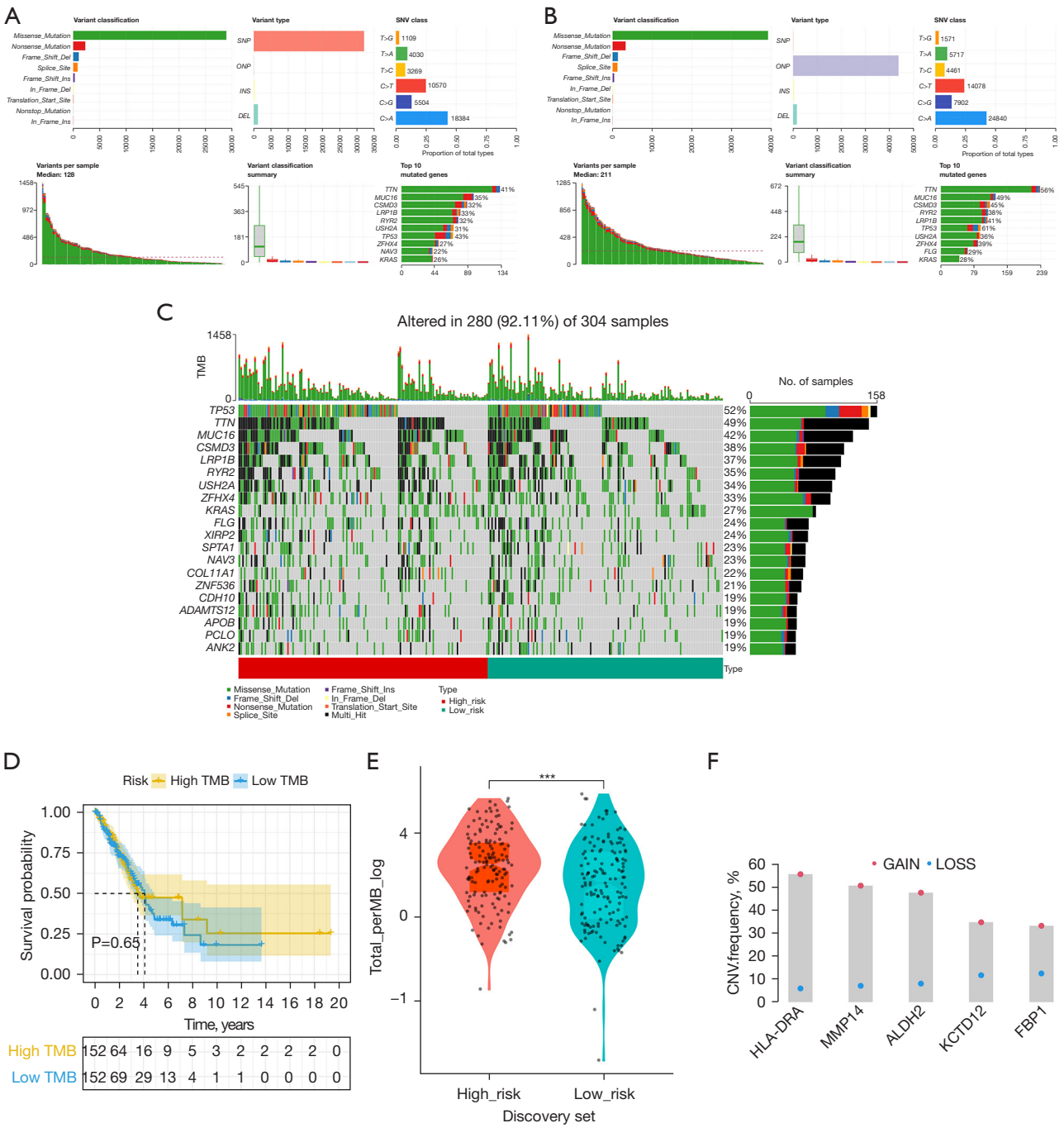

Figure 7 The predictive power of ICDRS for immunization status. (A) Mutations in the high-risk group. (B) Mutations in the low-risk group. (C) Mutation types of genes with higher mutation frequency in 30 patients with NSCLC. (D) Association of TMB with prognosis (P=0.65). (E) Differences of TMB in high- and low-risk subsets. (F) CNV alteration frequency of five key genes of ICDRS. ***, P<0.001. CNV, copy number variation; ICDRS, immunogenic cell death-related signature; NSCLC, non-small cell lung cancer; SNV, single nucleotide variant; TMB, tumor mutational burden.

