



Developing a prognostic signature: identifying differentially expressed genes in cardia and non-cardia gastric cancer for immunity and therapeutic sensitivity analysis

Xianmin Li^{1,2#}, Chong Zhou^{3#}, Yindi Zhu^{4#}, Wenjie Wang⁵, Shuguang Han⁶, Yicen Zou², Lian Lian², Kai Chen¹

¹Department of Oncology, The First Affiliated Hospital of Soochow University, Suzhou, China; ²Department of Oncology, Suzhou Xiangcheng People's Hospital, Suzhou, China; ³Department of Radiation Oncology, Xuzhou Central Hospital, Xuzhou, China; ⁴Department of Gynecology and Obstetrics, The First Affiliated Hospital of Soochow University, Suzhou, China; ⁵Department of Radiation Oncology, The Affiliated Suzhou Hospital of Nanjing Medical University, Suzhou, China; ⁶Department of Gastrointestinal Surgery, Suzhou Xiangcheng People's Hospital, Suzhou, China

Contributions: (I) Conception and design: X Li, C Zhou, Y Zhu, W Wang, L Lian; (II) Administrative support: K Chen, L Lian; (III) Provision of study materials or patients: X Li, C Zhou, Y Zhu, S Han; (IV) Collection and assembly of data: W Wang, S Han, Y Zou; (V) Data analysis and interpretation: X Li, C Zhou, W Wang; (VI) Manuscript writing: All authors; (VII) Final approval of manuscript: All authors.

#The authors contributed equally to this work.

Correspondence to: Kai Chen, PhD. Department of Oncology, The First Affiliated Hospital of Soochow University, No. 188 Shizi Street, Suzhou 215006, China. Email: kaichen@suda.edu.cn; Lian Lian, PhD. Department of Oncology, Suzhou Xiangcheng People's Hospital, No. 1060 Huayuan Road, Xiangcheng District, Suzhou 215131, China. Email: dr_lianlian@163.com.

Background: Gastric cancer (GC) can be anatomically categorized into two subtypes; that is, cardia gastric cancer (CGC) and non-cardia gastric cancer (NCGC), which have distinct molecular mechanisms and prognoses. At present, the majority of pharmacological interventions for GC adhere to non-specific treatment regimens. The stratification of GC based on molecular disparities between CGC and NCGC has important clinical guidance value and could help in the development of precision therapies tailored to individual patient needs. Nevertheless, research in this specialized field remains notably limited. This study aims to investigate the molecular differences between CGC and NCGC and to leverage these differences to develop a prognostic risk scoring model (PRSM).

Methods: We used patient data from The Cancer Genome Atlas (TCGA) and performed a differentially expressed gene (DEG) analysis between CGC and NCGC. A PRSM was developed from the prognosis-associated DEGs identified through Cox regression analyses and was well validated using Gene Expression Omnibus (GEO) data.

Results: A total of 339 DEGs were identified between CGC and NCGC, and four prognosis-associated genes were used to construct the PRSM. Using the risk coefficients and expression levels of signature genes, a median risk score (RS) was calculated to classify patients into high- and low-risk groups. The high-risk group had a significantly worse prognosis than the low-risk group. An in-depth analysis revealed that *TP53* mutations were more prevalent in the high-risk group, and *MUC16* mutations were more prevalent in the low-risk group. A gene set enrichment analysis (GSEA) and the CIBERSORT algorithm were used to assess the differences in the significantly enriched pathways and immune microenvironment in the high- and low-risk groups, respectively. The inhibitory concentration (IC₅₀) values of the chemotherapy drugs for GC also varied between the two groups.

Conclusions: This study elucidated the unique molecular characteristics of GC subtypes based on the anatomical site and provided a preliminary contribution for the development of precision medicine for GC.

Keywords: Gastric cancer (GC); cardia gastric cancer (CGC); non-cardia gastric cancer (NCGC); prognostic risk scoring model (PRSM); therapeutic sensitivity

Submitted Jul 16, 2024. Accepted for publication Aug 20, 2024. Published online Aug 28, 2024.

doi: 10.21037/jgo-24-541

View this article at: <https://dx.doi.org/10.21037/jgo-24-541>

Introduction

Gastric cancer (GC) is a significant global health issue due to its high prevalence and mortality rate, and accounts for over 750,000 cancer-related deaths annually (1). Anatomically, GC can be categorized into the following two sub-types: cardia gastric cancer (CGC) and non-cardia gastric cancer (NCGC). The two subtypes display distinct disparities in terms of their geographical distribution and risk factors. CGC is more common in economically developed regions and is primarily associated with factors, such as obesity, gastroesophageal reflux, and the consumption of hot foods (2). Conversely, NCGC is more prevalent in developing countries and is often associated with high-salt diets and *Helicobacter pylori* infection (3). Notably, the survival outcomes and cancer-specific mortality rates of CGC samples are significantly worse than those of NCGC samples (4,5).

Investigations into gene expression profiles have revealed distinct variations in gene expression patterns between

CGC and NCGC, primarily encompassing cell cycle regulation, cell proliferation, and cell death (6,7). Genome-wide association studies have also discovered different susceptibility sites for single-nucleotide polymorphisms in CGC and NCGC. For example, the rs4072037 variant of the *MUC1* gene is associated with CGC, while the rs2294693 variant of the *UNC5CL* gene and the rs2294008 variant of the *PSCA* gene are predominantly linked to NCGC (8). These findings suggest distinct pathogeneses and genetic backgrounds of CGC and NCGC. Therefore, when establishing the suitable treatment approach, meticulous assessment grounded in the molecular distinctions of the anatomical site of GC is imperative.

In this study, we conducted an analysis of genes that exhibited differential expression in The Cancer Genome Atlas (TCGA) database with a specific focus on CGC and NCGC. We developed a prognostic risk scoring model (PRSM) using these differentially expressed genes (DEGs) and validated it in the Gene Expression Omnibus (GEO) database. Patients were categorized into either the high- or low-risk group based on their risk score (RS). Ultimately, the key prognostic DEGs were identified. Our findings provide a molecular basis for identifying sensitivities and have significant clinical value in guiding personalized treatment for GC. We present this article in accordance with the TRIPOD reporting checklist (available at <https://jgo.amegroups.com/article/view/10.21037/jgo-24-541/rc>).

Methods

Data collection and preprocessing

We used two distinct data sets obtained from separate sequencing platforms, including a total of 443 GC samples from TCGA (<http://portal.gdc.cancer.gov>) and 433 GC samples from the GEO (GSE84437) (<http://www.ncbi.nlm.nih.gov/geo/>). To facilitate subsequent analysis, we first excluded 36 adjacent normal samples and 50 cases lacking complete survival information and RNA sequencing data from TCGA. Consequently, the final number of cases included in the analysis was 357. TCGA samples were then classified according to the anatomical site and the 10th edition of the International Classification of Diseases using the clinical data. The classification details are shown in *Table 1*. Next, we

Highlight box

Key findings

- Identifying pivotal genes that distinguish the expression profiles of cardia and non-cardia gastric cancers (NCGC) for prognostic prediction, and developing a risk prognosis model, will be beneficial for enhancing gastric cancer risk stratification and support the advancement of precision therapies for gastric cancer.

What is known and what is new?

- Classifying gastric cancer into cardia and non-cardia subtypes based on anatomical location has long been a standard clinical practice. Although these two subtypes have different pathogenesis and genetic backgrounds, there is still no fundamental difference in their pharmacological treatment.
- We utilized the expression profile differences derived from the anatomical distinctions between CGC and NCGC to construct a prognostic risk model for gastric cancer, aiming to explore new directions in guiding clinical treatment.

What is the implication, and what should change now?

- Based on the observed differential gene expression patterns within the context of medical anatomy, it is feasible to construct a prognostic signature for gastric cancer using a straightforward methodology.

Table 1 Anatomical classification details of gastric cancer samples in TCGA

Diagnosis/anatomical site	ICD code	Number of samples	Total
CGC			89
Gastroesophageal junction	C16.0	41	
Cardia	C16.0	48	
NCGC			268
Fundus/body	C16.1/C16.2	130	
Antrum	C16.3	138	

TCGA, The Cancer Genome Atlas; CGC, cardia gastric cancer; NCGC, non-cardia gastric cancer; ICD, International Classification of Diseases.

transformed the probe matrix of the GSE84437 data set into a gene matrix and performed batch correction. The clinical information for GSE84437 was shown in Table S1. TCGA data were employed as the training set, while the GEO data served as the verification group. The study was conducted in accordance the Helsinki Declaration (as revised in 2013).

Clinical analysis of CGC and NCGC

We conducted an analysis of the clinicopathologic features between CGC and NCGC using the “tableOne” package in R (9). The features examined included age, gender, race, a family history of cancer, histological type, residual tumor, and stage.

Differential gene analysis between CGC and NCGC

We employed the “limma” package to identify the differential expression of the messenger RNA with a threshold of a $|\log_2 \text{fold change}| > 1$ and a false discovery rate < 0.05 (10). This analysis allowed us to identify the DEGs between CGC and NCGC. The results were visualized using the “EnhancedVolcano” package and the “heatmap” package in R through the creation of volcano plots and heatmap diagrams, respectively (11,12).

