Yuan Gao, Huxiong Zhang* and Xiaoxuan Tian*

Integrated analysis of TCGA data identifies endoplasmic reticulum stress-related lncRNA signature in stomach adenocarcinoma

https://doi.org/10.1515/oncologie-2023-0394
Received September 20, 2023; accepted November 29, 2023;
published online January 1, 2024

Abstract

Objectives: To investigaed the role of endoplasmic reticulum stress (ERS)-related long non-coding RNAs (lncRNAs) in stomach adenocarcinoma (STAD) using TCGA data.
Methods: This study integrated clinical, transcriptomic, and tumor data from the Cancer Genome Atlas (TCGA). The expression of ERS genes was evaluated, alongside their association with identified lncRNAs. Gene set enrichment analysis and immune cell infiltration analysis were performed to elucidate the biological pathways influenced by these lncRNAs.
Results: The study identified five lncRNAs – AC012055.1, LINC01235, LINC00571, LINC02073, and CFAP61-AS1 – strongly correlated with ERS pathways and cancer prognosis. A prognostic model based on these lncRNAs was developed and validated across low- and high-risk groups. Potential biological pathways associated with these lncRNAs were uncovered through immune cell infiltration and GSEA. Additionally, screening identified drugs potentially effective against STAD, highlighting co-expressed genes as probable therapeutic targets.
Conclusions: This research offers detailed insights into the molecular mechanisms of STAD, enhancing understanding of potential therapeutic targets and showing promise for clinical applications.

Keywords: stomach adenocarcinoma; lncRNA; endoplasmic reticulum stress; prognostic model

Introduction

Stomach adenocarcinoma (STAD) is a highly prevalent global malignancy, with approximately one million cases annually [1]. It is known to be a highly aggressive cancer that often results in poor patient outcomes [2]. Although the exact cause of STAD remains unclear, emerging evidence suggests the involvement of endoplasmic reticulum stress (ERS) and its associated signaling pathways in its development and progression. For instance, Ogawa et al. observed a significant correlation between ERS markers and STAD progression [3], and Koike et al. identified that STAD tissues exhibit heightened ERS signaling, suggesting its potential role in tumor malignancy [4]. The endoplasmic reticulum (ER) plays an important role in lipid biosynthesis [5], protein folding [6], and trafficking [7]. Homeostatic disturbances can result in the buildup of misfolded proteins and initiating ERS [8]. Persistent and severe ERS can ultimately lead to cell death. Consequently, cells activate the unfolded protein response (UPR) to restore cellular balance [9, 10]. Beyond STAD, ERS has been implicated in the pathogenesis of various cancers [11], reinforcing its potential role in oncogenesis. Notably, the expression of several UPR-related genes has been reported to be increased in STAD tissues, marking the activation of the UPR signaling pathway [12]. Moreover, recent studies highlight the aberrant expression of long non-coding RNAs (lncRNAs) in STAD [13]. Intriguingly, emerging literature, such as the findings of Cui et al. [14], suggests that lncRNAs can modulate ERS and UPR signaling, offering a novel avenue of exploration in STAD's molecular landscape.

LncRNAs, a class of non-coding RNAs exceeding 200 nucleotides in length, are transcribed from the genome similarly to protein-coding genes [15]. They regulate transcription, post-transcriptional processes, and chromatin modification [16]. Previous studies have implicated dysregulated lncRNAs in the pathogenesis of several diseases, including cancer [17, 18]. Additionally, Pan and Xie reported that lncRNAs regulate ERS and UPR signaling in cancer [19]. Numerous studies have pinpointed distinct lncRNAs involved in regulating ERS across different cancer types, including STAD [20]. For example, the lncRNA XIST has been shown to inhibit ERS-induced apoptosis in STAD

***Corresponding authors: Huxiong Zhang**, Inner Mongolia Autonomous Region Institute of Ultrasound Imaging, Ordos Central Hospital, Ordos City, Inner Mongolia Autonomous Region, 017000, China,
E-mail: zhanghuxiong821124@163.com; **Xiaoxuan Tian**, State Key Laboratory of Component-based Chinese Medicine, Tianjin University of Traditional Chinese Medicine, Tianjin, 301617, China; and Haihe Laboratory of Modern Chinese Medicine, Tianjin, 301617, China,
E-mail: tian_xiaoxuan@tjutcm.edu.cn
Yuan Gao, State Key Laboratory of Component-based Chinese Medicine, Tianjin University of Traditional Chinese Medicine, Tianjin, China; and Haihe Laboratory of Modern Chinese Medicine, Tianjin, China

cells [21]. Another lncRNA, UCA1, has been reported to regulate ERS and promote cell proliferation in STAD [22]. Hence, it is essential to further investigate the role of lncRNAs in STAD-associated ERS.

In this study, by integrating datasets from The Cancer Genome Atlas (TCGA) – including expression, mutation, and clinical data – the focus was on identifying prognostic lncRNAs and potential therapeutic targets for STAD. Our exploration led us to discern a subset of lncRNAs intricately linked to both ERS pathways and prognosis. This research led to the development of a prognostic model based on the identified lncRNAs, which was further validated across various clinical subgroups. The study involved evaluating Gene Set Enrichment Analysis (GSEA), immune cell infiltration, and the tumor microenvironment to understand the biological pathways potentially influenced by these lncRNAs. Moreover, our screening identified potential drugs against STAD and their affiliated co-expressed genes as therapeutic contenders. Notably, our drug sensitivity analysis unveiled that certain compounds manifested efficacy specifically in the C3 or C2 gene subtypes. Hence, gene subtype-based therapeutic strategies might usher in enhanced clinical outcomes for distinct prognostic risk profiles. Collectively, our research furnishes a holistic grasp of STAD's molecular intricacies and unveils potential therapeutic avenues for clinical deployment.

Materials and methods

Rationale for utilizing TCGA data analysis methods

This study utilized data from TCGA due to its comprehensive and well-validated nature, rendering it an ideal resource for cancer research. TCGA offers extensive genomic data, including transcriptomic, mutational, and clinical datasets, essential for identifying potential biomarkers and therapeutic targets in cancer research. Specifically, our focus on lncRNAs in the context of ERS in STAD necessitated a database with broad and detailed genomic information. The integrative bioinformatics approach adopted in this study, which combined gene expression profiles with clinical data, facilitated a comprehensive understanding of the molecular mechanisms underlying STAD. This method also aided in identifying potential prognostic lncRNAs and therapeutic targets. This approach also facilitated the development and validation of a prognostic model based on these lncRNAs, emphasizing the utility of TCGA data in clinical applications for STAD.

Data preprocessing

The study acquired 448 transcriptome expression datasets, 434 tumor mutation datasets, and 443 clinical datasets from the TCGA database, which is accessible at https://tcga-data.nci.nih.gov/tcga/. This clinical data included gender, age, and survival details. Using the Perl programming language, this data was consolidated into a matrix file. The research identified 295 ERS-related genes (Table S1), by querying the Molecular Signatures Database (https://www.gsea-msigdb.org/gsea/msigdb/index.jsp) using ERS-specific keywords (Table S2).

Evaluation of differentially expressed ERS-related lncRNAs

The "limma" R package was utilized to identify 1364 ERS-associated lncRNAs, applying cutoff values of $|r|>0.6$ and $p<0.001$ using Pearson's correlation coefficient. Differentially expressed ERS-related lncRNAs were further selected by applying a threshold of adjusted $p<0.05$ and $|\log_2$ (fold change, FC)$|>1$, in a comparison between STAD samples and non-carcinoma samples. The 'limma' R package was employed to identify 1364 ERS-associated lncRNAs, using cut-off values of $|r|>0.6$ and $p<0.001$ as determined by Pearson's correlation coefficient. The selection of differentially expressed ERS-related lncRNAs was further refined by applying a threshold of adjusted $p<0.05$ and $|\log_2$ (fold change, FC)$|>1$, in comparisons between STAD samples and non-carcinoma samples.

Development of a prognostic model for ERS-related lncRNAs

A prognostic model for ERS-related lncRNAs was constructed in the following manner: initially, survival-related lncRNAs were identified from the pool of differentially expressed ERS-related lncRNAs, employing univariate Cox regression analysis and taking into account clinical data from STAD patients. Subsequently, the identified prognostic lncRNAs were subjected to LASSO-Cox regression analysis to mitigate overfitting. This process involved one thousand cycles and was conducted with a significance level set at $p<0.05$. The risk scores were then calculated using the following formula:

$$\text{Risk score} = \sum_{i=1}^{n} (\text{coe}f_i * \exp r_i)$$

where Σ represents the sum of all included lncRNAs, coef_i represents coefficient index, and expr_i represents expression level.

Verification of the risk model

Initially, the entire TCGA cohort underwent random division into training and testing groups at a 1:1 ratio. Observations were allocated to either the low- or high-risk group based on their median risk scores, and the predictive efficacy of the risk model was assessed using the Chi-square test. Survival curves were generated according to the number of events in each group, utilizing the "survival" and "survminer" packages. The reliability of the risk scores was evaluated using the testing set. Furthermore, univariate Cox regressions, LASSO, and multivariate Cox regressions were conducted using STAD patients' clinical data, including ages, gender, and stage. The prognostic effectiveness was assessed using time-dependent Receiver Operating Characteristic (ROC) curve analysis and Harrell's concordance index (C-index) with the "survival", "survminer", and "timeROC" R packages.

Nomogram establishment

A nomogram was constructed to evaluate the reliability of the prediction outcomes, incorporating factors such as gender, age, grade, risk score, and tumor stage. The "rms" R package was utilized for this purpose, and Hosmer-Lemeshow goodness-of-fit test curves were created as line graphs for 1-, 2-, and 3-year overall survival (OS). The predictive performance of the nomogram was assessed using the C-index and time-dependent ROC curve, computed with the "survival" and "timeROC" R packages.

