



OPEN ACCESS

EDITED BY

Prashanth Ashok Kumar,
George Washington University Hospital,
United States

REVIEWED BY

Zhe Wang,
Shanghai University of Traditional Chinese
Medicine, China
Yang Pan,
Zhejiang Cancer Hospital, China

*CORRESPONDENCE

Zheyuan Xu,
✉ khyyxwky@163.com
Hao Peng,
✉ hao9375@163.com

[†]These authors have contributed equally to
this work

RECEIVED 12 July 2025

ACCEPTED 06 October 2025

PUBLISHED 16 October 2025

CITATION

Chen Y, Li P, Wang Y, Shen S, Chen N, Peng H
and Xu Z (2025) Identification and validation of
selenium metabolism-related genes in lung
adenocarcinoma prognosis using
bioinformatics analysis.
Front. Genet. 16:1655262.
doi: 10.3389/fgene.2025.1655262

COPYRIGHT

© 2025 Chen, Li, Wang, Shen, Chen, Peng and
Xu. This is an open-access article distributed
under the terms of the [Creative Commons
Attribution License \(CC BY\)](#). The use,
distribution or reproduction in other forums is
permitted, provided the original author(s) and
the copyright owner(s) are credited and that the
original publication in this journal is cited, in
accordance with accepted academic practice.
No use, distribution or reproduction is
permitted which does not comply with these
terms.

Identification and validation of selenium metabolism-related genes in lung adenocarcinoma prognosis using bioinformatics analysis

Yun Chen^{1†}, Ping Li^{2†}, Yang Wang¹, Shuai Shen¹, Ni Chen¹,
Hao Peng^{1*} and Zheyuan Xu^{1*}

¹Department of Thoracic Surgery, the First People's Hospital of Yunnan Province, The Affiliated Hospital of Kunming University of Science and Technology, Kunming, China, ²Department of Cardiovascular Medicine, The First People's Hospital of Kunming City & Ganmei Affiliated Hospital of Kunming Medical University, Kunming, China

Background: The disruption of selenium metabolism has been associated with tumor progression. However, the prognostic significance and underlying molecular mechanisms of selenium metabolism in lung adenocarcinoma (LUAD) remain inadequately understood. This study primarily aimed to identify and validate prognostic genes related to selenium metabolism in LUAD patients.

Methods: Transcriptomic datasets from patients diagnosed with LUAD were meticulously analyzed to identify differentially expressed genes associated with selenium metabolism. The genes selected for the prognostic risk model were determined through various analyses, including differential gene expression assessment, univariate and multivariate Cox proportional hazards regression analyses, as well as other relevant analytical methods. A systematic approach was employed for functional enrichment analysis, characterization of the immune microenvironment, somatic mutation analysis, and evaluation of drug sensitivity to elucidate the mechanisms linked to prognostic genes and risk categories. Finally, a reverse transcription quantitative PCR (RT-qPCR) assay was conducted to validate the expression levels of the identified prognostic genes.

Results: F2, GPX3, KMO, and KYNU were identified as prognostic genes for establishing a risk model. The functions of these LUAD prognostic genes were influenced by DNA replication pathways, cell cycle regulation, and quiescent CD4 memory T cells. In the high-risk group (HRG), KEAP1, TTN, and USH2A exhibited the highest mutation rate at 48%, while TTN had an even higher mutation rate of 52% in the low-risk group (LRG). Within the HRG cohort, both cisplatin and gemcitabine demonstrated significant sensitivity. Ultimately, RT-qPCR findings corroborated results obtained from bioinformatics analyses; specifically compared to normal samples: GPX3, KMO, KYNU showed significant downregulation in LUAD tissues while F2 was found to be upregulated in LUAD.

Conclusion: This study identified four prognostic genes in LUAD and examined their associated mechanisms of action, which may contribute to the

development of novel treatment strategies. The integration of immune characterization with drug sensitivity analysis offers valuable insights for stratified therapy.

KEYWORDS

lung adenocarcinoma, selenium metabolism, prognostic model, tumor microenvironment, immunotherapy response, drug sensitivity

1 Introduction

Lung cancer is one of the most prevalent forms of cancer worldwide, with lung adenocarcinoma (LUAD) being the most commonly diagnosed subtype (Huang et al., 2025). This disease represents a significant threat to human health, accounting for over 700,000 deaths annually. The high incidence and mortality rates associated with LUAD have made it a critical focus for global public health initiatives. Despite advancements in early detection methods and the continuous evolution of treatment modalities, including surgery, radiotherapy, chemotherapy, targeted therapy, and immunotherapy, the overall 5-year survival rate remains alarmingly low at approximately 25%. This is primarily due to elevated rates of recurrence and metastasis. Therefore, identifying new prognostic biomarkers and therapeutic targets is essential for improving patient outcomes in lung cancer (Goldberg and Herbst, 2023). In recent years, rapid progress in genomics and bioinformatics has enabled researchers to explore the relationship between gene expression patterns and lung cancer prognosis (Dizon and Kamal, 2024). By comparing gene expression profiles from tumor tissues with those from normal tissues, several genes associated with LUAD prognosis have been identified. Variations in their expression levels reflect both the biological behavior of the tumor and its clinical outcomes. A deeper understanding of these genes' functions as well as their roles in LUAD progression can facilitate the development of novel treatment strategies aimed at enhancing survival rates. Furthermore, studies indicate that the immune status of LUAD patients is closely linked to the degree of immune infiltration within tumors (Zhang et al., 2022), which may influence tumor growth and metastasis (Zhang et al., 2021). Consequently, characterizing the immune microenvironment of LUAD, particularly regarding the distribution and functional roles of immune cells within tumors, holds great promise for designing personalized treatment plans. These studies underscore the significance of public datasets and bioinformatics tools in the identification of clinically relevant molecular targets.

Selenium is a crucial trace element that plays an integral role in human metabolism, particularly within the antioxidant defense system (Zhang et al., 2023). Selenium metabolism facilitates the elimination of free radicals by forming the active center of glutathione peroxidase, thereby safeguarding cells against oxidative damage (Sun et al., 2023). Furthermore, selenium is closely linked to immune function and enhances the body's capacity to resist diseases (Luo et al., 2025). In cancer research, adequate selenium intake has been associated with a reduced risk of certain types of cancer. Consequently, elucidating the role and impact of selenium metabolism in cancer biology may yield new insights into prevention and treatment strategies. In the context of LUAD the anti-tumor effects of selenium have attracted increasing attention. Clinical and laboratory studies indicate that selenium

deficiency may correlate with a heightened risk of tumor development, while sufficient selenium intake could inhibit tumor cell proliferation and improve patient outcomes (Liang et al., 2024). Moreover, the anti-tumor effects attributed to selenium metabolism may modulate multiple signaling pathways such as PI3K/Akt and MAPK, influencing processes related to cancer cell proliferation, apoptosis, and migration (Guo et al., 2022; Xiao et al., 2025). Therefore, investigating the precise mechanisms underlying selenium metabolism in LUAD could open new avenues for early diagnosis and personalized treatment strategies, holding significant scientific and clinical implications.

In summary, while there has been notable progress in the treatment of LUAD, challenges persist regarding early diagnosis and prognosis. Gene expression and the immune microenvironment are critical factors influencing tumor initiation and progression. Selenium metabolism shows promise in modulating tumor development through mechanisms such as oxidative stress regulation and immune modulation. Investigating these elements may facilitate the identification of new prognostic biomarkers and therapeutic targets. This study primarily utilized transcriptome data from public databases pertaining to LUAD patients, employing bioinformatics techniques to construct and validate a novel selenium metabolism-related risk model aimed at predicting early recurrence in LUAD patients. Additionally, we analyzed the biological pathways associated with these genes, along with their relationships to clinical characteristics, somatic mutations, immune microenvironment dynamics, immunotherapy responses, and drug sensitivity. The expression levels of these prognosis-related genes were validated using clinical samples, thereby supporting the advancement of new chemotherapy regimens, immunotherapies, and targeted therapies for LUAD.

2 Materials and methods

2.1 Data acquisition and preprocessing

A transcriptomic dataset was obtained from The Cancer Genome Atlas (TCGA) database (<https://portal.gdc.cancer.gov/>, accessed on 15 January 2025), which includes RNA sequencing data along with clinical and pathological information for 530 LUAD samples and 59 normal samples (TCGA-LUAD) (Liu et al., 2019). Among these, a total of 517 LUAD samples contained complete survival data, while 222 samples had comprehensive clinical characteristics documented. An independent validation cohort, GSE26939, was retrieved from the Gene Expression Omnibus (GEO) database (<https://www.ncbi.nlm.nih.gov/geo/>), comprising 116 LUAD samples, 115 of which included survival information, profiled using the GPL9053 expression profiling by array platform (Song et al., 2022). Furthermore, a curated list of 86 selenium

metabolism-related genes (SMRGs) was compiled based on published literature (Fu et al., 2023).

2.2 Investigation of differential gene expression patterns in TCGA-LUAD

Differentially expressed genes (DEGs) that distinguish LUAD from non-tumor specimens within the TCGA-LUAD dataset were identified utilizing the R package “DESeq2” (v 1.4.2) (Love et al., 2014). For the downloaded raw data, it was recalibrated into a count matrix and filtered to retain genes exhibiting a total expression level greater than ten across all samples, applying criteria of adjusted $p < 0.05$ and absolute \log_2 -fold change (FC) > 1 . Visualization was conducted using the R package “ggVolcano” (v 0.0.2) (Wodrich et al., 2021), resulting in the generation of a volcano plot. Additionally, heatmap visualization was performed employing the R package “ComplexHeatmap” (v 2.14.0) (Gu et al., 2016).

2.3 Identifying candidate genes and analyzing their functions

To identify candidate genes related to selenium metabolism, we utilized the R package “VennDiagram” (version 1.7.3) (Chen and Boutros, 2011) to visualize the intersection analysis between DEGs and SMRGs, generating a Venn diagram. Subsequently, a functional enrichment assay was performed to investigate the biological roles of these candidate genes. Gene Ontology (GO) analysis was conducted using the `enrichGO` function via the “clusterProfiler” package (version 4.2.2) (Wu et al., 2021), employing the “SYMBOL” gene identifier sourced from the “org.Hs.e.g.,db” annotation database (version 3.18.0) (Qing et al., 2022) (adjusted $p < 0.05$). Furthermore, for Kyoto Encyclopedia of Genes and Genomes (KEGG) enrichment analysis, we applied the `enrichKEGG` function within the same R package (“clusterProfiler”), focusing on human genome data with organism set as “hsa,” while maintaining identical significance thresholds of adjusted $p < 0.05$.

2.4 Protein-protein interaction (PPI) network analysis

The STRING database was used to build PPI networks to investigate the functional relationships among candidate genes (confidence > 0.4). Following this, Cytoscape software version 3.8.2 (Smoot et al., 2011) was utilized for visualizing these interaction networks.

2.5 Development, assessment, and verification of the risk score model

Utilizing the “survival” package (version 3.7.0) (Lei et al., 2023), a univariate Cox regression analysis was conducted on LUAD samples with comprehensive survival data, using candidate genes as the basis for evaluation. Genes that met significance thresholds ($HR \neq 1$, $p < 0.05$) and satisfied the proportional hazards (PH)

assumption, assessed via the `cox.zph` function from the “survival” package, were retained ($p > 0.05$). Subsequently, we employed the “survival” package to construct a multivariable Cox regression model utilizing its built-in functions for Cox regression analysis. A bidirectional stepwise variable selection process was implemented based on the Akaike Information Criterion (AIC), employing the step function; additionally, an overall PH assumption test of the model was performed ($p > 0.05$). Through this methodology, prognostic genes were identified. Risk scores were then calculated by applying coefficients derived from this final model to determine risk associated with each prognostic gene in constructing our risk model. In this context, “coef” refers to the risk coefficient corresponding to each individual prognostic gene, while “expr” indicates the expression intensity of each respective prognostic gene.

$$\text{risk score} = \sum_{i=1}^n \text{coef}(\text{gene}_i) \times \text{expr}(\text{gene}_i)$$

The optimal cutoff value derived from the `surv_cutpoint` function in the R package “survminer” (version 0.4.9) (Liu et al., 2021) was utilized to classify patients into high-risk groups (HRG) and low-risk groups (LRG). Subsequently, principal component analysis (PCA) was performed using the `prcomp` function to visualize the separation of risk groups. Kaplan-Meier (KM) survival plots and log-rank tests, facilitated by the R package “survminer,” were employed to compare survival outcomes between these groups. The software package “timeROC” (version 1.18.0) (Blanche et al., 2013) was leveraged to generate receiver operating characteristic (ROC) curves at three time points for evaluating model precision. Following this, expression levels of prognostic genes across both cohorts were examined utilizing the “pheatmap” package (version 1.0.12) (Gu and Hübbschmann, 2022). For external validation, data from GSE26939 were analyzed with identical risk stratification methods to assess the robustness of the model.