Next, we used “clusterProfiler” package to perform the Gene Ontology (GO) and Kyoto Encyclopedia of Genes and Genomes (KEGG) enrichment analyses based on the DEGs (13). The GO terms and KEGG pathways were plotted using the “enrichplot”, “ggplot2”, and “GOplot” packages in R (14-16). The GO and KEGG gene sets were obtained from the MSigDB database ([\[msigdb.org/gsea/msigdb\]\(https://msigdb.org/gsea/msigdb\)\).](https://www.gsea-</p>
</div>
<div data-bbox=)

Construction of a risk model and validation model

A univariate Cox proportional hazard regression analysis (uniCOX) was employed to identify prognostic genes among the identified DEGs. The results are presented in a forest plot. We conducted a screening of the independent prognostic marker genes and developed RS models using a stepwise multivariate Cox analysis (multiCOX). The RS models were constructed using the following formula:

$$\text{Risk score}(RS) = \sum_{i=1}^n (\text{Coef}_i \times X_i) \quad [1]$$

where Coef_i refers to the risk coefficient of the signature genes, and X_i refers to the expression levels of the genes identified. The values for Coef_i were obtained from the multiCOX analysis. Using the median RS as the cut-off value, the samples were categorized into high- and low-risk groups. Survival curves were generated by a Kaplan-Meier (KM) analysis. Finally, we assessed the stability of the models in GSE84437. The survival receiver operating characteristic (ROC) curve was used to assess the prognostic signature’s performance, and the area under the curve (AUC) was calculated using the “timeROC” package in R software (17).

Gene set enrichment analysis (GSEA) of the two groups

To elucidate the biological functions distinguishing the two groups, we performed a GSEA of the high- and low-risk groups. This analysis was conducted using the “clusterProfiler” package in R (13), and the results were visualized using the “enrichplot” package (15). The gene sets used for the enrichment analysis were sourced from Molecular Signatures Database (MSigDB) (<http://software.broadinstitute.org/gsea/msigdb/>).

Single-gene mutation analysis of the two groups

Following the acquisition and processing of the GC mutation data from TCGA, we conducted a comparative analysis of the tumor mutational burden (TMB) between the high- and low-risk groups. Using the “maftools” package, we examined the prevalence of mutations and visualized the top 30 most frequently mutated genes in each group. These were represented as waterfall plots using the “oncplot” function in R (18).

Immune infiltration analysis of the different groups

We employed the CIBERSORT algorithm to calculate and compare the scores for 22 types of immune cells and immune-related functions in the high- and low-risk groups (19). The “estimate” package in R was used to analyze differences in the immune microenvironment between the two groups (20). This analysis included the computation of the immune, stromal, and estimation of stromal and immune cells in malignant tumor tissues using expression data (ESTIMATE) scores, and tumor purity. Additionally, we compared the tumor immune dysfunction and exclusion (TIDE) scores between the high- and low-risk groups to predict the potential therapeutic benefits of immune checkpoint inhibitors (21).

Drug sensitivity analysis

To evaluate the clinical applicability of our risk model in GC treatment, we used the Genomics of Drug Sensitivity in Cancer (<https://www.cancerrxgene.org/>) data set as a training set. The inhibitory concentration (IC₅₀) value, which represents the concentration of a drug necessary for 50% inhibition *in vitro*, was calculated for each sample using ridge regression. This was accomplished by employing the “calcPhenotype” function from the R package “oncoPredict” (22). Following this, we predicted the drug sensitivity of the GC patients in the high- and low-risk groups.

Statistical analysis

All the statistical analyses were conducted using R software version 4.2.2. The count data are presented as the number and rate (%), and comparisons between groups were performed using the χ^2 test. Univariate and multivariate Cox regression analyses were employed to examine independent prognostic factors incorporating clinical data. The KM method and log-rank tests were used to assess the survival rates of the patients in the high- and low-risk groups. Statistical significance was defined as a P value <0.05 for all the analyses.

Results

Differences in the clinical characteristics between CGC and NCGC

The CGC and NCGC data from TCGA database and the results of the χ^2 test for the clinical characteristics are presented in *Table 2*. These characteristics include age,

Table 2 Clinicopathologic features of gastric cancer in TCGA data set: comparison between CGC and NCGC patients

Clinicopathologic features	CGC	NCGC	P value
Age (years)			0.34
≤65	42 (47.2)	114 (42.5)	
>65	45 (50.6)	152 (56.7)	
N/A	2 (2.2)	2 (0.8)	
Gender			0.15
Male	63 (70.8)	165 (61.6)	
Female	26 (29.2)	103 (38.4)	
Race			0.002
Asian	7 (7.9)	67 (25.0)	
Black or African American	1 (1.1)	10 (3.7)	
White	66 (74.2)	159 (59.3)	
N/A	15 (16.9)	32 (12.0)	
Family history of cancer			0.13
No	60 (67.4)	205 (76.5)	
Yes	3 (3.4)	12 (4.5)	
N/A	26 (29.2)	51 (19.0)	
Histological type			0.04
Adenocarcinoma	44 (49.4)	83 (30.9)	
Diffuse type	9 (10.1)	51 (19.0)	
Intestinal type	16 (18.0)	50 (18.7)	
Mucinous type	3 (3.4)	16 (6.0)	
Papillary type	1 (1.1)	4 (1.5)	
Signet ring type	4 (4.5)	7 (2.6)	
Tubular type	12 (13.5)	57 (21.3)	
Grade			0.01
1	2 (2.2)	7 (2.6)	
2	44 (49.4)	86 (32.1)	
3	43 (48.3)	166 (61.9)	
N/A	0 (0)	9 (3.4)	
Residual tumor			0.32
R0	72 (80.9)	219 (81.7)	
R1	6 (6.7)	8 (3.0)	
R2	2 (2.2)	13 (4.9)	
N/A	9 (10.1)	28 (10.4)	

Table 2 (continued)

Table 2 (continued)

Clinicopathologic features	CGC	NCGC	P value
T stage			0.001
1	7 (7.9)	11 (4.1)	
2	26 (29.2)	46 (17.2)	
3	45 (50.6)	118 (44.0)	
4	10 (11.2)	87 (32.5)	
N/A	1 (1.1)	6 (2.2)	
N stage			0.87
0	27 (30.3)	78 (29.1)	
1	17 (19.1)	53 (19.8)	
2	20 (22.5)	55 (20.5)	
3	23 (25.8)	69 (25.7)	
N/A	2 (2.2)	13 (4.9)	
Metastasis			0.36
0	82 (92.1)	247 (92.2)	
1	7 (7.9)	16 (6.0)	
N/A	0 (0.0)	5 (1.9)	
TNM stage			0.16
I	17 (19.1)	32 (11.9)	
II	29 (32.6)	74 (27.6)	
III	34 (38.2)	132 (49.3)	
IV	7 (7.9)	16 (6.0)	
N/A	2 (2.2)	14 (5.2)	

Data are presented as n (%). TCGA, The Cancer Genome Atlas; CGC, cardia gastric cancer; NCGC, non-cardia gastric cancer; N/A, not available; T, tumor; N, nodes; M, metastasis.

gender, race, a family history of cancer, histological type, grade, residual tumor, and tumor stage (T), node stage (N), and metastasis (M), and TNM stage.

Significant differences were observed in race, histological type, grade, and T stage between the two groups. Consistent with previous related studies (23,24). CGC was found to be more prevalent among individuals of the white race, while NCGC was more common among those of the Asian race.

Interestingly, NCGC exhibited a notably high prevalence of T3 and T4 stages, and a significant portion of the pathological types consisted of diffuse and tubular adenocarcinoma, exhibiting poor differentiation (Grade 3: 61.9%). Conversely,

CGC primarily exhibited T2 and T3 stages, with moderate and low differentiation (Grade 2: 49.4%, Grade 3: 48.3%).

DEG analysis between CGC and NCGC

A total of 339 DEGs were identified, of which 270 were up-regulated in CGC and 69 were up-regulated in NCGC (for further details, see Table S2). The distribution of these DEGs was visualized in the volcano plot and heatmap depicted in Figure 1.

To further understand the biological function of these DEGs, GO terms and KEGG pathway analyses were conducted (Figure 2). The GO analysis primarily highlighted terms related to keratin function, such as “epidermis development”, “skin development”, “keratinocyte differentiation”, and “epidermal cell differentiation”. The most significant pathways identified in the KEGG analysis were the “*Staphylococcus aureus* infection pathway” and the “estrogen signaling pathway”.

Construction of the prognostic risk model based on the DEGs

To identify the genes related to prognosis, we screened 17 genes from the DEGs using the uniCOX. Subsequently, we employed a stepwise multiCOX to eliminate redundant factors, resulting in the final selection of the following four genes: *KRT17*, *PPP1R1C*, *SLC5A5*, and *SYT6*. The results are illustrated in the forest plots depicted in Figure 3A,3B.

We constructed a prognostic RS model based on the expression values of these four genes and their corresponding *Coefi* as detailed in Table S3. Using the Cox regression model, we calculated RS for each patient as follows: $RS = (0.106 \times \text{expression of } KRT17) + (0.536 \times \text{expression of } PPP1R1C) + (0.290 \times \text{expression of } SLC5A5) + (1.215 \times \text{expression of } SYT6)$. The GC samples from TCGA were then divided into high- and low-risk groups based on the median RS (Table S4).

