



OPEN ACCESS

EDITED BY

Qun Zhang,
Nanjing Medical University, China

REVIEWED BY

Qianshan Ding,
Northwest University, China
Min Fang,
Zhejiang Cancer Hospital, China

*CORRESPONDENCE

Qi Wang
✉ qdfy_wq@qdu.edu.cn

RECEIVED 24 April 2023

ACCEPTED 05 June 2023

PUBLISHED 19 June 2023

CITATION

Wang Q, Zhang B, Wang H, Hu M, Feng H, Gao W, Lu H, Tan Y, Dong Y, Xu M, Guo T and Ji X (2023) Identification of a six-gene signature to predict survival and immunotherapy effectiveness of gastric cancer.
Front. Oncol. 13:1210994.
doi: 10.3389/fonc.2023.1210994

COPYRIGHT

© 2023 Wang, Zhang, Wang, Hu, Feng, Gao, Lu, Tan, Dong, Xu, Guo and Ji. This is an open-access article distributed under the terms of the [Creative Commons Attribution License \(CC BY\)](https://creativecommons.org/licenses/by/4.0/). 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 of a six-gene signature to predict survival and immunotherapy effectiveness of gastric cancer

Qi Wang*, Biyuan Zhang, Haiji Wang, Mingming Hu, Hui Feng, Wen Gao, Haijun Lu, Ye Tan, Yinying Dong, Mingjin Xu, Tianhui Guo and Xiaomeng Ji

Department of Radiation Oncology, The Affiliated Hospital of Qingdao University, Qingdao, Shandong, China

Background: Gastric cancer (GC) ranks as the fifth most prevalent malignancy and the second leading cause of oncologic mortality globally. Despite staging guidelines and standard treatment protocols, significant heterogeneity exists in patient survival and response to therapy for GC. Thus, an increasing number of research have examined prognostic models recently for screening high-risk GC patients.

Methods: We studied DEGs between GC tissues and adjacent non-tumor tissues in GEO and TCGA datasets. Then the candidate DEGs were further screened in TCGA cohort through univariate Cox regression analyses. Following this, LASSO regression was utilized to generate prognostic model of DEGs. We used the ROC curve, Kaplan-Meier curve, and risk score plot to evaluate the signature's performance and prognostic power. ESTIMATE, xCell, and TIDE algorithm were used to explore the relationship between the risk score and immune landscape relationship. As a final step, nomogram was developed in this study, utilizing both clinical characteristics and a prognostic model.

Results: There were 3211 DEGs in TCGA, 2371 DEGs in GSE54129, 627 DEGs in GSE66229, and 329 DEGs in GSE64951 selected as candidate genes and intersected with to obtain DEGs. In total, the 208 DEGs were further screened in TCGA cohort through univariate Cox regression analyses. Following this, LASSO regression was utilized to generate prognostic model of 6 DEGs. External validation showed favorable predictive efficacy. We studied interaction between risk models, immunoscores, and immune cell infiltrate based on six-gene signature. The high-risk group exhibited significantly elevated ESTIMATE score, immunescore, and stromal score relative to low-risk group. The proportions of CD4⁺ memory T cells, CD8⁺ naive T cells, common lymphoid progenitor, plasmacytoid dendritic cell, gamma delta T cell, and B cell plasma were significantly enriched in low-risk group. According to TIDE, the TIDE scores, exclusion scores and dysfunction scores for low-risk group were lower than those for high-risk group. As a final step, nomogram was developed in this study, utilizing both clinical characteristics and a prognostic model.

Conclusion: In conclusion, we discovered a 6 gene signature to forecast GC patients' OS. This risk signature proves to be a valuable clinical predictive tool for guiding clinical practice.

KEYWORDS

gastric carcinoma, prognostic model, microenvironment, nomogram, bioinformatics

1 Introduction

Gastric cancer (GC) ranks as the fifth most prevalent malignancy and the second leading cause of oncologic mortality globally (1). There are several different types of GC, of which gastric adenocarcinoma accounts for 90% of total cases. The tumor, node, and metastasis (TNM) classification system and histological types are the most common methods for evaluating prognosis judgement and therapy guidance. Despite staging guidelines and standard treatment protocols, significant heterogeneity exists in patient survival and response to therapy for GC (2). Thus, an increasing number of research have examined prognostic models recently for screening high-risk GC patients. A large body of evidence suggests that immune cell infiltration in cancer has a critical function in carcinogenesis and progression, with much emphasis on predictive efficacy of immunotherapy (3, 4). A genetic analysis of Cancer Genome Atlas (TCGA) identified four distinct molecular subgroups of GC: Epstein-Barr virus (EBV) positive, microsatellite instability, genomic stability, and chromosomal instability (5). The EBV positive subtype presents with abundant PD-L1 expression, and has intensively described as a subset possibly profiting from immunotherapy (5). Infection with EBV triggers immune responses, and alters immune-related molecular components with immune cells recruitment (6). Although EBV positive GC patients are potentially eligible for immunotherapy theoretically, the efficacy of immune checkpoint inhibitors (ICIs) has been equivocal (7, 8). PD-L1 is a widely utilized prognostic biomarker for immunotherapy in variety of malignancies (9). Nevertheless, only around 20% of GC patients benefit from immunotherapy, and the immunological processes implicated in the processes are yet unknown (10). Due to high GC tumor heterogeneity, immune therapy can differ greatly from patient to patient. Furthermore, ICIs' high cost and limited availability significantly restrict their clinical application. Thus, the need to learn more about GC pathogenesis heterogeneity and to find new immunotherapeutic targets and prognostic markers has attracted increasing attention in recent years.

The current investigation utilized the Gene Expression Omnibus (GEO) and TCGA databases to formulate a prognostic signature of six genes for GC patients. At first, we explored differentially-expressed genes (DEGs) from TCGA and GEO databases across GC and adjacent non-tumor tissues. The gene ontology (GO) enrichment analysis and Kyoto Encyclopedia of Genes and Genomes (KEGG) indicated DEGs were potentially involved in modulation of tumor immune microenvironment (TME). We

further screened survival-related signatures and constructed a six-gene prognostic model among TCGA dataset. Based on the six-gene signature, we examined the link between risk models, immunoscores, immune cell infiltration, and cancer cell stemness. Overall, our study explored a six-gene risk model that can potentially identify GC patient's risk and predict immunotherapy response.

2 Methods and materials

2.1 Data source

Based on TCGA (<https://portal.gdc.cancer.gov/>), 375 GC samples and 32 non-tumor samples were downloaded, along with their mRNA expression profiles. Three independent datasets were obtained from GEO: GSE54129, GEO: GSE66229 (11), and GEO: GSE64951 (12). We obtained external validation data from GEO: GSE62254 (13). We examined the relationship across prognostic model and immunotherapy response in four immunotherapeutic cohorts: the IMvigor210 cohort (atezolizumab for locally advanced or metastatic urothelial cancer), the GSE78220 (melanoma with anti-PD-1 treatment), the GSE35640 (melanoma with MAGE-A3 immunotherapy), and GSE67501 (renal cell carcinoma with anti-PD-1 immunotherapy). The "IMvigor210CoreBiologies" R packages were utilized to retrieve the transcriptomic and clinical IMvigor210 variables. A log₂ transformation was conducted using limma Bioconductor package to transform the expression data from each database into fragments per kilobase of transcripts per million mapped reads (FPKM). Gene symbols identified by multiple probes were computed based on their average expression levels.

2.2 Identification ferroptosis-related DEGs

By using limma Bioconductor package, DEGs between GC and adjacent non-tumor tissues were detected among TCGA dataset, GSE54129, GSE62259, and GSE64951. The threshold values in the GSE54129, GSE66229, and GSE64951 were as follows: $\log_2|\text{fold change}| \geq 1$ and $P\text{-value} < 0.05$. The generation of heatmap and volcano plot was performed through the utilization of R "pheatmap" package (14). We analyzed GO and KEGG enrichment analyses to assess possible DEG functions (15, 16). BiNGO plugin for Cytoscape was used to analyze GO enrichment in DEGs.

2.3 Prognostic model construction and validation

Utilizing R “survival” package, a univariate Cox regression was done on TCGA cohort to assess overall survival (OS) related genes. Subsequently, we proceeded to generate a prognostic signature by means of the least absolute shrinkage and selection operator (LASSO) Cox regression method utilizing R “glmnet” package (17) in TCGA cohort. Each GC patient’s risk score was determined utilizing the following:

$$\text{Risk score} = \sum_i^n \text{coefficient} \times \text{DEGs expression}$$

The GC patient cohorts were segregated into low-risk and high-risk groups by means of the risk score median value. Subsequently, the assessment and comparison of OS times across groups was performed through Kaplan-Meier plot. Furthermore, prognostic model was evaluated for its sensitivity and specificity utilizing R “time ROC” package (18).

2.4 Immune landscape-risk score relationship

Through gene expression data transformation, ESTIMATE is capable of identifying the purity and activity of stromal and immune cells within TME. R packages “estimate” (19) was utilized to compare immunoscores abundance in high-risk and low-risk GC patients. By analyzing bulk samples using RNAseq profiles, the xCell can estimate 64 immune cell types abundance. R

“xCell” packages were utilized to analyze xCell scores (20). Each GC patient’s tumor-infiltrating immune cells and risk score were calculated relative to their abundance.