2.6 Development and validation of the nomogram

A univariate Cox regression analysis was conducted on risk scores alongside several clinical variables, including pathological stages, gender, age, and T/N/M stages ($p < 0.05$, $HR \neq 1$); additionally, a proportional hazards (PH) assumption test was performed with $p > 0.05$ as an acceptance criterion. Variables that met criteria in multivariate Cox regression analysis, specifically those with $p < 0.05$, and passed PH assumption testing were considered independent predictive factors for prognosis. Based on TCGA-LUAD data, we developed a nomogram to forecast 1-, 3-, and 5-year survival probabilities with the “regplot” package (version 1.1) (Sui et al., 2022). Thereafter, we generated a calibration curve employing bootstrap resampling techniques through two hundred iterations via the R package “rms” (version 6.8.1) (Sachs, 2017), aimed at assessing nomogram accuracy; proximity of slope values to one indicates enhanced predictive accuracy of our model’s estimates for patient outcomes over time.

Furthermore, we utilized the R package “ggDCA” version v (1.1) (Duo et al., 2023) to create a comprehensive nomogram forecasting 1-, 3-, and 5-year survival probabilities for patients diagnosed with LUAD.

2.7 Examination of the relationship between risk scores and the expression of prognostic genes within clinical variables

To investigate the diagnostic efficiency of risk scores in relation to clinical indicators, we analyzed correlations between risk scores and the aforementioned clinical variables. The Wilcoxon test (W-tests) was employed to compare varying risk scores and prognostic genes across different clinical subgroups ($p < 0.05$). Additionally, the survdiff function within the “survival” package was employed to assess survival differences among risk groups for each clinical indicator ($p < 0.05$).

2.8 Functional enrichment analysis

Within the TCGA-LUAD data collection, Gene Set Enrichment Analysis (GSEA) was conducted to pinpoint biological pathways linked to HRG and LRG. Initially, differential gene expression analysis was performed between the two risk groups via the “DESeq2” package (v 1.4.2), resulting in a ranked list of genes ordered by their \log_2FC values from largest to smallest. Subsequently, we retrieved a background gene set named “c2.cp.kegg.v7.4.symbols.gmt” from the MSigDB for pathway enrichment analysis utilizing the “clusterProfiler” package (v 4.2.2). Significant pathways were defined as those exhibiting a normalized enrichment score ($|NES| > 1$, $p < 0.05$, and false discovery rate (FDR) < 0.25).

For Gene Set Variation Analysis (GSVA), we employed the “GSVA” package (v 1.42.0) (Hänzelmann et al., 2013) to calculate pathway activity scores across samples denoted as *gsva_mat*. A design matrix was constructed using the *model.matrix* function from the “limma” package (v 3.54.0) (Ritchie et al., 2015), followed by an assessment of differential pathway activity between risk groups through linear model fitting via *lmFit* and empirical Bayes moderation using *eBayes* methods. Contrast matrices were generated with *makeContrasts* function, allowing us to extract significantly altered pathways with adjusted *p*-values < 0.05 through *topTable* function.

2.9 Analysis of the tumor immune microenvironment

The CIBERSORT method was employed to estimate the proportions of 22 distinct immune cell subtypes (Newman et al., 2015) within the tumor microenvironment of LUAD samples. Gene expression matrices underwent preprocessing and were subsequently submitted to the CIBERSORT platform, with 10 permutations and quantile normalization (QN) activated. To compare the levels of immune cell infiltration among different risk groups, W-tests were conducted ($p < 0.05$). The “psych” package (v

2.1.6) (Correction to Lancet Psych, 2022, 2023) was utilized to calculate Spearman correlation coefficients between differentially abundant immune cells and risk scores, as well as between these immune cells and prognostic genes ($|\text{correlation coefficient (cor)}| > 0.3$, $p < 0.05$).

2.10 Analysis of LUAD immunotherapy response

The Tumor Immune Dysfunction and Exclusion (TIDE) platform was used to assess mechanisms of immune evasion and predict responses to immunotherapy. Pre-treatment gene expression profiles from LUAD samples were standardized before being submitted to TIDE for generating Dysfunction, Exclusion, and TIDE scores. Dysfunction scores reflect the functional state of effector T lymphocytes within the tumor microenvironment, while Exclusion scores indicate barriers to immune cell infiltration.

Furthermore, a total of 47 immune checkpoint genes (e.g., LAG-3, CTLA-4, PD-1) were extracted from relevant literature (Xue et al., 2022). A differential expression analysis on these genes across risk groups was performed using W-tests ($p < 0.05$).

2.11 Somatic mutation profiling

The TCGA database was queried to obtain somatic mutation information for LUAD samples. Subsequently, the “Maftools” package (v 2.18.0) (Mayakonda et al., 2018) was employed to analyze the mutation annotation format (MAF) files and generate oncoplots (waterfall plots) that illustrate the mutation landscape of HRG and LRG. Entries with missing values were excluded following the annotation and filtering of variant types and classifications. The distribution of mutation types and allele-specific changes was assessed. Thereafter, using the R package “ggplot2” (v 3.4.1) (Gustavsson et al., 2022), we visualized the distribution of these mutation types across HRG and LRG populations. Tumor mutational burden (TMB), defined as the cumulative count of somatic mutations per megabase, was calculated with statistical significance set at $p < 0.05$. Meanwhile, the chi-square test was used to analyze the differences in the top 10 mutation classifications, different mutation types, and base changes between HRG and LRG ($p < 0.05$). In addition, the *somaticInteractions* function in the “Maftools” package (v 2.18.0) was utilized to perform co-occurrence and mutual exclusivity analysis on the mutation data of the high-risk and low-risk groups ($p < 0.05$).

2.12 Drug sensitivity analyses in risk groups

Drug sensitivity analyses were conducted utilizing the R package “pRRophetic” (v 0.5) (Geeleher et al., 2014). This method used the Genomics of Drug Sensitivity in Cancer (GDSC, also known as CGP 2016) as the reference training set to construct a drug sensitivity prediction model. Specifically, we input the gene expression matrix (FPKM) of the TCGA-LUAD cohort into the model, with parameters set as *tissueType* = “lung” and *dataset* = “cgp 2016”. Subsequently, the *pRRopheticPredict* () function was applied to

calculate the predicted half-maximal inhibitory concentration (IC₅₀) of each drug for each patient. Initially, gene expression data from the TCGA database were integrated with a well-established drug sensitivity dataset known as cgp 2016; drugs classified under LUAD by TCGA with an area under curve (AUC) greater than 0.98 were selected as candidates for further analysis. The half-maximal inhibitory concentration (IC₅₀) for each drug were calculated using the pRRopheticPredict function, followed by a W-test to identify disparities in IC₅₀ values among various drugs between HRG and LRG cohorts, establishing statistical significance at $p < 0.01$. Additionally, we utilized the R software package “pheatmap” to visually represent relationships between first-line chemotherapeutic agents and prognostic genes in LUAD.

2.13 Validation of the expression of prognostic genes in clinical samples

Drawing on TCGA-LUAD data, a W-test was first conducted to evaluate the expression of prognostic genes in LUAD samples relative to normal samples ($p < 0.05$). Boxplots were created with the “ggplot2” package.

Ethical approval was obtained from the ethics committee of the First People’s Hospital of Yunnan Province (approval number: KHLL2022-KY159) before this experiment was conducted. Five LUAD tissue samples and five adjacent non-tumor tissue samples were obtained from participants at the First People’s Hospital of Yunnan Province, all of whom provided informed consent. Initially, 50 mg of tissue was extracted from each sample and homogenized with 1 mL of TRIzol (Vazyme, R401-01, China) to ensure thorough mixing and grinding. After standing on ice for 10 min, 200 μ L of chloroform was added to facilitate RNA extraction from the aqueous phase. Subsequently, an equal volume of chilled isopropanol was incorporated for RNA precipitation. Following quantification, reverse transcription reactions commenced immediately thereafter. The cDNA synthesis reaction system was established strictly according to the manufacturer’s instructions for the SweScript First Strand cDNA Synthesis Kit (YEASEN, 11141ES60, China). Thereafter, qPCR amplification consisting of 40 cycles was conducted using a CFX96 real-time fluorescence quantitative PCR device (BIO-RAD, XLFZ006, United States). The primer sequences are detailed in [Supplementary Table S1](#). In this study, glyceraldehyde-3-phosphate dehydrogenase (GAPDH) was selected as the reference gene. In terms of experimental design, five biological replicates were set up to reduce the interference of individual sample differences, and three technical replicates were performed for each sample to ensure the repeatability of detection; after the completion of qPCR amplification, the cycle threshold (Ct values) of target genes and the reference gene were obtained simultaneously, and the melting curve was used to verify amplification specificity, while the amplification curve was used to evaluate amplification efficiency.

The expression levels of prognostic genes were evaluated utilizing the $2^{-\Delta\Delta Ct}$ method (Cheng et al., 2020), with specific steps as follows: first, the ΔCt value was calculated as the difference between the Ct value of the target gene and that of GAPDH ($\Delta Ct = Ct_{\text{target gene}} - Ct_{\text{GAPDH}}$); second, the $\Delta\Delta Ct$

value was calculated by taking the ΔCt value of the control group as a reference and finding the difference between the ΔCt value of the experimental group and that of the control group ($\Delta\Delta Ct = \Delta Ct_{\text{experimental group}} - \Delta Ct_{\text{control group}}$); finally, the relative expression level of the target gene was obtained through conversion using the $2^{-\Delta\Delta Ct}$ formula. All experimental data were subjected to statistical analysis and visualization using GraphPad Prism (v8.0) software (Chang et al., 2023): the t-test was used as the statistical method, and a $p < 0.05$ was considered to indicate a statistically significant difference.

2.14 Statistical analysis

All statistical analyses of data from public databases were performed using R software (version 4.2.2). The Wilcoxon test and chi-square test were used as the significance test method to compare differences between different groups, and a $p < 0.05$ was considered to indicate statistical significance. All network diagrams were constructed using Cytoscape software (v 3.8.2). For RT-qPCR experiments, the relative mRNA expression levels of prognostic genes were calculated using the $2^{-\Delta\Delta Ct}$ method, and the t-test was applied to compare differences in expression levels. A $p < 0.05$ was regarded as statistically significant.

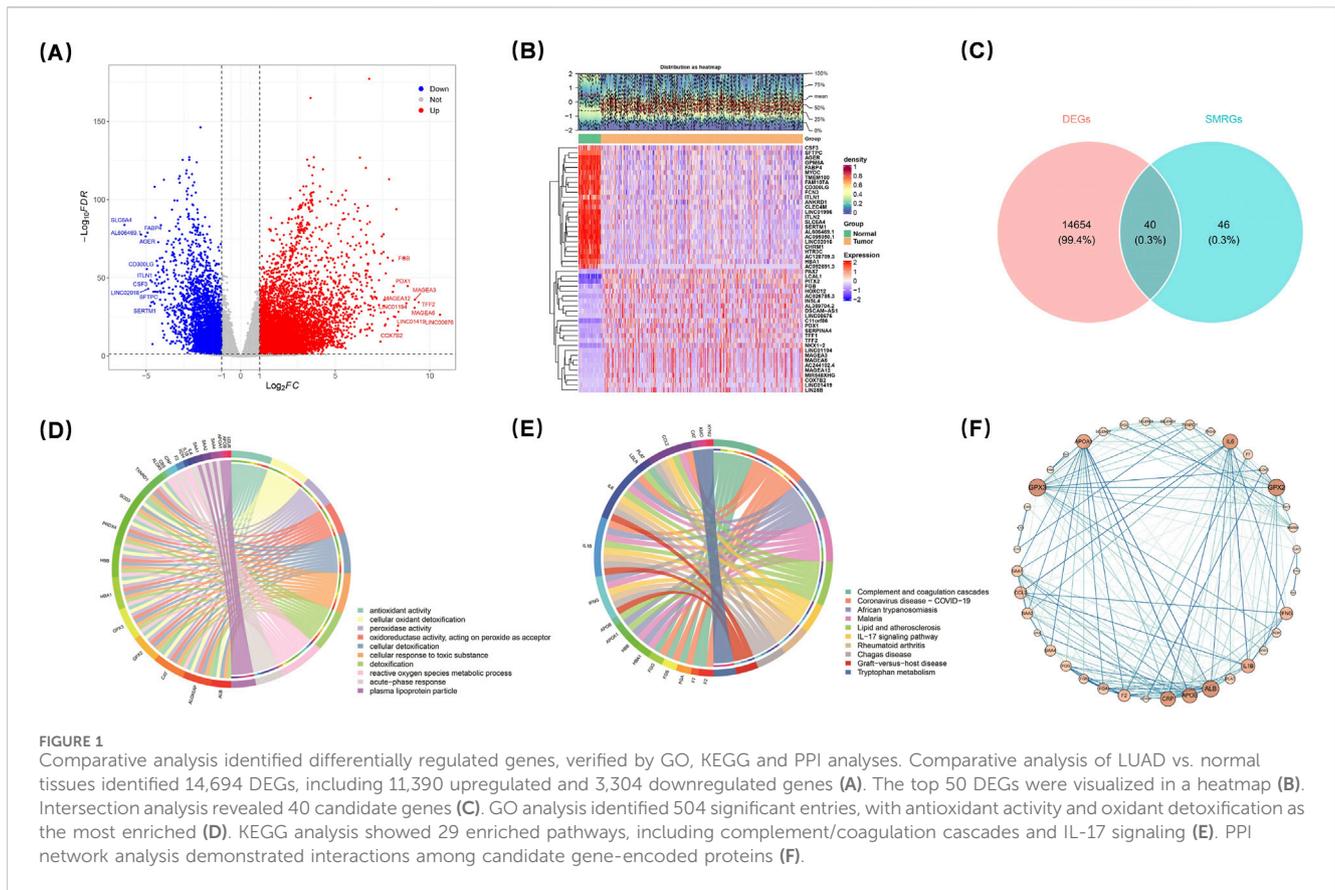
3 Results

3.1 Differential expression profiling

Comparative analysis of LUAD versus normal tissues revealed a total of 14,694 DEGs (adjusted $p < 0.05$ and $|\log_2 FC| > 1$), comprising 11,390 upregulated genes and 3,304 downregulated genes (Figure 1A). The top 50 DEGs exhibiting the largest $|\log_2 FC|$ values were visualized in a hierarchical clustering heatmap (Figure 1B), which highlighted distinct expression patterns between LUAD and normal tissue groups. Intersection analysis between DEGs and survival-related gene sets identified 40 candidate genes (Figure 1C) ([Supplementary Table S2](#)).