Validation of the prognostic risk model

To validate the predictive power of the prognostic risk model in TCGA data set, we plotted the survival curves for the high- and low-risk samples (Figure 3C). The results demonstrated that the survival rate of the high-risk group was significantly lower than that of the low-risk group. Further, we performed a ROC curve analysis based on the AUCs. The AUC values for the model's ROC curves were 0.71, 0.77, and 0.68, which were associated with the 1-, 3-,

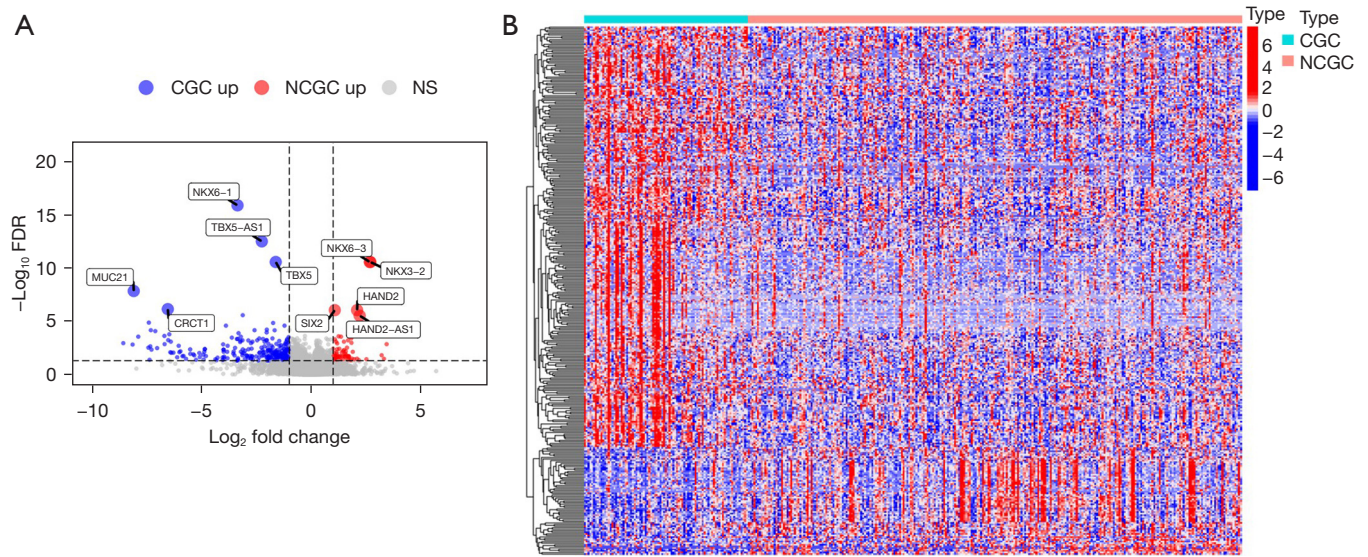


Figure 1 Volcano plot and heatmap of the DEGs between CGC and NCGC. (A) In the volcano plot, the blue points indicate the up-regulated genes in the CGC samples, and the red points indicate the up-regulated genes in the NCGC samples. (B) In the heatmap, each row represents a gene, and each column represents a sample. The depth of color indicates the strength of gene expression, with red indicating the up-regulation of gene expression. FDR, false discovery rate; CGC, cardia gastric cancer; NCGC, non-cardia gastric cancer; NS, not significant; DEGs, differentially expressed genes.

and 5-year survival rate forecasts, respectively. Notably, the prediction for the 3-year survival rate exhibited the highest accuracy (Figure 3D). The results suggested that the model could be effectively used to predict survival.

We conducted further validation of the prognostic model by utilizing the GES84437 dataset (Table S5). The findings also indicated that the survival outcomes for samples from the high-risk group were notably inferior to those from the low-risk group (Figure 3E). The AUC values of the model's ROC curves were 0.60, 0.61, and 0.58, which correspond to the predictions of survival rates at 1, 3, and 5 years, respectively (Figure 3F). Similarly, the prediction accuracy of the 3-year survival rate was relatively high.

Single-gene mutation landscape in the two groups

In both the high- and low-risk groups, the following mutations were the most prominent: the missense mutation, in-frame deletion, frameshift mutation, and nonsense mutation, all of which showed a consistent trend across all the GC samples from TCGA data set. Among these, the missense mutation was the most common.

The 30 most frequently mutated genes in each sample and the mutation frequency of different groups were

visualized in a waterfall plot (Figure 4A-4C). The mutation frequency of the low-risk group was similar to that of the high-risk group.

The overall TMB of the GC samples was not high, with the median cut-off being 1.94/MB (Figure 4D). A further comparison of the TMB differences between the groups showed that the TMB was higher in the low-risk group than the high-risk group (Figure 4E). The top five mutated genes in the high-risk group were *TTN*, *TP53*, *LRP1B*, *PCLO*, and *ARID1A*, while those in the low-risk group they were *TTN*, *TP53*, *MUC16*, *ARID1A*, and *DNAH5*. Notably, the *TP53* mutations were more pronounced in the high-risk group, while the *MUC16* mutations were significantly higher in the low-risk group. This indicated that the mutation frequency of the top 30 mutated genes was not entirely identical between the two groups. However, the correlation between TMB and RS in this study is statistically insignificant (Figure 4F).

GSEA of the two groups

Using the hallmark gene sets from the GSEA, we identified 12 pathways that were differentially involved in the biological functions between the high- and low-risk groups (Figure 5). Among these, pathways such as “coagulation”,

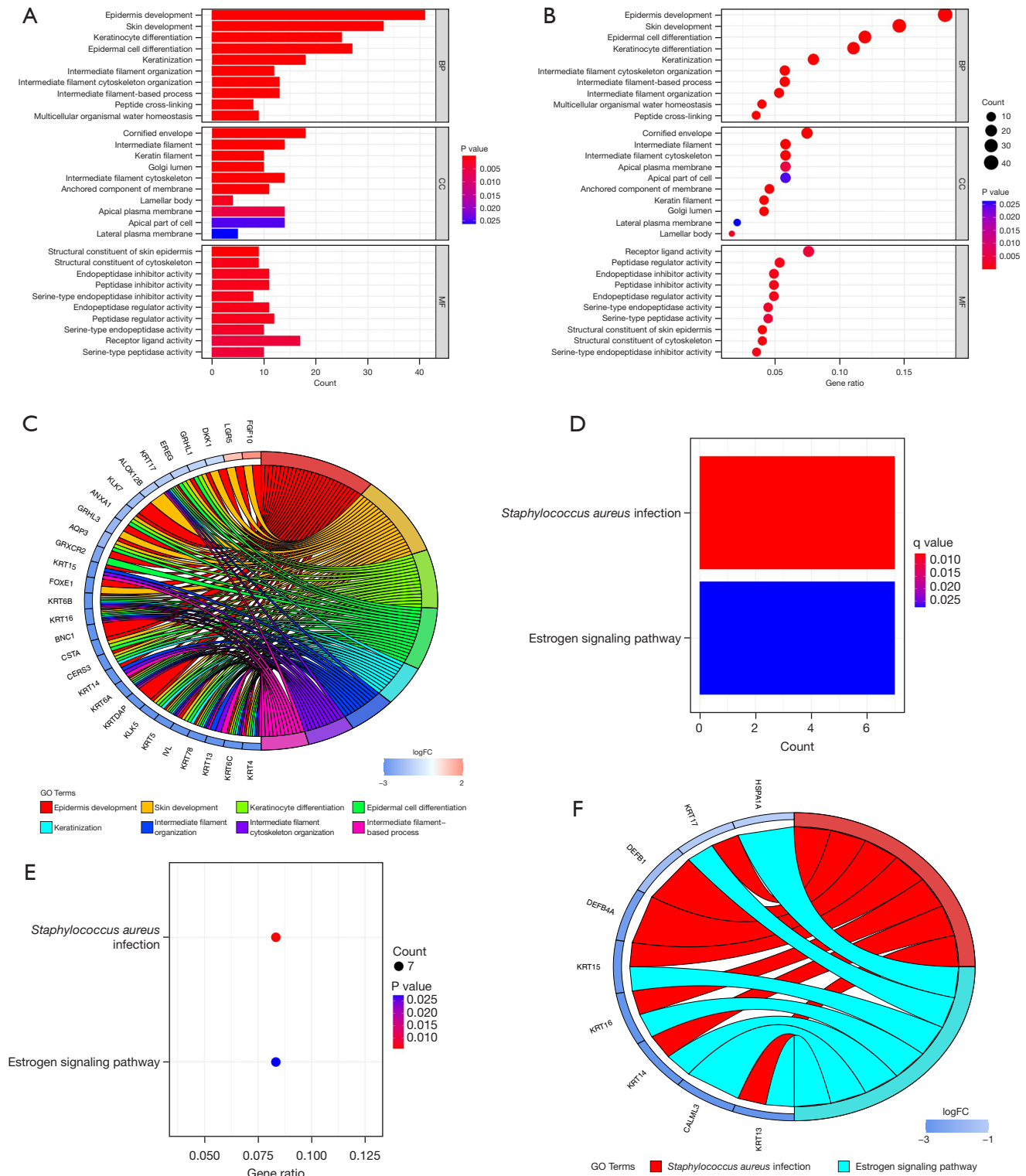


Figure 2 Functional enrichment analysis of the DEGs. (A,B) GO terms of the DEGs functional enrichment analysis (A: Bar chart; B: Bubble chart). (C) The proportion of each enriched GO term. (D,E) KEGG pathways of the DEGs functional enrichment analysis (D: Bar chart; E: Bubble chart). (F) The proportion of each enriched KEGG pathway. BP, biological process; CC, cellular component; MF, molecular function; FC, fold change; DEGs, differentially expressed genes; GO, Gene Ontology; KEGG, Kyoto Encyclopedia of Genes and Genomes.