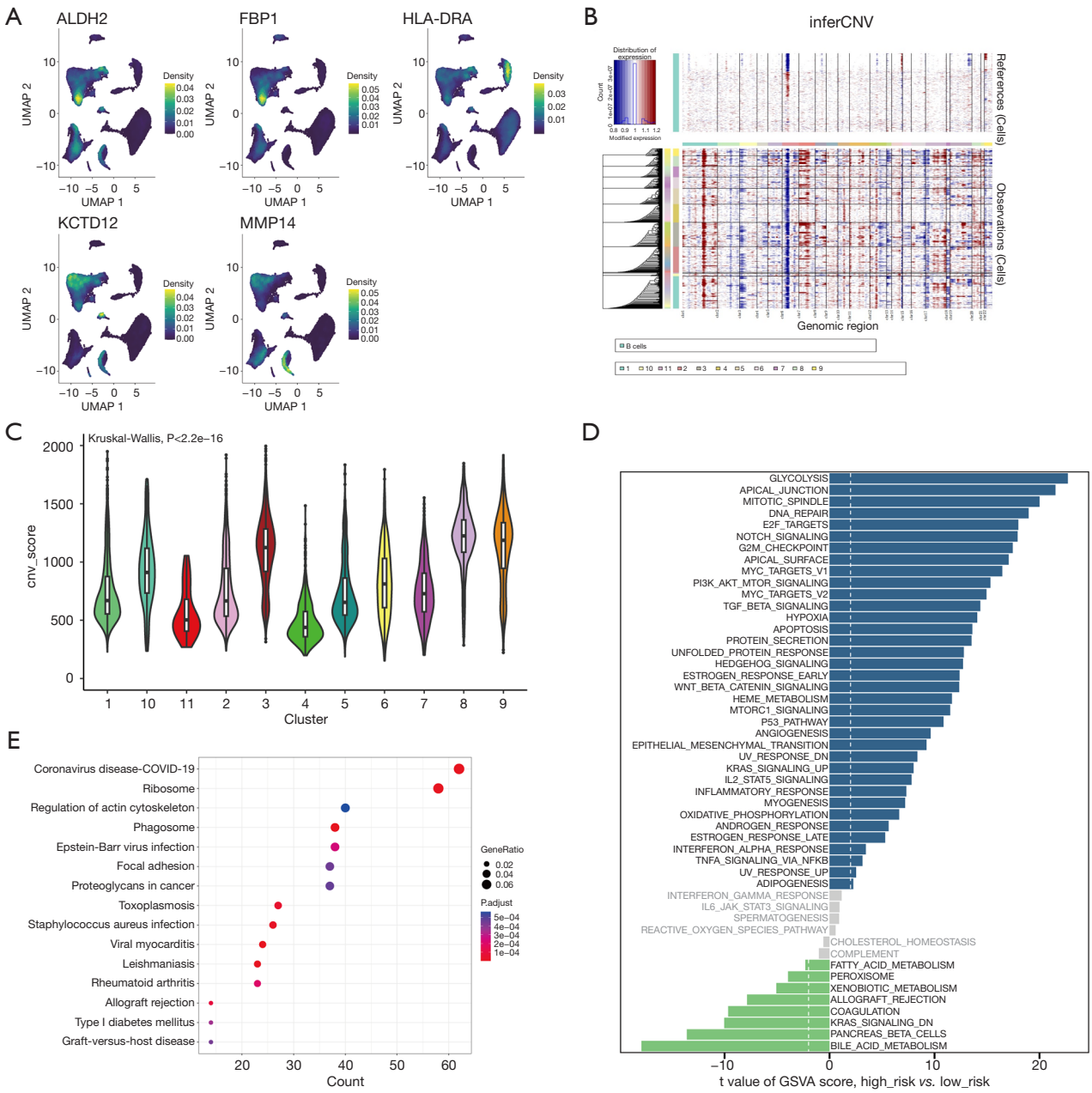

Figure 8 Manifestations of ICDRS in the single cell transcriptome. (A) The expression of key genes in the different types of cells. (B) Determination of the cancer status of 11 epithelial cell subsets. (C) CNV scores of 11 subsets of epithelial cells (Kruskal-Wallis, $P<2.2e^{-16}$). (D) GSVA of the enrichment pathway in the ICDRS subsets for 11 types of epithelial cells. (E) Point plots of enrichment for different subpopulations. CNV, copy number variation; GSVA, gene set variation analysis; ICDRS, immunogenic cell death-related signature; UMAP, uniform manifold approximation and projection.

diagnosed at advanced stages. Evidence suggests that there is a close correlation of ICD with cancer prognosis and sensitivity to ICI treatment (43).

In this study, we integrated various machine learning algorithms to establish 101 models and developed a nomogram for predicting NSCLC prognosis. Subsequently, we predicted the responses of the ICDRS subsets to ICI treatment and clarified the association with the prognosis

of patients with NSCLC and the related molecular mechanisms. We then observed that the mutation landscape of the high-risk group shared certain similarities with the low-risk group. This suggests that the prognostic disparities associated with different ICDRS subgroups may not be strongly correlated with genomic mutations. Additionally, investigations at the single-cell transcriptomic level also indicated the superior predictive performance of the ICDRS. In summary, the ICDRS can serve as an effective tool that can provide clinicians with critical evaluation information, which would enable improved and more convenient diagnosis and treatment.

In comparison with relevant studies, our methodology for patient selection from databases and identification of key genes exhibits certain similarities to those employed by other researchers (44-46). However, our study incorporates validation at both single-cell and transcriptome levels—an approach absents in most existing studies (47,48). Moreover, while our pathway enrichment analysis revealed differences