3.2 Functional enrichment characteristics

GO analysis of the candidate genes uncovered a total of 504 significant entries (adjusted $p < 0.05$), with antioxidant activity (GO:0016209, $p = 1.35e-15$) and cellular oxidant detoxification (GO:0098869, $p = 8.44e-15$) being the most enriched categories (Figure 1D) ([Supplementary Table S3](#)). Additionally, KEGG pathway analysis indicated significant enrichment across 29 pathways (adjusted $p < 0.05$). Notably among these pathways were the complement and coagulation cascades pathway (hsa04610, $p = 4.83e-07$) as well as the IL-17 signaling pathway (hsa04657, $p = 0.000311$) (Figure 1E) ([Supplementary Table S4](#)). Furthermore, PPI network analysis demonstrated interactions among proteins encoded by the identified candidate genes; specifically, ALB, GPX3, GPX2, IL6, APOB, APOA1, and CRP exhibited interactions with several other candidate gene products (Figure 1F). Collectively, this PPI



network provides valuable insights into the dynamic interactions among proteins corresponding to these relevant genes at the protein level.

3.3 Development and validation of the prognostic model

Cox regression analysis was performed on 517 LUAD samples from TCGA-LUAD to pinpoint genes significantly associated with overall survival. ALOX5AP, CRP, F2, GPX3, KMO, KYNU, SOD3, and TXNRD1 were ultimately retained in the analysis; among these, ALOX5AP, CRP, and F2 were identified as risk factors for LUAD ($HR > 1$) (Figure 2A). The results of the proportional hazards assumption test were illustrated using Schoenfeld residual plots; the trend lines of the residuals did not show significant differences from one another ($p > 0.05$) (Supplementary Figure S1A–H). Subsequently, a multivariate Cox regression model was constructed that ultimately screened four prognostic genes: F2, GPX3, KMO, and KYNU (Figure 2B). The overall proportional hazards assumption test for this model yielded $p = 0.228$, indicating compliance with the proportional risk assumption (Supplementary Figure S1I) (Supplementary Table S5). Risk scores were determined using the coefficients associated with these prognostic genes. As a result, the 517 LUAD patients were divided into a HRG ($n = 57$) and a LRG ($n = 460$), based on an optimal cut-off value of 1.705 (Figure 2C). Principal component

analysis demonstrated distinct separation between PC1 and PC2 for HRG and LRG groups (Figure 2D).

Conversely, mortality rates were found to be significantly higher within the HRG compared to LRG individuals (Figure 2E). The Kaplan-Meier curve revealed substantial disparities in survival rates between groups; those categorized as high-risk exhibited markedly reduced survival probabilities ($p < 0.001$) (Figure 2F). Furthermore, the ROC assessment of the risk model validated its effectiveness in forecasting survival probabilities at 1-, 3-, and 5-year for LUAD patients ($AUC > 0.6$) (Figure 2G). The heat map illustrating expression levels of prognostic genes indicated that F2 and KYNU had elevated expression levels in HRG compared to other groups (Figure 2H).

To assess the stability of the risk model, the same set of analyses were carried out using the dataset of the validation set. The obtained results were in line with those from the TCGA-LUAD. This consistency effectively demonstrated that the prognostic model developed in this research is capable of being utilized to predict the prognosis of patients afflicted with LUAD (Figures 2I–N).

3.4 Establishment of the nomogram

A comprehensive univariate Cox proportional hazards regression analysis was systematically conducted to examine the impact of individual factors on prognosis. The analyses revealed a significant association between specific factors and adverse clinical outcomes ($HR < 1, P < 0.05$). Concurrently, a proportional hazards

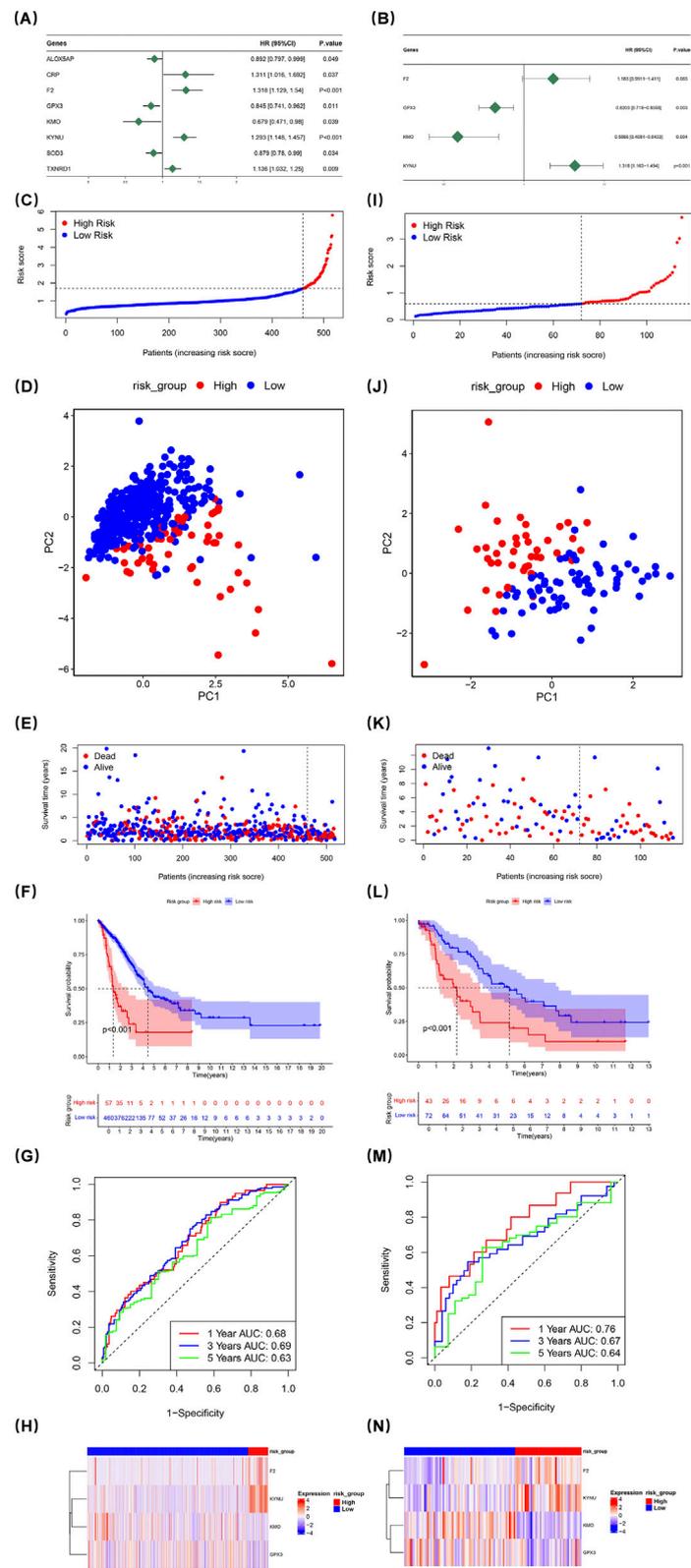


FIGURE 2 Univariate Cox regression analysis of 517 LUAD samples from TCGA-LUAD identified genes correlated with overall survival. ALOX5AP, CRP, F2, GPX3, KMO, KYNU, SOD3, and TXNRD1 were found as risk factors (A). Schoenfeld residual plots confirmed proportional hazards ($p > 0.05$) (Supplementary Figure S1A–H). A multifactorial Cox regression model screened four prognostic genes: GPX3, KMO, KYNU, and TXNRD1 (B). The model met the proportional hazards assumption ($p = 0.228$) (Supplementary Figure S1I) (Supplementary Table S5). Risk scores were calculated, classifying patients into high-risk (HRG) and low-risk (LRG) groups (C, I). PCA analysis showed distinct separation between HRG and LRG (D, J). Mortality rates were higher in (Continued)

FIGURE 2 (Continued)

HRG (E,K). The Kaplan-Meier curve revealed reduced survival in HRG ($p < 0.001$) (F,L). ROC analysis confirmed the model's predictive validity for one-, three-, and 5-year survival (AUC > 0.6) (G,M). The heat map showed elevated F2 and KYNU expression in HRG (H,N).

assumption test was performed, with the resulting p -value exceeding 0.05 indicating that the data met the proportional hazards assumption, thereby ensuring the reliability and validity of the univariate analysis results. Following this thorough evaluation, three variables, namely, risk assessment score, pathologic T stage, and pathologic N stage, were identified as potential prognostic indicators (Figure 3A).

Further multivariate Cox regression analyses were conducted to assess their independent predictive value. The multivariate analysis ($p < 0.05$) provided statistical evidence underscoring the importance of these variables. Additionally, another proportional hazards assumption test further validated the appropriateness of the model ($p > 0.05$). Collectively, these analyses robustly confirmed that risk score, pathologic T stage, and pathologic N stage are independent prognostic indicators (Figures 3B,C).

The nomogram illustrated that higher total points corresponded to poorer survival likelihood in LUAD patients (Figure 3D). A calibration plot was employed to assess the predictive accuracy of the nomogram; notably, at time points of 1-, 3-, and 5 years post-diagnosis, the slopes of these calibration plots closely aligned with reference lines, indicating an excellent fit for our model (Figure 3E). Furthermore, decision curve analysis curves demonstrated that the net benefit derived from using this nomogram exceeded zero and surpassed those associated with “all” or “none” options while generally outperforming both pathologic T/N staging and risk scores alone (Figure 3F).

3.5 Assessment of clinical features

Differences in risk scores were observed across various clinical features. Specifically, male patients exhibited a significantly higher risk score than female patients ($p = 0.0058$). Notably, box plots demonstrated significant differences in risk scores among different subgroups of clinical features (T1 and T3, stage I and stage III, N1 and N2, stage I and stage II, M0 and M1, N0 and N2, as well as between stages I and IV) ($p < 0.05$) (Figure 4A). Further analysis of prognostic genes within distinct clinical subgroups revealed that GPX3 and KYNU exhibited higher expression levels in male patients than in female patients; conversely, KMO and F2 showed the opposite trend. Additionally, the expression of KYNU displayed an increasing trend in cases of stage IV LUAD (Figures 4B–E). Remarkably, among LUAD patients at stages III–IV, significant survival disparities persisted between two groups. This was particularly evident among patients with pathologic N0 status as well as those with pathologic T1–2 status across all age groups (> 60 years old vs. ≤ 60 years old) ($p < 0.001$) (Figure 4F).

3.6 GSEA enrichment analysis of DEGs

To elucidate the signaling mechanisms underlying the DEGs between the two groups, GSEA was performed. A total of

50 pathway entries were identified through this analysis. Notably, in the LRG, several pathways associated with immune modulation and disorders were activated, including autoimmune thyroid disease, allograft rejection, asthma, and the intestinal immune network for IgA production. In contrast, the high-risk Group (HRG) exhibited significant enrichment in multiple pathways related to cell proliferation, metabolism, and gene regulation, specifically DNA replication, cell cycle progression, spliceosome activity, and ribosomal function (Figures 5A,B). Subsequently, GSVA was employed to further evaluate the activation or inhibition status of these pathways within both groups (Figure 5C). In particular, certain pathways in the HRG showed notable enrichment that suggests their critical role in intestinal immunomodulation. Core genes within these enriched pathways in the HRG, including MCM7, PCNA, and POLE, are recognized as essential players in DNA replication and repair processes. This observation implies a likely enhancement of cell proliferation activities within the HRG. Furthermore, genes such as CDK6, CCNB1, and CDK1 are known regulators of cell cycle progression; this finding further substantiates that the HRG promotes cellular proliferation.