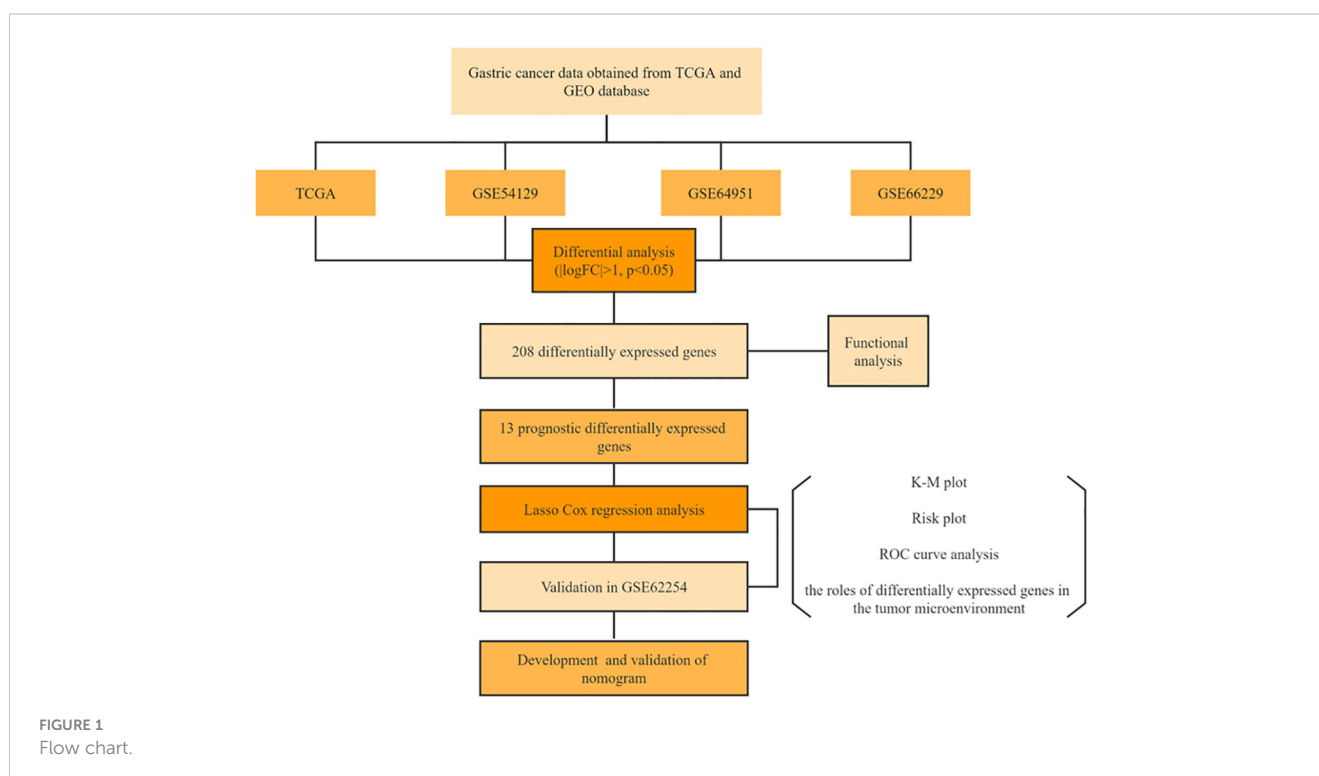
A total of eight transcripts were chosen for analysis based on their relevance to immunological checkpoints; as CD274, CTLA4, HAVCR2, LAG3, PDCD1, PDCD1LG2, TIGIT, and SIGLEC15. A heatmap illustrating risk scores and immune-checkpoint-relevant genes was generated utilizing R “pheatmap” package. We estimated potentiality of immunotherapy response utilizing TIDE algorithm (21).

2.5 Nomogram development and validation

To guide clinical decision-making, we developed a predictive nomogram combining predictive model risk score and clinical characteristics. A P-value<0.05 was used to screen survival-related clinical variables by univariate Cox analysis. After that, a nomogram was created using multivariate survival analysis. Calibration curves were utilized to plot nomogram predictions against measured rates. ROC curves were utilized to evaluate prognostic model’s specificity and sensitivity. We used R packages “rms”, “rmda”, and “time ROC” to plot the nomogram, calibration curve, and ROC curves.

2.6 Statistical analyses

The statistical analyses were done utilizing R Studio (V. 1.4). Students’ t-tests were used to determine the difference across normal and tumor samples. The spherical or Fisher’s tests were utilized when appropriate to determine if a correlation between risk score and clinical parameters existed. The Kaplan-Meier plot was



utilized to assess survival time. All P values were 2-tailed at 0.05 significance level.

3 Results

3.1 Identifying DEGs that are related to a worse prognosis

The study's procedure flowchart is depicted in Figure 1. From TCGA cohort, we obtained 3211 DEGs between GC tissues and adjacent non-tumor tissues. Among these, 2700 genes were

upregulated, while 511 genes were downregulated. The heatmap and the volcano plot of DEGs are depicted in Figures 2A, B. Genes involved in mismatch repair, IL-17 signaling pathway, cell cycle, and base excision repair, were primarily upregulated. Pathways including cAMP signaling, protein digestion and absorption, PPAR signaling, Gastric acid secretion, and chemical carcinogenesis were highly enriched in downregulated genes (Figure 2C). In the GO functional analysis, the upregulated DEGs were chiefly enriched in nuclear division, mitotic cell cycle checkpoint, and DNA replication. In response to zinc ion, alcohol, and positive regulation of ion transport, the downregulated DEGs were enriched (shown in Figure 2D).

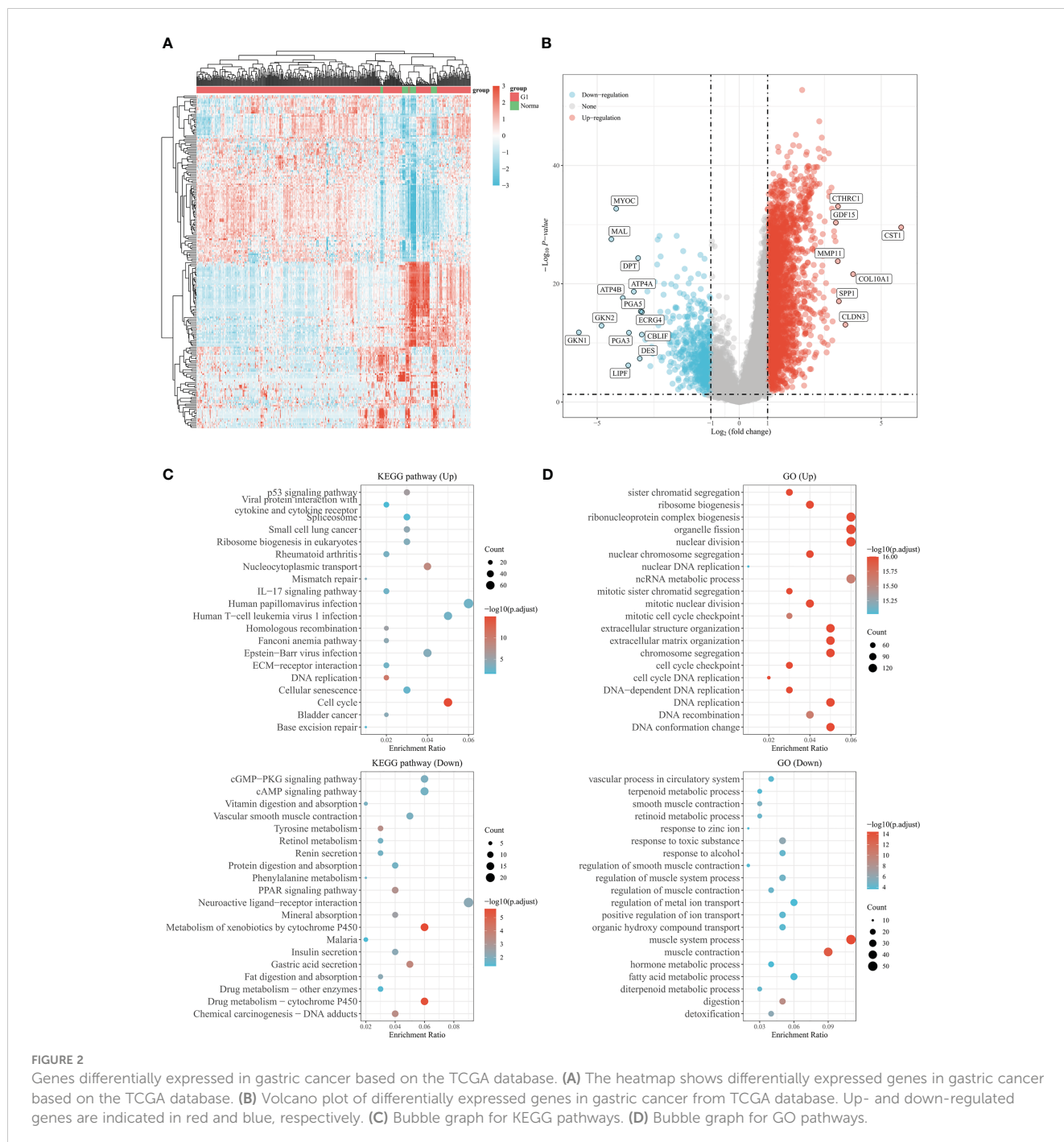
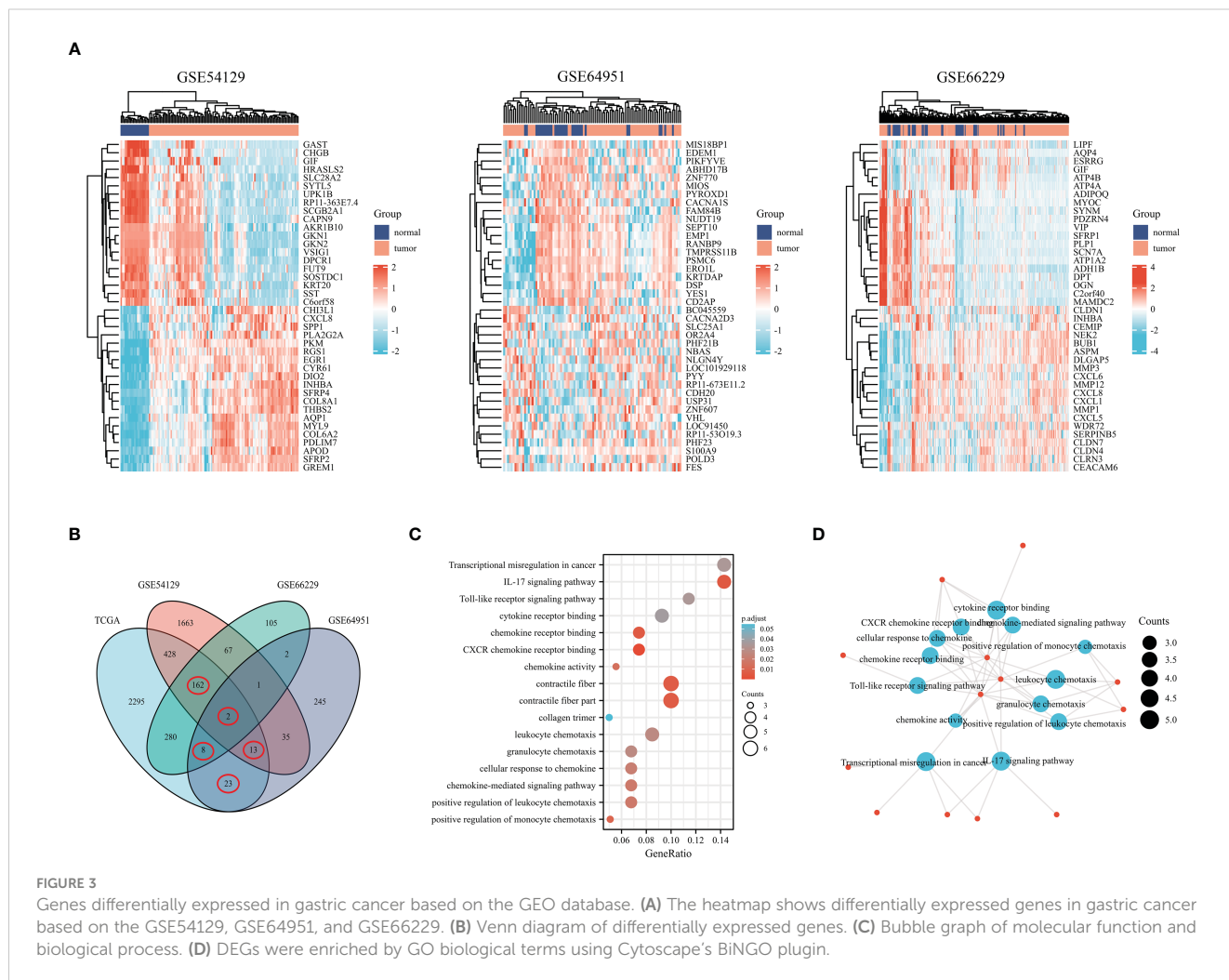


