ORIGINAL RESEARCH OPEN ACCESS

A Novel Lysosome-Related Gene Signature Predicts Lung Cancer Prognosis: A Bioinformatics-Driven Study

Wei Ye | Lin Sun | Cong Fu | Huajie Dong | Tong Zhou 🆔

Department of Oncology, Changzhou Tumor Hospital, Changzhou, China

Correspondence: Tong Zhou (zhoutong2930@163.com)

ABSTRACT

Background and Aims: The biological function of lysosomes has been increasingly appreciated in cancer. However, the relationship between lysosome and lung adenocarcinoma (LUAD) was not well understood. In this study, a lysosome-related signature was developed for LUAD risk stratification and prognosis prediction.

Methods: Download RNA-seq data of LUAD and clinical information of corresponding samples from the UCSC-Xena platform. GSE31210 databases is used as a validation cohort. The lysosome-related genes was obtained from molecular signature database. The differentially expressed genes (DEGs) as well as lysosome-associated prognosis signatures were identified by using univariate, multivariate cox, and Lasso regression. A nomogram was constructed and evaluated using ROC and DCA.

Results: A total of 109 lysosome-related DEGs were identified and 30 prognostic related DEGs were subsequently screened. Cluster analysis further divides the TCGA cohort into clusters 1 and 2. Patients in cluster 2 had a worse prognosis ($p = 0.016$), lower LYSOSOME score. Enrichment analysis showed that 21 significantly enriched gene sets in the cluster 2 were activated. And 10 pathways, such as E2F_TARGETS, G2M_CHECKPOINT were upregulated. Multivariate Cox regression analysis identified 17 best prognostic genes as risk signature. An independent prognostic factor, the risk signature, was identified. A prognostic nomogram including risk signature, age, TNM stage, and gender was constructed, and ROC and DCA curves proved its excellent performance. We examined CTSZ and AP3S2 protein expression in 48 stage 3–4 NSCLC samples. Low AP3S2 expression was associated with better prognosis (median overall survival: 37.87 vs. 8.53 months, $p = 0.0211$). Increased CTSZ expression also indicated better prognosis (median overall survival: 6.77 vs. 30.50 months, $p = 0.0306$).

Conclusion: We identified molecular subtypes and lysosomal-based prognostic signatures for LUAD patients, as well as 17 genes that serve as a biomarker for evaluating the prognosis of LUAD patients.

1 | Background

Although new cancer cases and overall death rates have declined globally in the past few decades [1], cancer-related deaths in China are primarily caused by lung cancer, which accounts for about 27.2% of deaths [2]. Screening with low-dose spiral computed tomography has improved early detection rates of this disease, but more than 70% of cases are still diagnosed at advanced stage [3]. Even if the early patients are successfully resected, a significant number of patients still experience local or distant recurrence [4]. Patients with non-small cell lung cancer (NSCLC), especially lung adenocarcinoma (LUAD), have significantly improved their survival with the development of small molecule tyrosine kinase inhibitors and immunotherapy.

This has also pushed the treatment of NSCLC to the forefront of other cancer treatments and deepened our understanding of the biology and pathogenesis of NSCLC. However, the problem of drug resistance has greatly limited the therapeutic effect, making the overall cure rate and survival rate of NSCLC still very low [5, 6]. Therefore, it is urgently needed to develop a novel and powerful prognostic stratification model to guide treatments and improve prognoses of NSCLC.

As a key component of cellular homeostasis, the lysosome plays an essential role in cell death, immune response, energy metabolism, cell signaling, and endocytosis [7]. In recent studies, it was found that lysosomes play an important role in the occurrence and development of tumors. In cells and tissues from patients and mouse models of renal cell carcinoma, pancreatic ductal adenocarcinoma, and melanoma, upregulation of the MiT/TFE gene triggers RagD-mediated mTORC1 induction, leading to excessive cell proliferation and tumor growth [8]. In prostate cancer, high expression of LAPTM4B-35 was shown to be significantly associated with poorer overall survival (OS), and subsequent multivariate COX analysis found LAPTM4B-35 to be an independent prognostic factor for OS [9]. In addition, the *SLC11A1* gene in the autophagy-lysosome pathway was associated with increased prostate cancer risk and prognosis [10]. In breast cancer, lysosome-associated genes were identified as having major potential roles in breast cancer development and drug resistance [11]. As for NSCLC, overexpression of *SLCA38A7* in lung squamous cell carcinoma predicts poor prognosis [12]. Additionally, advancements in nanotechnology have led to the exploration of nanoparticles as potential therapeutic agents for lung cancer, further emphasizing the need to understand the underlying molecular mechanisms governing tumor behavior [13]. However, the relationship between lysosome-related genes and LUAD has not been fully elucidated. In this study, we developed a lysosome-related risk and prognosis model for LUAD patients to assess risk factors for cancer clinical care through bioinformatics and survival analysis, and identify potential therapeutic targets.

2 | Methods

2.1 | Data Acquisition, Differential Expression Analysis, and Intersection Identification

Download the RNA-seq data of TCGA-LUAD from the UCSC-Xena platform (https://toil.xenahubs.net), as well as the clinical information (including age, sex, TNM stage, and more) and survival information of the corresponding samples. For the RNA-seq data, we implemented a rigorous normalization strategy using FPKM method. This comprehensive approach involved meticulously calculating expression count values and incorporating gene length to derive the FPKM measure. To further stabilize the variance and achieve a normal distribution of the data, we applied a $\log_2 (FPKM + 1)$ transformation. This step significantly enhanced the comparability and reliability of our results across different samples and studies. In addition, GSE31210 data set was downloaded from the GEO platform (http://www.ncbi.nlm.nih.gov/geo/), which included a total of 226 LUAD samples as a validated cohort for a subsequent

model. Using the reference's gene annotation information, we directly downloaded the processed and standardized probe expression matrix. A gene symbol was analyzed for a specific probe by taking the average value from all probes that corresponded to a particular gene symbol. The KEGG_LYSOSOME (M11266) and LYSOSOME (M13845) pathway genes were downloaded from the Molecular Signature Database (MsigDB database, http://www.gsea-msigdb.org/gsea/index.jsp) and were used as lysosome-related genes. A total of 163 lysosome related genes were selected by matching the genes in the TCGA data above (Table S1).

2.2 | Lysosome Related Differentially Expressed Genes (DEGS) Analysis

DEGs analysis between tumor tissue and adjacent tissue was performed in the TCGA-LUAD cohort. "Limma" package was used to obtain the corresponding information of the lysosomal gene, such as *p* value and logFC. The Benjamin & Hochberg method was further used to carry out multiple testing and correction, and the corrected *p* value was adjusted *P* value. The threshold of differential expression was set as follows: adjust $p < 0.05$.

2.3 | Cluster Analysis of Lysosome Related DEGS

Univariate Cox regression analysis of lysosomal-related DEGs was performed through the survival-V3.2.13 package (https://github.com/therneau/survival) to obtain candidate genes related to prognosis. Unsupervised cluster analysis of lysosomal prognostic-related gene expression profile data was performed using ConsensusClusterPlus (V-1.56.0).

For the evaluation of gene enrichment scores of lysosomes in two molecular subtypes, R's GSVA package (V-1.40.1) was used, along with the Wilcoxon rank test to assess the difference between subtypes. The prognosis of the two subtypes was compared using Kaplan-Meier analysis (KM). The Complex-Heatmap package (V-2.8.0) is used to show the distribution of clinical features across different subtypes. Further, 50 tumor hallmark gene sets were downloaded from MSigDB data, and the tumor hallmark enrichment score was calculated using the clusterProfiler package to evaluate the differential hallmarks between two subtypes. Functional enrichment outcomes were evaluated via normalized enrichment score (NES) and false discovery rate (FDR). In this analysis, an FDR cutoff of 0.05 was utilized to ascertain the statistical significance.

