



OPEN

Construction of an immune-related risk score signature for gastric cancer based on multi-omics data

Ying Wang^{1,5}, Wenting Huang^{2,5}, Shanshan Zheng¹, Liming Wang³, Lili Zhang⁴ & Xiaojuan Pei⁴

Early identification of gastric cancer (GC) is associated with a superior survival rate compared to advanced GC. However, the poor specificity and sensitivity of traditional biomarkers suggest the importance of identifying more effective biomarkers. This study aimed to identify novel biomarkers for the prognosis of GC and construct a risk score (RS) signature based on these biomarkers, with validation of its predictive performance. We used multi-omics data from The Cancer Genome Atlas to analyze the significance of differences in each omics data and combined the data using Fisher's method. Hub genes were subsequently subjected to univariate Cox and LASSO regression analyses and used to construct the RS signature. The RS of each patient was calculated, and the patients were divided into two subgroups according to the RS. The RS signature was validated in two independent datasets from the Gene Expression Omnibus and subsequent analyses were subsequently conducted. Five immune-related genes strongly linked to the prognosis of GC patients were obtained, namely CGB5, SLC10A2, THPO, PDGFRB, and APOD. The results revealed significant differences in overall survival between the two subgroups ($p < 0.001$) and indicated the high accuracy of the RS signature. When validated in two independent datasets, the results were consistent with those in the training dataset ($p = 0.003$ and $p = 0.001$). Subsequent analyses revealed that the RS signature is independent and has broad applicability among various GC subtypes. In conclusion, we used multi-omics data to obtain five immune-related genes comprising the RS signature, which can independently and effectively predict the prognosis of GC patients with high accuracy.

Early identification of stomach cancer is associated with a superior survival rate compared to that of advanced gastric cancer (GC)^{1–4}. As traditional biomarkers are not effective at predicting the prognosis of GC patients^{5,6}, novel therapeutic biomarkers are of crucial for improving prognoses.

The use of multi-omics data has the ability to uncover deeper insights⁷. Several recent studies have demonstrated that multi-omics data can be used to identify novel biomarkers for early diagnosis and treatment from new perspectives^{8–11} and that these biomarkers can improve the prognosis of cancer patients. Fisher's method is considered a classical method that can integrate information from multiple omics into one feature^{12,13}. A multi-omics study based on the TCGA database was performed. We utilized RNA sequencing (RNA-Seq) expression, copy number variation (CNV) and DNA methylation data, and tests for each OMC revealed distinct characteristics of the marker genotype¹⁴. The Fisher test was used to combine the p-values of each OMC, from which we combined the information from multiple views to screen for GC-associated genetic markers.

Tumor-infiltrating immune cells have been demonstrated to be linked to promoting and preventing cancer progression in distinct cancer types^{15,16}. Immune checkpoints are a class of components that are upregulated in the TME and inhibit antitumor T-cell responses¹⁷. Classifying the influence of genetic markers on the prognosis

¹Department of Oncology, National Cancer Center/National Clinical Research Center for Cancer/Cancer Hospital and Shenzhen Hospital, Chinese Academy of Medical Science and Peking Union Medical College, Shenzhen, Guangdong, China. ²Department of Pathology, National Cancer Center/National Clinical Research Center for Cancer/Cancer Hospital and Shenzhen Hospital, Chinese Academy of Medical Science and Peking Union Medical College, Shenzhen, Guangdong, China. ³Department of Gastrointestinal Surgery, National Cancer Center/National Clinical Research Center for Cancer/Cancer Hospital and Shenzhen Hospital, Chinese Academy of Medical Science and Peking Union Medical College, Shenzhen, Guangdong, China. ⁴Department of Pathology, Shenzhen Hospital, Southern Medical University, Shenzhen, Guangdong, China. ⁵These authors contributed equally: Ying Wang and Wenting Huang. ✉email: Drwangsz@yeah.net; peixiaojuan415@163.com

and diagnosis of tumor-infiltrating immune cells and immunological checkpoints might improve the treatment and survival of GC patients.

In this study, five immune-related genes related to GC patient prognosis that can serve as prospective biomarkers were found from multi-omics data in the TCGA database and utilized to construct a risk score (RS) signature for each patient and establish an RS signature. The results suggested that the RS signature created in this work accurately predicts the prognostic outcome of GC patients with greater predictive power than standard clinical indicators when validated using the Gene Expression Omnibus (GEO) database.