Gene ontology and Kyoto encyclopedia of genes and genomes enrichment analysis

In this study, an enrichment analysis was conducted to elucidate various signaling pathways and their biological effects, based on the findings obtained. This analysis was facilitated using the 'clusterProfiler' package, an R-based tool that offers a streamlined way to perform statistical analysis and visualization of functional profiles for genes and gene clusters. The package can be accessed via the Comprehensive R Archive Network at CRAN URL for clusterProfiler. For the analysis, both the p-value and q-value thresholds were set at 0.05. Statistically significant enrichment was defined as cases where both the False Discovery Rate (FDR) and p-value were less than 0.05. For reference to the Gene Ontology (GO) and Kyoto Encyclopedia of Genes and Genomes (KEGG) databases, please visit:

GO: http://geneontology.org/; KEGG: https://www.genome.jp/kegg/.

Gene set enrichment analysis

Several R packages were employed to further explore changes in gene expression among different risk groups in the study. These include: 'org.Hs.e.g. db': An R package providing mappings between Entrez Gene IDs and various gene identifiers. This package is crucial for gene annotation in our analysis. Available at Bioconductor URL for org.Hs.e.g. db. 'limma': utilized for analyzing gene expression data from microarray or RNA-seq technologies. It uses linear models for differential expression in complex experiments. Available at Bioconductor URL for limma. 'enrichplot': this package assists in visualizing functional enrichment results, making it easier to interpret gene set enrichment analysis. Available at Bioconductor URL for enrichplot. 'clusterProfiler': as previously mentioned, this package was used for both statistical analysis and visualization of functional profiles. For Gene Set Enrichment Analysis (GSEA), refer to the official GSEA website: http://www.gsea-msigdb.org/gsea/index.jsp.

Tumor mutation burden analysis

The R package "maftools" was utilized to investigate the Mutation Annotation Format (MAF) data on somatic mutations, and the tumor mutation burden (TMB) score for each STAD sample was calculated using a specified methodology:

$$\mathrm{TMB} = \frac{\text{total mutation}}{\text{total covered bases}} \times 10^6$$

Immune cell infiltration analysis and immune checkpoints analysis

The R package 'ESTIMATE' assesses stromal cell and immune cell scores, enabling the evaluation of tumor purity. Immune cell infiltration in tumor samples was analyzed using single-sample GSEA for 29 different immune cell types. Furthermore, the 'ggpubr' R package was used to evaluate immune checkpoint molecules and visualize them through boxplots in R.

Drug sensitivity analysis

Chemosensitivity analysis was conducted to explore potential differences in the sensitivity to chemotherapeutic drugs between low- and high-risk groups among STAD patients. The analysis of drug sensitivity was performed using the 'limma' package with a p-value threshold of 0.001, and diagrams illustrating the results were created using 'ggplot2,' with the 'ggpubr' package aiding in the visualization.

Consensus clustering

Consensus clustering was conducted on patients from TCGA database using the Consensus Cluster Plus R package, with a focus on the five selected ERS-related lncRNAs. The optimal number of clusters was determined through the analysis of the cumulative distribution function and consensus matrix.

Statistical analysis

Pearson correlation analysis was determined to investigate the association between ERS-related genes and lncRNAs using R version 4.2.2. The Wilcoxon signed-rank test was applied to assess immune infiltrating cells and p-value less than 0.05 was considered significant.

Results

Identification of LncRNAs co-expressed with RES in STAD

As depicted in Figure 1, the analysis was initiated by acquiring transcriptome data from 412 STAD tissues and 36 normal tissues, followed by conversion of Ensembl ID into official gene names and categorization of transcripts as either mRNAs or lncRNAs. Subsequently, 295 ERS-related genes were identified based on their mRNA expression. Potential ERS-related lncRNAs were identified by filtering out 958 lncRNA genes that exhibited a correlation coefficient greater than 0.6 and a significance level of $p<0.001$.

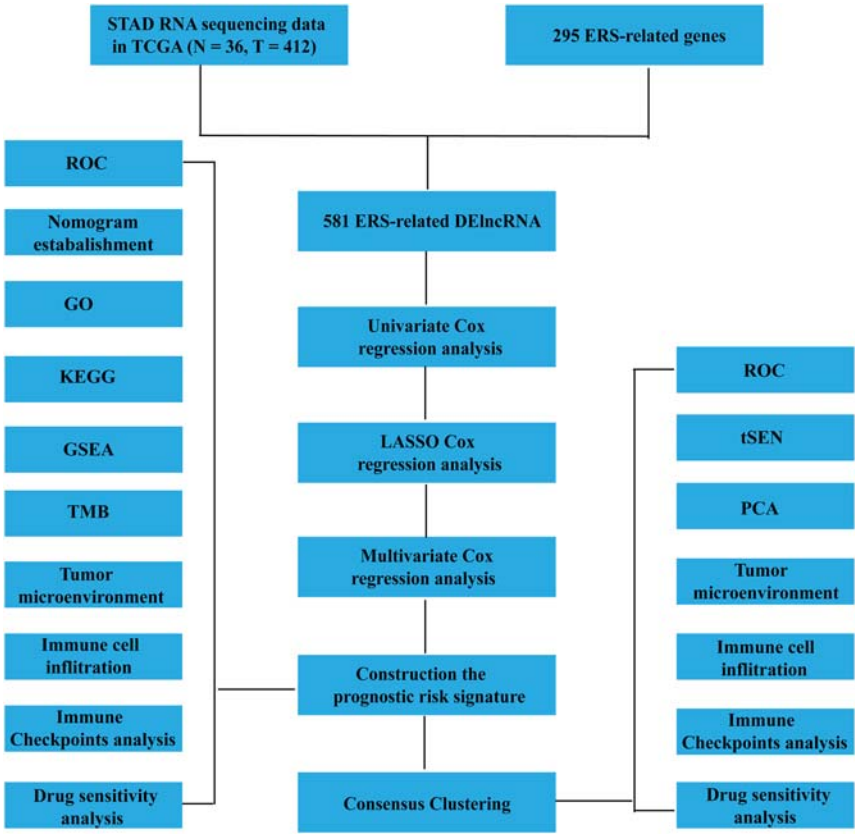

Figure 1: The workflow of this study.

Construction and the validation the prognostic risk signature

The WilcoxTest was employed to analyze the differential expression of lncRNAs co-expressed with ERS genes between tumor and normal groups in TCGA. In the tumor group, 523 genes exhibited upregulation, while 58 genes showed downregulation when compared to the normal group (Figure 2A and B). To construct a prognostic risk score model, clinical data of TCGA-STAD was downloaded and merged with the above differential expression data. The samples were divided into two groups: a training group for constructing the model and a testing group for validating the model. Univariate Cox regression analysis identified 32 lncRNAs (AL352979.4, AL137798.1, LINC02533, AC079467.1, LINC01354, AL022326.1, SNHG14, AC084880.3, AC100849.1, LINC01990, AC110491.1, AL161457.1, ADAMTS9-AS2, LINC01537, AC012055.1, AP006248.3, AC027243.3, RBMS3-AS3, AL033543.1, LINC01235, AC008808.1, GAS1RR, MRGPRF-AS1, LINC00571, AC073548.2, RMDN2-AS1, LINC02073, DIRC3, ARHGEF26-AS1, AL583829.1, AC109927.1 and CFAP61-AS1) that were associated with prognosis (Figure 2C and D). Using LASSO regression analysis, 12 lncRNAs (AL352979.4, LINC02533, AC110491.1, AL161457.1, AC012055.1, AP006248.3, AC027243.3, LINC01235, AC008808.1, LINC00571, LINC02073 and CFAP61-AS1) were selected (Figure 2E and F). Finally, five prognostic lncRNAs were identified using multivariate Cox regression analysis. The regression coefficients of the five lncRNAs were obtained using the R-package "Survival". The risk score was calculated based on the screened significant lncRNAs using the following formula: risk score = (1.2059 × AC012055.1) + (0.32184 × LINC01235) + (−1.3766 × LINC00571) + (1.0984 × LINC02073) + (0.3024 × CFAP61-AS1).

The prognostic power of the model was evaluated through several analyses, including Kaplan-Meier (KM) survival curve, independent prognostic, receiver operating characteristic (ROC) curve, and nomogram. The KM analysis showed that low-risk group patients had a better prognosis compared to the high-risk group ($p<0.001$), with an increasing death rate as the risk score increased (Figure 3A). Furthermore, a heatmap showed that LINC00571 was a protective lncRNA, while AC012055.1, LINC01235, LINC02073, and CFAP61-AS1 were risk lncRNAs (Figure 3B). ROC analysis revealed that the risk score exhibited superior predictive performance compared to other clinical indicators (Figure 3C), achieving an area under the ROC curve (AUC) of 0.656 for 1-year survival, 0.697 for 3-year survival, and 0.678 for 5-year survival (Figure 3D). Univariate and multivariate analyses of independent prognostic factors revealed that the risk score obtained from the prognostic prediction model was an independent prognostic factor (Figure 3E and F). A nomogram was developed to enhance