FIGURE 2

Genes differentially expressed in gastric cancer based on the TCGA database. (A) The heatmap shows differentially expressed genes in gastric cancer based on the TCGA database. (B) Volcano plot of differentially expressed genes in gastric cancer from TCGA database. Up- and down-regulated genes are indicated in red and blue, respectively. (C) Bubble graph for KEGG pathways. (D) Bubble graph for GO pathways.



Under the cut-off threshold, there were 2371 genes in GSE54129, 627 genes in GSE66229, and 329 genes in GSE64951 were selected as prospective genes and intersected with to obtain DEGs (shown in Figure 3A). In total, the 208 DEGs were further screened in TCGA dataset through univariate Cox regression analysis (shown in Figure 3B). Functional enrichment analysis was conducted to understand DEGs underlying mechanisms among GC. The functional analyses revealed that DEGs were largely enriched across IL-17 signaling pathways, cytokine receptor binding, chemokine activity, cellular response to chemokine and positive regulation of leukocyte chemotaxis (shown in Figures 3C, D). We included 353 comprehensive clinical data samples among TCGA cohort for subsequent analysis and 13 genes were identified as predictive genes.

3.2 Prognostic model construction and validation

The 353 GC patients from TCGA database were utilized as a training set. In order to verify prognostic signature accuracy and reliability, GEO GSE62254 was utilized as a validation cohort. Our

study utilized TCGA dataset to establish a prognostic signature utilizing LASSO regression. The above 13 genes were further narrowed to 6 genes, namely, CTHRC1, MAMDC2, HSPB8, EZH2, C7, and PSAPL1. Figures 4A–F shows the Kaplan-Meier plots for these 6 genes. Based on LASSO regression analysis, prognostic signature was developed (shown in Figures 5A, B). Accordingly, risk scores were determined for each patient: $0.19 \times \text{CTHRC1} + 0.06 \times \text{MAMDC2} + 0.0005 \times \text{HSPB8} - 0.13 \times \text{EZH2} + 0.03 \times \text{C7} + 0.15 \times \text{PSAPL1}$. Based on validation and training sets, a risk score has been assigned to each patient. Patients were classified into high-risk and low-risk categories according to median risk score. Figures 5C, D demonstrates that high-risk patients had a lower OS rate than low-risk patients. Figure 5E shows that the prognostic signature was well established at 1-, 3-, and 5-year AUCs of 0.62, 0.67, and 0.69. Further validation of the proposed 6-gene prognostic model was conducted. As shown in Figures 6A, B, the prognostic model could determine the level of risk for GC patients based on Kaplan-Meier survival plot of validation set. Survival times were significantly shorter for high-risk scores patients than for those with low-risk scores (HR 1.84, 95% CI 1.30-2.59, $p < 0.01$). In the validation set, the 1-, 3- and 5-year AUC values for risk score model were 0.57, 0.61 and 0.61, respectively (Figure 6C).

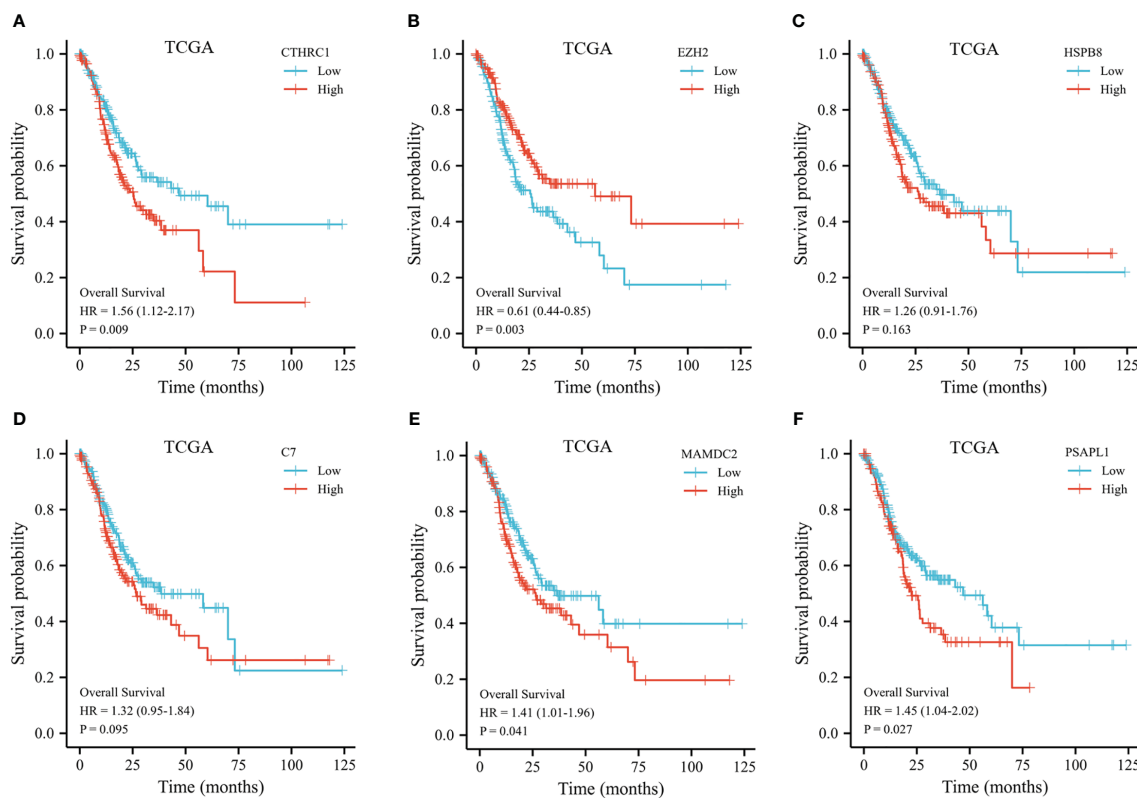


FIGURE 4 Kaplan-Meier plot of the selected genes from the TCGA dataset. **(A)** Kaplan-Meier plot of CTHRC1. **(B)** Kaplan-Meier plot of EZH2. **(C)** Kaplan-Meier plot of HSPB8. **(D)** Kaplan-Meier plot of C7. **(E)** Kaplan-Meier plot of MAMDC2. **(F)** Kaplan-Meier plot of PSAPL1.

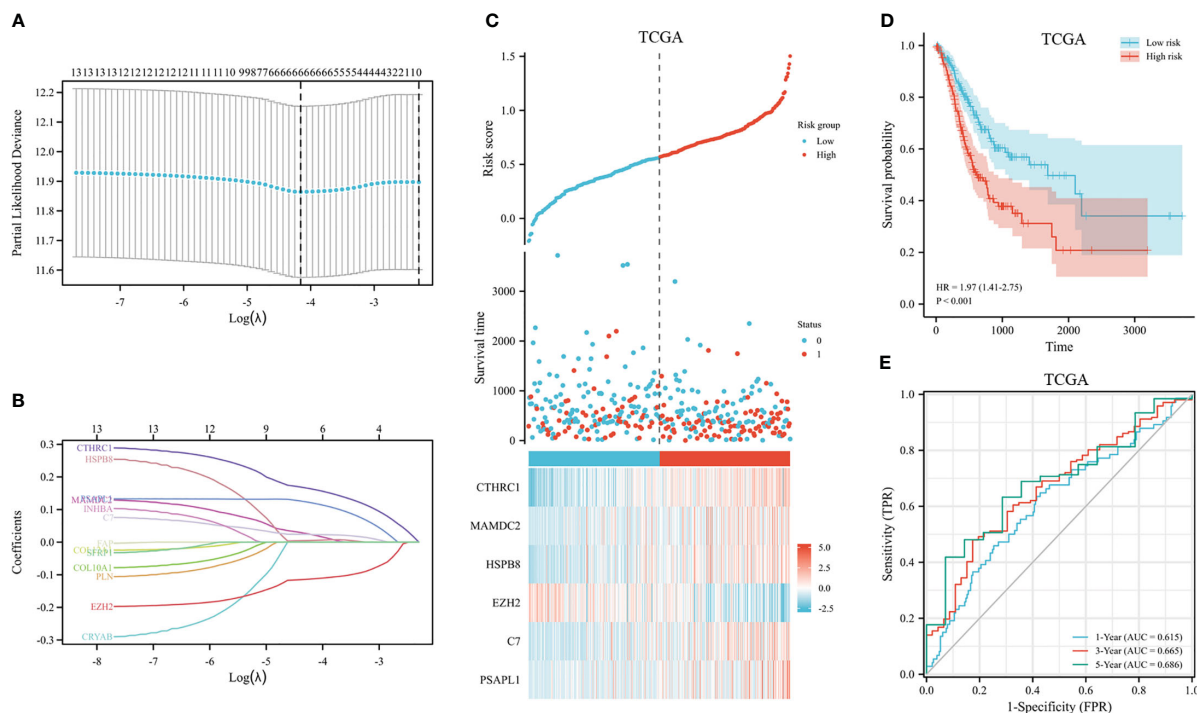


FIGURE 5 Construction of GC prognostic signature in the TCGA dataset. **(A)** The selection of optimal predictive variables by 10-fold cross-validation. **(B)** LASSO coefficients. **(C)** The risk plot between the high-risk and low-risk groups. **(D)** Analysis of overall survival between high-risk and low-risk groups. **(E)** The receiver operating curve for overall survival over time.