3.7 Tumor microenvironment characterization

A heatmap was generated to illustrate the enrichment ratios of 22 distinct immune cell subtypes across various risk levels (Figure 6A). Immune cells exhibiting diverse infiltration levels between the two groups were categorized as differentially infiltrating immune cells. Subsequently, a box-and-whisker plot was utilized to depict the percentage of these differentially infiltrating immune cells in both groups. Significant disparities were observed in the infiltration percentages of ten types of immune cells between the two groups, including CD8⁺ T cells and resting memory CD4⁺ T cells ($p < 0.05$) (Figure 6B).

Further analysis using Spearman's rank correlation revealed that plasma cells and naive B cells exhibited a robust positive correlation ($cor = 0.5825$, $p < 0.001$), while resting NK cells demonstrated a strong negative correlation with activated NK cells ($cor = -0.4887$, $p < 0.001$) (Figure 6C) (Supplementary Figure S1). Notably, a heatmap illustrating the correlation between risk scores and differential immune cell populations indicated that both resting dendritic cells and resting mast cells had positive correlations with monocytes ($cor = 0.33$; $cor = 0.30$; $p < 0.05$). In contrast, activated mast cells showed a negative correlation with resting mast cells ($cor = -0.44$; $p < 0.05$) (Figure 6D). This finding suggests that expression levels of prognostic genes are closely associated with the presence of resting mast cells, indicating that these genes may play a significant role in this relationship. These results could enhance our understanding of disease prognosis and underlying immunological mechanisms.

TIDE analysis showed that the dysfunction indicator in the HRG was notably reduced compared with the LRG, suggesting a greater potential for immune escape. Conversely, the value of Exclusion in the HRG was markedly greater than that in the LRG, suggesting a reduced

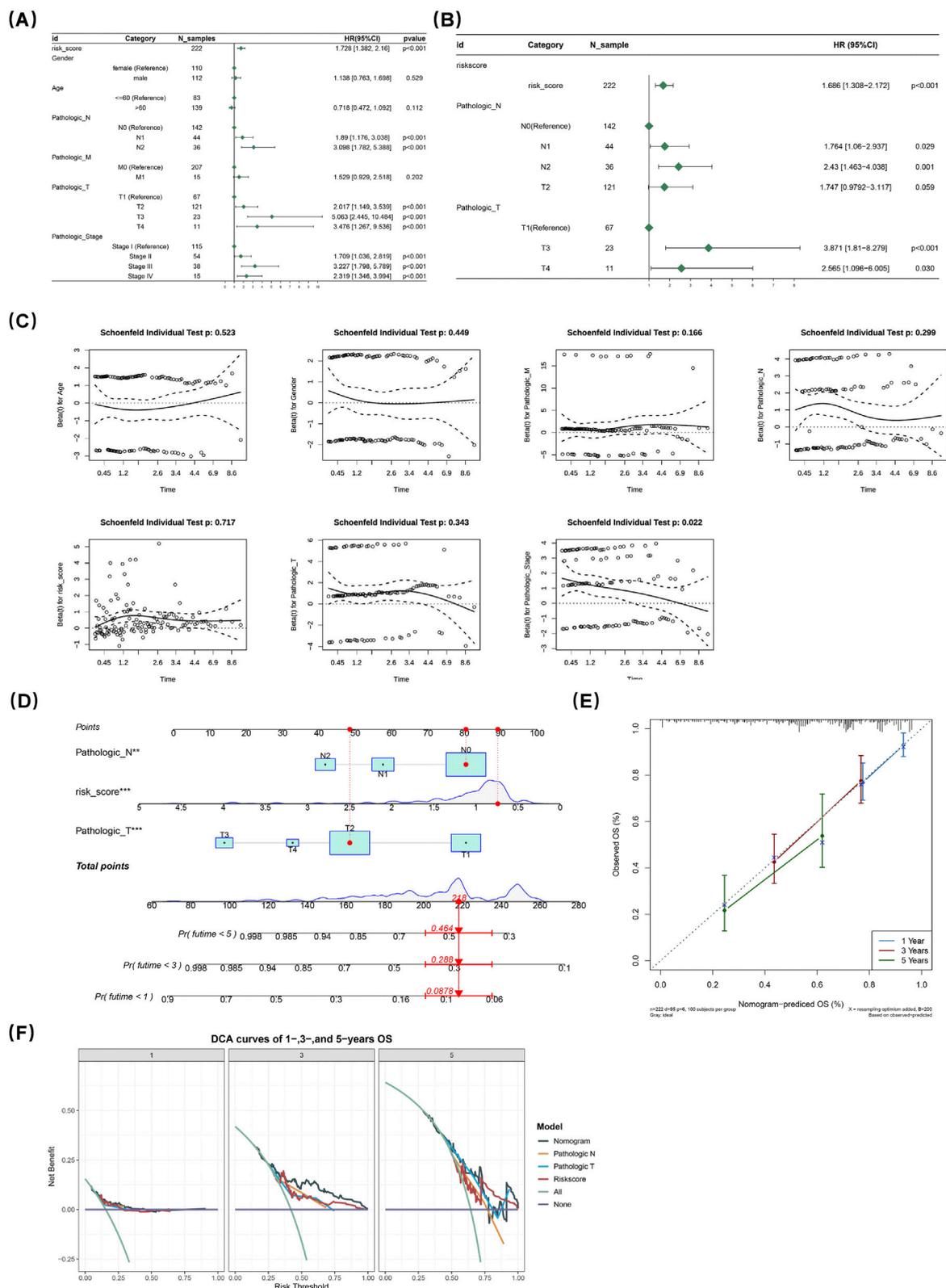
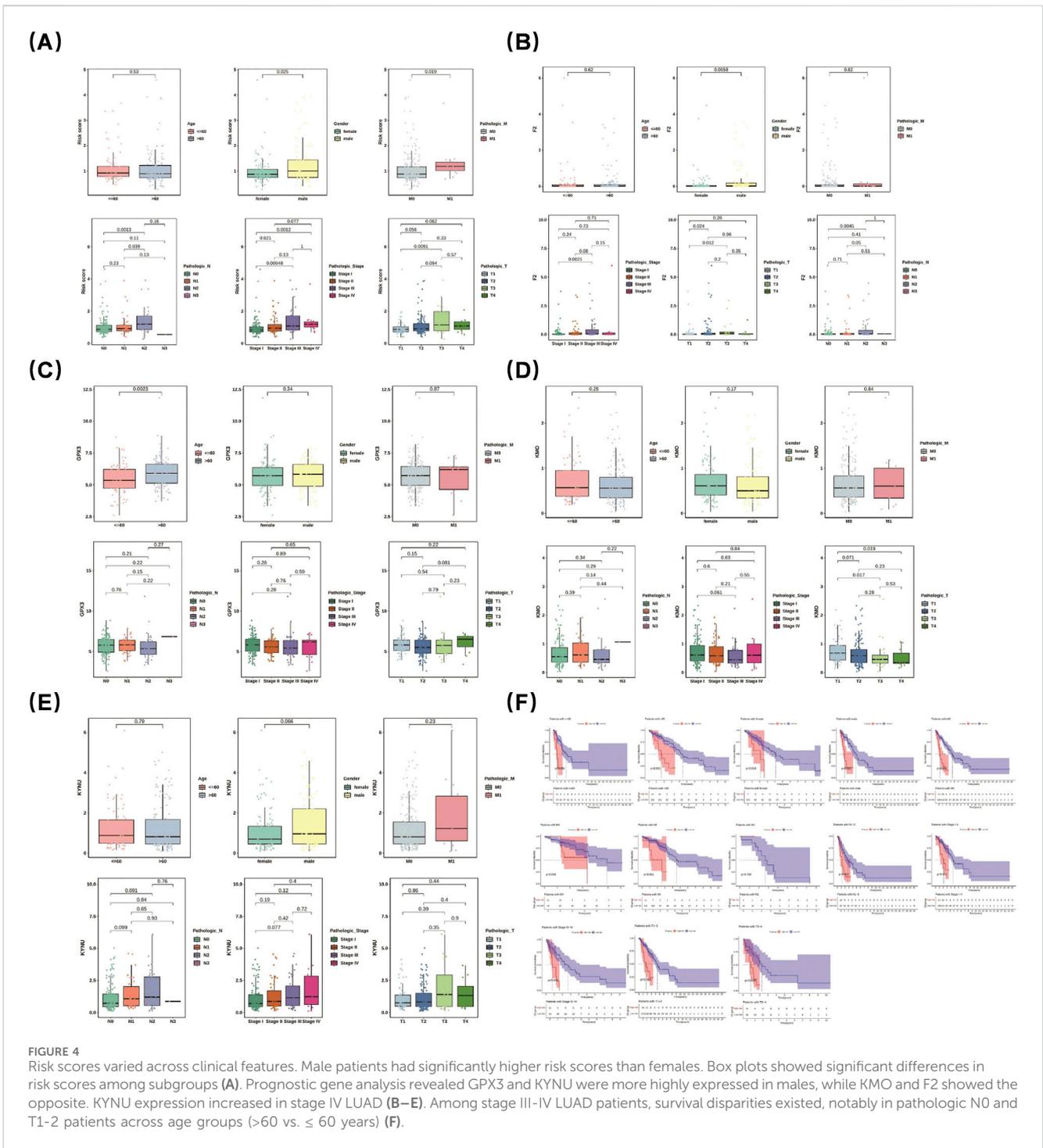


FIGURE 3
 A comprehensive univariate Cox regression analysis examined the impact of individual factors on prognosis, revealing significant associations between specific factors and adverse outcomes. A proportional hazards assumption test ($p > 0.05$) confirmed the reliability of the results. Three variables, risk assessment score, pathologic T stage, and pathologic N stage, were identified as potential prognostic indicators (A). Subsequent multivariate analysis ($p < 0.05$) underscored their independent prognostic significance, further validated by another proportional hazards test ($p > 0.05$) (B,C). The nomogram indicated higher total points correlated with poorer survival in LUAD patients (D). A calibration plot showed excellent model fit at 1, 3, and 5 years post-diagnosis (E). Decision curve analysis demonstrated the nomogram's net benefit exceeded zero and outperformed other options (F).

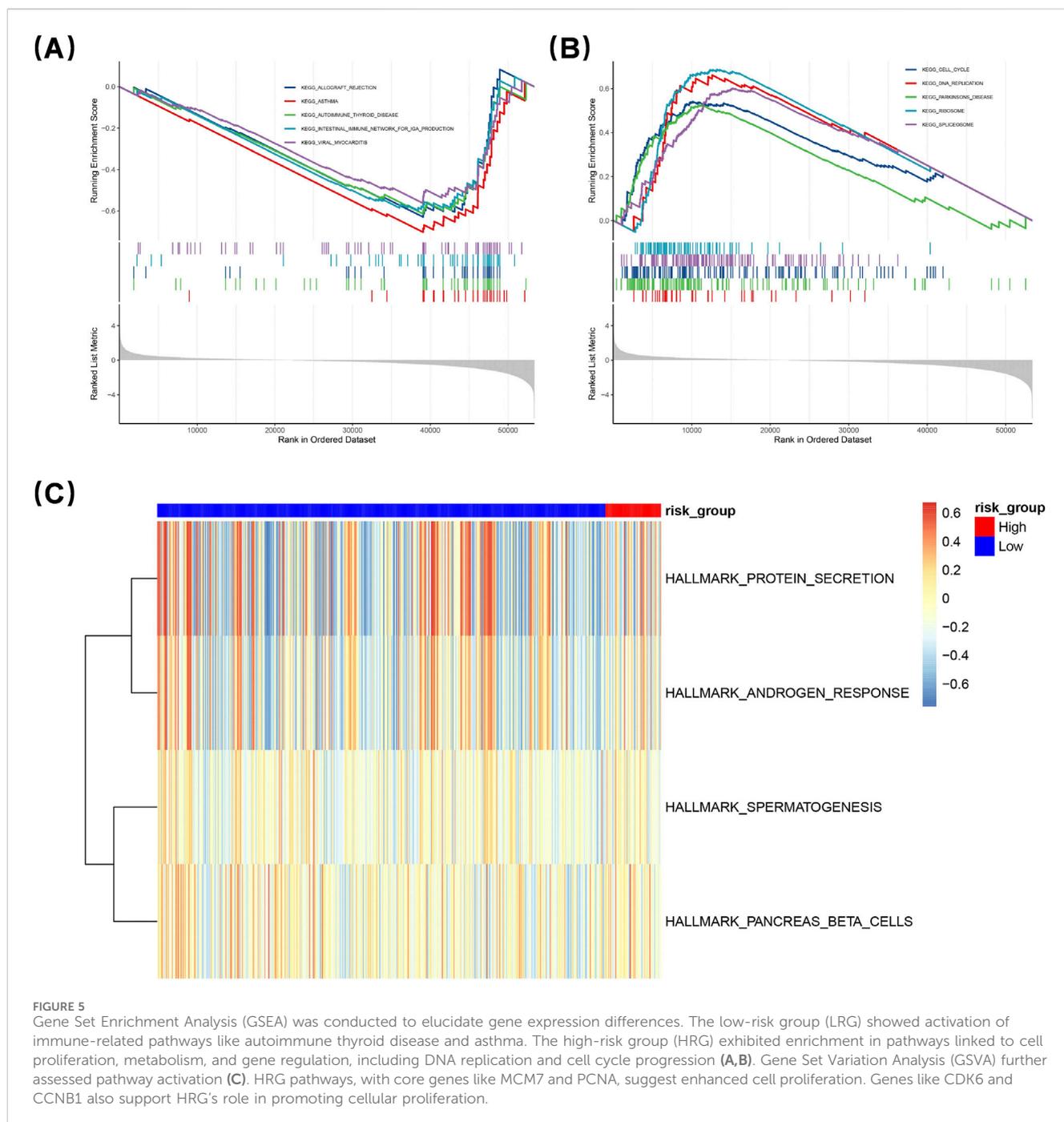


probability of immune escape (Figure 6E). The immune checkpoint mechanism plays a crucial role in evading detection and attack by T cells. We examined and compared the expression patterns of genes associated with immune checkpoints across different risk categories. Among the 47 immune checkpoint genes analyzed, 32 exhibited differential expression among distinct risk groups ($p < 0.05$) (Figure 6F). Notably, CD276 and TNFSF4 showed elevated expression levels in the HRG compared to those in the LRG. In contrast, the remaining 30 genes demonstrated higher expression levels within the LRG when compared to those in the HRG. The varying expressions of these genes

provide a foundation for categorizing LUAD patients who may be suitable candidates for immunosuppressive treatment.