2.4 | Lysosome Molecular Subtypes and Tumor Immune Microenvironment (TIME)

To further explore the relationship between lysosome molecular subtypes and TIME, the "ESTIMATE" package of R (V-1.0.13, https://bioinformatics.mdanderson.org/estimate/rpackage.html) was used to determine the TIME score of TCGA-LUAD samples, and CIBERSORT was used to determine the infiltration score of various immune cells. Finally, the Wilcoxon test was used to

evaluate the distribution difference of TIME score and immune cell infiltration ratio among different lysosome subtypes.

In addition, we analyzed expression data from the training cohort for genes related to immune checkpoints (PMID: 36249031) and the HLA family (reference PMID: 33724691). Lysosome subtypes were compared using the Wilcoxon test to see if immune checkpoint genes and HLA family genes differ in expression.

2.5 | Building and Validating the Prognostic Signature

Based on the prognostic lysosome-related genes mentioned above, LASSO Cox analysis was performed using glmnet package to screen characteristic genes for the construction of lysosome-related prognostic signature. For making the final prognostic score model, only genes with regression coefficients greater than 0 were included. The prognostic score is derived by using the following formula based on the regression coefficient (c):

Risk score = [c1* expression level of gene (1)] + [c2* expression level of gene (2)] + ... + [cn* expression level of gene (n)].

The risk score of each patient in the TCGA-LUAD data set was calculated according to the formula, and then divided into a high-risk group and a low-risk group according to the median value of risk score, and the disease-specific survival (DSS) of the two groups was compared by KM analysis. As part of the evaluation of the prognostic ability of the risk model, a time-related receiver operating characteristic (ROC) analysis of 1, 3, and 5 years was performed. In the above analysis, GSE31210 was selected as the external validated cohort through the GEO database.

The TCGA-LUAD and GSE31210 datasets were analyses using both univariate and multivariate Cox regression analysis. According to the constructed risk prognostic signature, as well as clinical characteristics (such as gender, age, TNM stage) of LUAD patients and by using the rams package in R version 6.2.0, build a nomogram model. We also evaluated the accuracy of the nomogram using the calibration curve, and the diagnostic accuracy of the model using the decision curve analysis (DCA). A KM curve was drawn after dividing patients into high-score and low-score groups based on their median total score in the column chart.

2.6 | Immunohistochemical (IHC)

To further validate our lysosomal central gene markers, we initiated additional experiments to closely examine the expression patterns of CTSZ and AP3S2 proteins, particularly in a clinical context. Using 48 stage 3–4 NSCLC samples (Table S5), we performed IHC staining to carefully assess CTSZ and AP3S2 expression levels. Data collection followed CONSORT guidelines, including TNM stage, sex, and age. The specific method is as follows: The tissue sample is fixed in formalin, embedded in paraffin, and cut into 4–5 μm slices. The slices were then

dewaxed, hydrated, and used for antigen retrieval using EDTA buffers. After being blocked with normal serum, the tissues were incubated overnight at 4°C with anti-CTSZ antibodies (Sigma-Aldrich, HPA049876, 1:100 dilution) and anti-AP3S2 antibodies (Sigma-Aldrich, HPA049270, 1:100 dilution), respectively. A peroxidase-conjugated secondary antibody is applied for 1 h, and staining is developed using a peroxidase substrate. Counterstaining with hematoxylin, dehydration, clearing in xylene, and mounting with mounting medium complete the process. The slides are examined under a microscope to identify CTSZ and AP3S2 protein staining, with negative and positive controls included for validation.

2.7 | Statistical Analysis

We used *R* for all statistical analyses conducted in this study. To ensure robustness and accuracy, we employed a variety of statistical tests and utilized relevant *R* packages tailored to our specific needs.

For comparing nonparametric data from two independent samples, we utilized the Wilcoxon test. This test was particularly useful for assessing differences in immune cell infiltration ratios, TIME scores, and expression levels of immune checkpoint and HLA family genes among different lysosome subtypes.

For parametric data, we employed both *t*-tests and one-way ANOVA. *T*-tests were used to compare means between two groups, such as comparing the lysosome scores or clinical characteristics between different clusters. One-way ANOVA was utilized when comparing means across three or more groups, such as analyzing the expression levels of lysosomal-related genes across various subtypes or stages of LUAD.

We considered a $p < 0.05$ as statistically significant, with $p < 0.01$ and < 0.001 denoted as ** and *** respectively, to indicate higher levels of significance. These thresholds helped us to identify which results were likely to be biologically meaningful and worthy of further exploration.

To facilitate our statistical analyses and data visualization, we utilized a range of R packages. We employed packages such as ggplot2 and ggpubr for creating visually appealing and informative graphs and plots. The survival and survminer packages were essential for performing survival analyses, including Kaplan-Meier curves and Cox regression models. Additionally, we utilized packages from the Bioconductor repository, such as ConsensusClusterPlus for cluster analysis and GSVA for gene set variation analysis.

3 | Results

3.1 | Identification of Lysosome-Related DEGS Between Tumor Tissue and Adjacent Sample

The lysosome-related DEGs between tumor tissue and adjacent tissue were compared. A total of 109 DEGs were found, including 68 downregulated genes and 41 upregulated genes

(Figure 1A, Table S2). Univariate Cox analysis was performed on the 109 DEGs obtained above, and 30 genes with significant prognostic correlation were identified (Figure 1B).

3.2 | Identification of Lysosome-Based Molecular Clusters Using Clustering Analysis

To identify lysosome-based molecular clusters, cluster analysis was performed on 30 prognostic-related DGEs. The maximum number of clusters, K, was set to 6. Based on the CDF curve and Delta area, two lysosome-based molecular clusters were constructed (Figure 2A,B). The LYSOSOME score of three LYSO-SOME gene sets (KEGG_LYSOSOME, LYSOSOME, and total LYSOSOME gene sets) was calculated respectively, and it was found that in cluster 2, the LYSOSOME score was significantly reduced (Figure 2C). More importantly, cluster 2 had a worse prognosis (median DSS: 1492 vs 1288 days, HR: 1.61, 95% CI: 1.09–2.39, $p = 0.016$, Figure 2D). The clinical characteristics of cluster 1 and cluster 2 are shown in Figure 2E, and there were no statistical differences in age, gender and stage (Table S3).

3.3 | Cluster-Based Analysis of Tumor Immune Microenvironment

Based on immune score, stromal score, and tumor purity scores, we assessed the differences in immune characteristics among the subtypes using the "ESTIMATE" R package. Compared to cluster 1, cluster 2 had lower immunological and stromal scores, but higher tumor purity levels (Figure 3A). According to these findings, immune and stromal components had a negative correlation with prognosis in LUAD. We utilized the CI-BERSORTx online tool for further exploration of immune cell distribution in the microenvironment of tumors. Compared to cluster 1, memory B cell, resting-memory CD4 T cells, mono-cytes, M2-type macrophages, dendritic cells and mast cells in cluster 2 were significantly reduced. Activated-memory CD4 T cells, M0-type macrophages, and M1-type macrophages, how-ever, showed significant increases (Figure 3B,C). Additionally, we analyzed the immune checkpoint genes and HLA genes and found that the expression of *HAVCR2*, *CD274*, *PDCD1LG2*, *CTLA4*, *CD80* and *CD86* were downregulated in cluster 2, and all HLA genes showed the same trend (Figure 3D,E, all $p < 0.05$). Obviously, patients in cluster 2 were im-munosuppressed in a more severe way.