to previous reports, the expression levels of each component of the 5-gene signature in our study were validated in actual patients, which adds to the validity of our results (49,50). Notably, although prior studies identified Toll-like receptor 4 (TLR4) and cluster of differentiation 4 (CD4) as key genes that differ from our candidates, their reported genes also exhibited strong macrophage-related characteristics (44,51). This observation further reinforces the association between macrophages and immunotherapy efficacy. Despite the promising results, certain limitations to this study should be noted. We validated our findings only in the GSE13213, GSE31210, GSE72094, internal, and training sets without conducting multicenter clinical studies. Additionally, issues such as the nonspecific side effects of ICD inducers, should be subjected to more in-depth investigation (52).

Conclusions

In this study, we integrated 10 machine learning algorithms and evaluated their effects at the level of single cells and transcriptomes to design an ICDRS with robustness and wide applicability. This is of great value for the prognostic evaluation of NSCLC and the response to drug treatment, and provides valuable tools and options for clinical treatment.

Acknowledgments

We would like to thank J. Iacona, PhD, from Liwen Bianji (Edanz) (www.liwenbianji.cn) for editing the English text of a draft of this manuscript.

Footnote

Reporting Checklist: The authors have completed the TRIPOD reporting checklist. Available at https://tlcr.amegroups.com/article/view/10.21037/tlcr-2025-769/rc

Data Sharing Statement: Available at https://tlcr.amegroups.com/article/view/10.21037/tlcr-2025-769/dss

Peer Review File: Available at https://tlcr.amegroups.com/article/view/10.21037/tlcr-2025-769/prf

Funding: This work was supported by the National Natural Science Foundation of China (No. 82173252).

Conflicts of Interest: All authors have completed the ICMJE uniform disclosure form (available at https://tlcr.amegroups.com/article/view/10.21037/tlcr-2025-769/coif). The authors have no conflicts of interest to declare.