The predictive signature was stratified based on histopathological grade, age, and TNM stage. Low-histologic grade patients with high-risk groups had a HR of 1.68 (95% CI 0.96-2.92) ($p=0.07$, shown in Figure 7A). Despite this, prognostic risk model correctly identified short- and long-term survival groups for high-grade GC patients (HR 1.92, 95% CI 1.25-2.96, $p=0.003$, shown in Figure 7B). We further conducted stratification analysis based on age, T, N, and clinical staging. There was a significant correlation of risk scores with survival was found in both groups (shown in Figures 7C-J). Among patients with late M stage disease, the relationship was not noteworthy (HR 1.30, 95% CI 0.43-3.96, $p=0.64$, shown in Figures 7K, L).

3.3 Risk score-immune landscape relationship

ESTIMATE, immune, and stromal scores were compared between the two groups to investigate possible biological mechanisms. The high expression of CTHRC1, MAMDC2, HSPB8, C7 and PSAPL1 was correlated with higher ESTIMATE, stromal and immune score, than low expression group (shown in Figures 8A, C-F). The opposite results were observed in EZH2

(Figure 8B). Moreover, our results showed high-risk patients had significantly higher immune, ESTIMATE, and stromal scores than low-risk (Figures 8G-I).

A heatmap shows immune cell infiltrate across TCGA cohort groups (shown in Figure 9A). EZH2 expression had a significant positive relationship with CD4⁺ memory T cell, CD8⁺ naive T cell, CD8⁺ effector memory T cell, common lymphoid progenitor, mast cell, gamma delta T cell, CD4⁺ Th1 and Th2 T cell, and a negative relationship with common myeloid progenitor, eosinophil, hematopoietic stem cell, and T cell NK. Research has shown significant differences in CTHRC1, MAMDC2, HSPB8, and C7 expression levels among the activated myeloid dendritic cell, CD4⁺ memory T cell, CD8⁺ naive T cell, hematopoietic stem cell, monocyte, gamma delta T cell, and B cell plasma. As shown in Figure 9B, the proportions of activated myeloid dendritic cell, CD4⁺ memory T cell, myeloid dendritic cell, eosinophil, granulocyte-monocyte progenitor, hematopoietic stem cell, macrophage, macrophage M1, and monocyte were significantly elevated among high-risk. We also noted a significant rise in relative fractions of CD4⁺ memory T cells, CD8⁺ naive T cells, common lymphoid progenitor, mast cell, plasmacytoid dendritic cell, regulatory T cells, gamma delta T cell, and B cell plasma among low-risk group.

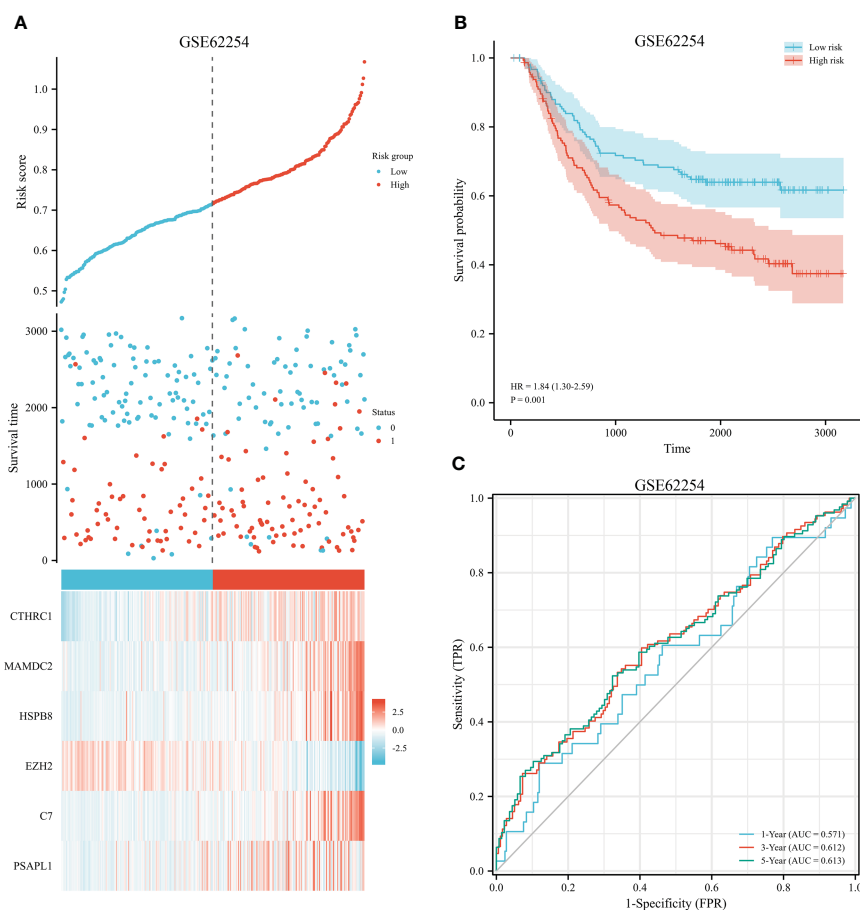


FIGURE 6

Validation of prognostic signature for GC. (A) Risk plot between the high-risk and low-risk groups in the external validation dataset. (B) High-risk versus low-risk survival analysis in the GSE62254. (C) Overall survival prediction curve based on GEO validation data.

3.4 Survival-related gene signature related to immune-checkpoint-relevant genes and immunotherapeutic response among GC patients

A heatmap illustrates survival-related gene expressions and immune-checkpoint-relevant genes (As depicted in Figure 10A). A notable association was observed across risk score and the expression of CTLA4, HAVCR2, PDCD1, PDCD1LG2, and TIGIT, indicating risk scores represent tumor-induced immunosuppression.

The TIDE score was used as a predictor of clinical outcome after immune checkpoint blockade. Risk groups differed significantly from low- and high-risk groups (depicted in Figures 10B–D). The TIDE scores, exclusion scores and dysfunction scores for low-risk were lower than high-risk group. In addition, we identified prognostic signature for immune checkpoint therapy response in the GSE78220, GSE35640, GSE67501, and IMvigor210 cohort. As shown in Figures 11A–D, there was a tendency for non-responders with higher risk scores than responders.

3.5 Nomogram validation and construction

According to survival-related gene signatures and clinical characteristics, a prognostic nomogram was developed. Univariate Cox regression analyses indicated that risk score, gender, age, clinical stage, and T stage are independent prognostic factors (depicted in Supplementary Table 1). Therefore, all of these factors were incorporated into a nomogram for the purpose of predicting the 1- and 3-year survival rates. Summing risk score and clinical parameters based on nomogram, the survival rate was calculated (Figure 12A). The nomogram calibration curves showed that 1-year and 3-year OS were in accordance (Figure 12B). Nomogram results indicate AUCs of 0.77 and 0.83 over 1- and 3-year periods (Figure 12C). Prognostic nomogram demonstrated greater accuracy in survival outcome predictions for GC patients.

4 Discussion

Over the past few years, the prognosis for GC patients has primarily been determined by clinical parameters, such as TNM