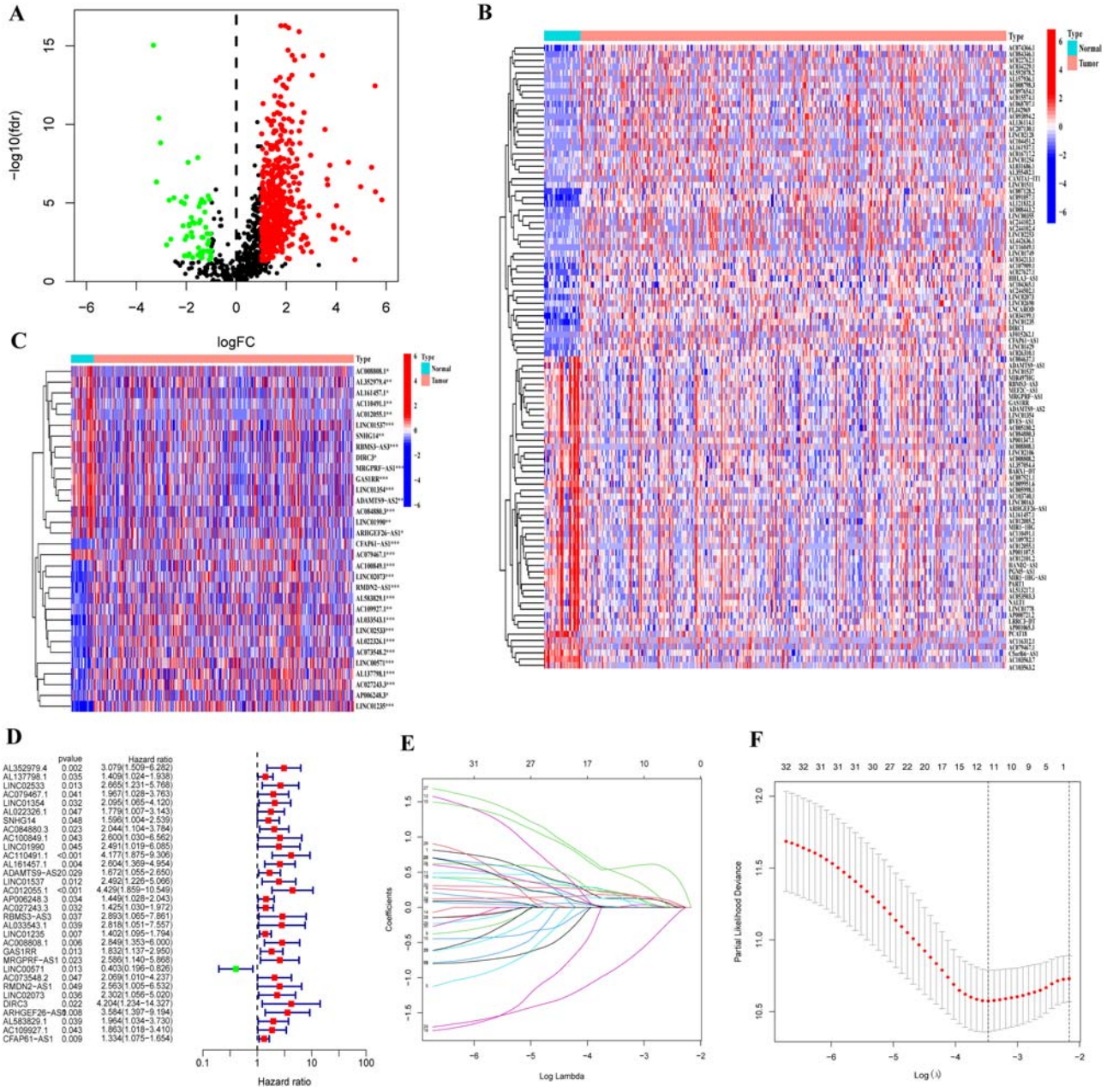

Figure 2: Identifying differentially expressed ERS-related lncRNAs in STAD. (A) Volcano plot depicting differential expression of ERS-related lncRNAs in STAD tissues compared to those in normal tissues; (B) heatmap illustrating the expression levels of ERS-associated differentially expressed lncRNAs; (C) heatmap displaying the expression levels of the 32 prognostic lncRNAs in all patients; (D) univariate Cox regression analysis of the 32 lncRNAs associated with prognosis; (E) distribution of LASSO coefficients for the ERS-related lncRNAs; (F) cross-validation procedure for optimizing LASSO regression parameters.

surgical prediction for STAD patients. The calibration curves for 1-, 2-, and 3-year OS demonstrated strong agreement between the nomogram's predictions and actual observations (Figure 3G and H). The C-index for this risk model was 0.695 (95 % CI: 0.671–0.719). These findings indicate that the newly constructed risk model is a more accurate predictor of prognosis (Figure S1A–F).

Identify the differentially expressed genes (DEGs) and functional enrichment analysis between high- and low-risk groups

Differential expression analysis was conducted to investigate the biological functions and pathways associated with the risk score. This analysis compared the high- and low-risk

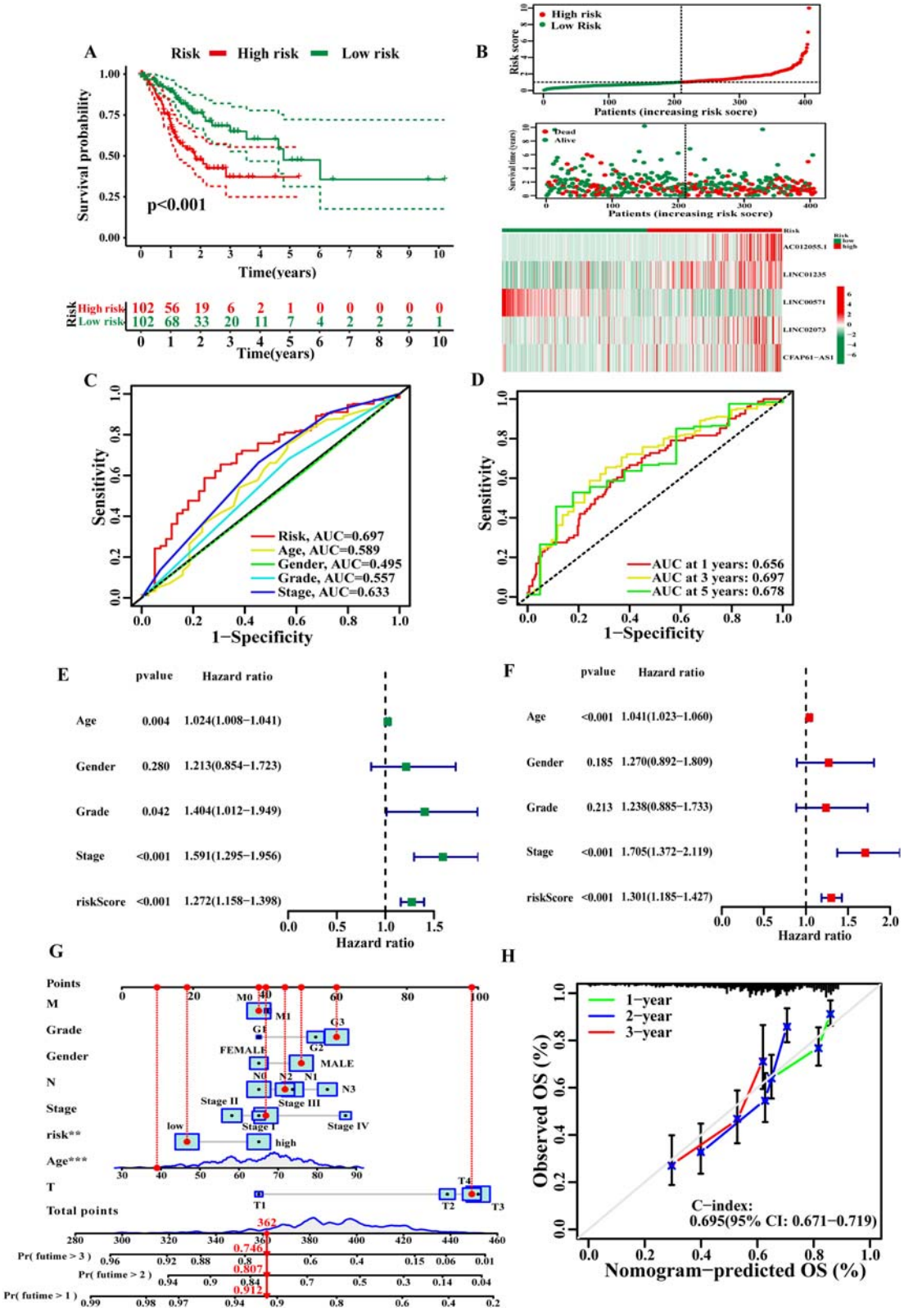

Figure 3: Prognostic value of the five ERS-related lncRNAs model. (A) KM survival curves comparing the OS of patients in these two groups in the training cohort; (B) the reference risk curve and heat map depicting the expression levels of the five ERS-related lncRNAs in the training cohort; (C) comparison of the predictive accuracy of the clinical characteristics and the risk model using ROC curves; (D) time-dependent ROC curve indicating the overall survival rates at 1, 2, and 3 years; (E) univariate Cox regression analysis of clinical features and the risk model for prognosis; (F) multivariate Cox regression analysis of clinical features and the risk model for prognosis; (G) nomogram integrating the risk score and other clinical factors to predict overall survival; (H) calibration curves for 1-, 2-, and 3-year survival predictions demonstrating the high agreement between predictions and observations.

groups, applying a cutoff value of |logFC|>1 and an FDR<0.05. Our analysis revealed 696 up-regulated genes and 54 down-regulated genes (Figure 4A and B). Subsequently, GO, KEGG enrichment analysis, and GSEA were conducted to identify enriched biological functions and pathways. Our GO enrichment analysis revealed that the DEGs were significantly enriched in collagen-containing extracellular matrix and receptor ligand activity (Figure 4C and D). KEGG analysis showed that the DEGs were significantly enriched in neuroactive ligand-receptor interaction and dilated cardiomyopathy (Figure 4E and F). Furthermore, GSEA analysis revealed that amino, nucleotide, and lipid metabolism were enriched in the low-risk group. In contrast, the high-risk group was enriched in vascular smooth muscle contraction, dilated cardiomyopathy, hypertrophic cardiomyopathy, focal adhesion, and neuroactive ligand receptor interaction (Figure 4G and H).

Analysis of mutation pattern between low- and high-risk groups

The relationship between the risk score and mutation patterns was investigated by conducting TMB analysis on 434 STAD samples obtained from TCGA. Significantly mutated genes associated with ERS were identified using Fisher's exact test with a p-value threshold of 0.01. It was observed that there were diverse somatic mutation profiles in both the low- and high-risk groups. The genes with the most mutations were TTN, TP53, MUC16, LRP1B, ARID1A, CSMD3, SYNE1, FAT4, FLG and PCLO, with TTN ranked first (Figure 5A and B). A significant negative correlation was observed between the risk score and TMB (R=−0.12, p=0.013) (Figure 5C and D). Additionally, patients with higher TMB scores displayed improved overall survival in the STAD cohort. Within the high-risk category, individuals with low TMB scores exhibited the poorest overall survival (Figure 5E). Moreover, patients in the low-risk group with higher TMB scores (Figure 5F).