Ethical Statement: The authors are accountable for all aspects of the work in ensuring that questions related to the accuracy or integrity of any part of the work are appropriately investigated and resolved. The study was conducted in accordance with the Declaration of Helsinki and its subsequent amendments. The study was approved by the Ethics Committee of Tangdu Hospital of Air Force Medical University (ethical approval No. TDLL-202411-07), and informed consent was taken from all the patients.

References

1. Ross HJ, Kozono D, Wang XF, et al. Atezolizumab Before and After Chemoradiation for Unresectable Stage III Non-Small Cell Lung Cancer: A Phase II Nonrandomized

Controlled Trial. JAMA Oncol 2024;10:1212-9.
2. Siegel RL, Giaquinto AN, Jemal A. Cancer statistics, 2024. CA Cancer J Clin 2024;74:12-49.
3. Lahiri A, Maji A, Potdar PD, et al. Lung cancer immunotherapy: progress, pitfalls, and promises. Mol Cancer 2023;22:40.
4. Spigel DR, Faivre-Finn C, Gray JE, et al. Five-Year Survival Outcomes From the PACIFIC Trial: Durvalumab After Chemoradiotherapy in Stage III Non-Small-Cell Lung Cancer. J Clin Oncol 2022;40:1301-11.
5. Cheng Y, Spigel DR, Cho BC, et al. Durvalumab after Chemoradiotherapy in Limited-Stage Small-Cell Lung Cancer. N Engl J Med 2024;391:1313-27.
6. Reck M, Rodríguez-Abreu D, Robinson AG, et al. Updated Analysis of KEYNOTE-024: Pembrolizumab Versus Platinum-Based Chemotherapy for Advanced Non-Small-Cell Lung Cancer With PD-L1 Tumor Proportion Score of 50% or Greater. J Clin Oncol 2019;37:537-46.
7. Campbell MT, Matin SF, Tam AL, et al. Pilot study of Tremelimumab with and without cryoablation in patients with metastatic renal cell carcinoma. Nat Commun 2021;12:6375.
8. Gadgeel S, Rodríguez-Abreu D, Speranza G, et al. Updated Analysis From KEYNOTE-189: Pembrolizumab or Placebo Plus Pemetrexed and Platinum for Previously Untreated Metastatic Nonsquamous Non-Small-Cell Lung Cancer. J Clin Oncol 2020;38:1505-17.
9. Paz-Ares L, Vicente D, Tafreshi A, et al. A Randomized, Placebo-Controlled Trial of Pembrolizumab Plus Chemotherapy in Patients With Metastatic Squamous NSCLC: Protocol-Specified Final Analysis of KEYNOTE-407. J Thorac Oncol 2020;15:1657-69.
10. Kroemer G, Galassi C, Zitvogel L, et al. Immunogenic cell stress and death. Nat Immunol 2022;23:487-500.
11. Zitvogel L, Kepp O, Kroemer G. Decoding cell death signals in inflammation and immunity. Cell 2010;140:798-804.
12. Røssevold AH, Andresen NK, Bjerre CA, et al. Atezolizumab plus anthracycline-based chemotherapy in metastatic triple-negative breast cancer: the randomized, double-blind phase 2b ALICE trial. Nat Med 2022;28:2573-83.