3.8 Examination of somatic cell mutations and assessment of drug sensitivity

The results from the somatic mutation analysis revealed a higher mutation percentage in the HRG compared to the LRG) with rates of 98.21% versus 92.94%, respectively. Notably, KEAP1, TTN, and



USH2A exhibited the highest mutation rates within the HRG at 48%, while TTN displayed the highest mutation rate in the LRG at 52%. This suggests that TTN may be one of the genes most frequently mutated in patients with LUAD (Figures 7A,B). A detailed analysis of the mutation data indicated that synonymous variants and missense variants occurred more frequently in the LRG. Furthermore, single nucleotide polymorphisms (SNPs) were also more prevalent in this group. Interestingly, G > T and C > A base changes were observed to be more common in the LRG as well. These mutations showed statistical significance between HRG and LRG. (Figures 7C–E). In addition, most of the top 20 mutated

genes showed significant co-occurrence, and in particular, KRAS and TP53 exhibited mutual exclusivity ($p < 0.05$) (Figure 7F).

Moreover, drug sensitivity analysis identified a total of 124 compounds exhibiting differential IC50 values between these two groups (Supplementary Table S7). From this screening process, five first-line chemotherapeutic agents for LUAD treatment were highlighted: cisplatin, gemcitabine, docetaxel, etoposide, and vinorelbine. The findings demonstrated a significant correlation between KYNU and Docetaxel ($p < 0.01$) (Figure 7G). The respective IC50 values for these drugs were lower in the HRG cohort, indicating that individuals within this group exhibited greater sensitivity to these therapeutic agents. Additionally,

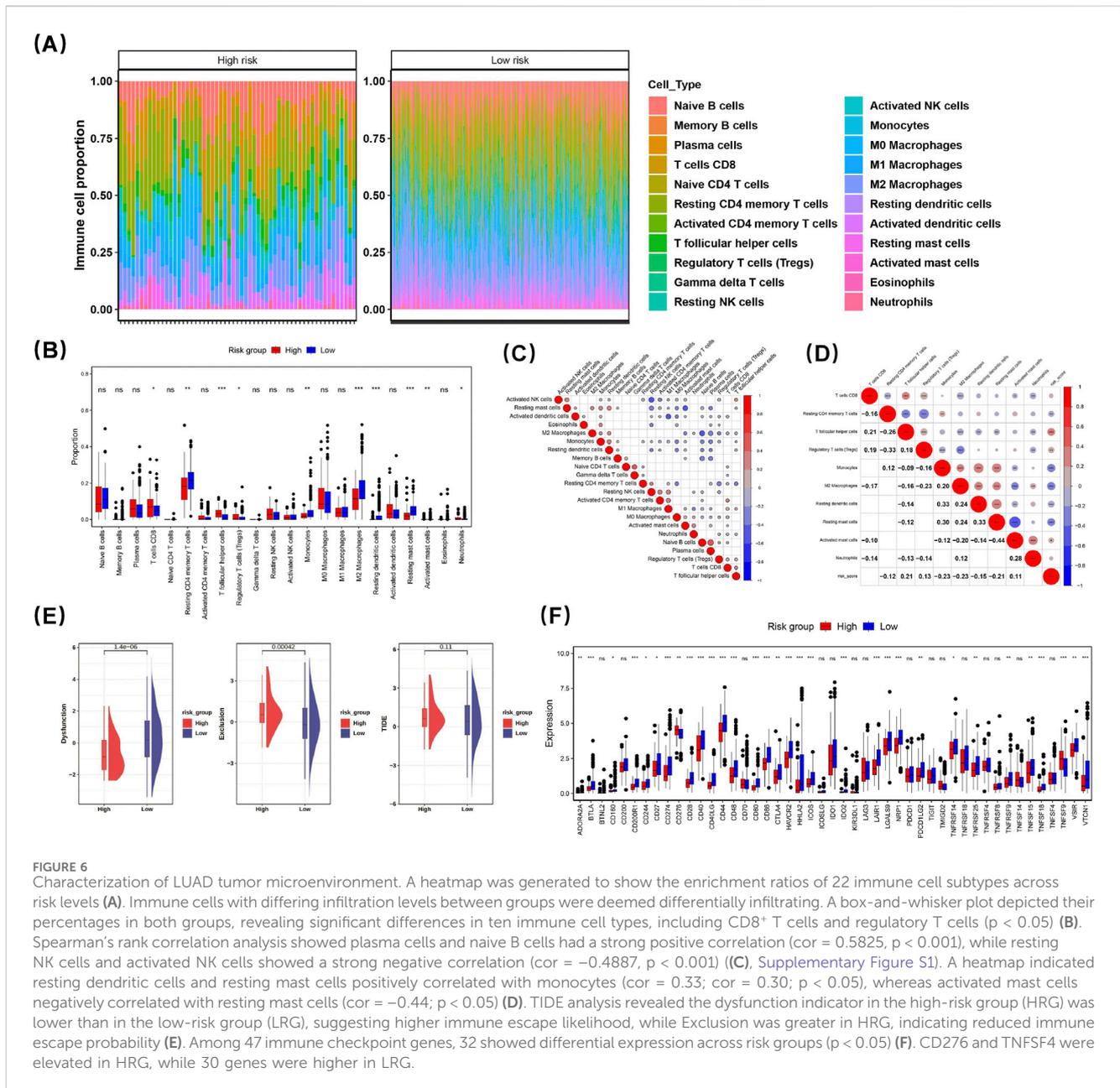


FIGURE 6 Characterization of LUAD tumor microenvironment. A heatmap was generated to show the enrichment ratios of 22 immune cell subtypes across risk levels (A). Immune cells with differing infiltration levels between groups were deemed differentially infiltrating. A box-and-whisker plot depicted their percentages in both groups, revealing significant differences in ten immune cell types, including CD8⁺ T cells and regulatory T cells ($p < 0.05$) (B). Spearman's rank correlation analysis showed plasma cells and naive B cells had a strong positive correlation ($cor = 0.5825$, $p < 0.001$), while resting NK cells and activated NK cells showed a strong negative correlation ($cor = -0.4887$, $p < 0.001$) (C, Supplementary Figure S1). A heatmap indicated resting dendritic cells and resting mast cells positively correlated with monocytes ($cor = 0.33$; $cor = 0.30$; $p < 0.05$), whereas activated mast cells negatively correlated with resting mast cells ($cor = -0.44$; $p < 0.05$) (D). TIDE analysis revealed the dysfunction indicator in the high-risk group (HRG) was lower than in the low-risk group (LRG), suggesting higher immune escape likelihood, while Exclusion was greater in HRG, indicating reduced immune escape probability (E). Among 47 immune checkpoint genes, 32 showed differential expression across risk groups ($p < 0.05$) (F). CD276 and TNFSF4 were elevated in HRG, while 30 genes were higher in LRG.

further confirmation was provided by observing that IC50 values for these drugs remained consistently lower among those classified within HRG; thus reinforcing their heightened responsiveness to such treatments.

3.9 Expression validation

Expression validation confirmed the differential patterns of prognostic genes. Compared to normal samples, GPX3 ($p < 0.0001$), KYNU ($p < 0.05$), and KMO ($p < 0.0001$) were found to be downregulated in LUAD, while F2 exhibited significant upregulation ($p < 0.0001$) (Figure 8A). The experimental results obtained from RT-PCR demonstrated a high degree of consistency with the predictions derived from bioinformatics analyses. In

comparison to normal samples, GPX3, KMO, and KYNU were significantly downregulated in LUAD, whereas F2 was upregulated in this context (Figures 8B–E). The amplification curves (obtained after 20 cycles) and melting curves (showing smooth, single peaks) demonstrated good primer specificity and favorable amplification efficiency (Supplementary Figure S2). These findings are consistent with expression profiles derived from TCGA, thereby reinforcing the biological relevance of the prognostic signature.

4 Discussion

Lung adenocarcinoma (LUAD), the predominant subtype of non-small cell lung cancer, continues to present significant challenges due to its heterogeneity and resistance to treatment

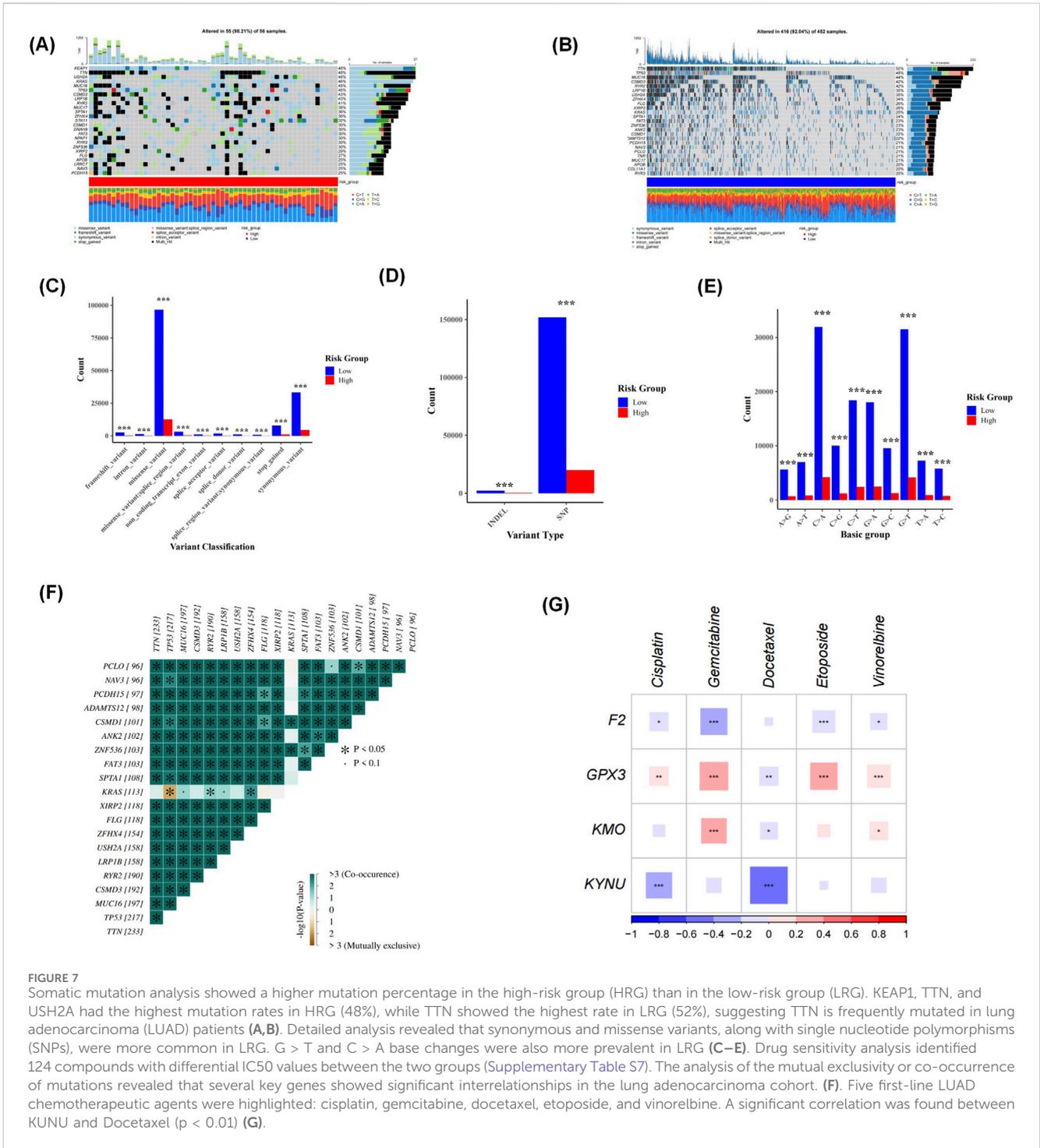


FIGURE 7 Somatic mutation analysis showed a higher mutation percentage in the high-risk group (HRG) than in the low-risk group (LRG). KEAP1, TTN, and USH2A had the highest mutation rates in HRG (48%), while TTN showed the highest rate in LRG (52%), suggesting TTN is frequently mutated in lung adenocarcinoma (LUAD) patients (A,B). Detailed analysis revealed that synonymous and missense variants, along with single nucleotide polymorphisms (SNPs), were more common in LRG. G > T and C > A base changes were also more prevalent in LRG (C–E). Drug sensitivity analysis identified 124 compounds with differential IC50 values between the two groups (Supplementary Table S7). The analysis of the mutual exclusivity or co-occurrence of mutations revealed that several key genes showed significant interrelationships in the lung adenocarcinoma cohort. (F). Five first-line LUAD chemotherapeutic agents were highlighted: cisplatin, gemcitabine, docetaxel, etoposide, and vinorelbine. A significant correlation was found between KUNU and Docetaxel ($p < 0.01$) (G).