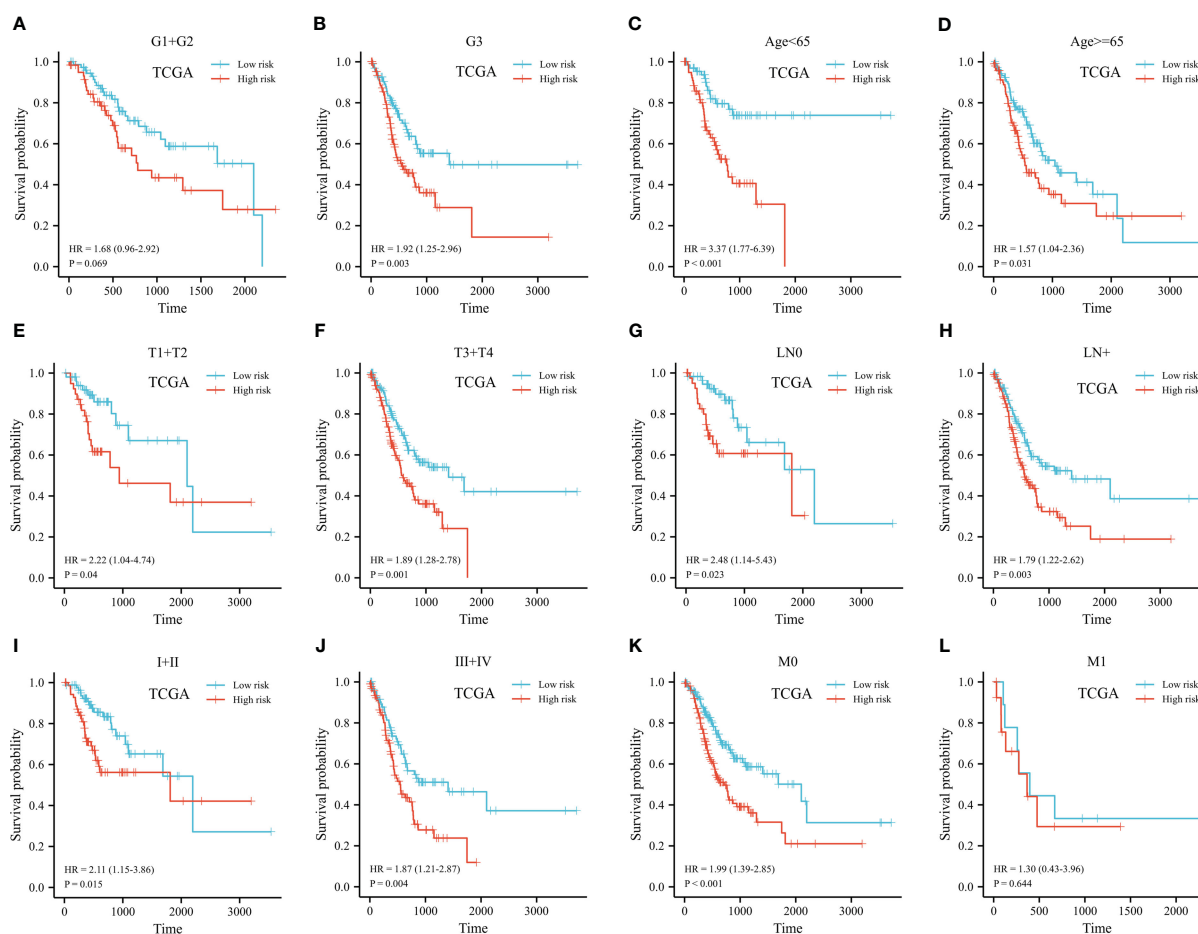


FIGURE 7

Kaplan-Meier plot of stratified analyses of the prognostic signature for associations with clinical characteristics. (A) OS plot in G1+G2 patients. (B) OS plot in G3 patients. (C) OS survival plot in patients older than 65 year-old. (D) OS survival plot in patients younger than 65 year-old. (E) OS survival plot in T1+T2 stage. (F) OS survival plot in T3+T4 stage. (G) OS survival plot in N0 stage. (H) OS survival plot in N+ stage. (I) OS survival plot in stage I+stage II. (J) OS survival plot in stage III+stage IV. (K) OS survival plot in M0 stage. (L) OS survival plot in M1 stage. (OS, overall survival; G, grade; T, tumor; N, lymph node; M, metastasis).

stage, serum tumor biomarkers, and pathological types. Nevertheless, these factors are not useful for clinical decision-making due to their limited predictive efficacy. As a result, discovery of more efficacious biomarkers could make it easier for physicians to make individual treatment decisions. With the continuous evolution of sequencing technology, genomics could potentially play a significant role in identifying predictive biomarkers for various malignancies. There is little predictive value in a single gene for the outcomes of GC patients. In comparison to single gene models, multigene models are much more predictive than single gene models.

We studied DEGs between GC tissues and adjacent non-tumor tissues in GEO and TCGA datasets. There were 3211 DEGs in TCGA, 2371 DEGs in GSE54129, 627 DEGs in GSE66229, and 329 DEGs in GSE64951 selected as candidate genes and intersected with to obtain DEGs. In total, the 208 DEGs were further screened in TCGA cohort through univariate Cox regression analyses. Following this, LASSO regression was utilized to generate prognostic model of 6 DEGs. External validation showed favorable predictive efficacy. As a final step, nomogram was developed in this study, utilizing both clinical characteristics and a prognostic model.

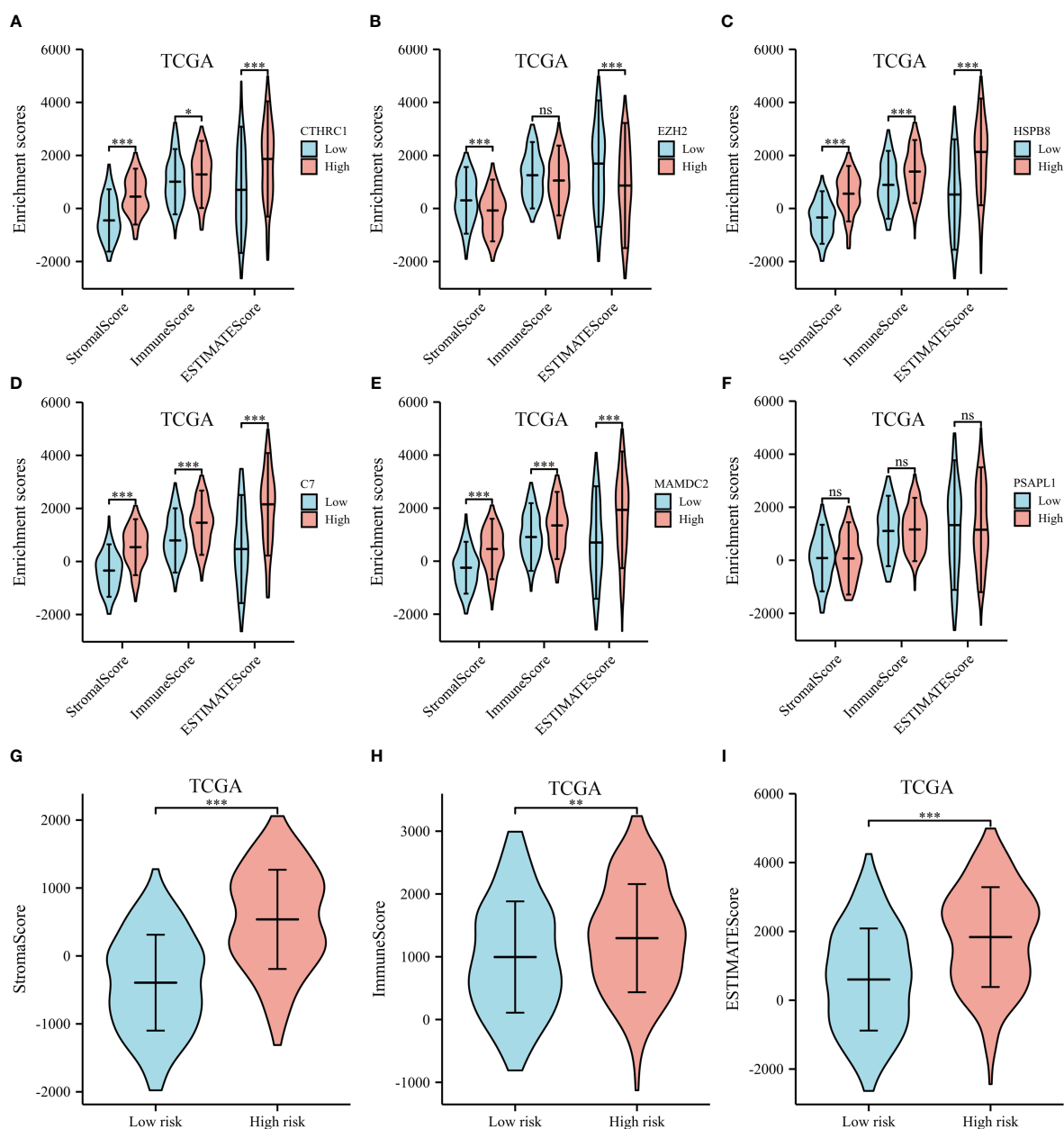
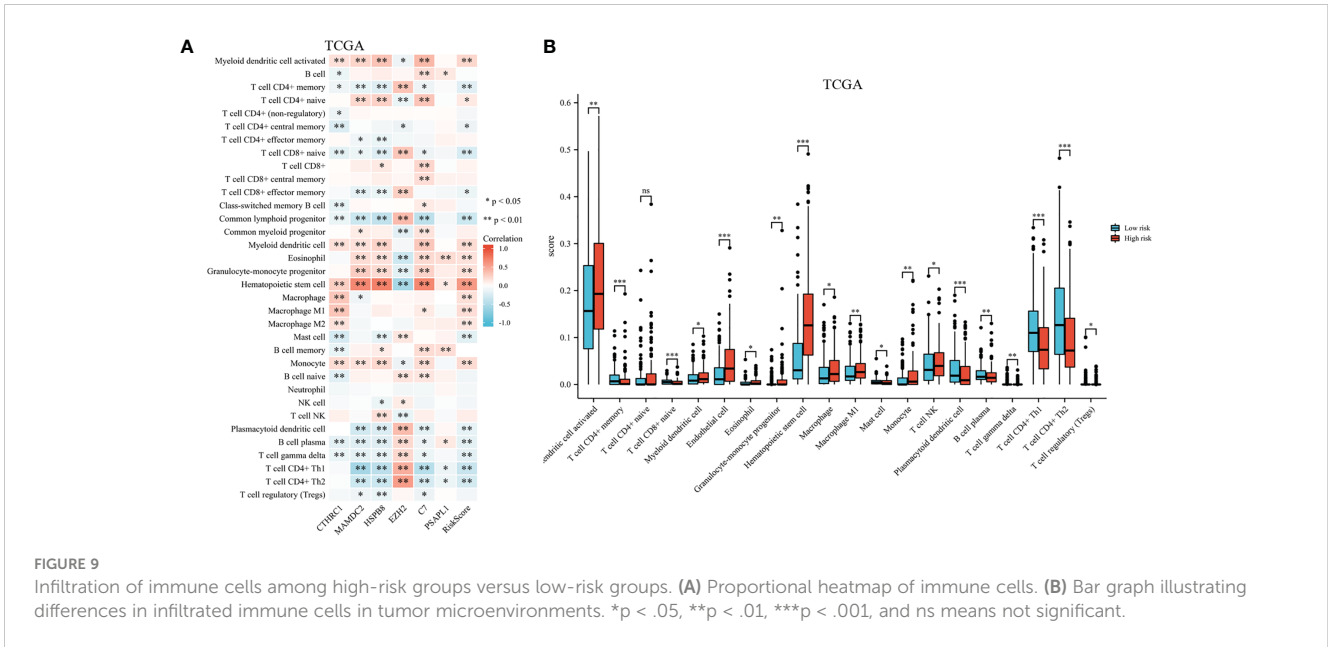
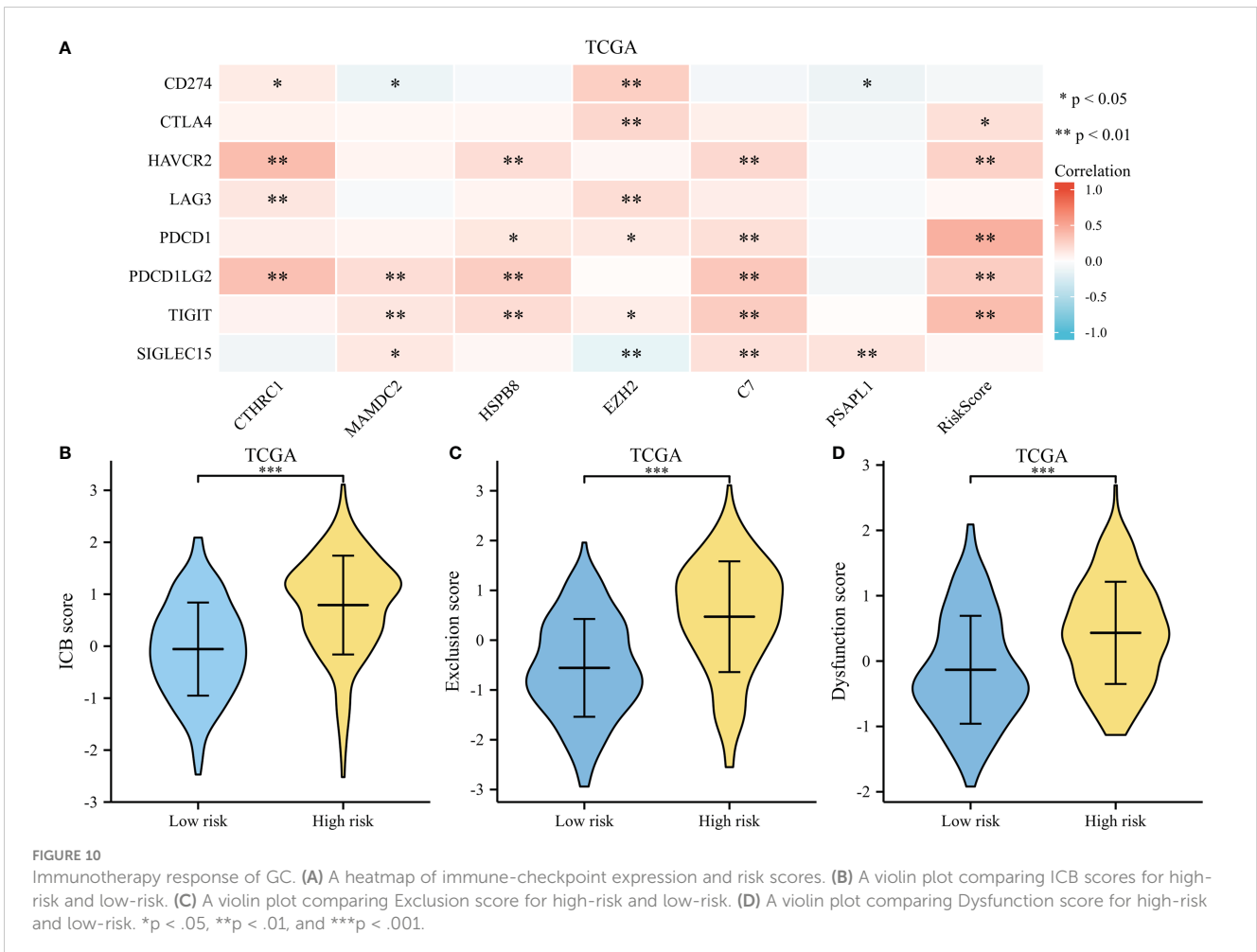