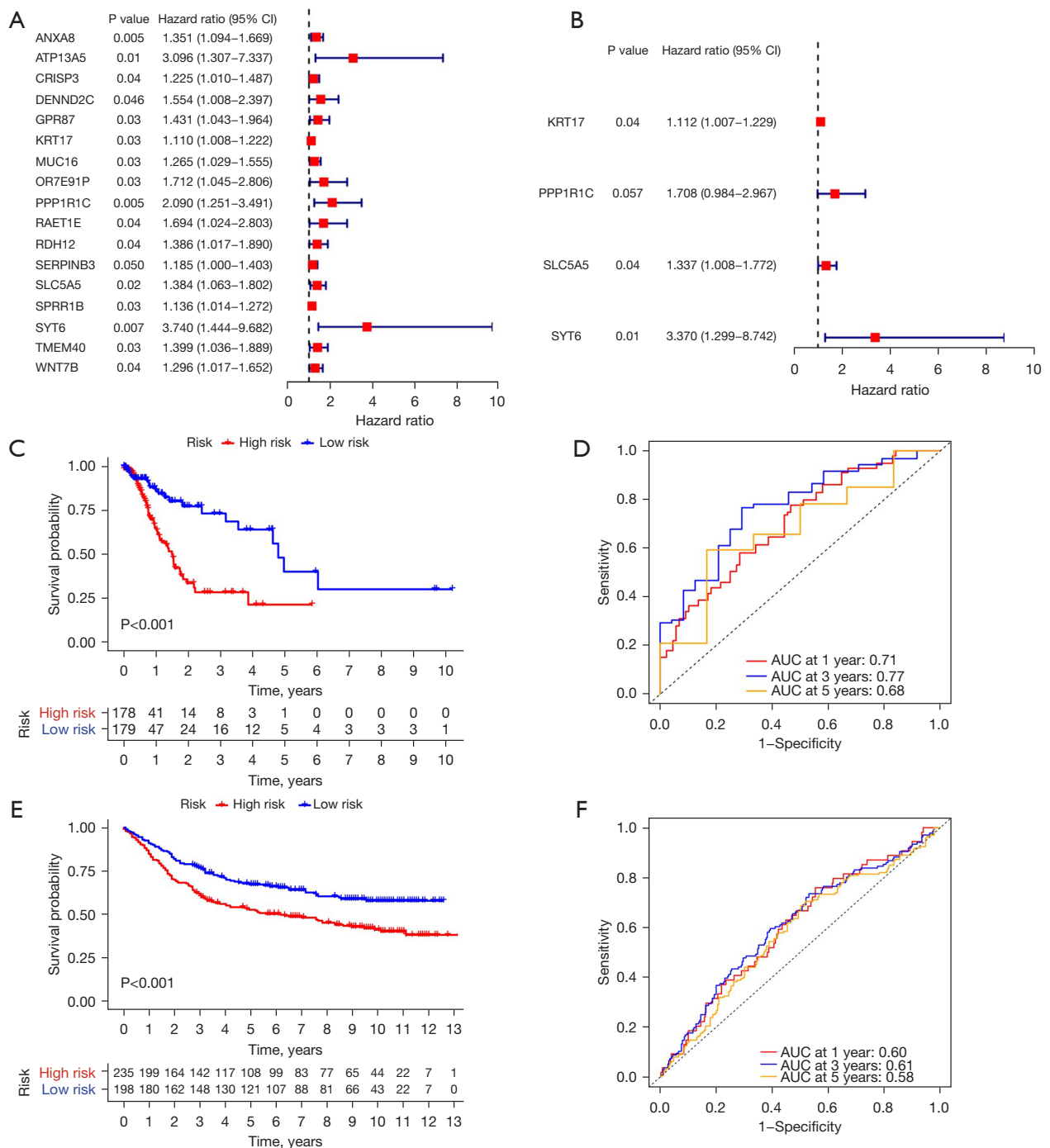


Figure 3 Construction and validation of a PRSM. (A) Forest plot showing the 17 prognosis-related genes that were selected from the DEGs between CGC and NCGC based on the uniCOX. (B) Forest plot showing the four genes (*KRT17*, *PPP1R1C*, *SLC5A5*, and *SYT6*) that were identified to construct the PRSM based on the multiCOX. (C) Internal validation of the PRSM in TCGA data. (D) Survival-dependent ROC curve validation of the model's ability to predict patient prognosis in TCGA data. (E) External validation of the PRSM in the GEO data. (F) Survival-dependent ROC curve validation of the model's ability to predict patient prognosis in the GEO data. CI, confidence interval; AUC, area under the curve; PRSM, prognostic risk scoring model; DEGs, differentially expressed genes; CGC, cardia gastric cancer; NCGC, non-cardia gastric cancer; uniCOX, univariate Cox proportional hazard regression analysis; multiCOX, multivariate Cox analysis; TCGA, The Cancer Genome Atlas; ROC, receiver operating characteristic; GEO, Gene Expression Omnibus.

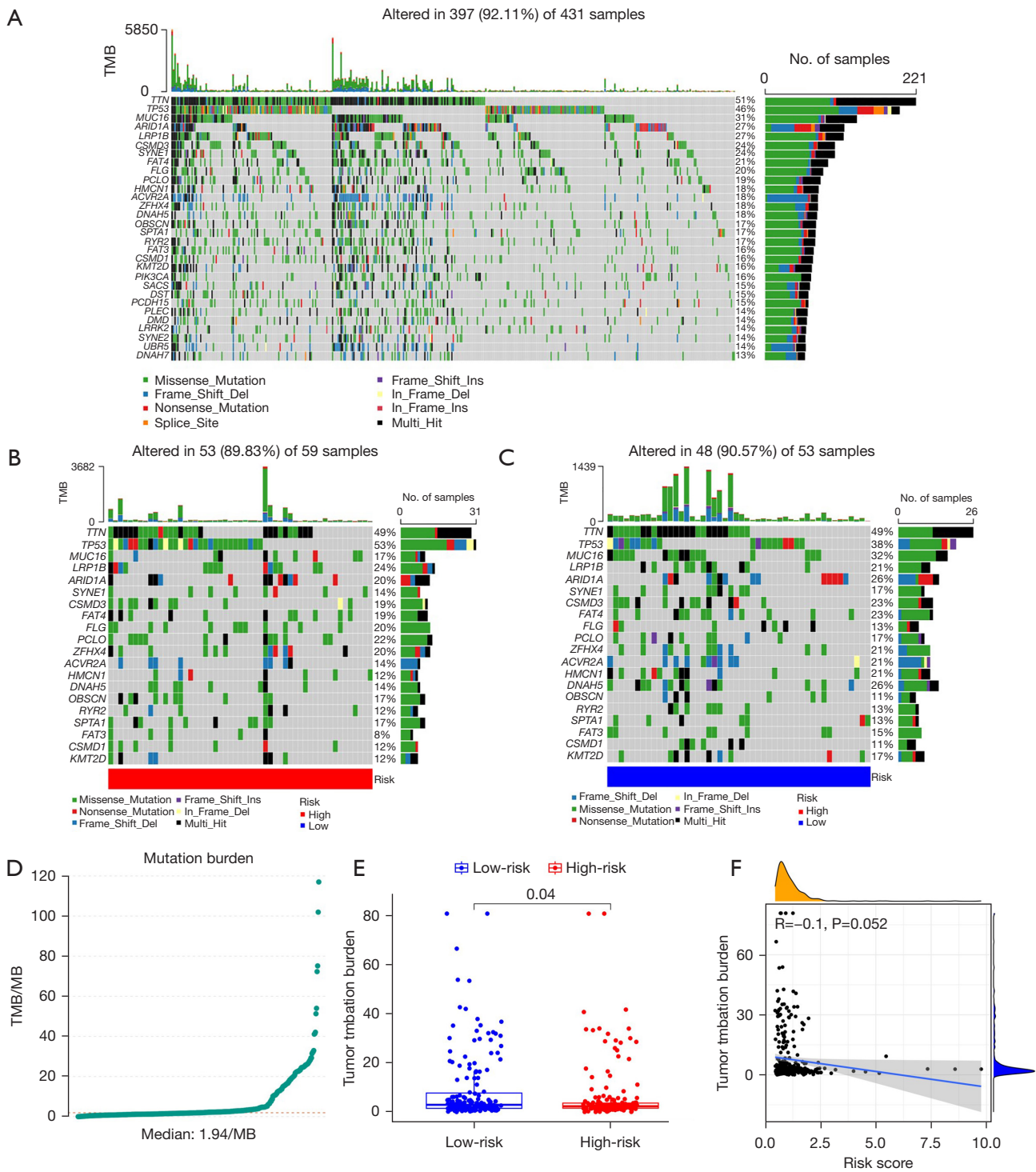


Figure 4 Single-gene mutation landscape of GC in the high- and low-risk groups. (A) Waterfall plot of single-gene mutation landscape in TCGA GC samples. (B) Waterfall plot of the single-gene mutation landscape in the high-risk group. (C) Waterfall plot of the single-gene mutation landscape in the low-risk group. Each row represents a gene, and each column represents a sample. Different colors represent different types of mutations. (D) The TMB in TCGA samples. (E) Differences in the TMB between the high- and low-risk groups. (F) The correlation between the TMB and RS. MB, million bases; GC, gastric cancer; TCGA, The Cancer Genome Atlas; TMB, tumor mutational burden; RS, risk score.