Drug sensitivity analysis and immune cell infiltration analysis

Our analysis of the tumor microenvironment has revealed significant differences in immune and stromal scores between high- and low-risk groups (Figure 6A–C; p<0.001). Correlations between immune cell populations and risk scores were also investigated, with T cells and NK cells demonstrating significant negative correlations with risk scores, while other immune cell populations showed significant positive correlations (Figure 6D). Additionally, survival analyses performed on samples divided by immune cell content showed that cancer-associated fibroblast_MCPCOUNTER, Endothelial cell_XCELL, and Macrophage_TIMER had a higher prognosis in the low-risk group (p<0.05), while T cell follicular helper_CIBERSORT and uncharacterized cell_EPIC had a higher prognosis in the high-risk group (p<0.05) (Figure 6E–I). Our analysis of immune cell and function scores using ssGSEA further demonstrated significant differences between low- and high-risk groups (Figure 6J and H). Immune function scores were also significantly higher in the high-risk group (Figure 6L; p<0.05). Specifically, HLA (-DRA, -DOA, -DPB1, -DPA1, -DPB2, -DQA1, and -DRB6) expressions were significantly higher in the high-risk group than in the low-risk group (Figure 6L; p<0.05). These results highlight the important role of the tumor microenvironment and immune cell regulation in STAD and provide valuable insights for future research in this area.

Our tumor microenvironment analysis revealed that individuals in the high-risk group displayed a pronounced immune-hot phenotype, whereas those in the low-risk group exhibited immune-cold characteristics. These findings suggest that the high-risk group may benefit more from immune checkpoint inhibitor (ICI) treatment. Drug sensitivity results demonstrate that AZD8055, AZD8186, BMS-75807, CZC24832, Dactolisisb, AMG-319, AT13148, AZ960, AZDBB2, AZD2014, Dasatinib, Entospletinib, Foretinib, GNE-317, Luminespib, GSK2606414, IGFIR-3801, JAK-8517, JAKI_8709, Mirin, Nilotinib, NU7441, PCI-34051, PD173074, Pictilisib, PLX-4720, PRIMA-IMET, RVX-208, Staurosporine, Uprosertib, WZ4003 and XAV939 were effective in the low-risk group (Figures 7, S2, and S3). Conversely, Afatinib, Dabrafenib, Dihydrorotenone, Erlotinib, Gefitinib, Lapatinib, Osimertinib, PF-4708671, Ribociclib, Sapitinib and TAFI-5496 were more effective in the high-risk group (Figures 7, S2, and S3).

Using an unsupervised consensus clustering algorithm to validate the prognostic risk signature

This study employed a combination of statistical analyses and clustering algorithms to identify and characterize prognostic risk signatures in a cohort of cancer patients. Specifically, the approach involved the utilization of univariate Cox regression analysis, LASSO, and multivariate Cox regression analysis to develop the risk signature. This signature was then validated using unsupervised consensus clustering. This approach led us to identify three distinct gene subtypes, referred to as gene C1-3 (Figure 8A), with gene C3 showing the best survival prognosis (Figure 8B). To gain additional insights into the

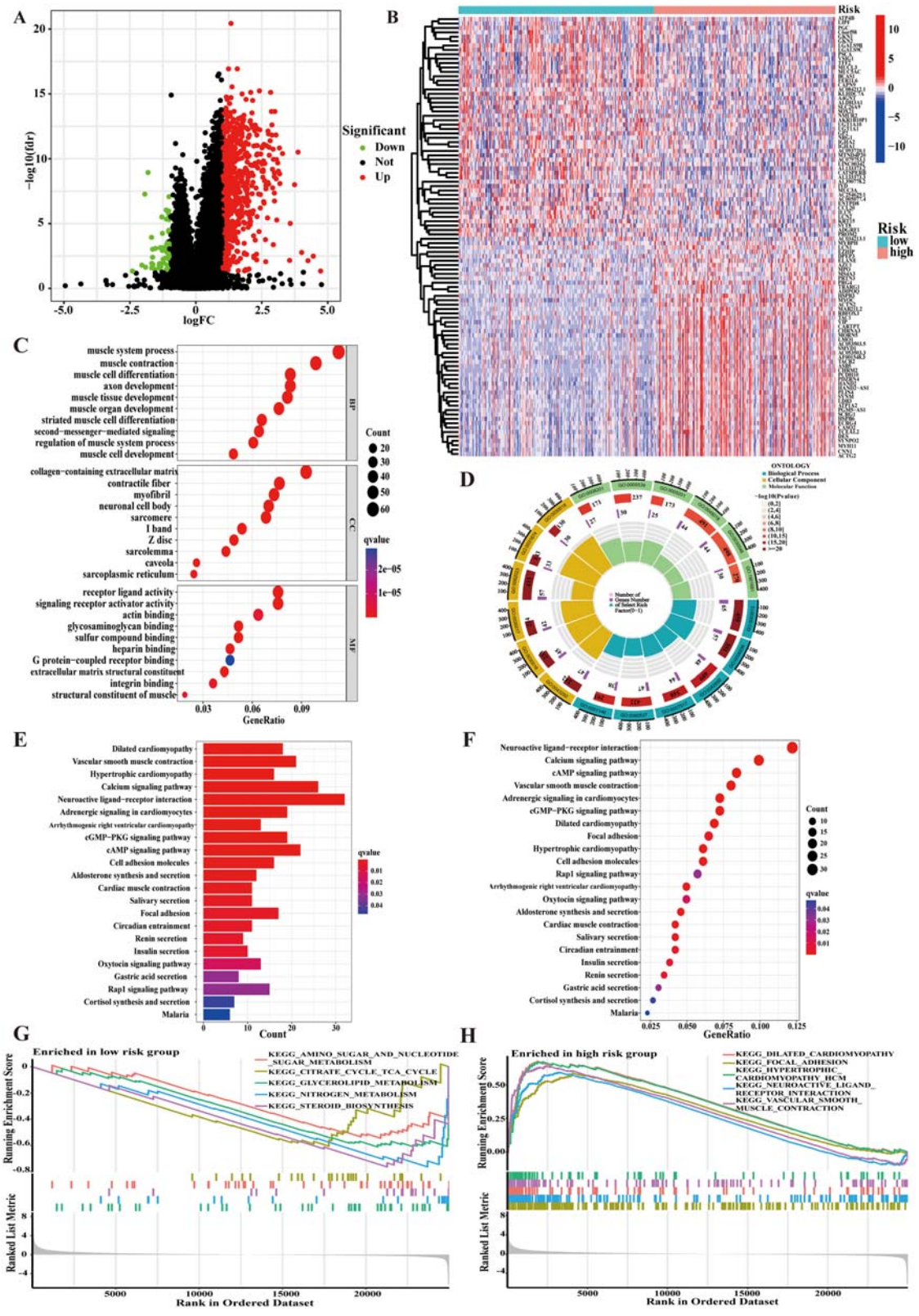

Figure 4: Identification of DEGs and functional enrichment analysis. (A) Volcano plot displaying the DEGs between the high- and low-risk groups; (B) heatmap of the top 50 DEGs; (C) GO analysis of DEGs enriched in biological processes; (D) GO analysis of DEGs enriched in molecular functions; (E) KEGG analysis of DEGs; (F) KEGG pathway analysis of DEGs; (G) GSEA showing enriched pathways in low-risk group; (H) GSEA showing enriched pathways in high-risk group.

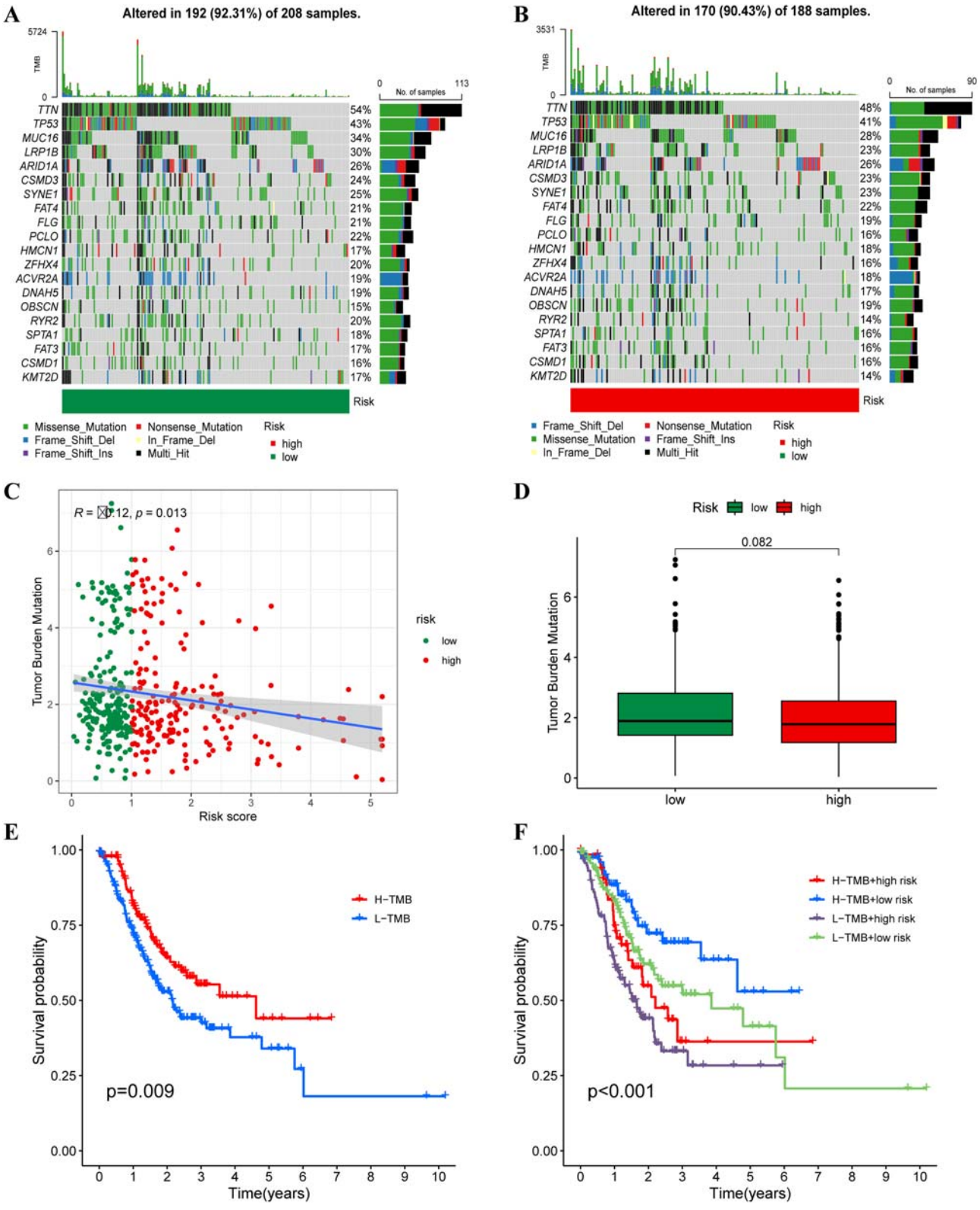

Figure 5: Illustrates the TMB analysis between the low- and high-risk groups. (A, B) The somatic mutations in each group; (C) the relationship between the risk score and TMB; (D) TMB distribution in STAD cases categorized by risk groups; (E, F) the correlation between TMB, which includes different risk scores, and survival probability.