FIGURE 8 ESTIMATE scores of the selected genes from the TCGA dataset. (A) ESTIMATE scores of CTHRC1. (B) ESTIMATE scores of EZH2. (C) ESTIMATE scores of HSPB8. (D) ESTIMATE scores of C7. (E) ESTIMATE scores of MAMDC2. (F) ESTIMATE scores of PSAPL1. (G) Stromal score between the high- and low-risk groups. (H) Immune score between the high- and low-risk groups. (I) ESTIMATE score between the high- and low-risk groups. *p < .05, **p < .01, ***p < .001, and ns means not significant.



The prognostic signature contains 6 biomarkers and categorizes GC patients into high- and low-risk group. Among 6 genes in the prognostic model, CTHRC1 was recognized as a novel gene involved in tissue remodeling and found to be overexpressed in

carcinogenesis and metastasis of several solid cancers, as breast cancer (BC) (22) and non-small cell lung cancer (23). Gu (24) et al. investigated the relation of CTHRC1 expression and clinical features among GC patients. Patients with high CTHRC1



expression displayed poorer OS and disease-free survival (DFS) than low CTHRC1 expression patients. CTHRC1 has been reported to promoting cell migration and invasion through HIF-1 α /CXCR4 signaling pathway in GC (25). Additionally, CTHRC1 overexpression induced tumor associated macrophage infiltration *via* AnxA1/FPR1 and GRN/TNFRSF1A signaling pathway, indicating CTHRC1 might be a promising predictive factor for immunotherapy (26).

MAMDC2 was differentially expressed between normal tissues and several solid tumors, including BC (27), head and neck squamous cell carcinoma (28), and GC (29). Meng (30) et al. identified MAMDC2 overexpression was significantly linked to poor DFS of BC. There have been controversial results reported in the literature (31). MAMDC2 was down-regulated in the BC

cells. MAMDC2 Overexpression significantly suppressed proliferation and induced cell apoptosis *in vitro* and *in vivo*. There were contradictory reports on the involvement of MAMDC2 in tumor progression in the literature. In this study, high MAMDC2 expression patients experienced shorter OS than low MAMDC2 expression. MAMDC2 expression was inversely associated with CD4⁺ memory T cell, CD4⁺ effector memory T cell, CD8⁺ effector T cells, plasmacytoid dendritic cell, B cell plasma, CD4⁺ Th1 T cells, and gamma delta T cells. According to immune-checkpoint–relevant transcripts, MAMDC2 was strongly correlated to expression of PDCD1LG2, TIGIT, and SIGLEC15. Our findings suggest that MAMDC2 overexpression may contribute to immunological suppression in GC patients as well as a poor prognosis.

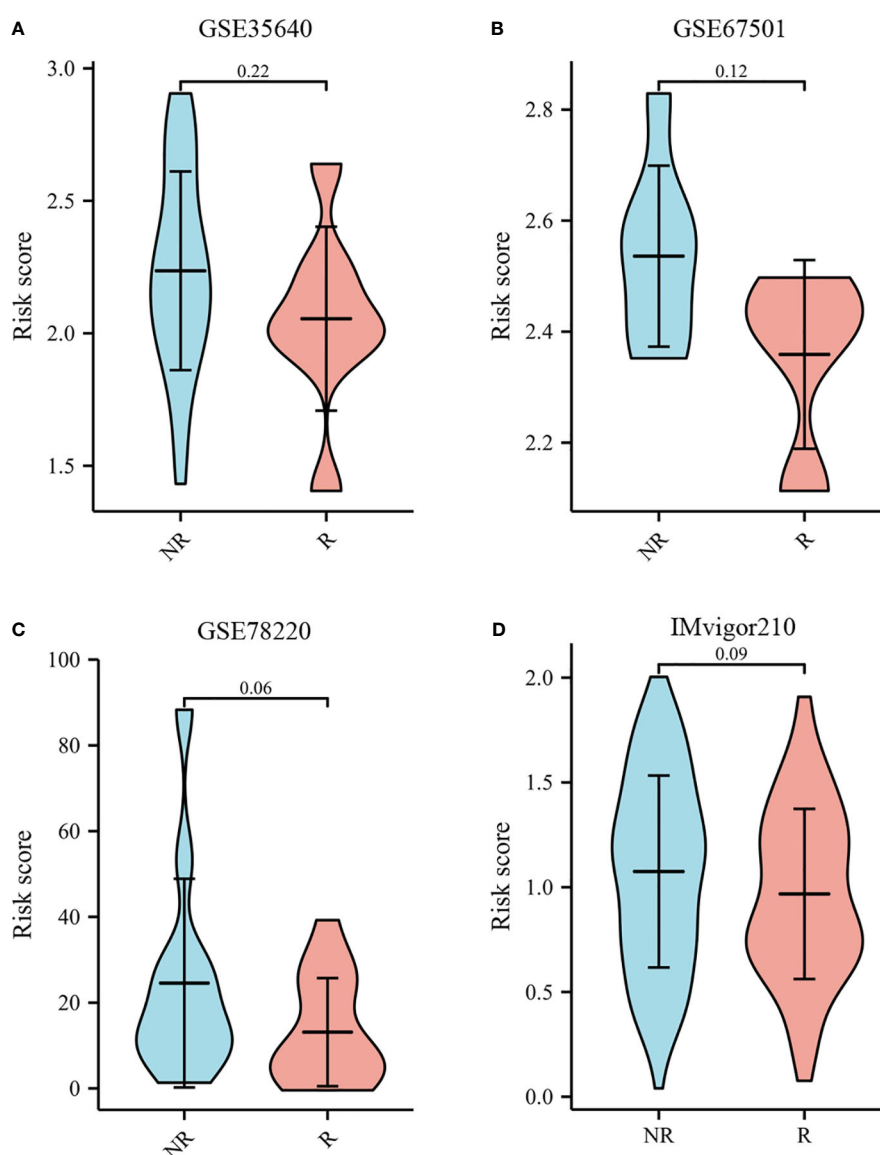


FIGURE 11

Immunotherapy response to immune checkpoints in the GSE35640, GSE67501, GSE78220, and IMvigor210 cohort. (A) The violin plot comparing responders and non-responders to immunotherapy in GSE35640. (B) The violin plot comparing responders and non-responders to immunotherapy in GSE67501. (C) The violin plot comparing responders and non-responders to immunotherapy in GSE78220. (D) The violin plot comparing responders and non-responders to immunotherapy in IMvigor210.