To investigate the potential biological function of cluster 2, we conducted gene set enrichment analysis of upregulated and downregulated genes of cluster 2. The leading 10 significantly enriched gene sets in the upregulated cluster 2 were E2F_TAR GETS, G2M_CHECKPOINT, MYC_TARGETS_V1, MYC_TAR GETS_V2, MTORC1_SIGNALING, MITOTIC_SPINDLE, UNFOL DED_PROTEIN_RESPONSE, GLYCOLYSIS, DNA_REPAIR and HYPOXIA ($p < 0.05$, Figure 3F). In addition, 21 pathways including INFLAMMATORY_RESPONSE, IL6_JAK_STAT3_ SIGNALING, TNFA_SIGNALING_VIA_NFKB, METABOLISM, and P53_PATHWAY were also activated in cluster 2 (Supplementary Figure S1).
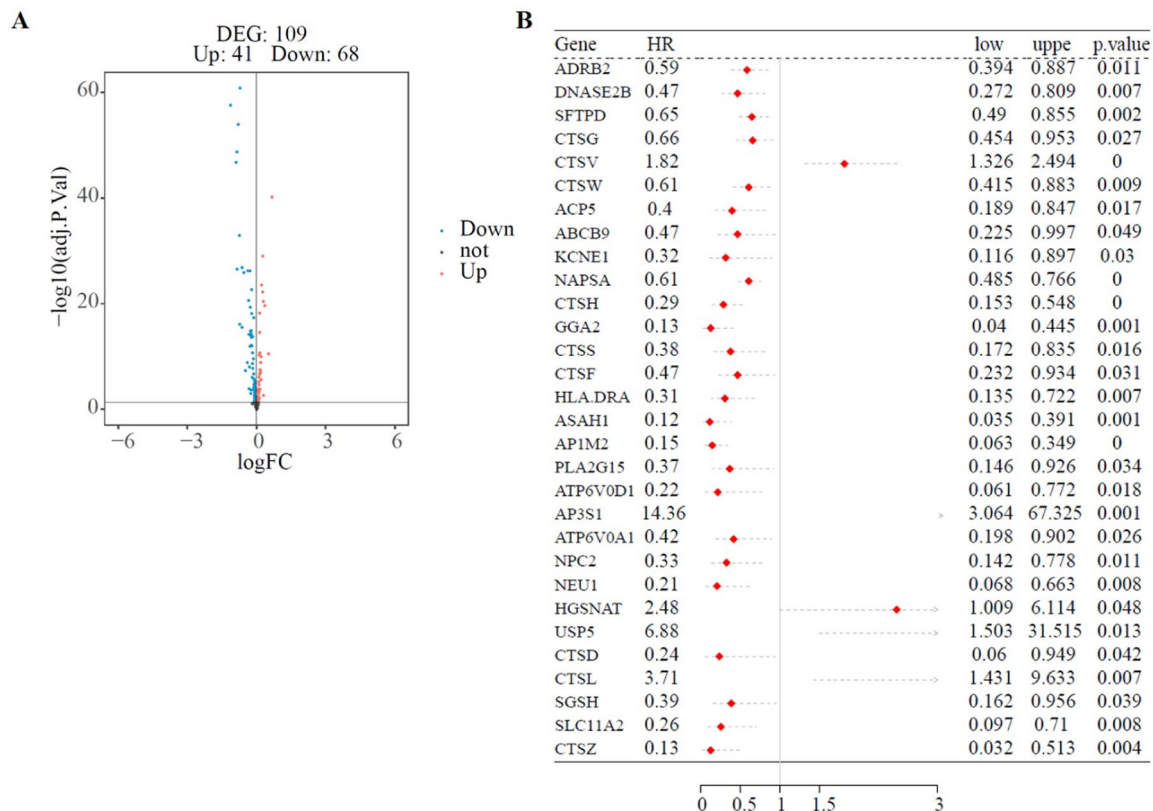

A

DEG: 109
Up: 41 Down: 68

B

Gene	HR		low	uppe	p.value
ADRB2	0.59		0.394	0.887	0.011
DNASE2B	0.47		0.272	0.809	0.007
SFTPD	0.65		0.49	0.855	0.002
CTSG	0.66		0.454	0.953	0.027
CTSV	1.82		1.326	2.494	0
CTSW	0.61		0.415	0.883	0.009
ACP5	0.4		0.189	0.847	0.017
ABCB9	0.47		0.225	0.997	0.049
KCNE1	0.32		0.116	0.897	0.03
NAPSA	0.61		0.485	0.766	0
CTSH	0.29		0.153	0.548	0
GGA2	0.13		0.04	0.445	0.001
CTSS	0.38		0.172	0.835	0.016
CTSF	0.47		0.232	0.934	0.031
HLA.DRA	0.31		0.135	0.722	0.007
ASAH1	0.12		0.035	0.391	0.001
AP1M2	0.15		0.063	0.349	0
PLA2G15	0.37		0.146	0.926	0.034
ATP6V0D1	0.22		0.061	0.772	0.018
AP3S1	14.36		3.064	67.325	0.001
ATP6V0A1	0.42		0.198	0.902	0.026
NPC2	0.33		0.142	0.778	0.011
NEU1	0.21		0.068	0.663	0.008
HGSNAT	2.48		1.009	6.114	0.048
USP5	6.88		1.503	31.515	0.013
CTSD	0.24		0.06	0.949	0.042
CTSL	3.71		1.431	9.633	0.007
SGSH	0.39		0.162	0.956	0.039
SLC11A2	0.26		0.097	0.71	0.008
CTSZ	0.13		0.032	0.513	0.004