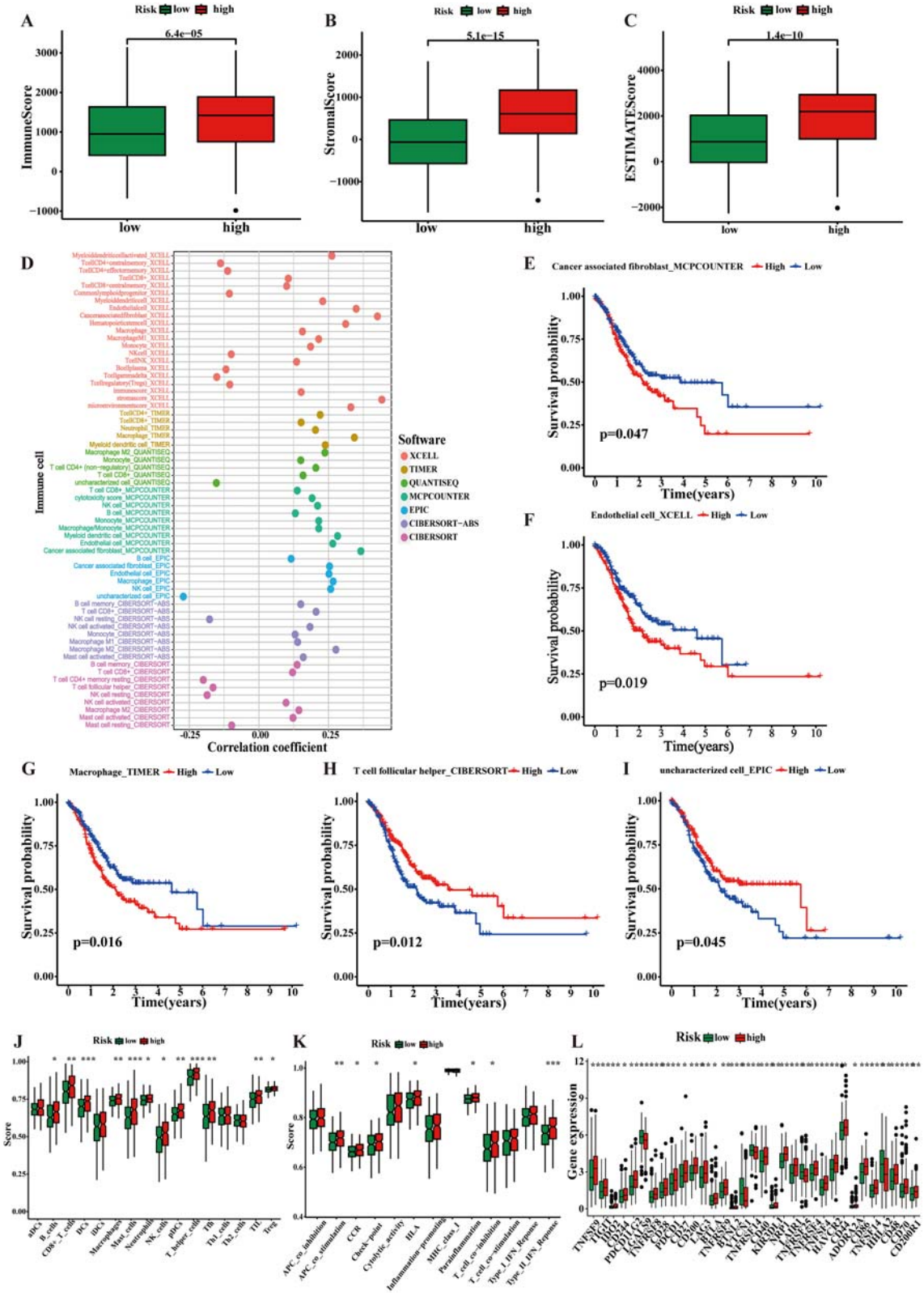

Figure 6: Presents the results of the immune cell infiltration analysis. (A–C) Box plots reveal a statistically significant difference in immune score, stromal score, and estimat score; (D) correlation analysis indicates a significant relationship between immune cell infiltration and risk score; (E–I) the correlation between immune cell infiltration and survival probability is demonstrated through KM curves; (J) ssGSEA enrichment analysis suggests the involvement of various immune cells in the tumor microenvironment; (K) functional enrichment analysis identifies immune-related pathways that are significantly enriched; (L) box plots show differential expression of immune checkpoint genes in the high-risk group compared to the low-risk group. *p<0.05, **p<0.01, ***p<0.001 vs. low-risk group.

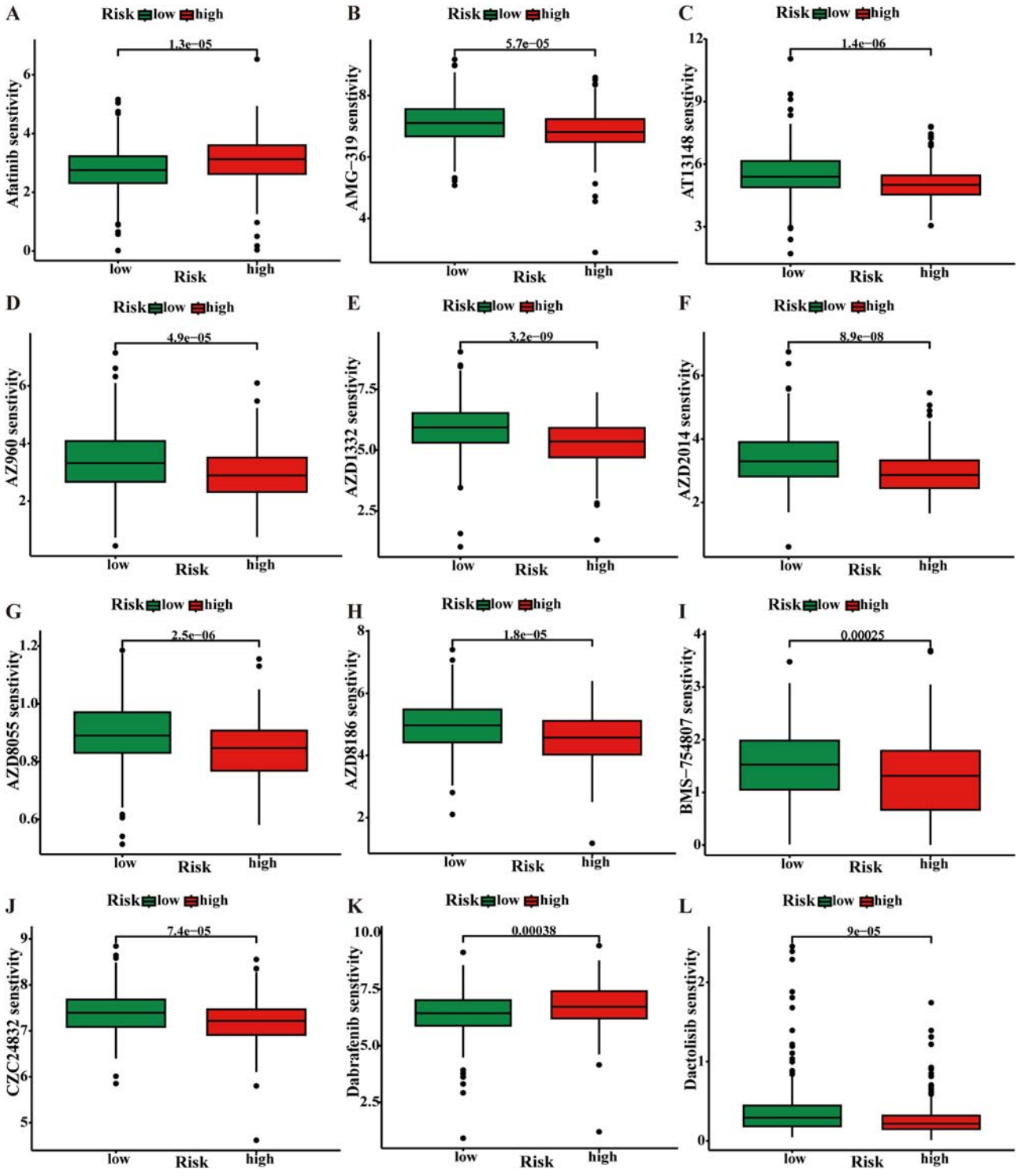

Figure 7: Analysis of drug sensitivity across high- and low-risk groups. (A) Sensitivity of Afatinib differentiated between high- and low-risk cohorts; (B) comparative sensitivity of AMG-319 in high-vs. low-risk groups; (C) differential sensitivity to AT13148 between the risk categories; (D) assessment of AZ960 sensitivity across the risk strata. (E) Sensitivity variation of AZD1332 between high- and low-risk groups; (F) AZD2014 sensitivity contrasted between the risk groups; (G) sensitivity profile of AZD8055 in the two risk classifications; (H) disparity in AZD8186 sensitivity between high- and low-risk patients; (I) evaluation of BMS-754807 sensitivity across high- and low-risk groups; (J) sensitivity differentiation of CZC24832 among the risk cohorts; (K) dabrafenib sensitivity comparison between high- and low-risk classifications; (L) examination of dactolisib sensitivity between the identified risk groups.