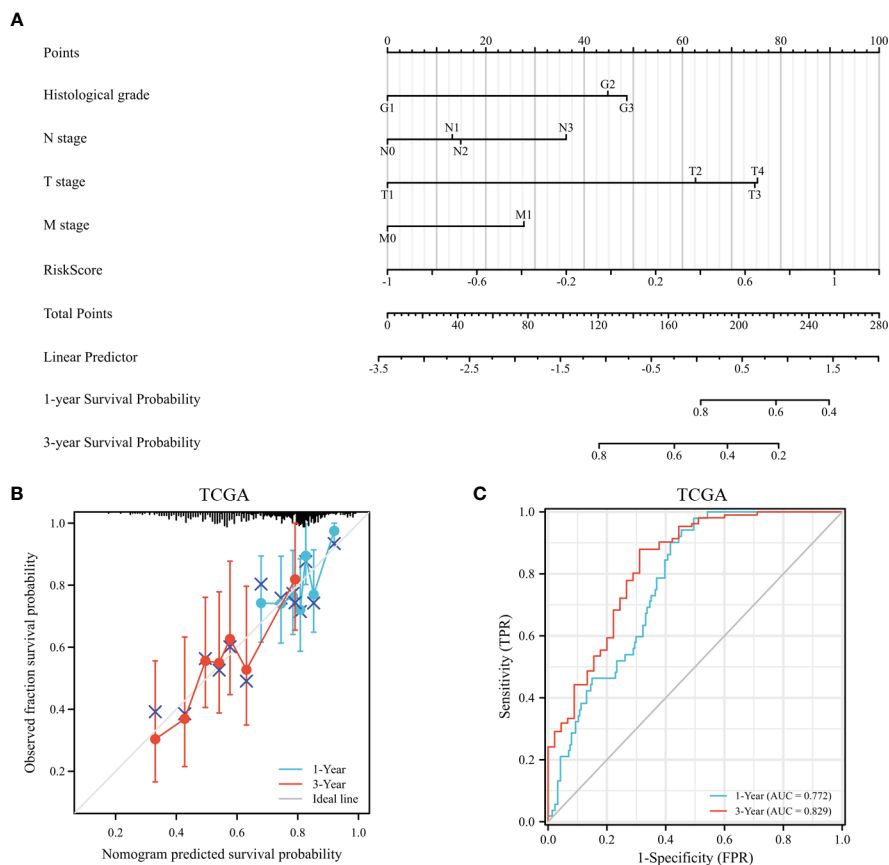


FIGURE 12
 The assessment of a nomogram based on clinical characteristics and risk scores. **(A)** A nomogram based on the clinical parameters and risk model. **(B)** An analysis of the calibration curves for the TCGA dataset at 1-year survival and 3-year survival. **(C)** Time-dependent receiver operating curve that predicted overall survival.

The stress-related protein HSPB8 was first discovered in human melanoma cells as a kinase of the H11 protein. HSPB8 Overexpression promoted proliferation, migration, and invasion (32, 33). HSPB8 suppressed mitochondrial impairment and aggravated proliferation and migration of A549 lung adenocarcinoma cells (34). In GC, HSPB8 expression was significantly linked to worse OS and recurrence-free survival (35). Furthermore, immune cell infiltrate analysis indicated B cells, CD4⁺T cells, and CD8⁺T cells were significantly different across high HSPB8 expression group and low HSPB8 expression group in bladder cancer (36).

The EZH2 methyltransferase is a core catalytic subunit of the polycomb repressive complex 2. The EZH2 protein mediates modifications in histone methylation that repress several tumor suppressor genes, as DKK1, CDH1, and DAB2IP (37). Zhao (38) et al. found EZH2 mediated EphB3 transcription through H3K27me3 modification, and inhibited tumor proliferation and metastasis by regulation E-cadherin and vimentin expression. Our findings indicate that high EZH2 expression patients had significantly longer OS than those with low EZH2 expression. The proportions of CD8⁺ effector memory T cells, CD8⁺ naive T cells, plasmacytoid dendritic cell, B cell plasma, CD4⁺ Th1 T cells, and

gamma delta T cells were significantly elevated in low EZH2 expression patients, indicating EZH2 represented the activation of TME.

C7 was the final component of the complement cascade and essential for complement activation. Seol (39) et al. found C7 overexpression induced tumorsphere formation, and maintain stemness of liver cancer cells. However, several studies reported C7 displayed as a potential tumor suppressor and was related to tumor progression and prognosis for certain cancers (40, 41). Our findings suggested the high C7 expression patients had the trend to live a longer OS. Additionally, there was significant differences in the C7 expression levels across common lymphoid progenitor, plasmacytoid dendritic cell, CD4⁺ memory T cell, CD8⁺ naive T cell, gamma delta T cell, CD4⁺ Th1 T cell, and CD4⁺ Th2 T cell. The C7 expression had notable direct correlation with the expression of HAVCR2, PDCD1, PDCD1LG2, SIGLEC15, and TIGIT, indicating risk score was indicative of the extent of immunosuppression induced by tumors.

There was some literature about the oncogenic role of PSAPL1 in various solid cancers, including GC (42), BC (43), hepatocellular carcinoma (44). Our study indicated high expression PSAPL1 patients had a worse OS than those of low expression PSAPL1.

According to immune scores and immune cell infiltration, there was no significant difference of stroma score, immune score, ESTIMATE score, and immune cell infiltration regard to PSAPL1 expression in GC patients. Further studies are required to elucidate PSAPL1 mechanisms in GC.

We studied interaction between risk models, immunoscores, and immune cell infiltrate based on six-gene signature. The high-risk group exhibited significantly elevated ESTIMATE score, immunoscore, and stromal score relative to low-risk group. Utilizing abundance of tumor-infiltrating immune cells in each GC patient, risk scores were calculated. As a result, we identified that CD4⁺ memory T cells, CD8⁺ naive T cells, common lymphoid progenitor, mast cell, plasmacytoid dendritic cell, B cell plasma, regulatory T cells, gamma delta T cell, and B cell plasma relative fractions were significantly enriched in low-risk group. A notable correlation was observed between risk score and expression of CTLA4, HAVCR2, PDCD1, PDCD1LG2, and TIGIT, indicating risk scores represent tumor-induced immunosuppression. According to TIDE, the TIDE scores, exclusion scores and dysfunction scores for low-risk group were lower than those for high-risk group. TIDE algorithm and immune-checkpoint-relevant transcripts were consistent with xCell, suggesting immunosuppressive microenvironments were more common among high-risk GC patients.

This study may have major implications for patients with GC as far as prognosis and treatment are concerned. To help with clinical practice and risk classification, we offered a new signature. Low-risk individuals had longer life periods and were more likely to benefit from ICIs. In addition, we found a number of crucial genes that could serve as GC treatment targets. Several previous studies (45–47) constructed mRNA prognosis signatures in GC patients. The AUC values of these studies ranged from 0.54 to 0.62, which were inferior to the current study. We examined prognostic model significance for other solid cancer in addition to researching immunotherapy response and prognosis among GC patients.

Several limitations were identified in our study. Firstly, data on mRNA expression and clinical information were downloaded from an open-source database. The findings of these studies have not yet been confirmed in clinical trials. The environment, genetics, and epigenetics are also factors influencing GC. Further molecular biological studies are necessary to verify involvement of the 6 DEGs in GC progression. Finally, due to TCGA data unavailability, the risk model was unable to provide the predictive value compared to a number of widely used predictors, such as pathological grade and treatment approach.

In conclusion, we discovered a 6 gene signature to forecast GC patients' OS. This risk signature proves to be a valuable clinical predictive tool for guiding clinical practice. Moreover, the profile revealed discernible variations in immune cell infiltration levels and immunotherapy response among low- and high-risk groups. This prognostic model thus offers a precise and impartial basis for directing unique therapy choices for GC.

Data availability statement

Publicly available datasets were analyzed in this study. This data can be found here: TCGA repository (<https://portal.gdc.cancer.gov/>), GEO database (<https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi>), and IMvigor210 (<http://research-pub.gene.com/IMvigor210CoreBiologies/>).

Author contributions

QW and BZ designed the study. HW, HF and WG collected the data. MH, TG and HL performed the research. QW wrote the paper. XJ, MX and YD analyzed the data. All authors contributed to the article and approved the submitted version.

Funding

This work was supported by the Qingdao Postdoctoral Sustentation Fund (RZ2100001380) and National Natural Science Foundation of China (82003224). The funder had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

Acknowledgments

We thank Home for Researchers editorial team (www.home-for-researchers.com) for language editing service.

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.

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/fonc.2023.1210994/full#supplementary-material>