FIGURE 1 | The role of lysosome-related genes in tumor and adjacent tissue. (A) DEGs of lysosome-related genes between tumor and adjacent tissue. (B) Univariate Cox analysis.
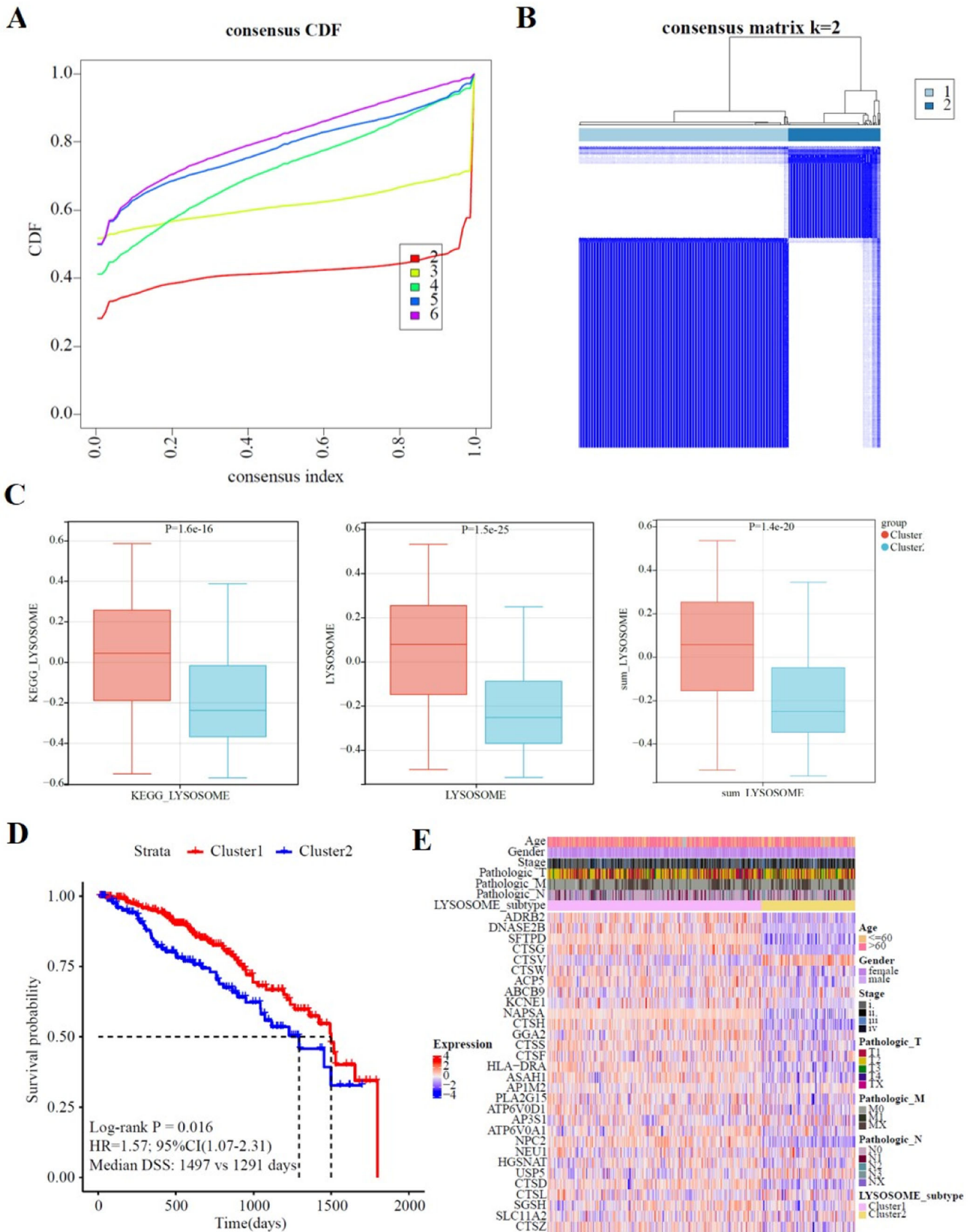
FIGURE 2 | Consensus clustering of molecular subgroups based on prognostic related DEGs. (A) CDF curve. (B) Consensus clustering matrix with K as 2. (C). LYSOSOME score in KEGG_LYSOSOME, LYSOSOME, and total LYSOSOME gene sets, respectively. (D) KM survival curve. (E) Distribution of clinical characteristics between two clusters. ****$p < 0.0001$.
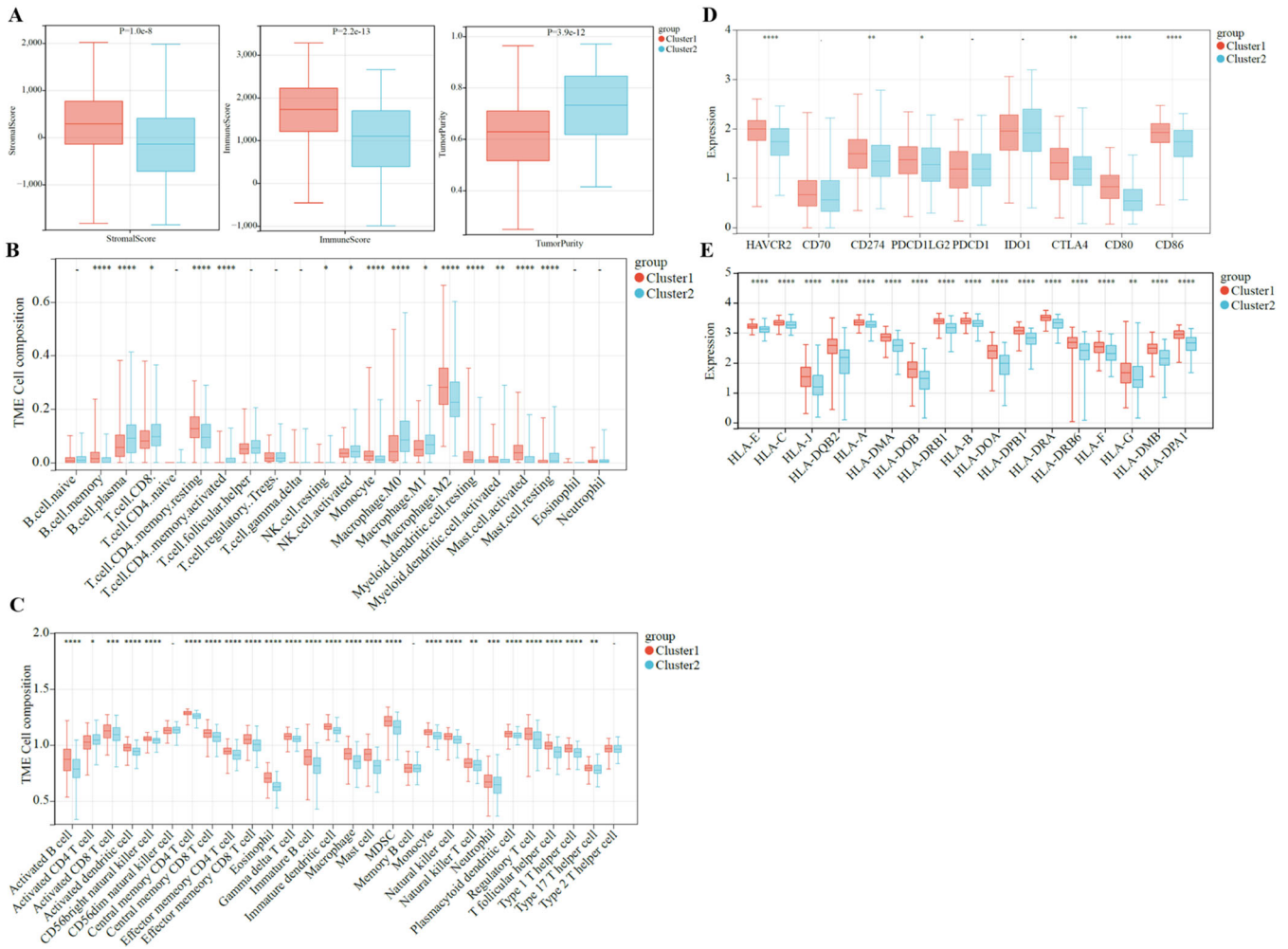
FIGURE 3 | Cluster-based analysis of tumor immune microenvironment. (A) Comparison of tumor immune microenvironment components. (B) Box plots present differential immune infiltration. (C) ssGSEA results of differential immune infiltration. (D) Immune checkpoint genes expression. (E) HLA family genes expression. (F) Significantly enriched pathways with GSEA between cluster 1 and cluster 2. *$p < 0.05$. ***$p < 0.01$. ****$p < 0.0001$.

## 3.4 | Construction and Validation of Lysosome-Related Prognostic Signature

The DEGs-based univariate Cox analysis among three subtypes identified 17 prognostic genes (Table S4, Figure 4A). By using univariate Cox analysis, prognostic genes were analyzed with LASSO regression. Figure 4B shows the locus of each independent variable. As the lambda (λ) value increases, the number of independent variables approaching zero also increases (Figure 4C).