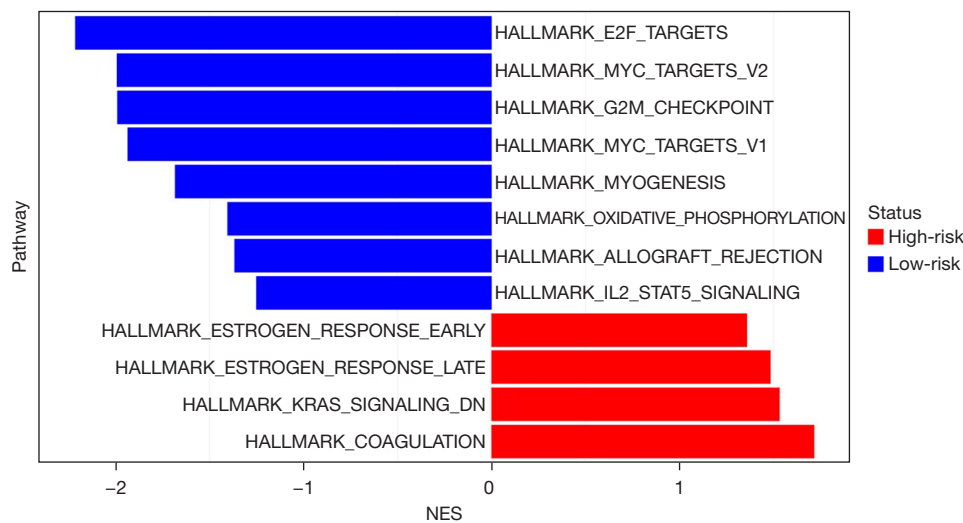


Figure 5 GSEA results of the high- and low-risk groups based on the hallmark gene sets. NES, normalized enrichment score; GSEA, gene set enrichment analysis.

“estrogen response early/late”, and “KRAS signaling DN” were enriched in the high-risk group, while “E2F targets”, “MYC targets”, and the “G2M checkpoint” were primarily enriched in the low-risk group.

In addition, we performed a GSEA based on the KEGG gene sets and identified 36 differentially enriched pathways between the two groups. Several of these pathways mirrored the results from the hallmark gene sets, such as the enrichment of the “complement and coagulation cascades” pathway in the high-risk group, and the enrichment of the “cell cycle” pathway in the low-risk group. Notably, the “ribosomes” pathway was the most enriched pathway in the low-risk group, suggesting a link between ribosome biosynthesis and the biological function of the low-risk group.

Further, pathways such as “tryptophan metabolism”, “glycerolipid metabolism”, “JAK-STAT signaling pathway”, and the “VEGF signaling pathway” were also significantly enriched in the low-risk group. Interestingly, apart from the aforementioned pathways, we identified several immune-related pathways enriched in the low-risk group, including the “FcεRI signaling pathway”, the “intestinal immune network for IgA production”, and the “T cell receptor signaling pathway”.

Immune microenvironment landscape within different groups

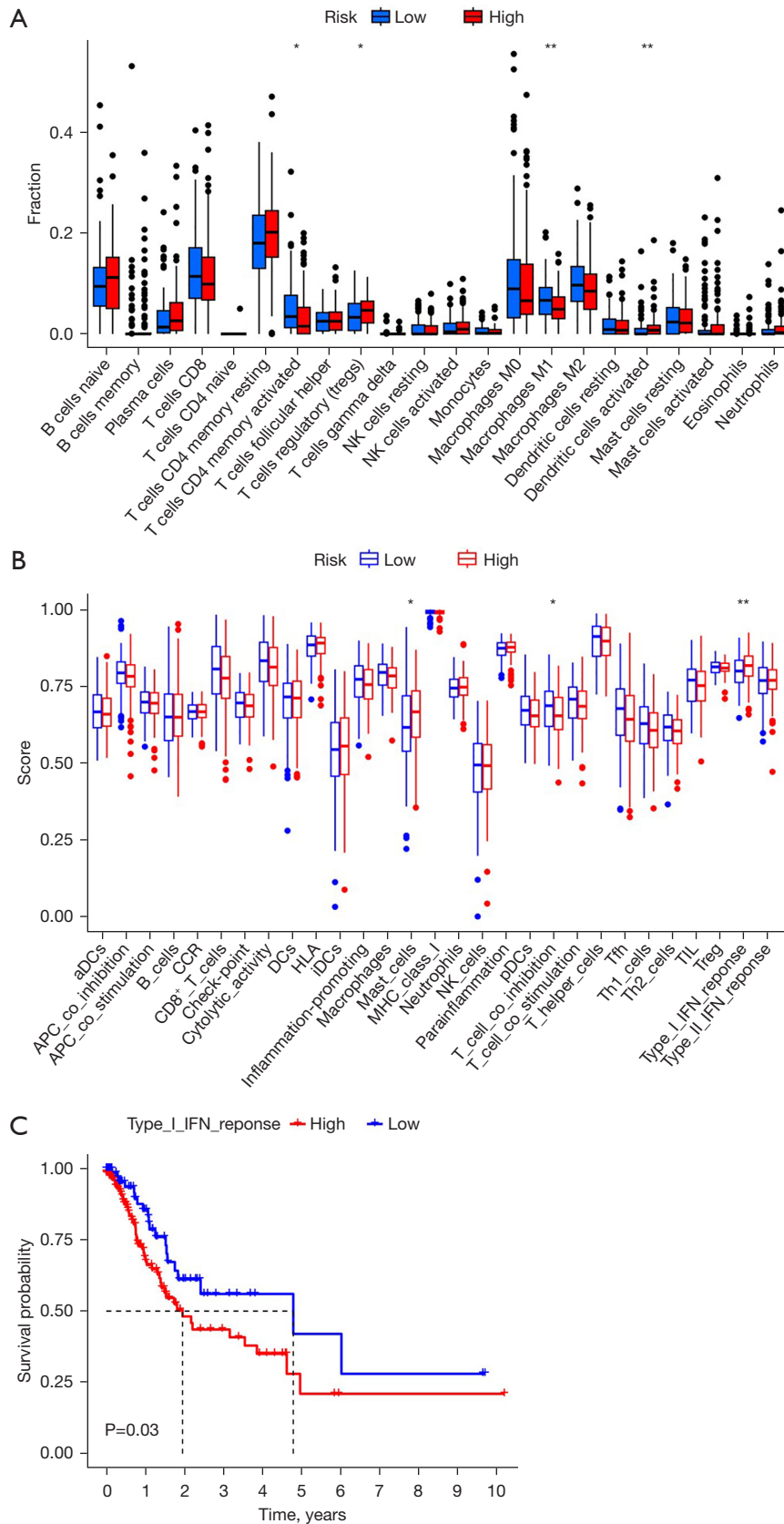
We analyzed the immune cell infiltration and immune

cell function of the high-risk and low-risk groups using the “CIBERSORT” and Gene Set Variation Analysis algorithms, respectively. The results indicated that the levels of infiltrating T cells (CD4 memory activated) and macrophages (M1) were higher in the low-risk group than the high-risk group (Figure 6A). Conversely, the levels of the activated dendritic cells and regulatory T cells were lower in the low-risk group than the high-risk group.

When comparing the immune cell functions between the two groups, the high-risk group exhibited a stronger type I interferon (IFN) response than the low-risk group, which is associated with a poor prognosis (Figure 6B,6C). There was no significant difference between the groups in terms of the ESTIMATE score; however, a higher number of patients in the low-risk group exhibited microsatellite instability (Figure 6D). Additionally, the high-risk group had a higher TIDE prediction score than the low-risk group, which is associated with immune escape (Figure 6E). Therefore, these results suggest that the high-risk group had a lower probability of benefiting from immune therapies than the low-risk group.