13. Sistigu A, Yamazaki T, Vacchelli E, et al. Cancer cell-autonomous contribution of type I interferon signaling to the efficacy of chemotherapy. Nat Med 2014;20:1301-9.
14. Antonia SJ, Villegas A, Daniel D, et al. Overall Survival with Durvalumab after Chemoradiotherapy in Stage III NSCLC. N Engl J Med 2018;379:2342-50.
15. Ma Y, Adjemian S, Mattarollo SR, et al. Anticancer chemotherapy-induced intratumoral recruitment and differentiation of antigen-presenting cells. Immunity 2013;38:729-41.
16. Thibaudin M, Fumet JD, Chibaudel B, et al. First-line durvalumab and tremelimumab with chemotherapy in RAS-mutated metastatic colorectal cancer: a phase 1b/2 trial. Nat Med 2023;29:2087-98.
17. Bezu L, Kepp O, Kroemer G. Immunogenic chemotherapy sensitizes RAS-mutated colorectal cancers to immune checkpoint inhibitors. Oncoimmunology 2023;12:2272352.
18. Li X, Li Y, Tuerxun H, et al. Firing up "cold" tumors: Ferroptosis causes immune activation by improving T cell infiltration. Biomed Pharmacother 2024;179:117298.
19. Zhang X, Shang C, Qiao X, et al. Role and clinical significance of immunogenic cell death biomarkers in chemoresistance and immunoregulation of head and neck squamous cell carcinoma. Biomed Pharmacother 2023;167:115509.
20. Tian J, Wan S, Yang Z, et al. PDL1/HER2-Targeted Lipid-Encapsulated Oxygen Nanobubbles Combined with Photodynamic Therapy for HER2(+) Breast Cancer Immunotherapy. Adv Healthc Mater 2024;13:e2400030.
21. Stuart T, Butler A, Hoffman P, et al. Comprehensive Integration of Single-Cell Data. Cell 2019;177:1888-1902.e21.
22. Liu J, Li H, Wang L, et al. Spatial transcriptome and single-cell reveal the role of nucleotide metabolism in colorectal cancer progression and tumor microenvironment. J Transl Med 2024;22:702.
23. Liu J, Shi Y, Zhang Y. Multi-omics identification of an immunogenic cell death-related signature for clear cell renal cell carcinoma in the context of 3P medicine and based on a 101-combination machine learning computational framework. EPMA J 2023;14:275-305.
24. Langfelder P, Horvath S. WGCNA: an R package for weighted correlation network analysis. BMC Bioinformatics 2008;9:559.
25. Sabah A, Tiun S, Sani NS, et al. Enhancing web search result clustering model based on multiview multirepresentation consensus cluster ensemble (mmcc) approach. PLoS One 2021;16:e0245264.
26. Huan Q, Cheng S, Ma HF, et al. Machine learning-derived identification of prognostic signature for improving prognosis and drug response in patients with ovarian cancer. J Cell Mol Med 2024;28:e18021.