Each patient's risk score and survival time are shown in Figure 4D,E. Patients were grouped into high-risk groups based on their risk scores. As shown by the KM curve, low-risk patients have significantly better prognoses than high-risk patients (median DSS: not reached vs. 1288 days, HR: 0.33, 95% CI: 0.22–0.50, $p < 0.001$, Figure 4F). The area under the curves (AUCs) of time-dependent ROC curves for 1-, 3-, and 5-year DSS were 0.78, 0.72, and 0.69, respectively (Figure 4G), indicating a good predictive performance. A similar phenomenon also appeared in the GSE31210 cohort (Figure 4H,I). A significant difference was found between low-risk and high-risk

patients in their prognosis (median DSS: not reached vs not reached, HR: 0.35, 95% CI: 0.17–0.73, $p = 0.0033$, Figure 4J), and the ROC curves of 1-year, 3-year, and 5-year DSS were 0.65, 0.64, and 0.70, respectively (Figure 4K).

3.5 | Developing a Predictive Nomogram for DSS Prediction

Cox regression was used to analyze clinical information and risk scores to identify prognostic factors. It was found that risk scores and TNM stage were associated with a poor prognosis ($p < 0.05$, Figure 5A,B). To help predict clinical outcomes accurately, a prediction nomogram was developed (Figure 5C). A nomogram calibration map showed a good match between predicted and observed DSS results, and 1-year calibration curves showed good consistency, but as time went on, some deviation occurred (Figure 5D). At the same time, we also plotted DCA and ROC curves of nomogram, demonstrating a better discriminative ability than signature (Figure 5E,F). According to the median value of the nomogram score, the high-score group and low-score group were distinguished, and the
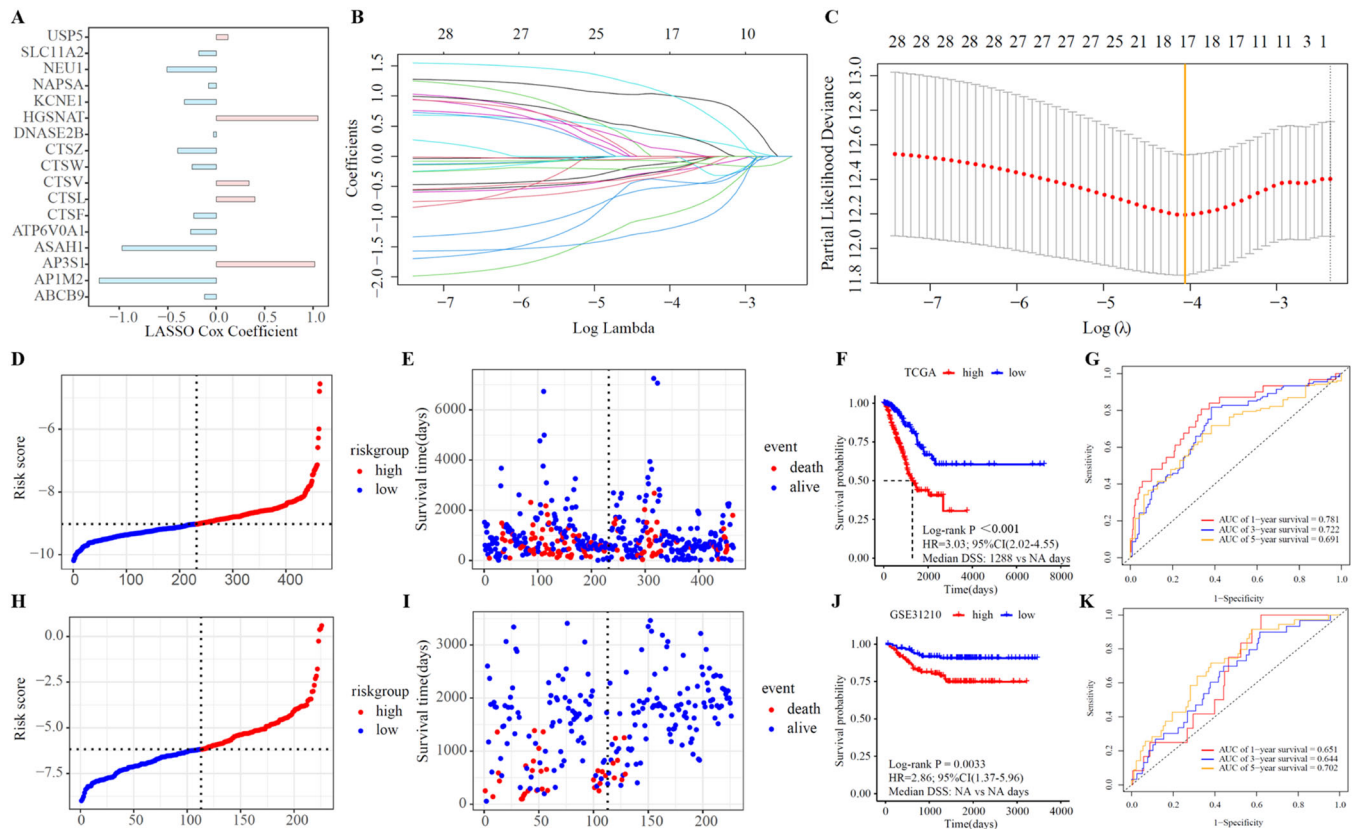
F
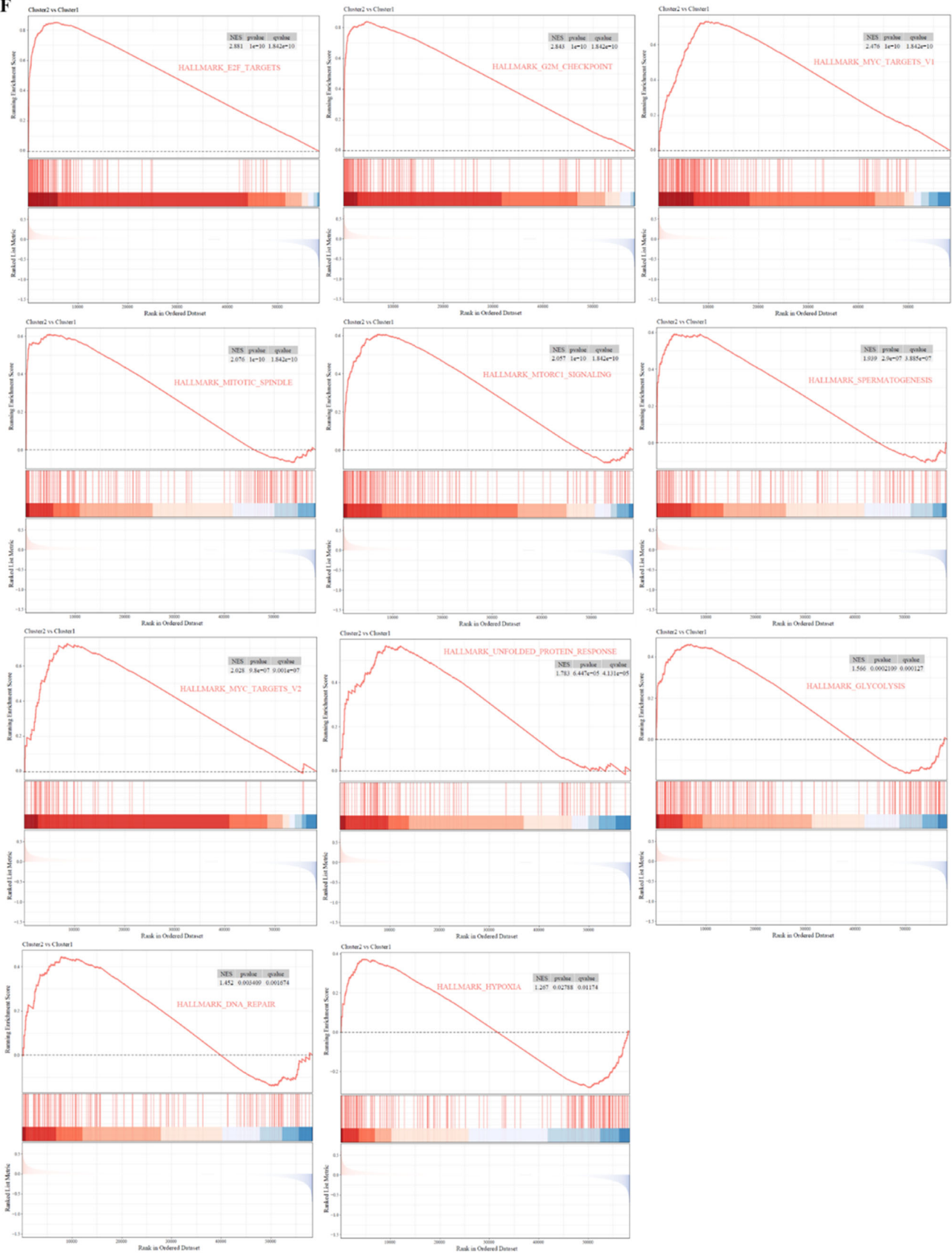
FIGURE 3 | (Continued)