Sensitivity of different chemotherapy drugs in the two groups

Chemotherapy still remains the primary therapeutic strategy in the systemic management of advanced GC. To evaluate drug sensitivity, we conducted an analysis to determine the half-maximal IC₅₀ values of commonly used



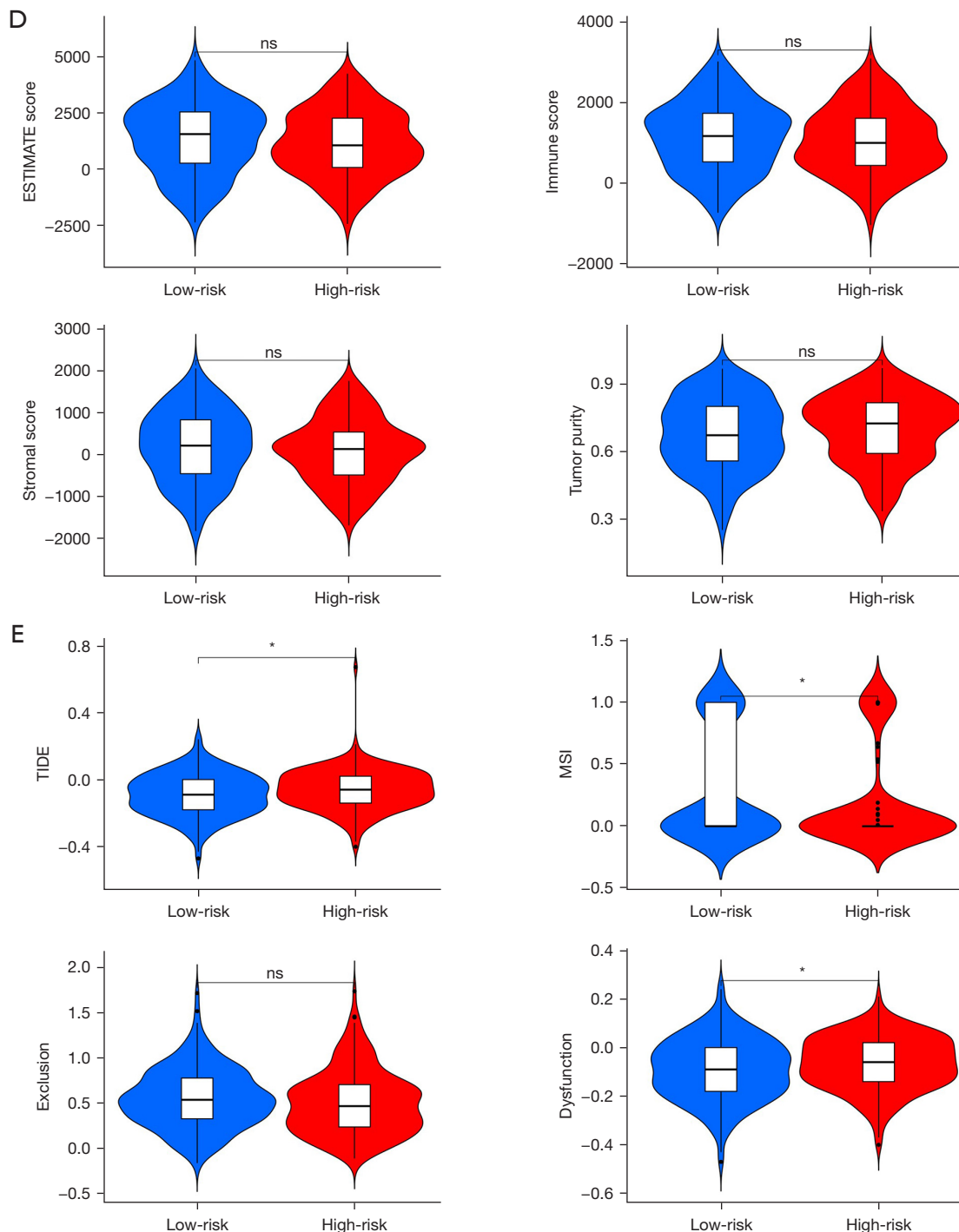


Figure 6 Immune microenvironment landscapes of different groups. (A) The levels of immune cell infiltration in the high- and low-risk groups. (B) The immune function scores of the high- and low-risk groups. (C) Discrepancies in the survival outcomes between the high- and low-risk groups in terms of IFN-1 response. (D) ESTIMATE score results of the high- and low-risk groups. (E) TIDE score results of the high- and low-risk groups. *, $P < 0.05$; **, $P < 0.01$; ns, not significant. NK, natural killer; aDCs, activated dendritic cells; APC, antigen-presenting cells; CCR, chemokine receptor; DCs, dendritic cells; HLA, human leukocyte antigen; iDCs, immature dendritic cells; MHC, major histocompatibility complex; TIL, tumor-infiltrating lymphocytes; IFN-1, type I interferon; ESTIMATE, estimation of stromal and Immune cells in malignant tumours using expression data; TIDE, tumor immune dysfunction and exclusion.

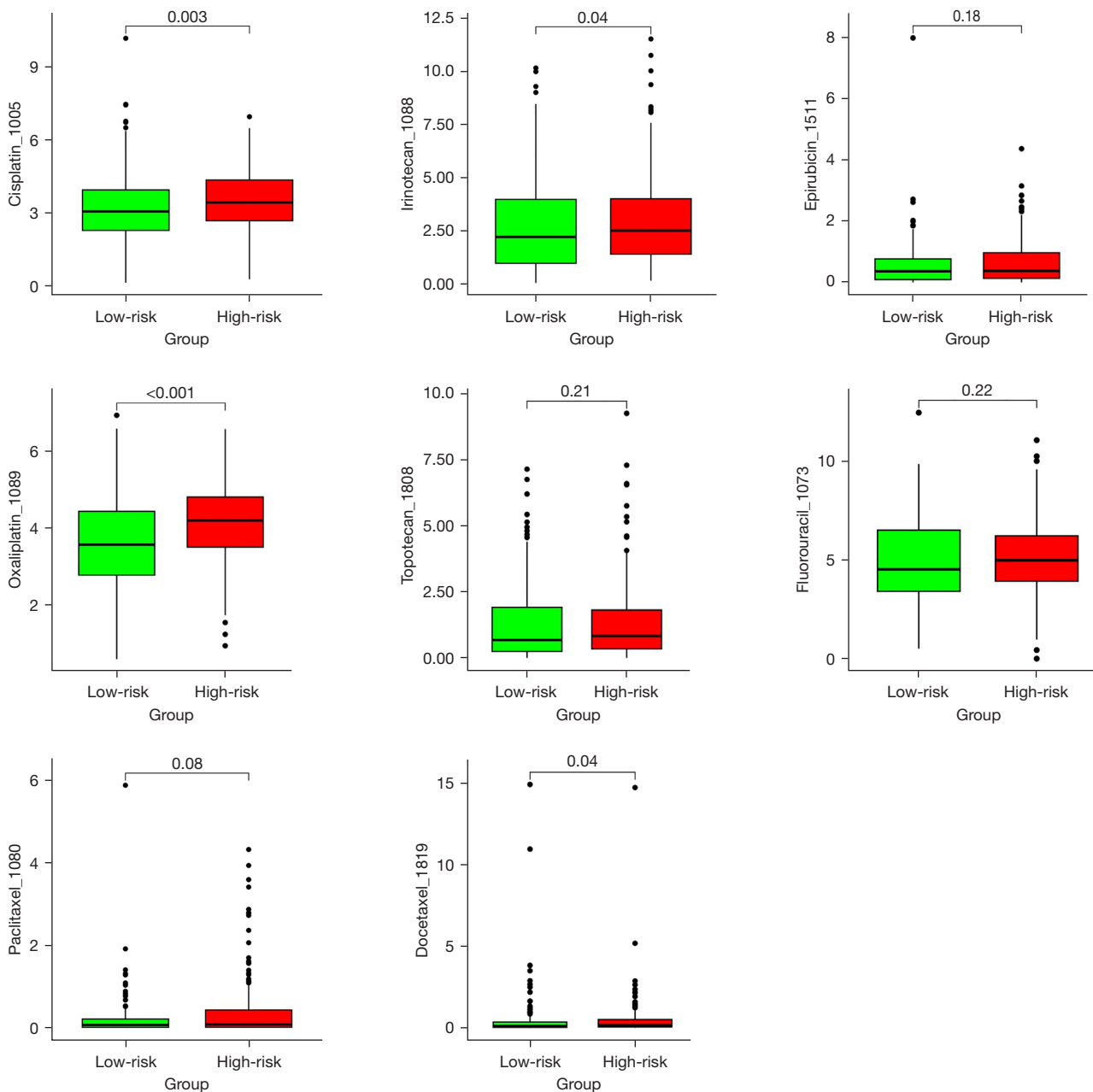


Figure 7 Sensitivity of different chemotherapy drugs in the two groups.

chemotherapy drugs for GC in both the high- and low-risk groups. The IC₅₀ values for most chemotherapeutic agents, including cisplatin, oxaliplatin, docetaxel, and irinotecan, were observed to be significantly reduced in the low-risk cohort compared to the high-risk cohort. This implied a greater sensitivity to these specific drugs in the low-risk group. However, no significant difference was noted in the IC₅₀ values of paclitaxel and 5-fluorouracil between the two

groups (Figure 7). The high-risk group did not demonstrate a distinct advantage in the selection of chemotherapeutic agents, further emphasizing the more severe prognosis-associated with this group.

Discussion

GC can be categorized into CGC and NCGC based on the

primary sites. Clinical statistics indicate that CGC patients who have undergone R0 resection have inferior disease-free survival and overall survival (OS) than NCGC patients (25). Data analyses from the Surveillance, Epidemiology, and End Results (SEER) database also corroborate the view that CGC patients have a poorer prognosis than NCGC patients; thus CGC can serve as an independent prognostic factor for GC (5,26). These findings suggest that CGC and NCGC may possess distinct biological characteristics that influence prognosis.

In recent years, numerous studies have begun to explore the differences in expression profiles between CGC and NCGC. Wang *et al.* were the first to report the global gene expression of CGC and NCGC through tumor-normal paired patient testing (6). Further enriched functionality results revealed that CGC-unique DEGs and NCGC-unique DEGs were enriched in different biological processes, such as cell cycle, cell proliferation, and cell death (7). However, the previous two studies were primarily based on the differential analysis of tumor-normal paired tissues, which while advantageous in exploring the etiological molecular features and screening the early diagnosis markers of CGC and NCGC, do not adequately address the different treatment outcomes caused by the intrinsic molecular differences between them. In this study, we analyzed the GC data from TCGA and directly compared the expression profiles of CGC and NCGC, and thus identified the key genes that affect the prognosis of GC patients. Based on these findings, we developed a PRSM and classified GC into different subtypes for predicting therapeutic sensitivity.