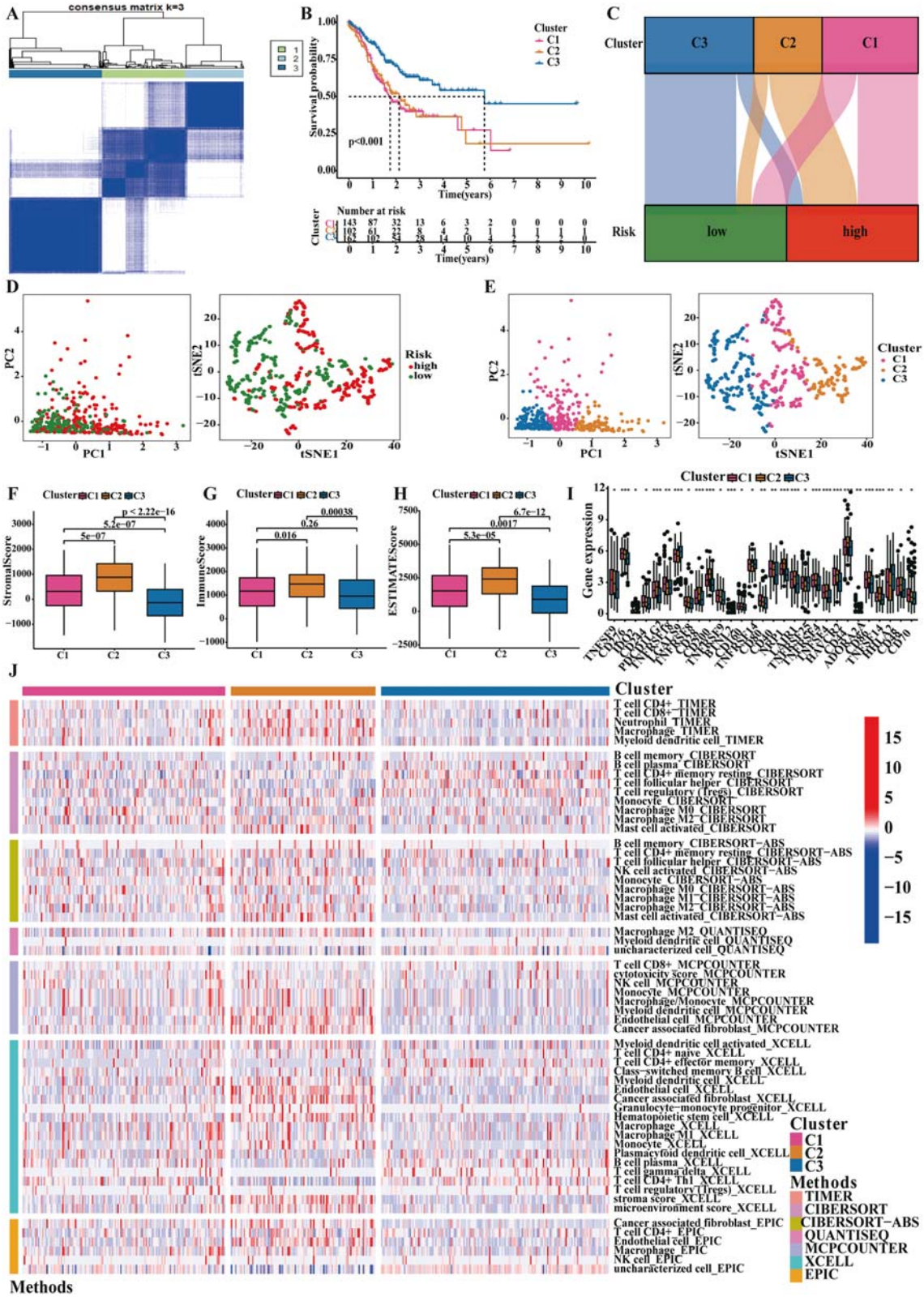

Figure 8: The relationship between gene subtypes and various aspects of the TME, including immune cell infiltration, immune checkpoint expression, TMB, and drug sensitivity. (A) The consensus clustering of STAD patients with K=3; (B) overall survival curves for three different gene subtypes; (C) the Sankey diagram illustrates the distribution of different risk groups and gene subtypes; (D) PCA and tSNE analysis of the high- and low-risk groups; (E) PCA and tSNE analysis of the different gene subtypes; (F–H) boxplots depicting the statistical differences in immune score, stromal score, and estimat score; (I) boxplots showing the differential expression of immune checkpoint genes; (J) the expression levels of immune cells in different gene subtypes.

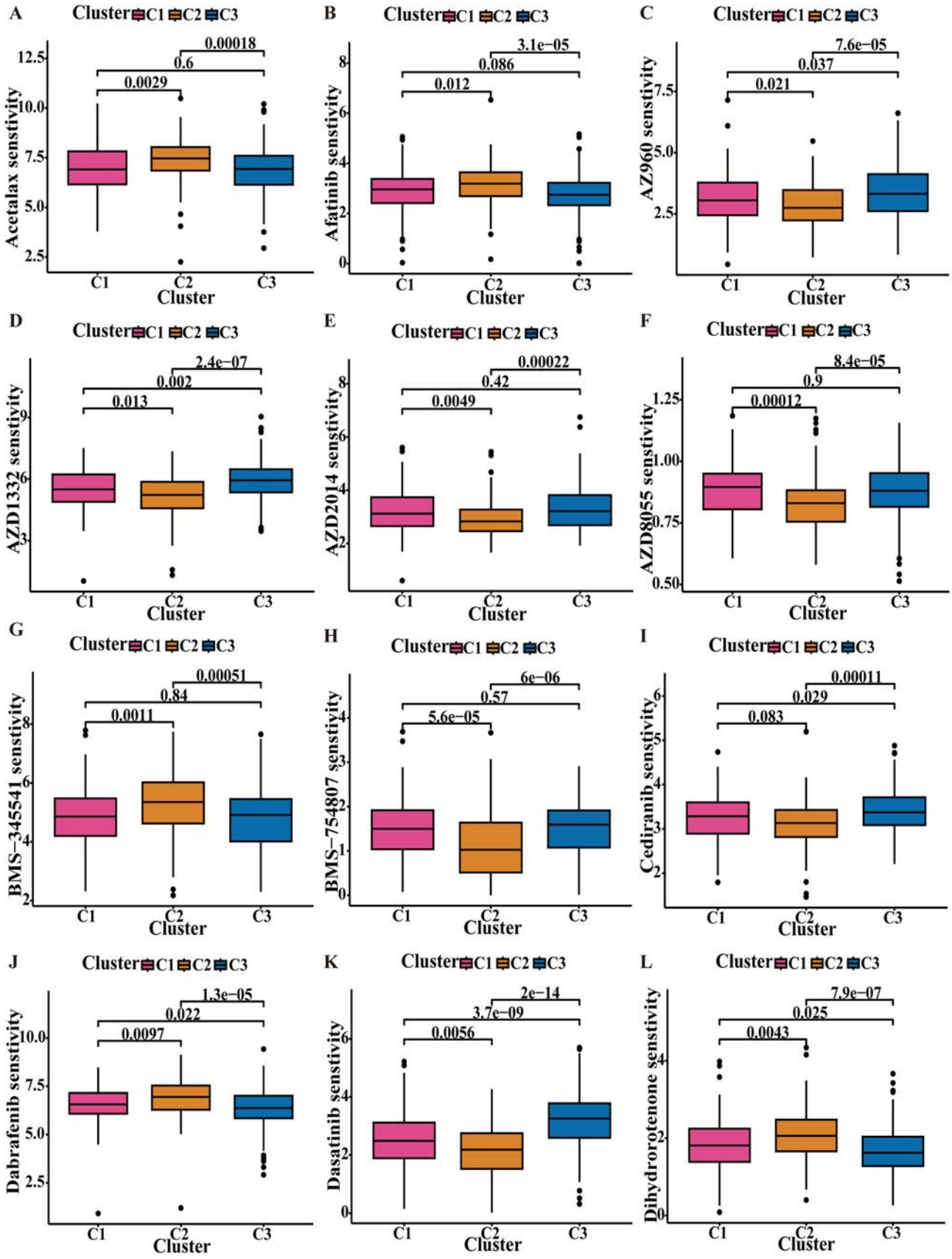

Figure 9: Analysis of drug sensitivity in different gene subtypes. (A) Sensitivity of Acetalax differentiated among C1, C2, and C3 gene subtypes; (B) comparative sensitivity of Afatinib in these groups; (C) differential sensitivity to AZ960 among the gene subtypes; (D) assessment of AZD1332 sensitivity across the gene strata. (E) Sensitivity variation of AZD2014; (F) AZD8055 sensitivity contrasted among groups; (G) sensitivity profile of BMS-345541 in the three gene classifications; (H) disparity in BMS-754807 sensitivity among C1, C2, and C3 patients; (I) evaluation of Cediranib sensitivity; (J) sensitivity differentiation of dabrafebib among the gene cohorts; (K) dasatinib sensitivity; (L) examination of dihydrorotenone sensitivity.

relationship between the risk score and genotypes, a Sankey diagram was employed. This diagram revealed that gene C3 was closely associated with the low-risk group, while the other two clusters were more closely associated with the high-risk group (Figure 8C). Dimensionality reduction analyses using PCA and t-SNE further confirmed these findings, with clear clustering observed between high and low-risk groups based on risk score, as well as between the three gene subtypes (Figure 8D and E). The study also investigated the relationship between the risk signature and the tumor microenvironment, including the evaluation of immune cell expression and scores for immune infiltration, stromal cells, and ESTIMATE scores. Notably, gene C3 exhibited significantly lower immunescores, stromalscores, and ESTIMATEscores compared to C1 and C2, suggesting a less immunogenic and stromal-rich tumor microenvironment in this group (Figure F-H; $p<0.001$). In line with this, the study observed that stroma score_XCELL, endothelial cell_XCELL, cancer-associated fibroblast_MCPCOUNTER, cancer-associated fibroblast_XCELL, and endothelial cell_MCPCOUNTER were all lower in C3 when compared to the other two gene subtypes (Figure 8J). The study's analysis of immune checkpoints revealed significant differences in the expression of several key immune checkpoint genes. These included ADORA2A, BTNL2, BTLA, CTLA4, C10orf54, CD (27, 28, 40, 40LG, 44, 48, 70, 80, 86, 160, 200, 200R1, 244, 274, 276), IDO (1, 2), ICOS, ICOSLG, LAG3, LAIR1, TNFRSF (4, 8, 9, 18), TNFSF (4, 9, 14, 15, 18), TIGIT, TMIGD2, PDCD1LG2, PDCD1, KIR3DL1, HAVCR2, VTCN1, and NRP1 (Figure 8I). Notably, gene C3 exhibited lower expression of several immune checkpoint genes, including CTLA4, LAG3, IDO1, and PDCD1LG2, which could potentially make it more responsive to immune checkpoint inhibitor therapy (Figure 8I). Overall, our findings highlight the potential clinical utility of our prognostic risk signature in guiding personalized treatment decisions for cancer patients, particularly those in the low-risk and gene C3 subgroups.