(Zheng et al., 2025). Therefore, it is crucial to promptly identify patients at risk for recurrence and develop personalized treatment strategies tailored to their needs. As a trace element, selenium plays complex regulatory roles in various immune cells. In recent years, there has been growing interest in the role of selenium metabolism-related genes (SMRGs) in modulating tumor oxidative stress and reshaping the immune microenvironment. However, their prognostic significance and underlying molecular mechanisms remain largely unexplored (Fu et al., 2023; Ahmed et al., 2025). This study employed bioinformatics analysis to establish a risk

model based on selenium metabolism in LUAD and further examined its relationships with the tumor microenvironment, somatic mutations, and drug sensitivity. These findings enhance our understanding of LUAD progression and provide valuable insights for developing more precise therapeutic strategies.

In this research, we identified four hub genes associated with selenium metabolism through regression analysis. Additionally, we developed a risk model for calculating the selenium metabolism score. The four SMRGs identified in this study represent potential vulnerabilities in cancer cells, providing functional targets for novel

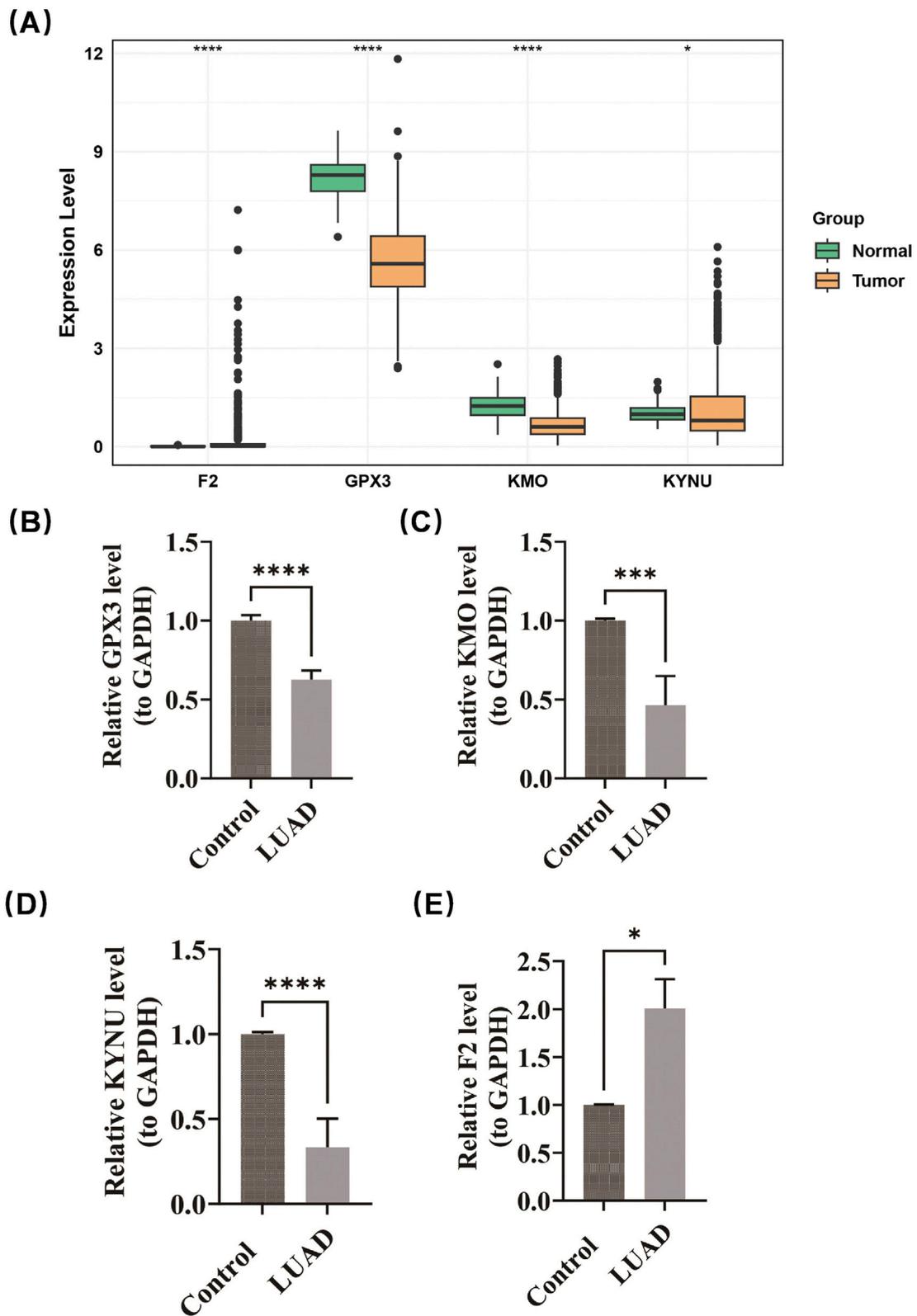


FIGURE 8 Expression validation confirmed the differential patterns of prognostic genes. GPX3 ($p < 0.0001$), KYNU ($p < 0.05$), and KMO ($p < 0.0001$) were found to be downregulated in LUAD, while F2 exhibited significant upregulation ($p < 0.0001$) (A). RT-PCR demonstrated that GPX3, KMO, and KYNU were significantly downregulated in LUAD, whereas F2 was upregulated in this context (B–E).

therapies against LUAD. GPX3, a member of the glutathione peroxidase family, inhibits ferroptosis by neutralizing lipid peroxides. Its downregulation is correlated with poor survival outcomes in high-risk patients characterized by elevated tumor mutational burden (TMB) and KEAP1 mutation enrichment (Shimada et al., 2022; Metlay et al., 1995; Zhang F. et al., 2025). Enrichment analysis revealed that GPX3 is significantly associated with antioxidant and peroxidase activities, suggesting that dysregulated oxidative stress contributes to tumor progression. Studies indicate that GPX3 may serve as a diagnostic biomarker for oxidative stress-induced encephalitis; furthermore, GPX4, another family member, regulates reactive oxygen species (ROS) levels in breast cancer cells to resist ferroptosis (Lee et al., 2021). In lung cancer specifically, GPX3 expression is silenced via methylation, a phenomenon linked to metastasis and chemotherapy resistance in LUAD. It is plausible that GPX3 expression correlates with clinical indicators such as disease stage and prognosis. Moreover, GPX3 emerges as a promising prognostic marker and therapeutic target for both LUAD and lung squamous cell carcinoma (LUSC) (Zhang et al., 2025b; Chełchowska et al., 2025). KMO functions as kynurenine 3-monooxygenase and plays a pivotal role in the kynurenine metabolic pathway by catalyzing the conversion of kynurenine to 3-hydroxykynurenine. This pathway has been implicated in inflammation, oxidative stress, and neurotoxicity (Zhang M. et al., 2025). Aberrant expression of KMO in tumors may significantly influence immune regulation within the tumor microenvironment. Research indicates that alterations in the kynurenine pathway are associated with mechanisms of immune evasion in cancer cells. Elevated KMO expression may modulate immune cell activity by affecting metabolite levels, thereby impacting patient outcomes (Chen et al., 2025). In stroke research, KMO has been shown to inhibit mitochondrial autophagy, facilitating brain repair following a stroke (Wang et al., 2024). The differential expression of KMO between high- and low-risk groups in this study may correlate with adverse prognoses; furthermore, KMO has been identified as a protective factor against recurrence in LUAD. Investigating compounds involved in selenium metabolism could elucidate the role of KMO in tumor immune escape and disease progression, potentially identifying it as a target for combinational therapy. KYNU, also known as kynureninase, is a key enzyme within the kynurenine pathway responsible for converting kynurenine into anthranilic acid. Dysregulation of KYNU has been observed across various cancers and is closely linked to tumor initiation and progression (Xu et al., 2025). Its metabolites can influence cellular redox states, thereby affecting both tumor cell proliferation and apoptosis (Zhang et al., 2025d). Additionally, changes in KYNU activity may impair immune cell function, contributing to tumor immune escape (Xiang et al., 2024). Within the risk model developed in this study, variations in KYNU expression may serve as critical determinants of patient prognosis. Exploring upstream and downstream regulatory pathways related to KYNU could uncover new therapeutic avenues for treating LUAD. The F2 gene encodes coagulation factor II (prothrombin), which plays a central role in the coagulation cascade. Beyond its classical function in hemostasis, emerging evidence links abnormalities involving coagulation factor

II and related pathways to tumor metastasis (Teofilov et al., 2025), angiogenesis (Kvasnička et al., 2024), and disease severity (PMID: 39497411). Activation of the coagulation system by tumor cells facilitates angiogenesis and metastasis. Altered expression of F2 may influence these processes, potentially elucidating the poorer outcomes observed in high-risk LUAD patients. For example, F2 may play a role in microthrombosis formation within tumors or regulate factors that drive angiogenesis. Clinically, evaluating coagulation parameters such as D-dimer levels in LUAD patients could provide insights into whether anticoagulant therapy enhances prognosis.

Subsequently, we undertook a more comprehensive exploration of the molecular characteristics associated with distinct risk populations through GSEA. Pathway enrichment analysis in the high-risk group revealed significant involvement of pathways related to cell proliferation, metabolism, and gene regulation. Notably, these included DNA replication, the cell cycle, spliceosome activity, and ribosomal pathways. These findings are consistent with previous studies that have documented uncontrolled cell cycle progression and aberrant DNA replication in LUAD (Wu et al., 2025). In contrast, the low-risk group demonstrated an enrichment of immune-related pathways. These encompassed those linked to asthma, autoimmune thyroid disease, transplant rejection, and immunoglobulin A (IgA) production within the intestinal immune network. Selenium intake has been associated with asthma prevalence (Zajac, 2021) and plays a role in selenium metabolism. Furthermore, serum selenium levels correlate with thyroid disorders; specifically, selenium deficiency is known to elevate the risk of autoimmune thyroid conditions (Troshina et al., 2024).

Immune cell infiltration analysis revealed an increased proportion of M2 macrophages and resting CD4 memory T cells in high-risk tumors, whereas low-risk tumors exhibited elevated levels of CD8⁺ T cells and resting dendritic cells. These findings corroborate previous studies regarding the roles of immune cells within the tumor microenvironment. M2 macrophages and neutrophils are known to promote tumor growth and facilitate immune evasion (Zeng et al., 2025), while CD8⁺ T cells and dendritic cells are pivotal in driving anti-tumor immunity. The differential expression of immune checkpoint genes across risk groups indicates varying efficacies of immunotherapy (Jiang et al., 2025). In the realm of cancer immunity, selenium has been shown to enhance lysosomal activity and cytotoxicity in CD8⁺ T cells (Chen et al., 2019). These results underscore the significance of the immune microenvironment in LUAD progression and advocate for personalized immunotherapeutic strategies. The development of selenium nanoparticles or selenium-containing compounds may improve selenium protein status in LUAD cells, thereby synergistically enhancing the efficacy of immunotherapy.

For LUAD cases that lack driver gene mutations, chemotherapy remains the primary treatment option, often in combination with immunotherapy or anti-angiogenic agents. The first-line regimen typically consists of platinum-based drugs alongside pemetrexed. In this study, we analyzed the differences in chemosensitivity between high-risk and low-risk LUAD patients. Notably, cisplatin demonstrated significantly lower IC50 values in the high-risk group compared to the low-risk group, indicating enhanced efficacy among high-risk patients.