27. Yoshihara K, Shahmoradgoli M, Martínez E, et al. Inferring tumour purity and stromal and immune cell admixture from expression data. Nat Commun 2013;4:2612.
28. Zhao P, Zhen H, Zhao H, et al. Identification of hub genes and potential molecular mechanisms related to radiotherapy sensitivity in rectal cancer based on multiple datasets. J Transl Med 2023;21:176.
29. Subramanian A, Tamayo P, Mootha VK, et al. Gene set enrichment analysis: a knowledge-based approach for interpreting genome-wide expression profiles. Proc Natl Acad Sci U S A 2005;102:15545-50.
30. Rong F, Xiang H, Qian L, et al. Machine Learning for Prediction of Outcomes in Cardiogenic Shock. Front Cardiovasc Med 2022;9:849688.
31. Deo RC. Machine Learning in Medicine. Circulation 2015;132:1920-30.
32. Hong K, Cen K, Chen Q, et al. Identification and validation of a novel senescence-related biomarker for thyroid cancer to predict the prognosis and immunotherapy. Front Immunol 2023;14:1128390.
33. Li H, Sha X, Wang W, et al. Identification of lysosomal genes associated with prognosis in lung adenocarcinoma. Transl Lung Cancer Res 2023;12:1477-95.
34. Shi Y, Wang Y, Dong H, et al. Crosstalk of ferroptosis regulators and tumor immunity in pancreatic adenocarcinoma: novel perspective to mRNA vaccines and personalized immunotherapy. Apoptosis 2023;28:1423-35.
35. D'Agostino RB Sr, Vasan RS, Pencina MJ, et al. General cardiovascular risk profile for use in primary care: the Framingham Heart Study. Circulation 2008;117:743-53.
36. Wu J, Zhang H, Li L, et al. A nomogram for predicting overall survival in patients with low-grade endometrial stromal sarcoma: A population-based analysis. Cancer Commun (Lond) 2020;40:301-12.
37. Li S, Ding B, Weng D. Characterization of prognostic signature related with twelve types of programmed cell death in lung squamous cell carcinoma. J Cardiothorac Surg 2024;19:569.
38. Mayakonda A, Lin DC, Assenov Y, et al. Maftools: efficient and comprehensive analysis of somatic variants in cancer. Genome Res 2018;28:1747-56.
39. Bhattarai A, Shah S, Abu Serhan H, et al. Genomic profiling for non-small cell lung cancer: Clinical relevance in staging and prognosis. Medicine (Baltimore) 2023;102:e36003.
40. Hansen O, Schytte T, Nielsen M, et al. Age dependent prognosis in concurrent chemo-radiation of locally advanced NSCLC. Acta Oncol 2015;54:333-9.
41. Zhang M, Wen F, Zhao K. Circular RNA_0001946 is insufficiently expressed in tumor tissues, while its higher expression correlates with less lymph node metastasis, lower TNM stage, and improved prognosis in NSCLC patients. J Clin Lab Anal 2021;35:e23625.
42. Mestre-Farrera A, Bruch-Oms M, Peña R, et al. Glutamine-Directed Migration of Cancer-Activated Fibroblasts Facilitates Epithelial Tumor Invasion. Cancer Res 2021;81:438-51.
43. Galluzzi L, Buqué A, Kepp O, et al. Immunogenic cell death in cancer and infectious disease. Nat Rev Immunol 2017;17:97-111.
44. Li G, Chen K, Dong S, et al. Immunogenic cell death-related genes predict prognosis and response to immunotherapy in lung squamous cell carcinoma. Biotechnol Appl Biochem 2025;72:138-49.
45. Sun J, Guo H, Zhang S, et al. Machine learning-based integration develops an immunogenic cell death-derived lncRNA signature for predicting prognosis and immunotherapy response in lung adenocarcinoma. Sci Rep 2024;14:11724.
46. Matboli M, Abdelbaky I, Khaled A, et al. Machine learning based identification potential feature genes for prediction of drug efficacy in nonalcoholic steatohepatitis animal model. Lipids Health Dis 2024;23:266.
47. Li X, Zhang D, Guo P, et al. Identifying an immunogenic cell death-related gene signature contributes to predicting prognosis, immunotherapy efficacy, and tumor microenvironment of lung adenocarcinoma. Aging (Albany NY) 2024;16:6290-313.
48. Dong Y, Yu X, Song H, et al. Identification of molecular subtypes and prognostic model to reveal immune infiltration and predict prognosis based on immunogenic cell death-related genes in lung adenocarcinoma. Cell Cycle 2023;22:2566-83.
49. Wang S, Zhang Y. Construction of an immunogenic cell death-related LncRNA signature to predict the prognosis of patients with lung adenocarcinoma. BMC Med Genomics 2024;17:277.
50. Sun D, Zhang C. Immunogenic cell death-related long noncoding RNA influences immunotherapy against lung adenocarcinoma. Oncol Res 2023;31:753-67.
51. Zhang J, Li H, Zhang X, et al. The landscape of immunogenic cell death-related genes predicts the overall survival and immune infiltration status of non-small-cell lung carcinoma. Heliyon 2025;11:e40869.

52. Elzoghby AO, Samir O, Emam HE, et al. Engineering nanomedicines for immunogenic eradication of cancer cells: Recent trends and synergistic approaches. Acta Pharm Sin B 2024;14:2475-504.

(English Language Editor: J. Gray)

Cite this article as: Sun Y, Chen H, Wang Z, Jiao R, Zehentmayr F, Tabbò F, Wu C, Zhang T, Yan H, Wang J, Yan X. The role of immunogenic cell death in the prognosis and development of treatment strategies for non-small cell lung cancer: a multiomics and machine learning approach for predictive and personalized treatment. Transl Lung Cancer Res 2025;14(8):3107-3125. doi: 10.21037/tlcr-2025-769