Materials and methods

Data preparation

This study followed the workflow shown in Fig. 1. The University of California Santa Cruz (UCSC) database (<https://xenabrowser.net/datapages/>) was used to obtain gene expression RNA-Seq data (n = 442), DNA methylation 27k array data (n = 142), gene-level CNV data (n = 413), and clinical information. GSE62254 (n = 300), GSE26942 (n = 126), and GSE13861 (n = 65) RNA-Seq data, as well as corresponding patient clinical information, were downloaded from the GEO (<http://www.ncbi.nlm.nih.gov/geo/>). The present study used TCGA datasets for training. The GSE62254 dataset served as the first independent validation dataset. Due to the limited sample size, GSE13861 was combined with GSE26942 to create a second independent validation dataset (n = 191).

Hub gene screening

All the data analyses were performed in R (version 4.1.2). First, RNA-Seq data, methylation 27k array data, and CNV data from the TCGA were filtered to retain only genes present in all three datasets (11,246 genes in total). The DESeq2¹⁸ package (version 1.34.0) was used to calculate *p* values (p_{RNAs}) for the significance of differential expression between tumor samples (n = 389) and normal tissue samples (n = 33) for each gene in the training dataset. Similarly, *p*-values (p_{methy}) were calculated for each gene in methylation 27 k array data between tumor

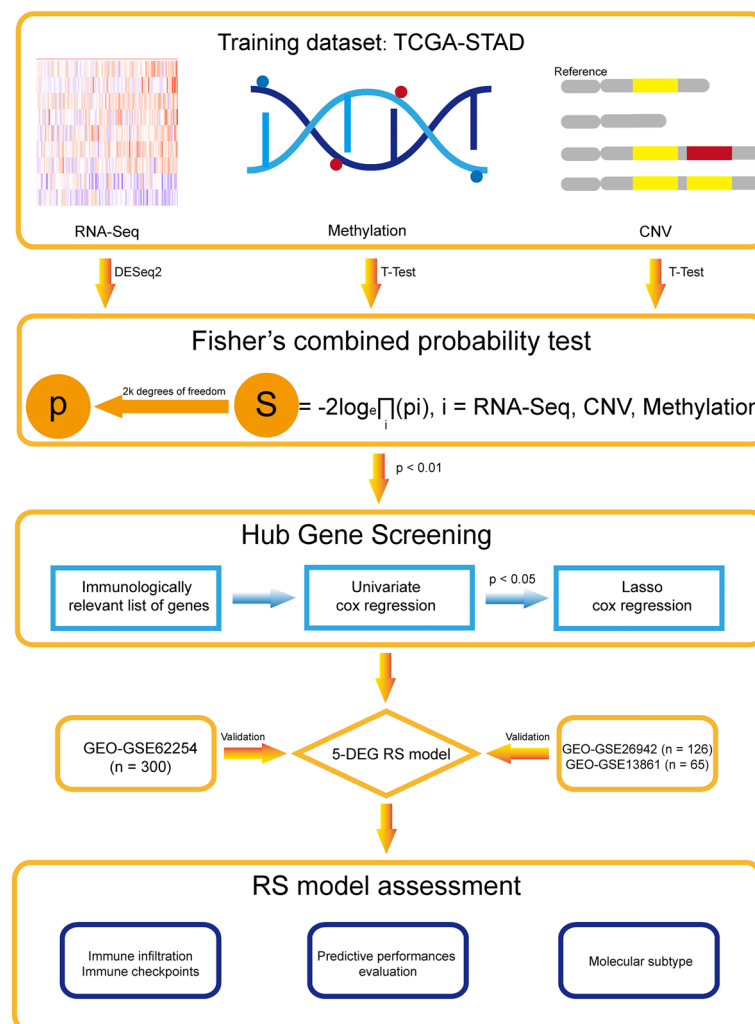


Figure 1. Workflow of this study.

samples and normal tissue samples using the Student's *t* test¹⁹. On the basis of CNV data, patients in the training dataset were divided into copy number variation and nonvariation groups. The *p*-values (p_{CNVs}) of the DEGs between these two groups in the RNA-Seq dataset were subsequently calculated using Student's *t* test. After obtaining the *p* value for each gene in the three omics analyses, we calculated the *S* statistic using the Fisher's method¹³(1); three independent *p*-values with 2k degrees of freedom were then used to transform the *S* statistic into the null hypothesis *p* value ($p_{combined}$). The $p_{combined}$ value was considered to represent the significance of the gene for the prognostic profile of GC patients according multi-omics to data. Genes with a $p_{combined}$ value less than 0.010 were considered to be significant.

$$S = -2\log_e \prod_i p(i) \quad (1)$$

In Eq. (1), *i* represents the p_{RNA} , p_{methy} , or p_{CNV} of each gene. Using the IMMPORT database (<http://www.immport.org>), we downloaded a list of immune-related genes (IRGs) and retained only significant IRGs. The remaining genes in the RNA-Seq and methylation data combined with clinical information were subsequently analyzed via univariate Cox regression analysis. Genes with a *p* value less than 0.050 in both sets of results were considered significant. Further screening was then performed using LASSO regression analysis, the results of which revealed candidate genes that correlated strongly with the prognosis of GC patients. The pheatmap²⁰ package (version 1.0.12) was used to plot heatmaps showing differences in candidate gene expression between tumor and normal tissue samples.