7 of 11

FIGURE 4 | Prognostic model of LUAD based on prognostic genes. (A) Histogram of coefficient distribution of feature genes. (B) LASSO coefficient profiles. (C) A plot of the error rates from 10-fold cross-validation. Distribution of risk score (D), survival time (E) for each patient, KM survival curve (F) and ROC curve (G) of the predictive value of the risk model in the TCGA cohort. Distribution of risk score (H), survival time (I) for each patient, KM survival curve (J) and ROC curve (K) of the predictive value of the risk model in the GSE31210 cohort.
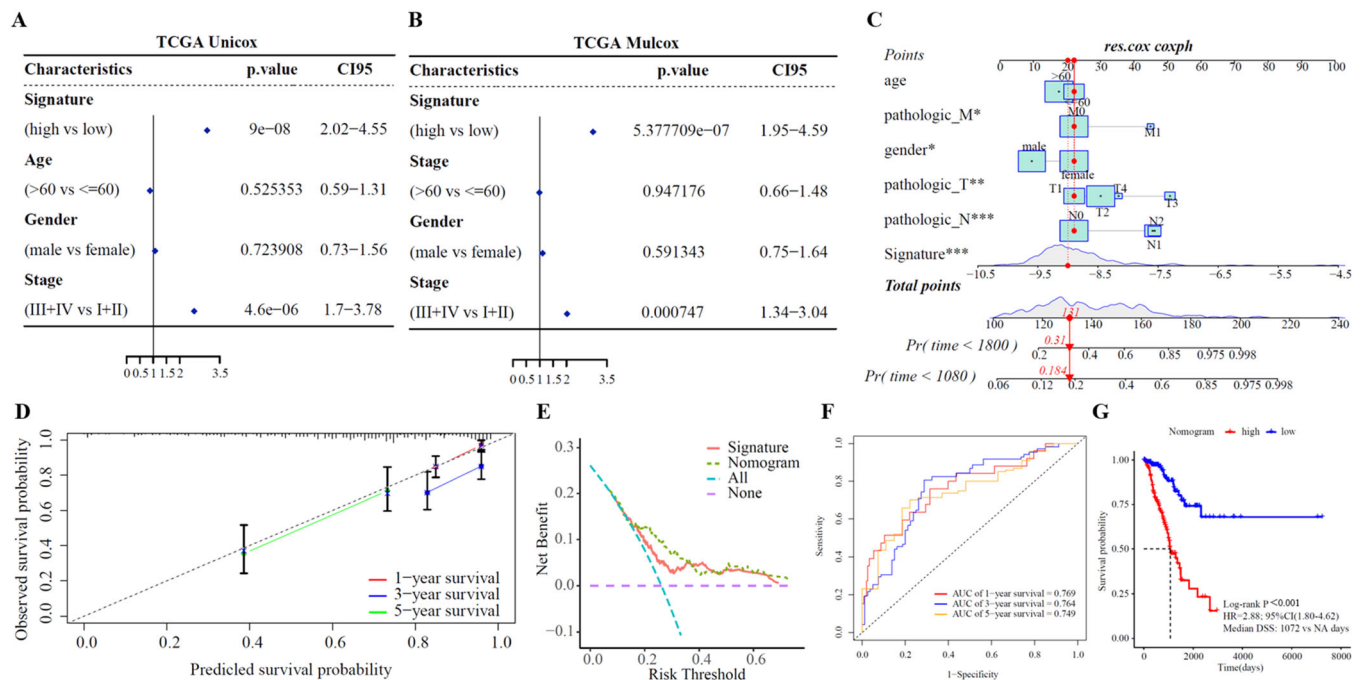

FIGURE 5 | Construction of the nomogram for predicting DSS of LUAD patients. Univariate forest plots (A) and multivariate forest plots (B) of the risk score model and clinicopathological characteristics associated with DSS. (C) The nomogram was constructed. (D) Calibration plot of the nomogram. (E) DCA curve. (F) ROC curve. (G) KM survival curve.

prognostic survival curve was drawn. The results showed that patients with a low score had a better prognosis ($p < 0.001$, Figure 5G).

3.6 | Expression of CTSZ and AP3S2 Proteins

We examined the expression of CTSZ and AP3S2 proteins in 48 stage 3–4 NSCLC samples. Results showed that low AP3S2 expression was a favorable prognostic factor (median OS: 37.87 vs. 8.53 months, HR: 0.51, 95% CI: 0.24–1.05, $p = 0.0211$, Figure S2A). Increased CTSZ expression was a favorable prognostic indicator (median OS: 6.77 vs 30.50 months, HR: 2.66, 95% CI: 0.65–10.87, $p = 0.0306$, Figure S2B,C).

4 | Discussion

The lysosome is a crucial component of the cell membrane and participates in a wide range of biological processes within the cell, including intracellular transport, autophagy, metabolism, cell division, cell migration, and gene expression [14–16]. In recent years, increasing evidence has shown that lysosomes play an important role in the occurrence and progression of cancer [17–19]. For example, in cholangiocarcinoma, *PTEN* deficiency impairs lysosomal biogenesis in a TFEB phosphorylation-dependent manner, promoting exosome secretion and cancer metastasis [18]. Gefitinib selectively degrades COX6A1, an important antiapoptotic factor in the autophagy-lysosome pathway, and then activates apoptosis, thereby causing liver injury [20]. Moreover, some studies have found that constructing a lysosome-related gene signature, combined with Gleason score, could well predict the prognosis of prostate cancer [21]. Although the role of lysosomes in cancer initiation and progression is becoming increasingly apparent, their exact role in predicting the prognosis of LUAD remains unclear. In recent years, TIME has gained widespread attention in the cancer research arena. Stromal cells, as well as tumor cells, contribute to cancer's initiation, development, and metastasis. Some researchers even believe that stroma, as a mutagenic agent, plays a more direct role in tumorigenesis, potentially slowing or reversing tumor progression by "normalizing" tumor stroma [22].

In light of these characteristics, we established a prognostic model of lysosome-related genes that can provide potential therapeutic targets for clinical treatment and prognosis in LUAD patients through bioinformatics technology.

We analyzed immune scores between class 1 and class 2 subgroups and found significant differences in tumor-infiltrating lymphocytes (TILs), immune checkpoint genes, and major histocompatibility complex. In general, the efficacy of clinical treatment depends on the quantity and nature of TILs, which is a favorable prognostic factor for NSCLC [23]. TILs have been shown to be associated with a higher neoadjuvant pathologic response or better prognosis [24, 25]. In HLA, it was found that the infiltration of T cells and NK cells was reduced in patients with selective downregulation of HLA-related genes in cancer cells, and the OS was shortened [26]. This is consistent with our

results that patients in the cluster 2 subgroup had a lower lysosome score, a more severe immunosuppressive state, and a poorer prognosis.