References

- Joshi SS, Badgwell BD. Current treatment and recent progress in gastric cancer. *CA Cancer J Clin* (2021) 71:264–79. doi: 10.3322/caac.21657
- Park SH, Lim DH, Sohn TS, Lee J, Zang DY, Kim ST, et al. A randomized phase iii trial comparing adjuvant single-agent s1, s-1 with oxaliplatin, and postoperative chemoradiation with s-1 and oxaliplatin in patients with node-positive gastric cancer after d2 resection: the artist 2 trial(). *Ann Oncol* (2021) 32:368–74. doi: 10.1016/j.annonc.2020.11.017
- Schreiber RD, Old LJ, Smyth MJ. Cancer immunoediting: integrating immunity's roles in cancer suppression and promotion. *Science* (2011) 331:1565–70. doi: 10.1126/science.1203486
- Mlecnik B, Bindea G, Kirilovsky A, Angell HK, Obenauf AC, Tosolini M, et al. The tumor microenvironment and immunoscore are critical determinants of dissemination to distant metastasis. *Sci Transl Med* (2016) 8:326r–7r. doi: 10.1126/scitranslmed.aad6352
- Cancer Genome Atlas Research Network. Comprehensive molecular characterization of gastric adenocarcinoma. *Nature* (2014) 513:202–9. doi: 10.1038/nature13480
- Ouaguia L, Mrizak D, Renaud S, Morales O, Delhem N. Control of the inflammatory response mechanisms mediated by natural and induced regulatory t-cells in hcv-, htlv-1-, and ebv-associated cancers. *Mediators Inflamm* (2014) 2014:564296. doi: 10.1155/2014/564296
- Kim ST, Cristescu R, Bass AJ, Kim KM, Odegaard JL, Kim K, et al. Comprehensive molecular characterization of clinical responses to pd-1 inhibition in metastatic gastric cancer. *Nat Med* (2018) 24:1449–58. doi: 10.1038/s41591-018-0101-z
- Wang F, Wei XL, Wang FH, Xu N, Shen L, Dai GH, et al. Safety, efficacy and tumor mutational burden as a biomarker of overall survival benefit in chemo-refractory gastric cancer treated with toripalimab, a pd-1 antibody in phase ib/ii clinical trial nct02915432. *Ann Oncol* (2019) 30:1479–86. doi: 10.1093/annonc/mdz197
- Shen X, Zhao B. Efficacy of pd-1 or pd-l1 inhibitors and pd-l1 expression status in cancer: meta-analysis. *Bmj* (2018) 362:k3529. doi: 10.1136/bmj.k3529
- Akin TT, Bregni G, Camera S, Deleporte A, Hendlitz A, Sclafani F. Pd-1 and pd-l1 inhibitors in oesophago-gastric cancers. *Cancer Lett* (2020) 469:142–50. doi: 10.1016/j.canlet.2019.10.036
- Oh SC, Sohn BH, Cheong JH, Kim SB, Lee JE, Park KC, et al. Clinical and genomic landscape of gastric cancer with a mesenchymal phenotype. *Nat Commun* (2018) 9:1777. doi: 10.1038/s41467-018-04179-8
- Li F, Yoshizawa JM, Kim KM, Kanjanapangka J, Grogan TR, Wang X, et al. Discovery and validation of salivary extracellular rna biomarkers for noninvasive detection of gastric cancer. *Clin Chem* (2018) 64:1513–21. doi: 10.1373/clinchem.2018.290569
- Cristescu R, Lee J, Nebozhyn M, Kim KM, Ting JC, Wong SS, et al. Molecular analysis of gastric cancer identifies subtypes associated with distinct clinical outcomes. *Nat Med* (2015) 21:449–56. doi: 10.1038/nm.3850
- Ritchie ME, Phipson B, Wu D, Hu Y, Law CW, Shi W, et al. Limma powers differential expression analyses for rna-sequencing and microarray studies. *Nucleic Acids Res* (2015) 43:e47. doi: 10.1093/nar/gkv007
- Yu G, Wang LG, Han Y, He QY. Clusterprofiler: an r package for comparing biological themes among gene clusters. *Omic* (2012) 16:284–7. doi: 10.1089/omi.2011.0118
- Walter W, Sanchez-Cabo F, Ricote M. Gplot: an r package for visually combining expression data with functional analysis. *Bioinformatics* (2015) 31:2912–4. doi: 10.1093/bioinformatics/btv300
- Friedman J, Hastie T, Tibshirani R. Regularization paths for generalized linear models via coordinate descent. *J Stat Software* (2010) 33:1–22. doi: 10.18637/jss.v033.i01
- Blanche P, Dartigues JF, Jacqmin-Gadda H. Estimating and comparing time-dependent areas under receiver operating characteristic curves for censored event times with competing risks. *Stat Med* (2013) 32:5381–97. doi: 10.1002/sim.5958
- Yoshihara K, Shahmoradgol M, Martinez E, Vegesna R, Kim H, Torres-Garcia W, et al. Inferring tumour purity and stromal and immune cell admixture from expression data. *Nat Commun* (2013) 4:2612. doi: 10.1038/ncomms3612
- Aran D. Cell-type enrichment analysis of bulk transcriptomes using xcell. *Methods Mol Biol* (2020) 2120:263–76. doi: 10.1007/978-1-0716-0327-7_19
- Jiang P, Gu S, Pan D, Fu J, Sahu A, Hu X, et al. Signatures of t cell dysfunction and exclusion predict cancer immunotherapy response. *Nat Med* (2018) 24:1550–8. doi: 10.1038/s41591-018-0136-1
- Lai YH, Chen J, Wang XP, Wu YQ, Peng HT, Lin XH, et al. Collagen triple helix repeat containing-1 negatively regulated by microRNA-30c promotes cell proliferation and metastasis and indicates poor prognosis in breast cancer. *J Exp Clin Cancer Res* (2017) 36:92. doi: 10.1186/s13046-017-0564-7
- Ke Z, He W, Lai Y, Guo X, Chen S, Li S, et al. Overexpression of collagen triple helix repeat containing 1 (cthrcl) is associated with tumour aggressiveness and poor prognosis in human non-small cell lung cancer. *Oncotarget* (2014) 5:9410–24. doi: 10.18632/oncotarget.2421
- Gu L, Liu L, Zhong L, Bai Y, Sui H, Wei X, et al. Cthrc1 overexpression is an independent prognostic marker in gastric cancer. *Hum Pathol* (2014) 45:1031–8. doi: 10.1016/j.humpath.2013.12.020
- Ding X, Huang R, Zhong Y, Cui N, Wang Y, Weng J, et al. Cthrc1 promotes gastric cancer metastasis via hif-1alpha/cxcr4 signaling pathway. *BioMed Pharmacother* (2020) 123:109742. doi: 10.1016/j.biopha.2019.109742
- Hu Y, Huang L, Zhao K, Li Y, Givens NT, Heslin AJ, et al. Cthrc1 is associated with immune escape and poor prognosis in gastric cancer. *Anticancer Res* (2023) 43:115–26. doi: 10.21873/anticancer.16140
- Sultan G, Zubair S, Tayubi IA, Dahms HU, Madar IH. Towards the early detection of ductal carcinoma (a common type of breast cancer) using biomarkers linked to the ppar(gamma) signaling pathway. *Bioinformatics* (2019) 15:799–805. doi: 10.6026/97320630015799
- Darda L, Hakami F, Morgan R, Murdoch C, Lambert DW, Hunter KD. The role of hoxb9 and mir-196a in head and neck squamous cell carcinoma. *PLoS One* (2015) 10:e122285. doi: 10.1371/journal.pone.0122285
- Xie R, Liu L, Lu X, He C, Li G. Identification of the diagnostic genes and immune cell infiltration characteristics of gastric cancer using bioinformatics analysis and machine learning. *Front Genet* (2022) 13:1067524. doi: 10.3389/fgene.2022.1067524
- Meng L, Xu Y, Xu C, Zhang W. Biomarker discovery to improve prediction of breast cancer survival: using gene expression profiling, meta-analysis, and tissue validation. *Oncotargets Ther* (2016) 9:6177–85. doi: 10.2147/OTT.S113855
- Lee H, Park BC, Soon KJ, Cheon Y, Lee S, Jae MP. Mam domain containing 2 is a potential breast cancer biomarker that exhibits tumour-suppressive activity. *Cell Prolif* (2020) 53:e12883. doi: 10.1111/cpr.12883
- Shen J, Li M, Min L. Hspb8 promotes cancer cell growth by activating the erk-creb pathway and is indicative of a poor prognosis in gastric cancer patients. *Oncol Rep* (2018) 39:2978–86. doi: 10.3892/or.2018.6376
- Piccolella M, Crippa V, Cristofani R, Rusmini P, Galbiati M, Cicardi ME, et al. The small heat shock protein b8 (hspb8) modulates proliferation and migration of breast cancer cells. *Oncotarget* (2017) 8:10400–15. doi: 10.18632/oncotarget.14422
- Yu LL, Wang Y, Xiao ZK, Chen SS. Heat shock protein b8 promotes proliferation and migration in lung adenocarcinoma a549 cells by maintaining mitochondrial function. *Mol Cell Biochem* (2021) 476:187–97. doi: 10.1007/s11010-020-03896-3
- Li XS, Xu Q, Fu XY, Luo WS. Heat shock protein 22 overexpression is associated with the progression and prognosis in gastric cancer. *J Cancer Res Clin Oncol* (2014) 140:1305–13. doi: 10.1007/s00432-014-1698-z
- Tan Z, Fu S, Huang Y, Duan X, Zuo Y, Zhu X, et al. Hspb8 is a potential prognostic biomarker that correlates with immune cell infiltration in bladder cancer. *Front Genet* (2022) 13:804858. doi: 10.3389/fgene.2022.804858
- Chang CJ, Hung MC. The role of eh2 in tumour progression. *Br J Cancer* (2012) 106:243–7. doi: 10.1038/bjc.2011.551
- Zhao K, He J, Wang YF, Jin SD, Fan Y, Fang N, et al. Eh2-mediated epigenetic suppression of ephb3 inhibits gastric cancer proliferation and metastasis by affecting e-cadherin and vimentin expression. *Gene* (2019) 686:118–24. doi: 10.1016/j.gene.2018.11.015
- Seol HS, Lee SE, Song JS, Rhee JK, Singh SR, Chang S, et al. Complement proteins c7 and cfh control the stemness of liver cancer cells via lsf-1. *Cancer Lett* (2016) 372:24–35. doi: 10.1016/j.canlet.2015.12.005
- Ying L, Zhang F, Pan X, Chen K, Zhang N, Jin J, et al. Complement component 7 (c7), a potential tumor suppressor, is correlated with tumor progression and prognosis. *Oncotarget* (2016) 7:86536–46. doi: 10.18632/oncotarget.13294
- Chen Z, Yan X, Du GW, Tuoheti K, Bai XJ, Wu HH, et al. Complement c7 (c7), a potential tumor suppressor, is an immune-related prognostic biomarker in prostate cancer (pc). *Front Oncol* (2020) 10:1532. doi: 10.3389/fonc.2020.01532
- Li J, Han T, Wang X, Wang Y, Chen X, Chen W, et al. H19 may regulate the immune cell infiltration in carcinogenesis of gastric cancer through mir-378a-5p/serpinh1 signaling. *World J Surg Oncol* (2022) 20:295. doi: 10.1186/s12957-022-02760-6
- Rakha EA, Alsalem M, ElSharawy KA, Toss MS, Raafat S, Mihai R, et al. Visual histological assessment of morphological features reflects the underlying molecular profile in invasive breast cancer: a morphomolecular study. *Histopathology* (2020) 77:631–45. doi: 10.1111/his.14199
- Hymel E, Fisher KW, Farazi PA. Differential methylation patterns in lean and obese non-alcoholic steatohepatitis-associated hepatocellular carcinoma. *BMC Cancer* (2022) 22:1276. doi: 10.1186/s12885-022-10389-7
- Wang N, Liu D. Identification and validation a necroptosis-related prognostic signature and associated regulatory axis in stomach adenocarcinoma. *Oncotargets Ther* (2021) 14:5373–83. doi: 10.2147/OTT.S342613
- Cho JY, Lim JY, Cheong JH, Park YY, Yoon SL, Kim SM, et al. Gene expression signature-based prognostic risk score in gastric cancer. *Clin Cancer Res* (2011) 17:1850–7. doi: 10.1158/1078-0432.CCR-10-2180
- Wu X, Jian A, Tang H, Liu W, Liu F, Liu S, et al. A multi-omics study on the effect of helicobacter pylori-related genes in the tumor immunity on stomach adenocarcinoma. *Front Cell Infect Microbiol* (2022) 12:880636. doi: 10.3389/fcimb.2022.880636