Cisplatin induces cytotoxicity by damaging tumor cell DNA and activates the immune system through immunogenic cell death (ICD). In high-risk patients, cisplatin may enhance antigen presentation by dendritic cells via the release of immune-stimulatory molecules such as calreticulin (CRT) and HMGB1, thereby activating T-cell-mediated anti-tumor responses (Ma et al., 2025; Yasuda et al., 2025). Similarly, docetaxel, gemcitabine, and etoposide exhibit superior efficacy in high-risk patients; this is likely attributable to their ability to inhibit tumor angiogenesis, reverse epithelial-mesenchymal transition (EMT), and exert immunomodulatory effects such as reducing Treg activity (Massa et al., 2025; Obradovic et al., 2023) while promoting M1 polarization of tumor-associated macrophages (TAMs) (Jiménez-Cortegana et al., 2021). These agents also regulate the tumor microenvironment to suppress angiogenesis (Al-Omar et al., 2025), providing a rationale for combining chemotherapy with immunotherapy in high-risk LUAD patients. Our analysis of prognostic gene-drug associations revealed that GPX3 expression significantly correlates with drug IC50 values. It is crucial to conduct further research on how GPX3 interacts with selenium metabolism and chemotherapy drugs to influence both the occurrence and progression of LUAD.

This project systematically evaluated the pivotal role of SMRGs in LUAD by integrating transcriptome data analysis. We developed a risk score model based on these genes to distinguish high-risk patients and predict their prognosis. At the mRNA level, we conducted a preliminary validation of the four prognostic genes through RT-qPCR in five pairs of LUAD and adjacent non-tumor tissues. The expression difference patterns of the genes were basically consistent with the trends of the TCGA-LUAD data. However, it should be noted that the current validation only focused on the “expression differences of the genes between tumor and normal tissues”, and could only serve as preliminary exploratory evidence that the genes screened out by bioinformatics have real expression differences. The results suggest that patients in the high-risk group may experience a more complex immunosuppressive microenvironment. Future research could further explore the related pathways enriched by these core genes through cell and animal experiments to elucidate their specific mechanisms in tumor progression and immune evasion. Additionally, these genes hold potential as diagnostic and therapeutic targets, which may facilitate the development of novel selenium metabolism therapies or precise nutritional supply models, thereby providing new avenues for personalized treatment and precision medicine in LUAD. However, certain limitations must be acknowledged. Firstly, this study mainly relies on bioinformatics analysis of public datasets. There are two major deficiencies in the existing experimental validation: on the one hand, the expression of prognostic genes was only verified through RT-qPCR experiments on five pairs of LUAD and adjacent non-tumor tissues. Although the trend is consistent with the TCGA dataset and the authenticity of gene expression was preliminarily explored, the clinical application value of the prognostic model was not touched upon. On the other hand, due to the limitations of clinical sample acquisition conditions, the current validation sample size is only five pairs. Although technical repetition has ensured the reliability of the results, the small sample size may still lead to insufficient statistical power and cannot be widely promoted. Secondly, there is a lack of *in vitro* and *in vivo* functional

validation experiments, which makes the mechanism explanation of the model insufficient. In addition, as this study is a retrospective study, bias may be difficult to avoid; therefore, it is necessary to further verify the findings of this study through prospective studies.

5 Conclusion

In conclusion, we have developed a risk model associated with selenium metabolism genes to predict recurrence in patients diagnosed with LUAD. The accuracy of this model was further validated using an external validation cohort. This signature demonstrates a robust prognostic predictive capability and can be utilized to characterize the tumor microenvironment of LUAD. The novel methodologies and key genes identified in our study may offer valuable insights for advancing precision oncology in LUAD.

Data availability statement

The datasets presented in this study can be found in online repositories. The names of the repository/repositories and accession number(s) can be found in the article/Supplementary Material.

Author contributions

YC: Supervision, Formal Analysis, Resources, Writing – original draft, Software, Funding acquisition, Project administration, Validation, Methodology, Conceptualization, Investigation, Visualization, Writing – review and editing, Data curation. PL: Investigation, Writing – original draft, Software, Visualization, Project administration, Validation, Formal Analysis, Data curation, Writing – review and editing, Methodology. YW: Validation, Data curation, Methodology, Investigation, Writing – review and editing, Writing – original draft, Formal Analysis, Visualization. SS: Validation, Data curation, Formal Analysis, Visualization, Methodology, Writing – review and editing, Investigation, Resources, Writing – original draft. NC: Formal Analysis, Writing – original draft, Data curation, Methodology, Visualization, Validation, Investigation, Writing – review and editing. HP: Writing – review and editing, Funding acquisition, Formal Analysis, Visualization, Writing – original draft, Data curation, Validation, Supervision, Conceptualization, Project administration, Methodology. ZX: Resources, Writing – original draft, Project administration, Funding acquisition, Validation, Formal Analysis, Visualization, Writing – review and editing, Investigation, Supervision, Data curation, Conceptualization, Software, Methodology.

Funding

The author(s) declare that financial support was received for the research and/or publication of this article. This study was supported by the Yunnan Provincial Basic Research Program (Grant No. 202501AT070123), Joint Fund of Yunnan Science and

Technology Department and Kunming Medical University (Grant Nos. 202301AY070001-231 and 202401AY070001-257) and Yunnan Provincial Clinical Research Center for Respiratory Diseases (No. 2023YJZX-HX01). The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

Conflict of interest

The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Generative AI statement

The author(s) declare that no Generative AI was used in the creation of this manuscript.

Any alternative text (alt text) provided alongside figures in this article has been generated by Frontiers with the support of artificial intelligence and reasonable efforts have been made to ensure accuracy, including review by the authors wherever possible. If you identify any issues, please contact us.

References

- Ahmed, M. M., Mukheed, M., Tariq, T., Hasan, M., Shaaban, M., Mustafa, G., et al. (2025). Physio-biochemical insights into arsenic stress mitigation regulated by selenium nanoparticles in *Gossypium hirsutum* L. *BMC Plant Biol.* 25 (1), 482. doi:10.1186/s12870-025-06514-w
- Al-Omar, A., Asadi, M., Mert, U., Muftuoglu, C., Karakus, H. S., Goksel, T., et al. (2025). Effects of vinorelbine on M2 macrophages in non-small cell lung cancer. *Int. J. Mol. Sci.* 26 (5), 2252. doi:10.3390/ijms26052252
- Blanche, P., Dartigues, J. F., and Jacqmin-Gadda, H. (2013). Estimating and comparing time-dependent areas under receiver operating characteristic curves for censored event times with competing risks. *Stat. Med.* 32 (30), 5381–5397. doi:10.1002/sim.5958
- Chang, J., Wu, H., Wu, J., Liu, M., Zhang, W., Hu, Y., et al. (2023). Constructing a novel mitochondrial-related gene signature for evaluating the tumor immune microenvironment and predicting survival in stomach adenocarcinoma. *J. Transl. Med.* 21 (1), 191. doi:10.1186/s12967-023-04033-6
- Chelchowska, M., Gajewska, J., Szczepanik, E., Mazur, J., Cychol, A., Kuźniar-Palka, A., et al. (2025). Oxidative stress indicated by nuclear transcription factor Nrf2 and glutathione status in the blood of young children with autism spectrum disorder: pilot study. *Antioxidants (Basel)* 14 (3), 320. doi:10.3390/antiox14030320
- Chen, H., and Boutros, P. C. (2011). VennDiagram: a package for the generation of highly-customizable venn and euler diagrams in R. *BMC Bioinforma.* 12, 35. doi:10.1186/1471-2105-12-35
- Chen, Q., Liu, T., Chen, S., Luo, Y., Ma, M., Xue, F., et al. (2019). Targeted therapeutic-immunomodulatory nanoplatform based on noncrystalline selenium. *ACS Appl. Mater. Interfaces* 11 (49), 45404–45415. doi:10.1021/acsami.9b15774
- Chen, N., Zong, Y., Yang, C., Yi, Y., Zhao, J., et al. (2025). KMO-Driven metabolic reconfiguration and its impact on immune cell infiltration in nasopharyngeal carcinoma: a new avenue for immunotherapy. *Cancer Immunol. Immunother.* 74 (3), 75. doi:10.1007/s00262-024-03928-7
- Cheng, F., Luk, A. O., Tam, C. H. T., Fan, B., Wu, H., Yang, A., et al. (2020). Shortened relative leukocyte telomere length is associated with prevalent and incident cardiovascular complications in type 2 diabetes: analysis from the Hong Kong diabetes register. *Diabetes Care* 43 (9), 2257–2265. doi:10.2337/dc20-0028
- Correction to lancet psych (2023). *Lancet Psychiatry.* 10(4):e10. doi:10.1016/S2215-0366(22)00411-4
- Dizon, D. S., and Kamal, A. H. (2024). Cancer statistics 2024: all hands on deck. *CA Cancer J. Clin.* 74 (1), 8–9. doi:10.3322/caac.21824
- Duo, L., Chen, L., Zuo, Y., Guo, J., He, M., Zhao, H., et al. (2023). Machine learning model to estimate probability of remission in patients with idiopathic membranous nephropathy. *Int. Immunopharmacol.* 125 (Pt A), 111126. doi:10.1016/j.intimp.2023.111126
- Fu, X., Deng, Y., Xu, H., Shu, Y., and Chen, H. N. (2023). Selenium metabolism heterogeneity in pan-cancer: insights from bulk and single-cell RNA sequencing. *J. Cancer Res. Clin. Oncol.* 149 (17), 15535–15551. doi:10.1007/s00432-023-05333-6
- Geeleher, P., Cox, N., and Huang, R. S. (2014). pRRophetic: an R package for prediction of clinical chemotherapeutic response from tumor gene expression levels. *PLoS One* 9 (9), e107468. doi:10.1371/journal.pone.0107468
- Goldberg, S. B., and Herbst, R. S. (2023). The end of the beginning: progress and next steps in KRAS-Mutant non-small-cell lung cancer. *Lancet* 401 (10378), 706–707. doi:10.1016/S0140-6736(23)00288-X
- Gu, Z., and Hübschmann, D. (2022). Make interactive complex heatmaps in R. *Bioinformatics* 38 (5), 1460–1462. doi:10.1093/bioinformatics/btab806
- Gu, Z., Eils, R., and Schlesner, M. (2016). Complex heatmaps reveal patterns and correlations in multidimensional genomic data. *Bioinformatics* 32 (18), 2847–2849. doi:10.1093/bioinformatics/btw313
- Guo, C. H., Shih, M. Y., Chung, C. H., Lin, Y. C., Fan, C. T., Peng, C. L., et al. (2022). Fish oil and selenium with doxorubicin Modulates expression of Fatty acid receptors and selenoproteins, and targets multiple anti-cancer signaling in triple-negative breast cancer tumors. *Int. J. Med. Sci.* 19 (14), 2044–2057. doi:10.7150/ijms.75848
- Gustavsson, E. K., Zhang, D., Reynolds, R. H., Garcia-Ruiz, S., and Ryten, M. (2022). Ggtranscript: an R package for the visualization and interpretation of transcript isoforms using ggplot2. *Bioinformatics* 38 (15), 3844–3846. doi:10.1093/bioinformatics/btac409
- Hänzelmann, S., Castelo, R., and Guinney, J. (2013). GSEA: gene set variation analysis for microarray and RNA-Seq data. *BMC Bioinforma.* 14, 7. doi:10.1186/1471-2105-14-7
- Huang, C., Qiu, H., Xu, C., Tan, Z., Jin, M., Hu, J., et al. (2025). Downregulation of tropomyosin 2 promotes the progression of lung adenocarcinoma by regulating neutrophil infiltration through neutrophil elastase. *Cell Death Dis.* 16 (1), 264. doi:10.1038/s41419-025-07531-1
- Jiang, F., Yang, X., Shan, L., Miao, H., and Shi, C. (2025). 15-Deoxy- Δ -12,14-Prostaglandin J2 represses immune escape of lung adenocarcinoma by polarizing macrophages through epidermal growth factor Receptor/Ras/Raf pathway. *J. Immunother.* 48 (4), 119–126. doi:10.1097/CJI.0000000000000546
- Jiménez-Cortegana, C., Sánchez-Martínez, P. M., Palazón-Carrión, N., Nogales-Fernández, E., Henao-Carrasco, F., Martín García-Sancho, A., et al. (2021). Lower survival and increased circulating suppressor cells in patients with Relapsed/Refractory diffuse large B-Cell lymphoma with deficit of vitamin D levels using R-GDP plus lenalidomide (R2-GDP): results from the R2-GDP-GOTEL trial. *Cancers (Basel)* 13 (18), 4622. doi:10.3390/cancers13184622
- Kvasnička, J., Jansa, P., Cífková, R., Dušková, D., Bobčíková, P., Ševčík, M., et al. (2024). The incidence of the thrombophilic SNPs rs6025, rs1799963, rs2066865,

Publisher's note

All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

Supplementary material

The Supplementary Material for this article can be found online at: <https://www.frontiersin.org/articles/10.3389/fgene.2025.1655262/full#supplementary-material>

SUPPLEMENTARY FIGURE S1

The results of the proportional hazards assumption test were illustrated using Schoenfeld residual plots. The trend lines of the residuals did not show significant differences from one another ($p > 0.05$) (A–H). The overall proportional hazards assumption test for this model yielded $p = 0.228$, indicating compliance with the proportional risk assumption (I).