In this study, we further found that 17 lysosome-related genes were significantly associated with LUAD prognosis through Cox analysis. Notably, most of them have rarely been studied for cancer-related diseases, and some genes have been the subject of several studies. Firstly, adapter related protein complex 1 subunit mu 2 (AP1M2) has been shown to be significantly overexpressed in a variety of solid tumors (e.g., breast cancer, liver cancer, lung cancer, cholangiocarcinoma, prostate cancer, gastric cancer, thyroid cancer, and common genital tumors) and associated with cancer cell death [27]. High expression of AP1M2 was linked to poor prognosis in breast-infiltrating carcinoma ($p = 0.039$) and cutaneous melanoma ($p = 0.0015$) [28]. Regarding adapter related protein complex 3 subunit sigma 1 (*AP3S1*), researchers found that AP3S1 was overexpressed in the majority of tumors and significantly associated with lower survival. To explain this phenomenon, its expression is positively correlated with the level of infiltration of immunosuppressive cells (tumor-associated macrophages, cancer-associated fibroblasts, Tregs) and negatively correlated with the level of immune killer cells (including NK cells and CD8+ T cells) [29]. Similarly, in patients with LUAD, AP3S1 promotes the growth and migration of LUAD cells in vitro [30]. As for cathepsin Z Gene (CTSZ), Liu et al. demonstrated that the natural compound deguelin inhibits the migration, invasion and metastasis of NSCLC cells in vitro and in vivo by inhibiting CTSZ expression [31]. To further validate our lysosome-centric gene signature, we embarked on additional experiments, particularly scrutinizing the expression patterns of CTSZ and AP3S2 proteins in a clinical context. Our comprehensive IHC analysis unveiled a significant correlation between CTSZ protein expression and the prognosis of NSCLC patients, as illustrated in Supplementary Figure S2. Notably, elevated CTSZ expression emerged as a favorable prognostic indicator, aligning with previous findings across diverse cancer types such as kidney renal papillary cell carcinoma, glioblastoma multiforme, and childhood acute myeloid leukemia, where similar trends have been reported for CTSZ [32–34]. However, we acknowledge the existence of contrasting reports, notably in kidney renal clear cell carcinoma [34], emphasizing the intricate and disease-specific nature of gene expression patterns and their prognostic implications. These inconsistencies underscore the paramount importance of adopting multi-gene signatures to mitigate the limitations inherent in single-gene predictors, thereby enhancing the accuracy and reliability of prognosis in cancer.

Our study offers significant prognostic insights for LUAD; however, it does have some limitations. First, the data utilized in our research is retrospective, and incorporating prospective samples will be essential to validate our findings. Second, while we emphasize clinical prognosis, further investigation into the specific molecular mechanisms is needed. Third, as our study primarily involves bioinformatics, further research is needed for direct experimental validation.

In summary, our study classified NSCLC patients into two clusters based on lysosome-related genes. Functional analysis

and immune scoring indicated that a low lysosome score was correlated with a suppressed immune status and poor prognosis. Additionally, we developed a corresponding risk model aimed at enhancing clinical treatment strategies and providing theoretical support for future research.

Author Contributions

Conceptualization: Wei Ye and Tong Zhou. Methodology, software, validation: Wei Ye and Cong Fu. Writing–original draft preparation: Wei Ye. Writing–review and editing: Tong Zhou. Formal analysis, investigation, resources: Lin Sun. Data curation: Huajie Dong. Visualization, supervision, project administration: Tong Zhou. All authors have read and approved the final version of the manuscript (corresponding author or manuscript guarantor) had full access to all of the data in this study and takes complete responsibility for the integrity of the data and the accuracy of the data analysis.

Ethics Statement

TCGA and GEO belong to public databases. The patients involved in the database have obtained ethical approval. Users can download relevant data for free for research and publish relevant articles. The studies involving NSCLC participants were reviewed and approved by Changzhou Tumor Hospital Affiliated to Soochow University. The patients/participants provided their written informed consent to participate in this study.

Conflicts of Interest

The authors declare no conflicts of interest.

Data Availability Statement

The data used in this study are available in the manuscript and Supplementary File. For additional inquiries, please contact the corresponding author directly via email.

References

1. R. L. Siegel, K. D. Miller, N. S. Wagle, and A. Jemal, "Cancer Statistics, 2023," *CA: A Cancer Journal for Clinicians* 73, no. 1 (2023): 17–48, https://doi.org/10.3322/caac.21763.

2. R. Zheng, S. Zhang, H. Zeng, et al., "Cancer Incidence and Mortality in China, 2016," *Journal of the National Cancer Center* 2, no. 1 (2022): 1–9, https://doi.org/10.1016/j.jncc.2022.02.002.

3. R. S. Herbst, D. Morgensztern, and C. Boshoff, "The Biology and Management of Non-Small Cell Lung Cancer," *Nature* 553 (2018): 446–454, https://doi.org/10.1038/nature25183.

4. F. Lou, J. Huang, C. S. Sima, J. Dycoco, V. Rusch, and P. B. Bach, "Patterns of Recurrence and Second Primary Lung Cancer in Early-Stage Lung Cancer Survivors Followed With Routine Computed Tomography Surveillance," *The Journal of Thoracic and Cardiovascular Surgery* 145 (2013): 75–82, https://doi.org/10.1016/j.jtcvs.2012.09.030.

5. Z. Piotrowska, H. Isozaki, J. K. Lennerz, et al., "Landscape of Acquired Resistance to Osimertinib in EGFR-Mutant NSCLC and Clinical Validation of Combined EGFR and RET Inhibition With Osimertinib and BLU-667 for Acquired RET Fusion," *Cancer Discovery* 8, no. 12 (2018): 1529–1539, https://doi.org/10.1158/2159-8290.CD-18-1022.

6. J. F. Gainor, L. Dardaei, S. Yoda, et al., "Molecular Mechanisms of Resistance to First- and Second-Generation ALK Inhibitors in ALK-Rearranged Lung Cancer," *Cancer Discovery* 6, no. 10 (2016): 1118–1133, https://doi.org/10.1158/2159-8290.CD-16-0596.

7. H. Appelqvist, P. Wäster, K. Kågedal, and K. Öllinger, "The Lysosome: From Waste Bag to Potential Therapeutic Target," *Journal of Molecular Cell Biology* 5, no. 4 (2013): 214–226, https://doi.org/10.1093/jmcb/mjt022.

8. C. Di Malta, D. Siciliano, A. Calcagni, et al., "Transcriptional Activation of RagD GTPase Controls mTORC1 and Promotes Cancer Growth," *Science* 356, no. 6343 (2017): 1188–1192, https://doi.org/10.1126/science.aag2553.

9. H. Zhang, Q. Wei, R. Liu, et al., "Overexpression of LAPTM4B-35: A Novel Marker of Poor Prognosis of Prostate Cancer," *PLoS One* 9, no. 3 (2014): e91069, https://doi.org/10.1371/journal.pone.0091069.

10. Q. Zhu, Y. Meng, S. Li, et al., "Association of Genetic Variants in Autophagy-Lysosome Pathway Genes With Susceptibility and Survival to Prostate Cancer," *Gene* 808 (2022): 145953, https://doi.org/10.1016/j.gene.2021.145953.