Initially, we identified 339 DEGs between CGC and NCGC and conducted a functional analysis. The significantly enriched GO terms from the DEGs included epidermis development, intermediate filament organization, and intermediate filament cytoskeleton organization (27). Multiple genes from the keratin family were involved in these processes, such as *KRT3*, *KRT5*, and *KRT17*. Under physiological conditions, these genes play a crucial role in organizing the cytoskeleton and maintaining epithelial integrity. In epithelial malignant tumors, the abnormal expression of keratins plays a significant biological role in tumor metastasis, angiogenesis, immune evasion, and resistance to immune checkpoint blockades (ICBs) (28-30).

Additionally, the KEGG results showed that the DEGs were significantly enriched in the following two pathways: *Staphylococcus aureus* infection, and the estrogen signaling pathway. Recent studies have suggested that

Staphylococcus aureus infection might promote tumor growth and metastasis by affecting the immune system (31,32). However, there have been no similar reports in GC. Estrogen and estrogen receptors participate in the regulation of the body's metabolism, either in physiological or pathological states (33). The estrogen signaling pathway significantly enriched in the DEGs of gastric signet ring cell carcinoma can interact with the mitogen-activated protein kinase and promote tumor progression (34). Further, estrogen can polarize macrophages toward an immunosuppressive phenotype through estrogen receptors, leading to abnormal CD8 T cell function and affecting the response of melanoma cells to ICBs (35). Collectively, these findings suggest that differences in the expression profiles between CGC and NCGC are involved in tumor growth, metastasis, angiogenesis, and immune regulation. Subsequently, we used uniCOX and multiCOX to identify four key prognostic genes (i.e., *KRT17*, *SLC5A5*, *PPP1R1C*, and *SYT6*), among which the first three genes were up-regulated in CGC, while *SYT6* was up-regulated in NCGC.

KRT17, a type I keratin intermediate filament, is abnormally expressed in multiple types of tumors and has prognostic value (36,37). Our study found a correlation between the high expression of *KRT17* and a poor prognosis in GC. *PPP1R1C*, also known as inhibitor-5 of protein phosphatase 1, primarily affects the cell cycle of tumor cells (38,39). *SLC5A5* plays an important role in thyroid hormone synthesis and radioactive iodine therapy for thyroid tumors (40,41). In our study, the high expression of *SLC5A5* indicated a poor prognosis in GC. *SYT6*, a member of the synaptotagmin family of membrane transport proteins, was found to have significantly lower expression in CGC than NCGC. Moreover, the high expression of *SYT6* in GC was associated with a shorter survival time.

Based on these four key genes, we developed a PRSM to differentiate the GC samples into high- and low-risk groups. The model demonstrated good predictive performance with AUC values of 0.71, 0.77, and 0.68 at 1, 3, and 5 years, respectively. We further validated its prognostic value in the GEO database. These findings provide valuable insights for the development of personalized therapeutic strategies for GC.

According to the GSEA, the genes in the low-risk group were primarily enriched in pathways related to the cell cycle and cell proliferation, such as E2F targets, the G2M checkpoint, and MYC targets. Conversely, the genes in the high-risk group showed significant enrichment in pathways, such as early estrogen response, late estrogen response, and

coagulation. Keenan *et al.* reported that patients with high GSEA scores for an early and late estrogen response might exhibit lower expressions of antigen presentation genes, potentially impacting the effectiveness of immunotherapy (42). The activation of the coagulation system can also suppress the therapeutic efficacy of ICBs via platelets, leukocytes, and the complement system (43). These results suggest that the pathways enriched in the high-risk group were mainly associated with tumor immune suppression.

In the immune landscape analysis, the low-risk group was characterized by the infiltration of cells promoting anti-tumor effects, such as activated CD4 memory T cells and M1 macrophages. Conversely, the primary infiltrating cells in the high-risk group were dendritic cells and regulatory T cells, which play roles in antigen presentation and immune suppression, respectively. Meanwhile, the TMB of the low-risk group was higher than that of the high-risk group. Therefore, while there was no significant difference in the estimate scores between the two groups, the low-risk group appeared to have a better immune response than the high-risk group, as reflected by the higher TIDE scores in the low-risk group.

To ensure the better use of the PRSM in clinical applications, we compared the molecular characteristics specific to the high- and low-risk groups, and conducted a drug sensitivity analysis. *TP53* is a well-known tumor suppressor gene, and the wild-type P53 can inhibit tumor development through various pathways. Mutations in *TP53* and the resulting p53 inactivation can enable tumor cells to evade death, progress rapidly, and promote the suppression of the immune microenvironment, leading to a poor prognosis (44,45). Defects in the p53 signaling pathway are one of the predictive features for fluorouracil and oxaliplatin chemoresistance (45,46). In this study, the frequency of the *TP53* mutations was significantly increased in the high-risk group, consistent with their more aggressive biological behavior, poor immune environment, and prognosis.

Another gene with a higher mutation rate in the low-risk group was *MUC16*, which encodes the carbohydrate antigen 125, a common clinical biomarker for ovarian cancer. Recent studies have found that mutations in *MUC16* not only decrease the Warburg effect, affecting metabolic reprogramming (47), but are also associated with a high TMB and improved OS in various solid tumors, including GC (48-50). This suggests that the *MUC16* mutations served as a marker for favorable prognosis in the low-risk group.

The IC₅₀ values for commonly administered chemotherapeutic agents in GC, including cisplatin, oxaliplatin, docetaxel,

and irinotecan, were generally lower in the low-risk group than the high-risk group. This suggests a greater sensitivity to these drugs in the low-risk group. The sensitivity to other drugs, such as paclitaxel and fluorouracil, appeared to be similar between the two groups. Generally, the high-risk group did not demonstrate distinct advantages in the selection of chemotherapeutic and immunotherapeutic agents. Patients in this group may need more potent combined treatment strategies to enhance their therapeutic outcomes and survival.

It should be noted that the data used in this study primarily originated from public databases, and thus might be subject to the selection biases and limitations inherent to such sources. Future research should be conducted to validate these findings with broader clinical data sets and animal models, and to delve deeper into the molecular mechanisms underlying the differences between CGC and NCGC and their effects on treatment responses. This would better inform the development of personalized therapeutic strategies.

Conclusions

In conclusion, this study provided valuable insights into the molecular characteristics and immune landscape of high- and low-risk groups, which could potentially guide the development of more effective therapeutic strategies. However, further experimental research is needed to validate these findings and explore their clinical implications.

Acknowledgments

Funding: This work was supported by the Health Commission Research Project of Jiangsu (No. Z2022039); the Gusu Health Talent Plan of Suzhou (No. GSWS2022115); the Basic Research in Medical Applications of Suzhou (No. SKY2023034); and the Health Youth Backbone Talents “National Mentor System” Training Project of Suzhou (No. Qngg2022061).

Footnote

Reporting Checklist: The authors have completed the TRIPOD reporting checklist. Available at <https://jgo.amegroups.com/article/view/10.21037/jgo-24-541/rc>

Peer Review File: Available at <https://jgo.amegroups.com/article/view/10.21037/jgo-24-541/prf>

Conflicts of Interest: All authors have completed the ICMJE uniform disclosure form (available at <https://jgo.amegroups.com/article/view/10.21037/jgo-24-541/coif>). The authors have no conflicts of interest to declare.

Ethical Statement: The authors are accountable for all aspects of the work in ensuring that questions related to the accuracy or integrity of any part of the work are appropriately investigated and resolved. The study was conducted in accordance the Helsinki Declaration (as revised in 2013).

Open Access Statement: This is an Open Access article distributed in accordance with the Creative Commons Attribution-NonCommercial-NoDerivs 4.0 International License (CC BY-NC-ND 4.0), which permits the non-commercial replication and distribution of the article with the strict proviso that no changes or edits are made and the original work is properly cited (including links to both the formal publication through the relevant DOI and the license). See: <https://creativecommons.org/licenses/by-nc-nd/4.0/>.