SUPPLEMENTARY FIGURE S2

The amplification curves and melting curves for significant genes identified in this study. (A–E) present the amplification curves for F2 (A), GPX3 (B), KMO (C), KYNU (D) and GAPDH (E), respectively. (F–J) present the melting curves for F2 (F), GPX3 (G), KMO (H), KYNU (I) and GAPDH (J), respectively.

- rs2289252, and rs176719 in chronic thromboembolic pulmonary hypertension. *Clin. Appl. Thromb. Hemost.* 30, 10760296241271369. doi:10.1177/10760296241271369
- Lee, N., Carlisle, A. E., Peppers, A., Park, S. J., Doshi, M. B., Spears, M. E., et al. (2021). xCT-Driven expression of GPX4 determines sensitivity of breast cancer cells to ferroptosis inducers. *Antioxidants (Basel)* 10 (2), 317. doi:10.3390/antiox10020317
- Lei, J., Qu, T., Cha, L., Tian, L., Qiu, F., Guo, W., et al. (2023). Clinicopathological characteristics of pheochromocytoma/paraganglioma and screening of prognostic markers. *J. Surg. Oncol.* 128 (4), 510–518. doi:10.1002/jso.27358
- Liang, B., Lin, W., Tang, Y., Li, T., Chen, Q., Zhang, W., et al. (2024). Selenium supplementation elevated SELENBP1 to inhibit fibroblast activation in pulmonary arterial hypertension. *iScience* 27 (10), 111036. doi:10.1016/j.isci.2024.111036
- Liu, L., Lin, J., and He, H. (2019). Identification of potential crucial genes associated with the pathogenesis and prognosis of endometrial cancer. *Front. Genet.* 10, 373. doi:10.3389/fgene.2019.00373
- Liu, T. T., Li, R., Huo, C., Li, J. P., Yao, J., Ji, X. L., et al. (2021). Identification of CDK2-Related immune forecast model and ceRNA in lung adenocarcinoma, a pan-cancer analysis. *Front. Cell Dev. Biol.* 9, 682002. doi:10.3389/fcell.2021.682002
- Love, M. I., Huber, W., and Anders, S. (2014). Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. *Genome Biol.* 15 (12), 550. doi:10.1186/s13059-014-0550-8
- Luo, Q., Bai, X., Li, X., and Liu, C. (2025). The role and mechanism of selenium in the prevention and progression of hepatocellular carcinoma. *Front. Oncol.* 15, 1557233. doi:10.3389/fonc.2025.1557233
- Ma, Y., Lai, P., Sha, Z., Li, B., Wu, J., Zhou, X., et al. (2025). TME-responsive nanocomposite hydrogel with targeted capacity for enhanced synergistic chemoimmunotherapy of MYC-amplified osteosarcoma. *Bioact. Mater.* 47, 83–99. doi:10.1016/j.bioactmat.2025.01.006
- Massa, A., Vita, F., Peraldo-Neia, C., Varamo, C., Basiricò, M., Raggi, C., et al. (2025). Doxycycline restores gemcitabine sensitivity in preclinical models of multidrug-resistant intrahepatic cholangiocarcinoma. *Cancers (Basel)* 17 (1), 132. doi:10.3390/cancers17010132
- Mayakonda, A., Lin, D. C., Assenov, Y., Plass, C., and Koeffler, H. P. (2018). Maftools: efficient and comprehensive analysis of somatic variants in cancer. *Genome Res.* 28 (11), 1747–1756. doi:10.1101/gr.239244.118
- Metlay, M. P., Johnson, J. L., Canterbury, J. D., Cottle, P., Nestor, C., Jr, Raman, S., et al. (1995). Are octupole vibrations harmonic? *Phys. Rev. C Nucl. Phys.* 52 (4), 1801–1806. doi:10.1103/physrevc.52.1801
- Newman, A. M., Liu, C. L., Green, M. R., Gentles, A. J., Feng, W., Xu, Y., et al. (2015). Robust enumeration of cell subsets from tissue expression profiles. *Nat. Methods* 12 (5), 453–457. doi:10.1038/nmeth.3337
- Obrovic, A., Ager, C., Turunen, M., Nirschl, T., Khosravi-Maharlooee, M., Iuga, A., et al. (2023). Systematic elucidation and pharmacological targeting of tumor-infiltrating regulatory T cell master regulators. *Cancer Cell* 41 (5), 933–949.e11. doi:10.1016/j.ccell.2023.04.003
- Qing, J., Li, C., Hu, X., Song, W., Tirichen, H., Yaigoub, H., et al. (2022). Differentiation of T helper 17 cells may mediate the abnormal humoral immunity in IgA nephropathy and inflammatory bowel disease based on shared genetic effects. *Front. Immunol.* 13, 916934. doi:10.3389/fimmu.2022.916934
- Ritchie, M. E., Phipson, B., Wu, D., Hu, Y., Law, C. W., Shi, W., et al. (2015). Limma powers differential expression analyses for RNA-sequencing and microarray studies. *Nucleic Acids Res.* 43 (7), e47. doi:10.1093/nar/gkv007
- Sachs, M. C. (2017). plotROC: a tool for plotting ROC curves. *J. Stat. Softw.* 79, 2. doi:10.18637/jss.v079.c02
- Shimada, B. K., Swanson, S., Toh, P., and Seale, L. A. (2022). Metabolism of selenium, selenocysteine, and selenoproteins in ferroptosis in solid tumor cancers. *Biomolecules* 12 (11), 1581. doi:10.3390/biom12111581
- Smoot, M. E., Ono, K., Ruschinski, J., Wang, P. L., and Ideker, T. (2011). Cytoscape 2.8: new features for data integration and network visualization. *Bioinformatics* 27 (3), 431–432. doi:10.1093/bioinformatics/btq675
- Song, P., Li, W., Guo, L., Ying, J., Gao, S., and He, J. (2022). Identification and validation of a novel signature based on NK cell marker genes to predict prognosis and immunotherapy response in lung adenocarcinoma by integrated analysis of single-cell and bulk RNA-Sequencing. *Front. Immunol.* 13, 850745. doi:10.3389/fimmu.2022.850745
- Sui, Z., Wu, X., Du, L., Wang, H., Yuan, L., Zhang, J. V., et al. (2022). Characterization of the immune cell infiltration landscape in esophageal squamous cell carcinoma. *Front. Oncol.* 12, 879326. doi:10.3389/fonc.2022.879326
- Sun, Z., Wu, K., Feng, C., and Lei, X. G. (2023). Selenium-dependent glutathione peroxidase 1 regulates transcription of elongase 3 in murine tissues. *Free Radic. Biol. Med.* 208, 708–717. doi:10.1016/j.freeradbiomed.2023.09.010
- Teofilov, S., Miljanović, O., Vuckovic-Filipovic, J., and Djordjevic, N. (2025). F5 6665A>G polymorphism is associated with increased risk of venous thromboembolism in females. *Int. J. Mol. Sci.* 26 (6), 2403. doi:10.3390/ijms26062403
- Troshina, E. A., Platonova, N. M., Senyushkina, E. S., Ioutsy, V. A., Smolin, E. S., Nikankina, L. V., et al. (2024). Involvement of essential trace elements in the pathogenesis of thyroid diseases: diagnostic markers and analytical methods for determination. *Probl. Endokrinol.* 71 (1), 10–19. doi:10.14341/probl13402
- Wang, Y., Bai, Y., Cai, Y., Zhang, Y., Shen, L., Xi, W., et al. (2024). Circular RNA SCMH1 suppresses KMO expression to inhibit mitophagy and promote functional recovery following stroke. *Theranostics* 14 (19), 7292–7308. doi:10.7150/thno.99323
- Wodrich, M. D., Sawatlon, B., Busch, M., and Corminboeuf, C. (2021). The Genesis of molecular Volcano plots. *Acc. Chem. Res.* 54 (5), 1107–1117. doi:10.1021/acs.accounts.0c00857
- Wu, T., Hu, E., Xu, S., Chen, M., Guo, P., Dai, Z., et al. (2021). clusterProfiler 4.0: a universal enrichment tool for interpreting omics data. *Innov. (Camb)* 2 (3), 100141. doi:10.1016/j.xinn.2021.100141
- Wu, X., Xu, Q., Aiyiti, P., Meng, D., Zhang, J., Citarella, F., et al. (2025). The identification and prediction of lung adenocarcinoma prognosis using a novel gene signature associated with DNA replication. *Transl. Cancer Res.* 14 (3), 1971–1981. doi:10.21037/tcr-2024-2536
- Xiang, H., Kasajima, R., Azuma, K., Tagami, T., Hagiwara, A., Nakahara, Y., et al. (2024). Multi-omics analysis-based clinical and functional significance of a novel prognostic and immunotherapeutic gene signature derived from amino acid metabolism pathways in lung adenocarcinoma. *Front. Immunol.* 15, 1361992. doi:10.3389/fimmu.2024.1361992
- Xiao, X., Huang, G., Yu, X., and Tan, Y. (2025). Advances in selenium and related compounds inhibiting multi-organ fibrosis. *Drug Des. Devel. Ther.* 19, 251–265. doi:10.2147/DDDT.S488226
- Xu, Y., Lin, P., Zhu, Y., Zhang, Q., and Zhou, J. (2025). Applying integrated transcriptome and single-cell sequencing analysis to develop a prognostic signature based on M2-like tumor-associated macrophages for breast cancer. *Discov. Oncol.* 16 (1), 389. doi:10.1007/s12672-025-02161-7
- Xue, C., Gu, X., Zhao, Y., Jia, J., Zheng, Q., Su, Y., et al. (2022). Prediction of hepatocellular carcinoma prognosis and immunotherapeutic effects based on tryptophan metabolism-related genes. *Cancer Cell Int.* 22 (1), 308. doi:10.1186/s12935-022-02730-8
- Yasuda, K., Kiyota, N., Matsuura, K., Saito, S., Honma, Y., Imamura, Y., et al. (2025). Corrigendum: indications for chemoradiotherapy in older patients with locally advanced head and neck cancer in Japan: a questionnaire survey in the JCOG head and neck cancer study group. *Front. Oncol.* 15, 1558600. doi:10.3389/fonc.2025.1558600
- Zajad, D. (2021). Mineral micronutrients in asthma. *Nutrients* 13 (11), 4001. doi:10.3390/nul13114001
- Zeng, L., Wu, S., Li, Z., Tang, Y., Tan, Y., Liang, R., et al. (2025). A novel mitochondrial quality regulation gene signature for anticipating prognosis, TME, and therapeutic response in LUAD by multi-omics analysis and experimental verification. *Cancer Cell Int.* 25 (1), 138. doi:10.1186/s12935-025-03764-4
- Zhang, C., Yin, K., Liu, S. Y., Yan, L. X., Su, J., Wu, Y. L., et al. (2021). Multiomics analysis reveals a distinct response mechanism in multiple primary lung adenocarcinoma after neoadjuvant immunotherapy. *J. Immunother. Cancer* 9 (4), e002312. doi:10.1136/jitc-2020-002312
- Zhang, H., Zhang, X., Wang, X., Sun, H., Hou, C., Yu, Y., et al. (2022). Comprehensive analysis of TRP channel-related genes in patients with triple-negative breast cancer for guiding prognostic prediction. *Front. Oncol.* 12, 941283. doi:10.3389/fonc.2022.941283
- Zhang, T., Qi, M., Wu, Q., Xiang, P., Tang, D., and Li, Q. (2023). Recent research progress on the synthesis and biological effects of selenium nanoparticles. *Front. Nutr.* 10, 1183487. doi:10.3389/fnut.2023.1183487
- Zhang, F., Zhang, C., Sun, W., Xie, S., Wu, P., Zeng, G., et al. (2025a). Proteomic profiling and therapeutic targeting of oxidative stress in autoimmune encephalitis. *J. Mol. Neurosci.* 75 (2), 38. doi:10.1007/s12031-025-02332-9
- Zhang, Y., Chen, F., Cao, Y., Zhang, H., Zhao, L., and Xu, Y. (2025b). Identifying diagnostic markers and establishing prognostic model for lung cancer based on lung cancer-derived exosomal genes. *Cancer Biomark.* 42 (2), 18758592251317400. doi:10.1177/18758592251317400
- Zhang, M., Sun, Q., Yao, Y., Chen, X., Li, J., Yuan, T., et al. (2025c). Identification of a prognostic signature based on tumor-infiltrating B lymphocyte mRNA in head and neck squamous cell carcinoma. *J. Immunol. Res.* 2025, 9375885. doi:10.1155/jimr/9375885
- Zhang, Y., Feng, S., Lv, L., Wang, C., Kong, R., Zhong, G., et al. (2025d). Amino acid metabolism-related gene kynureninase (KYNU) as a prognostic predictor and regulator of diffuse large B-Cell lymphoma. *Biochem. Genet.* doi:10.1007/s10528-025-11047-w
- Zheng, Y., Sun, B., and Qu, Z. (2025). Adverse predictive value of ASPM on lung adenocarcinoma overall survival depended on chemotherapy status. *Future Sci. OA* 11 (1), 2489328. doi:10.1080/20565623.2025.2489328