Our analysis on drug sensitivity underscored that distinct compounds manifested efficacy tailored to specific gene subtypes, either C3 or C2. For the C3 group, therapeutic potential was evident with compounds including AZ960, AZD1332, AZD2014, AZD8055, Dasatinib, BMS-754807, Cediranib, Doramapimod, Entospletinib, IGFIR_3801, JAK_8517, JQ1, NU7441, Nutlin-3a, PCI-34051, RO-3306, RVX-208, SB216763, Staurosporine, WZ4003, and XAV939 (Figures 9, S4, and S5). On the other hand, the C2 group exhibited responsiveness to Acetalax, Afatinib, BMS-345541, Dabrafebib, Dihydrorotenone, Erlotinib, Gefitinib, Lapatinib, Leflunomide, ML323, OSI-027, Oxaliplatin, Pevonedistat, PF-4708671, Sapitinib, and TAFFI_5496 (Figures 9, S4, and S5). These findings suggest that personalized therapeutic interventions based on gene subtypes may lead to better clinical outcomes for patients with different prognostic risk signatures.

Discussion

Adenocarcinomas of the esophagus, stomach, colon, or rectum collectively result in approximately 1.4 million global deaths each year [23]. Among them, STAD ranks fifth in incidence and third in cancer-related mortality, remaining a substantial global health concern [24]. The high incidence, aggressive nature, and drug resistance associated with STAD have compounded the challenges in therapy and prognostic management [25]. Hence, the quest for a novel molecular marker to assess prognosis and guide treatment in STAD patients assumes paramount importance.

Several studies have introduced prognostic signatures for assessing patient outcomes in various types of tumors [26–30]. Wu et al. [26] revealed the potential of ERS-related lncRNAs in predicting the prognosis of bladder cancer patients. Li et al. [27] showed that lncRNA MALAT1 may enhance the protective effect of DEX against acute lung injury by sequestering miR-135a-5p to suppress ERS. Lin et al. [28] identified that lncRNA DIRC1's expression significantly correlated with poor survival and immune infiltrations in STAD, suggesting its promise as a prognostic biomarker for this disease. Liao et al. [29] constructed a prognostic prediction model for STAD using seven lncRNAs -AL (353804.1, 355574.1, 161785.1), AC (010719.1, 009948.1, 005586.1), and TNFRSF10A-AS1-, which accurately predicted patient prognosis. Cai et al. [30] developed an immune signature to forecast both the prognosis and drug sensitivity of STAD patients. Furthermore, an immune-related lncRNA-based prognostic evaluation model has emerged as a promising therapeutic biomarker for STAD. However, the precise role of ERS-related lncRNAs in STAD necessitates further investigation.

Through the study, five lncRNAs associated with STAD were identified. Among them, AC012055.1, LINC01235, LINC02073, and CFAP61-AS1 exhibited a protective effect for prognosis, while LINC00571 demonstrated the opposite effect. Previous studies have identified LINC01235 (also known as FLJ41200; ENSG00000270547.1) as a cancer-associated gene located in the telomeric and centromeric regions of CD274 (PDL-1) at 9p23 in small-cell lung carcinoma [31]. This implies a potential association between LINC01235 and immune regulation in cancer. Additionally, LINC02073 has been implicated as a damage factor in cervical squamous cell carcinoma [32]. CFAP61-AS1, on the other hand, has been

found to be significantly associated with targeted therapy and prognosis in patients with depression [33]. Moreover, its expression was observed to be higher in STAD compared to normal samples and was shown to promote tumor growth and progression, indicating its potential as a prognostic predictor for STAD patients.

Subsequently, it was observed that the ERS-related lncRNAs identified in the study were enriched in a variety of biological pathways. Notably, amino sugar and nucleotide sugar metabolism are critical pathways involved in numerous cellular processes, such as glycosylation, glycosaminoglycan biosynthesis, and protein and lipid modifications. Aberrations in these metabolic pathways have been linked to the development and progression of several cancer types, including prostate [34], breast [35], liver [36], and lung cancer [37]. For instance, studies have demonstrated that changes in amino sugar metabolism can drive prostate cancer development [38]. Specifically, increased expression of the enzyme responsible for converting UDP-GlcNAc to UDP-GalNAc, a key step in glycosaminoglycan synthesis, was found to be associated with prostate cancer progression and poor patient prognosis [39]. In breast cancer, alterations in amino sugar metabolism have also been linked to disease progression [40]. Notably, increased expression of the enzyme responsible for converting UDP-GlcNAc to UDP-GalNAc was associated with poor patient prognosis and increased risk of metastasis [41]. Moreover, nucleotide sugar metabolism has been implicated in cancer development and progression [42], with alterations in the biosynthesis of sialic acid promoting tumor growth and metastasis in liver cancer [43]. Similarly, changes in the biosynthesis of UDP-GlcNAc have been linked to cancer cell proliferation and survival in lung cancer [44]. As the ER is the source of ERS, and the degradation of stress proteins aids in restoring cellular homeostasis, these findings suggest the need for further investigation into the molecular mechanisms underlying ERS-related lncRNAs in STAD [45].

Conclusions

This comprehensive bioinformatics analysis of TCGA data, centered on STAD, has brought the ERS pathway to the forefront, identifying key lncRNAs co-expressed with ERS-related genes. Our investigations into immune cell infiltration and the tumor microenvironment have revealed potential immunosuppressive agents impacting STAD prognosis. Further, GSEA enrichment has linked these lncRNAs to critical metabolic and cellular pathways. A prognostic model, derived from these lncRNAs, has been developed and validated across diverse clinical scenarios, affirming its reliability. Importantly, this study underscores the pivotal roles of lncRNAs in STAD's molecular landscape, offering new insights for diagnostic and therapeutic strategies. Looking ahead, the identified lncRNAs and their associated pathways present promising targets for future research. These findings pave the way for developing more personalized and effective treatments for STAD. Ongoing studies should focus on translating these bioinformatics findings into clinical trials, exploring the therapeutic potential of targeting these lncRNAs, and further unraveling their roles in STAD progression and response to therapy. This research forms a foundation for innovative approaches to combat this challenging malignancy, signifying a leap towards more targeted and precise oncological interventions.

Acknowledgment: The author thanks the TCGA network.
Research ethics: Not applicable.
Informed consent: Informed consent was obtained from all individuals included in this study.
Author contributions: X. X. Tian and H. X. Zhang design of the study; Y. Gao collected the data and performed the statistical analysis; Y. Gao wrote the first draft of the manuscript; X. X. Tian revised the manuscript and gave the final approval of the version to be submitted. All authors contributed to manuscript and approved the submitted version.
Competing interests: Authors state no conflict of interest.
Research funding: The study is funded from Foundation Science and Technology Program of Tianjin (No. 22ZYJDSS00040, 20ZYJDJC00120, China); Inner Mongolia Autonomous Region Health and Health Technology Plan (No. 202202368, China).
Data availability: This study's complete dataset, encompassing all data generated and analyzed, is fully accessible. This includes detailed information in the published article and comprehensive data in the supplementary files accompanying this publication.

References

1. Chang J, Wu H, Wu J, Liu M, Zhang W, Hu Y, et al. Constructing a novel mitochondrial-related gene signature for evaluating the tumor immune microenvironment and predicting survival in stomach adenocarcinoma. J Transl Med 2023;21:191.
2. Yang M, Sun X, Chen Y, Yang P. Twenty cases of gastric adenocarcinoma of the fundic gland type. Scand J Gastroenterol 2023;58:744–50.
3. Ogawa H, Kaira K, Takahashi K, Shimizu A, Altan B, Yoshinari D, et al. Prognostic role of BiP/GRP78 expression as ER stress in patients with gastric adenocarcinoma. Cancer Biomark 2017;20:273–81.
4. Koike H, Harada M, Kusamoto A, Xu Z, Tanaka T, Sakaguchi N, et al. Roles of endoplasmic reticulum stress in the pathophysiology of polycystic ovary syndrome. Front Endocrinol 2023;14:1124405.
5. Zhang Q, Li H, Wang S, Liu M, Feng Y, Wang X. Icariin protects rat cardiac H9c2 cells from apoptosis by inhibiting endoplasmic reticulum stress. Int J Mol Sci 2013;14:17845–60.