11. A. Shiralipour, B. Khorsand, L. Jafari, et al., "Identifying Key Lysosome-Related Genes Associated With Drug-Resistant Breast Cancer Using Computational and Systems Biology Approach," *Iranian Journal of Pharmaceutical Research* 21, no. 1 (2022): e130342, https://doi.org/10.5812/ijpr-130342.

12. N. Haratake, Q. Hu, T. Okamoto, et al., "Identification of SLC38A7 as a Prognostic Marker and Potential Therapeutic Target of Lung Squamous Cell Carcinoma," *Annals of Surgery* 274, no. 3 (2021): 500–507, https://doi.org/10.1097/SLA.0000000000005001.

13. A. Girigoswami and K. Girigoswami, "Potential Applications of Nanoparticles in Improving the Outcome of Lung Cancer Treatment," *Genes* 14, no. 7 (2023): 1370, https://doi.org/10.3390/genes14071370.

14. L. Yu, Y. Chen, and S. A. Tooze, "Autophagy Pathway: Cellular and Molecular Mechanisms," *Autophagy* 14, no. 2 (2018): 207–215, https://doi.org/10.1080/15548627.2017.1378838.

15. L. Bar-Peled, L. Chantranupong, A. D. Cherniack, et al., "A Tumor Suppressor Complex With Gap Activity for the RAG GTPases That Signal Amino Acid Sufficiency to mTORC1," *Science* 340, no. 6136 (2013): 1100–1106, https://doi.org/10.1126/science.1232044.

16. R. E. Lawrence and R. Zoncu, "The Lysosome as a Cellular Centre for Signalling, Metabolism and Quality Control," *Nature Cell Biology* 21, no. 2 (2019): 133–142, https://doi.org/10.1038/s41556-018-0244-7.

17. L. Brisson, P. Bański, M. Sboarina, et al., "Lactate Dehydrogenase B Controls Lysosome Activity and Autophagy in Cancer," *Cancer Cell* 30, no. 3 (2016): 418–431, https://doi.org/10.1016/j.ccell.2016.08.005.

18. T. Y. Jiang, Y. Y. Shi, X. W. Cui, et al., "PTEN Deficiency Facilitates Exosome Secretion and Metastasis in Cholangiocarcinoma by Impairing TFEB-Mediated Lysosome Biogenesis," *Gastroenterology* 164, no. 3 (2023): 424–438, https://doi.org/10.1053/j.gastro.2022.11.025.

19. M. Jacquet, M. Guittaut, A. Fraichard, and G. Despouy, "The Functions of Atg8-Family Proteins in Autophagy and Cancer: Linked or Unrelated?," *Autophagy* 17, no. 3 (2021): 599–611, https://doi.org/10.1080/15548627.2020.1749367.

20. P. Luo, H. Yan, J. Du, et al., "PLK1 (Polo Like Kinase 1)-Dependent Autophagy Facilitates Gefitinib-Induced Hepatotoxicity by Degrading COX6A1 (Cytochrome C Oxidase Subunit 6A1)," *Autophagy* 17, no. 10 (2021): 3221–3237, https://doi.org/10.1080/15548627.2020.

21. Y. Huang, F. Yang, W. Zhang, et al., "A Novel Lysosome-Related Gene Signature Coupled With Gleason Score for Prognosis Prediction in Prostate Cancer," *Frontiers in Genetics* 14 (2023): 1135365, https://doi.org/10.3389/fgene.2023.1135365.

22. M. M. Mueller and N. E. Fusenig, "Friends or Foes - Bipolar Effects of the Tumour Stroma in Cancer," *Nature Reviews Cancer* 4, no. 11 (2004): 839–849, https://doi.org/10.1038/nrc1477.

23. L. Federico, D. J. McGrail, S. E. Bentebibel, et al., "Distinct Tumor-Infiltrating Lymphocyte Landscapes Are Associated With Clinical Outcomes in Localized Non-Small-Cell Lung Cancer," *Annals of Oncology* 33, no. 1 (2022): 42–56, https://doi.org/10.1016/j.annonc.2021.09.021.

24. J. D. Fumet, C. Richard, F. Ledys, et al., "Prognostic and Predictive Role of CD8 and PD-L1 Determination in Lung Tumor Tissue of Patients Under anti-PD-1 Therapy," *British Journal of Cancer* 119, no. 8 (2018): 950–960, https://doi.org/10.1038/s41416-018-0220-9.

25. P. M. Forde, J. E. Chaft, K. N. Smith, et al., "Neoadjuvant PD-1 Blockade in Resectable Lung Cancer," *New England Journal of Medicine* 378, no. 21 (2018): 1976–1986, https://doi.org/10.1056/NEJMoa1716078.

26. I. J. Datar, S. C. Hauc, S. Desai, et al., "Spatial Analysis and Clinical Significance of Hla Class-I and Class-II Subunit Expression in Non-Small Cell Lung Cancer," *Clinical Cancer Research* 27, no. 10 (2021): 2837–2847, https://doi.org/10.1158/1078-0432.CCR-20-3655.

27. M. Muthu, S. Chun, J. Gopal, et al., "The MUDENG Augmentation: A Genesis in Anti-Cancer Therapy?," *International Journal of Molecular Sciences* 21, no. 15 (2020): 5583, https://doi.org/10.3390/ijms21155583.

28. Y. Yi, Q. Zhang, Y. Shen, et al., "System Analysis of Adaptor-Related Protein Complex 1 Subunit Mu 2 (AP1M2) on Malignant Tumors: A Pan-Cancer Analysis," *Journal of Oncology* 2022 (2022): 7945077, https://doi.org/10.1155/2022/7945077.

29. G. Wu, M. Chen, H. Ren, et al., "AP3S1 is a Novel Prognostic Biomarker and Correlated With an Immunosuppressive Tumor Microenvironment in Pan-Cancer," *Frontiers in Cell and Developmental Biology* 10 (2022): 930933, https://doi.org/10.3389/fcell.2022.930933.

30. C. Qian, Z. Jiang, T. Zhou, et al., "Vesicle-Mediated Transport-Related Genes are Prognostic Predictors and Are Associated With Tumor Immunity in Lung Adenocarcinoma," *Frontiers in Immunology* 13 (2022): 1034992, https://doi.org/10.3389/fimmu.2022.1034992.

31. W. Li, X. Yu, X. Ma, et al., "Deguelin Attenuates Non-Small Cell Lung Cancer Cell Metastasis Through Inhibiting the CtsZ/FAK Signaling Pathway," *Cellular Signalling* 50 (2018): 131–141, https://doi.org/10.1016/j.cellsig.2018.07.001.

32. S. S. Redekar, S. L. Varma, and A. Bhattacharjee, "Gene Co-Expression Network Construction and Analysis for Identification of Genetic Biomarkers Associated With Glioblastoma Multiforme Using Topological Findings," *Journal of the Egyptian National Cancer Institute* 35, no. 1 (2023): 22, https://doi.org/10.1186/s43046-023-00181-4.

33. N. Zhang, Y. Chen, S. Lou, Y. Shen, and J. Deng, "A Six-Gene-Based Prognostic Model Predicts Complete Remission and Overall Survival in Childhood Acute Myeloid Leukemia," *OncoTargets and Therapy* 12 (2019): 6591–6604, https://doi.org/10.2147/OTT.S218928.

34. F. Zhang, J. Liang, Y. Lu, et al., "Macrophage-Specific Cathepsin as a Marker Correlated With Prognosis and Tumor Microenvironmental Characteristics of Clear Cell Renal Cell Carcinoma," *Journal of Inflammation Research* 15 (2022): 6275–6292, https://doi.org/10.2147/JIR.S375250.

Supporting Information

Additional supporting information can be found online in the Supporting Information section.