References

- Sung H, Ferlay J, Siegel RL, et al. Global Cancer Statistics 2020: GLOBOCAN Estimates of Incidence and Mortality Worldwide for 36 Cancers in 185 Countries. *CA Cancer J Clin* 2021;71:209-49.
- Abdi E, Latifi-Navid S, Zahri S, et al. Risk factors predisposing to cardia gastric adenocarcinoma: Insights and new perspectives. *Cancer Med* 2019;8:6114-26.
- Kamangar F, Dawsey SM, Blaser MJ, et al. Opposing risks of gastric cardia and noncardia gastric adenocarcinomas associated with *Helicobacter pylori* seropositivity. *J Natl Cancer Inst* 2006;98:1445-52.
- Anderson LA, Tavilla A, Brenner H, et al. Survival for oesophageal, stomach and small intestine cancers in Europe 1999-2007: Results from EURO CARE-5. *Eur J Cancer* 2015;51:2144-57.
- Lv L, Liang X, Wu D, et al. Is cardia cancer a special type of gastric cancer? A differential analysis of early cardia cancer and non-cardia cancer. *J Cancer* 2021;12:2385-94.
- Wang G, Hu N, Yang HH, et al. Comparison of global gene expression of gastric cardia and noncardia cancers from a high-risk population in china. *PLoS One* 2013;8:e63826.
- Song B, Du J, Deng N, et al. Comparative analysis of gene expression profiles of gastric cardia adenocarcinoma and gastric non-cardia adenocarcinoma. *Oncol Lett* 2016;12:3866-74.
- Hu N, Wang Z, Song X, et al. Genome-wide association study of gastric adenocarcinoma in Asia: a comparison of associations between cardia and non-cardia tumours. *Gut* 2016;65:1611-8.
- Pollard TJ, Johnson AEW, Raffa JD, et al. tableone: An open source Python package for producing summary statistics for research papers. *JAMIA Open* 2018;1:26-31.
- Ritchie ME, Phipson B, Wu D, et al. limma powers differential expression analyses for RNA-sequencing and microarray studies. *Nucleic Acids Res* 2015;43:e47.
- Zhao S, Guo Y, Sheng Q, et al. Advanced heat map and clustering analysis using heatmap3. *Biomed Res Int* 2014;2014:986048.
- Blighe K, Rana S, Lewis M. enhancedVolcano: Publication-ready volcano plots with enhanced colouring and labeling. R package version 1.16.0 [Internet]. 2022 [cited 2024 Jan 24]. Available online: <https://doi.org/10.18129/B9.bioc.EnhancedVolcano>
- Yu G, Wang LG, Han Y, et al. clusterProfiler: an R package for comparing biological themes among gene clusters. *OMICS* 2012;16:284-7.
- Walter W, Sánchez-Cabo F, Ricote M. GOplot: an R package for visually combining expression data with functional analysis. *Bioinformatics* 2015;31:2912-4.
- Yu G. enrichplot: Visualization of Functional Enrichment Result_. R package version 1.18.4. 2023.
- Wickham H. Ggplot2: Elegant graphics for data analysis. 2nd ed. New York: Springer; 2016. Available online: <https://link.springer.com/book/10.1007/978-3-319-24277-4>
- 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.
- Mayakonda A, Lin DC, Assenov Y, et al. Maftools: efficient and comprehensive analysis of somatic variants in cancer. *Genome Res* 2018;28:1747-56.
- Chen T, Hua W, Xu B, et al. Robust rank aggregation and ciphersort algorithm applied to the identification of key genes in head and neck squamous cell cancer. *Math Biosci Eng* 2021;18:4491-507.
- Yoshihara K, Shahmoradgoli M, Martínez E, et al. Inferring tumour purity and stromal and immune cell admixture from expression data. *Nat Commun* 2013;4:2612.

21. Jiang P, Gu S, Pan D, et al. Signatures of T cell dysfunction and exclusion predict cancer immunotherapy response. *Nat Med* 2018;24:1550-8.
22. Maeser D, Gruener RF, Huang RS. oncoPredict: an R package for predicting in vivo or cancer patient drug response and biomarkers from cell line screening data. *Brief Bioinform* 2021;22:bbab260.
23. Liu J, Medina H, Reis IM, et al. Disadvantages for non-Hispanic whites in gastric carcinoma survival in Florida. *Cancer Causes Control* 2020;31:815-26.
24. Yao Q, Qi X, Cheng W, et al. A Comprehensive Assessment of the Racial and Ethnic Disparities in the Incidence of Gastric Cancer in the United States, 1992-2014. *Cancer Res Treat* 2019;51:519-29.
25. Zhao J, Zhao J, Du F, et al. Cardia and Non-Cardia Gastric Cancer Have Similar Stage-for-Stage Prognoses After R0 Resection: a Large-Scale, Multicenter Study in China. *J Gastrointest Surg* 2016;20:700-7.
26. Mo H, Li P, Jiang S. A novel nomogram based on cardia invasion and chemotherapy to predict postoperative overall survival of gastric cancer patients. *World J Surg Oncol* 2021;19:256.
27. Moll R, Divo M, Langbein L. The human keratins: biology and pathology. *Histochem Cell Biol* 2008;129:705-33.
28. Yin L, Li Q, Mrdenovic S, et al. KRT13 promotes stemness and drives metastasis in breast cancer through a plakoglobin/c-Myc signaling pathway. *Breast Cancer Res* 2022;24:7.
29. Xu Y, Zhang SZ, Huang CH, et al. Keratin 17 identified by proteomic analysis may be involved in tumor angiogenesis. *BMB Rep* 2009;42:344-9.
30. Wang W, Lozar T, Golfinos AE, et al. Stress Keratin 17 Expression in Head and Neck Cancer Contributes to Immune Evasion and Resistance to Immune-Checkpoint Blockade. *Clin Cancer Res* 2022;28:2953-68.
31. Zhao H, Teng D, Yang L, et al. Myeloid-derived itaconate suppresses cytotoxic CD8(+) T cells and promotes tumour growth. *Nat Metab* 2022;4:1660-73.
32. Qi JL, He JR, Liu CB, et al. Pulmonary *Staphylococcus aureus* infection regulates breast cancer cell metastasis via neutrophil extracellular traps (NETs) formation. *MedComm (2020)* 2020;1:188-201.
33. Kulkoyluoglu-Cotul E, Arca A, Madak-Erdogan Z. Crosstalk between Estrogen Signaling and Breast Cancer Metabolism. *Trends Endocrinol Metab* 2019;30:25-38.
34. Zhao W, Jia Y, Sun G, et al. Single-cell analysis of gastric signet ring cell carcinoma reveals cytological and immune microenvironment features. *Nat Commun* 2023;14:2985.
35. Chakraborty B, Byemerwa J, Shepherd J, et al. Inhibition of estrogen signaling in myeloid cells increases tumor immunity in melanoma. *J Clin Invest* 2021;131:e151347.
36. Merkin RD, Vanner EA, Romeiser JL, et al. Keratin 17 is overexpressed and predicts poor survival in estrogen receptor-negative/human epidermal growth factor receptor-2-negative breast cancer. *Hum Pathol* 2017;62:23-32.
37. Mockler D, Escobar-Hoyos LF, Akalin A, et al. Keratin 17 Is a Prognostic Biomarker in Endocervical Glandular Neoplasia. *Am J Clin Pathol* 2017;148:264-73.
38. Zeng Q, Huang Y, Zeng L, et al. IPP5, a novel inhibitor of protein phosphatase 1, suppresses tumor growth and progression of cervical carcinoma cells by inducing G2/M arrest. *Cancer Genet* 2012;205:442-52.
39. Wang X, Liu B, Li N, et al. IPP5, a novel protein inhibitor of protein phosphatase 1, promotes G1/S progression in a Thr-40-dependent manner. *J Biol Chem* 2008;283:12076-84.
40. Dohán O, Carrasco N. Advances in Na(+)/I(-) symporter (NIS) research in the thyroid and beyond. *Mol Cell Endocrinol* 2003;213:59-70.
41. Cazarin J, Dupuy C, Pires de Carvalho D. Redox Homeostasis in Thyroid Cancer: Implications in Na(+)/I(-) Symporter (NIS) Regulation. *Int J Mol Sci* 2022;23:6129.
42. Keenan TE, Guerriero JL, Barroso-Sousa R, et al. Molecular correlates of response to eribulin and pembrolizumab in hormone receptor-positive metastatic breast cancer. *Nat Commun* 2021;12:5563.
43. Bauer AT, Gorzelanny C, Gebhardt C, et al. Interplay between coagulation and inflammation in cancer: Limitations and therapeutic opportunities. *Cancer Treat Rev* 2022;102:102322.
44. Bykov VJN, Eriksson SE, Bianchi J, et al. Targeting mutant p53 for efficient cancer therapy. *Nat Rev Cancer* 2018;18:89-102.
45. Blagih J, Buck MD, Vousden KH. p53, cancer and the immune response. *J Cell Sci* 2020;133:jcs237453.
46. Na D, Chae J, Cho SY, et al. Predictive biomarkers for 5-fluorouracil and oxaliplatin-based chemotherapy in gastric cancers via profiling of patient-derived xenografts. *Nat Commun* 2021;12:4840.
47. Zhao H, Zhang L. MUC16 mutation predicts a favorable clinical outcome and correlates decreased Warburg effect in gastric cancer. *Biochem Biophys Res Commun* 2018;506:780-6.

48. Wang X, Yu X, Krauthammer M, et al. The Association of MUC16 Mutation with Tumor Mutation Burden and Its Prognostic Implications in Cutaneous Melanoma. *Cancer Epidemiol Biomarkers Prev* 2020;29:1792-9.
49. Zhang L, Han X, Shi Y. Association of MUC16 Mutation With Response to Immune Checkpoint Inhibitors in Solid Tumors. *JAMA Netw Open* 2020;3:e2013201.
50. Li X, Pasche B, Zhang W, et al. Association of MUC16 Mutation With Tumor Mutation Load and Outcomes in Patients With Gastric Cancer. *JAMA Oncol* 2018;4:1691-8.

Cite this article as: Li X, Zhou C, Zhu Y, Wang W, Han S, Zou Y, Lian L, Chen K. Developing a prognostic signature: identifying differentially expressed genes in cardia and non-cardia gastric cancer for immunity and therapeutic sensitivity analysis. *J Gastrointest Oncol* 2024;15(4):1446-1463. doi: 10.21037/jgo-24-541