6. El-Gazzar A, Voraberger B, Rauch F, Mairhofer M, Schmidt K, Guillemyn B, et al. Bi-allelic mutation in SEC16B alters collagen trafficking and increases ER stress. EMBO Mol Med 2023;15:e16834.
7. Hong L, Xu Y, Wang D, Zhang Q, Li X, Xie C, et al. Sulforaphane ameliorates bisphenol A-induced hepatic lipid accumulation by inhibiting endoplasmic reticulum stress. Sci Rep 2023;13:1147.
8. Yu H, Zhen J, Yang Y, Gu J, Wu S, Liu Q. Ginsenoside Rg1 ameliorates diabetic cardiomyopathy by inhibiting endoplasmic reticulum stress-induced apoptosis in a streptozotocin-induced diabetes rat model. J Cell Mol Med 2016;20:623–31.
9. Duan X-C, Wang W, Feng D-X, Yin J, Zuo G, Chen D-D, et al. Roles of autophagy and endoplasmic reticulum stress in intracerebral hemorrhage-induced secondary brain injury in rats. CNS Neurosci Ther 2017;23:554–66.
10. Wu M-Z, Fu T, Chen J-X, Lin Y-Y, Yang J-E, Zhuang S-M. LncRNA GOLGA2P10 is induced by PERK/ATF4/CHOP signaling and protects tumor cells from ER stress-induced apoptosis by regulating Bcl-2 family members. Cell Death Dis 2020;11:276.
11. Sun Z, He Z, Liu R, Zhang Z. Cation lipid-assisted PEG6-PLGA polymer nanoparticles encapsulated knocking down long ncRNAs reverse non-coding RNA of Xist through the support vector machine model to regulate the molecular mechanisms of gastric cancer cell apoptosis. J Biomed Nanotechnol 2021;17:1305–19.
12. Lin H, Wang J, Wang T, Wu J, Wang P, Huo X, et al. The LncRNA MIR503HG/miR-224-5p/TUSC3 signaling cascade suppresses gastric cancer development via modulating ATF6 branch of unfolded protein response. Front Oncol 2021;11:708501.
13. Li Q, Zhang H, Hu J, Zhang L, Zhao A, Feng H. Construction of anoikis-related lncRNAs risk model: predicts prognosis and immunotherapy response for gastric adenocarcinoma patients. Front Pharmacol 2023;14: 1124262.
14. Cui Y, Xu H, Yang Y, Zhao D, Wen Y, Lv C, et al. The regulation of miR-320a/XBP1 axis through LINC00963 for endoplasmic reticulum stress and autophagy in diffuse large B-cell lymphoma. Cancer Cell Int 2021;21:305.
15. Wen X, Han W, Liu C. Long non-coding RNA TTTY15 silencing inhibits gastric cancer progression by sponging microRNA-98-5p to down-regulate cyclin D2 expression. Bioengineered 2022;13:7380–91.
16. Yuan X, He Y, Wang W. ceRNA network-regulated COL1A2 high expression correlates with poor prognosis and immune infiltration in colon adenocarcinoma. Sci Rep 2023;13:16932.
17. Schmitz SU, Grote P, Herrmann BG. Mechanisms of long noncoding RNA function in development and disease. Cell Mol Life Sci 2016;73:2491–509.
18. Yang H, Chen Y, Zeng M, Wu H, Zou X, Fang T, et al. Long non-coding RNA LINC01480 is activated by Foxo3a and promotes hydroquinone-induced TK6 cell apoptosis by inhibiting the PI3K/AKT pathway. Ecotoxicol Environ Saf 2023;255:114786.
19. Pan K, Xie Y. LncRNA FOXC2-AS1 enhances FOXC2 mRNA stability to promote colorectal cancer progression via activation of Ca^{2+}-FAK signal pathway. Cell Death Dis 2020;11:434.
20. Zhao N, Guo M, Zhang C, Wang C, Wang K. Pan-cancer methylated dysregulation of long non-coding RNAs reveals epigenetic biomarkers. Front Cell Dev Biol 2022;10:882698.
21. Gu S, Xie R, Liu X, Shou J, Gu W, Che X. Long coding RNA XIST contributes to neuronal apoptosis through the downregulation of AKT phosphorylation and is negatively regulated by miR-494 in rat spinal cord injury. Int J Mol Sci 2017;18:732.
22. Dong H-T, Liu Q, Zhao T, Yao F, Xu Y, Chen B, et al. Long non-coding RNA LOXL1-AS1 drives breast cancer invasion and metastasis by antagonizing miR-708-5p expression and activity. Mol Ther Nucleic Acids 2020;19:696–705.
23. Xu H, Liu T, Li J, Chen F, Xu J, Hu L, et al. Roburic acid targets TNF to inhibit the NF-κB signaling pathway and suppress human colorectal cancer cell growth. Front Immunol 2022;13:853165.
24. Shang F, Wang Y, Shi Z, Deng Z, Ma J. Development of a signature based on eight metastatic-related genes for prognosis of GC patients. Mol Biotechnol 2023;65:1796–808.
25. Zheng W, Bian S, Qiu S, Bishop CE, Wan M, Xu N, et al. Placenta mesenchymal stem cell-derived extracellular vesicles alleviate liver fibrosis by inactivating hepatic stellate cells through a miR-378c/SKP2 axis. Inflamm Regen 2023;43:47.
26. Wu M, Wu Z, Yan J, Zeng J, Kuang J, Zhong C, et al. Integrated analysis of single-cell and Bulk RNA sequencing reveals a malignancy-related signature in lung adenocarcinoma. Front Oncol 2023;13:1198746.
27. Li P, Gu L, Bian Q, Jiao D, Xu Z, Wang L. Long non-coding RNA MALAT1 enhances the protective effect of dexmedetomidine on acute lung injury by sponging miR-135a-5p to downregulate the ratio of X-box binding proteins XBP-1S/XBP-1U. Bioengineered 2021;12:6377–89.
28. Lin Y, Li Y, Chen Y, Zhang Z. LncRNA ALMS1-IT1 is a novel prognostic biomarker and correlated with immune infiltrates in colon adenocarcinoma. Medicine (Baltim) 2022;101:e31314.
29. Nie K, Deng Z, Zheng Z, Wen Y, Pan J, Jiang X, et al. Identification of a 14-lncRNA signature and construction of a prognostic nomogram predicting overall survival of gastric cancer. DNA Cell Biol 2020;39: 1532–44.
30. Cai J, Hu Y, Ye Z, Ye L, Gao L, Wang Y, et al. Immunogenic cell death-related risk signature predicts prognosis and characterizes the tumour microenvironment in lower-grade glioma. Front Immunol 2022;13: 1011757.
31. Vishnubalaji R, Shaath H, Elkord E, Alajez NM. Long non-coding RNA (lncRNA) transcriptional landscape in breast cancer identifies LINC01614 as non-favorable prognostic biomarker regulated by TGFβ and focal adhesion kinase (FAK) signaling. Cell Death Discov 2019;5: 109.
32. Lv X, Liu L, Li P, Yuan Y, Peng M, Jin H, et al. Constructing a novel signature based on immune-related lncRNA to improve prognosis prediction of cervical squamous cell carcinoma patients. Reprod Sci 2022;29:800–15.
33. Cheng B, Qi X, Meng P, Cheng S, Yang X, Liu L, et al. Genome-wide association studies in non-anxiety individuals identified novel risk loci for depression. Eur Psychiatry 2022;65:e38.
34. Arredouani MS, Lu B, Bhasin M, Eljanne M, Yue W, Mosquera J-M, et al. Identification of the transcription factor single-minded homologue 2 as a potential biomarker and immunotherapy target in prostate cancer. Clin Cancer Res 2009;15:5794–802.
35. Smith A, Cao X, Gu Q, Kubi Amos-Abanyie E, Tolley EA, Vidal G, et al. Characterization of the metabolome of breast tissues from non-hispanic black and non-hispanic white women reveals correlations between microbial dysbiosis and enhanced lipid metabolism pathways in triple-negative breast tumors. Cancers 2022;14:4075.
36. Dalwadi DA, Torrens L, Abril-Fornaguera J, Pinyol R, Willoughby C, Posey J, et al. Liver injury increases the incidence of HCC following AAV gene therapy in mice. Mol Ther 2021;29:680–90.
37. Duan W, Zhang R, Zhao Y, Shen S, Wei Y, Chen F, et al. Bayesian variable selection for parametric survival model with applications to cancer omics data. Hum Genomics 2018;12:49.
38. Blessing AM, Rajapakshe K, Reddy Bollu L, Shi Y, White MA, Pham AH, et al. Transcriptional regulation of core autophagy and lysosomal

genes by the androgen receptor promotes prostate cancer progression. Autophagy 2017;13:506–21.
39. Fu BC, Tabung FK, Pernar CH, Wang W, Gonzalez-Feliciano AG, Chowdhury-Paulino IM, et al. Insulinemic and inflammatory dietary patterns and risk of prostate cancer. Eur Urol 2021;79:405–12.
40. Li X, Kong X, Huo Q, Guo H, Yan S, Yuan C, et al. Metadherin enhances the invasiveness of breast cancer cells by inducing epithelial to mesenchymal transition. Cancer Sci 2011;102:1151–7.
41. Gupta R, Sahu M, Srivastava D, Tiwari S, Ambasta RK, Kumar P. Post-translational modifications: regulators of neurodegenerative proteinopathies. Ageing Res Rev 2021;68:101336.
42. Wang L, Huang J, Jiang M, Lin H, Qi L, Diao H. Inhibited PTHLH downstream leukocyte adhesion-mediated protein amino acid N-linked glycosylation coupling Notch and JAK-STAT cascade to iron-sulfur cluster assembly-induced aging network in no-tumor hepatitis/cirrhotic tissues (HBV or HCV infection) by systems-theoretical analysis. Integr Biol 2012;4:1256–62.
43. Jarahian M, Marofi F, Maashi MS, Ghaebi M, Khezri A, Berger MR. Re-expression of poly/oligo-sialylated adhesion molecules on the surface of tumor cells disrupts their interaction with immune-effector cells and contributes to pathophysiological immune escape. Cancers 2021;13:5203.
44. Lin R, Elf S, Shan C, Kang H-B, Ji Q, Zhou L, et al. 6-Phosphogluconate dehydrogenase links oxidative PPP, lipogenesis and tumour growth by inhibiting LKB1-AMPK signalling. Nat Cell Biol 2015;17:1484–96.
45. Hoang CQ, Hale MA, Azevedo-Pouly AC, Elsässer HP, Deering TG, Willet SG, et al. Transcriptional maintenance of pancreatic acinar identity, differentiation, and homeostasis by PTF1A. Mol Cell Biol 2016;36:3033–47.

Supplementary Material: This article contains supplementary material (https://doi.org/10.1515/oncologie-2023-0394).