Functional enrichment analysis

Gene Ontology (GO) and Kyoto Encyclopedia of Genes and Genomes (KEGG) enrichment analyses were carried out on the genes screened out from multi-omics data to explore molecular mechanisms by the “clusterProfiler”²¹ R package (version 4.2.2) and the “org.Hs.eg.db”²² R package (version 3.14.0). *p* values were adjusted using the *fdr* method to control the false discovery rate (FDR).

RS signature development and validation

First, the risk score of each patient in the training dataset was calculated using Eq. (2).

$$Risk\ score = \sum \beta_{RNA} * Exp_{RNA} \quad (2)$$

β_{RNA} represents the coefficient of the candidate gene according to the univariate Cox regression analysis of the RNA-Seq data, and Exp_{RNA} represents expression of the candidate gene in RNA-Seq data. Patients were divided into high risk and low risk subgroups based on the median score.

We first observed the distribution of each clinical indicator in the different subgroups and assessed whether the distribution was significantly different using the chi-square test or Fisher's exact test. Furthermore, the survival (version 3.3.1)²³ and survminer (version 0.4.9)²⁴ packages were used to perform survival analysis on the high and low risk subgroups of the training dataset to analyze differences in overall survival (OS) between the two subgroups; the results of survival analysis were generated using Kaplan–Meier (KM) survival curves. To assess the accuracy of the RS in predicting the prognosis of GC patients, the survivalROC package²⁵ (version 1.0.3) was used to plot receiver operating characteristic (ROC) curves, and the area under the ROC curve (AUC) was used to determine the predictive accuracy of the RS signature. A similar approach was used for survival and ROC analyses for the two independent validation datasets from GEO cohort divided into high and low risk subgroups.

RS signature assessment

To explore the prognostic predictive capability of the RS signature in patients with GC in different subgroups, we performed survival and ROC analyses using the same approach as above for two subgroups of patients with stages I & II and III & IV in the TCGA training dataset. In 2014, TCGA classified gastric cancer patients into four molecular subtypes: Epstein–Barr virus (EBV) positive, microsatellite unstable (MSI), genomically stable (GS), and chromosomal unstable (CIN)²⁶. For the training dataset, violin plots were generated to illustrate differences in distributions of RSs among the four subtypes, and survival analysis was performed for each subtype. Afterward, univariate and multivariate Cox regression analyses were conducted to verify the superior predictive power of the RS signature over traditional clinical prognostic indicators. To assess the independence and predictive power of the RS signature, it was used as a prognostic indicator for GC patients in the training dataset in univariate and multivariate Cox regression analyses with other clinical indicators (age, sex, and stage). Moreover, a nomogram was drawn to better represent the predictive power of the RS and other clinical indicators.

Immune characteristics

To explore differences in immune characteristics between patients in the high and low risk subgroups, the CIBERSORT²⁷ algorithm and the LM22 gene signature were used to analyze differences in immune infiltration between patients in the high and low risk subgroups in the TCGA training dataset and the GSE62254 independent validation dataset. In the next step, we analyzed the differences in expression of 33 immune checkpoint molecules (Supplementary file) in the TCGA training cohort to investigate differences between immune mechanisms in high and low risk subgroups.

Molecular docking

By combining molecular docking analyses, we aim to comprehensively explore ligand-target interactions, ultimately advancing our understanding of molecular mechanisms and informing the development of novel therapeutic agents. Drugs that differed significantly between the high and low risk groups were first identified by calculating half maximal inhibitory concentration (IC50) values, after which the molecular operating environment (MOE) was used to predict interactions of the five constituent modeled genes with these drugs.

Results

Identification of prognostic genes in the TCGA dataset with multi-omics data

Based on $p_{combined}$ calculated using multi-omics data from the training dataset, 7787 genes were screened out as $p_{combined} < 0.010$, 798 of which are IRGs (Fig. 2A). Among these IRGs, 16 genes associated with GC prognosis were identified through univariate Cox analysis of RNA-Seq and DNA methylation data from the training dataset. Five genes were subsequently selected by LASSO regression; CGB5, THPO, and PDGFRB were upregulated in the tumor tissue, while SLC10A2 and APOD were downregulated (Fig. 2B). These 5 genes correlated positively with poor prognosis in GC patients (Table 1).

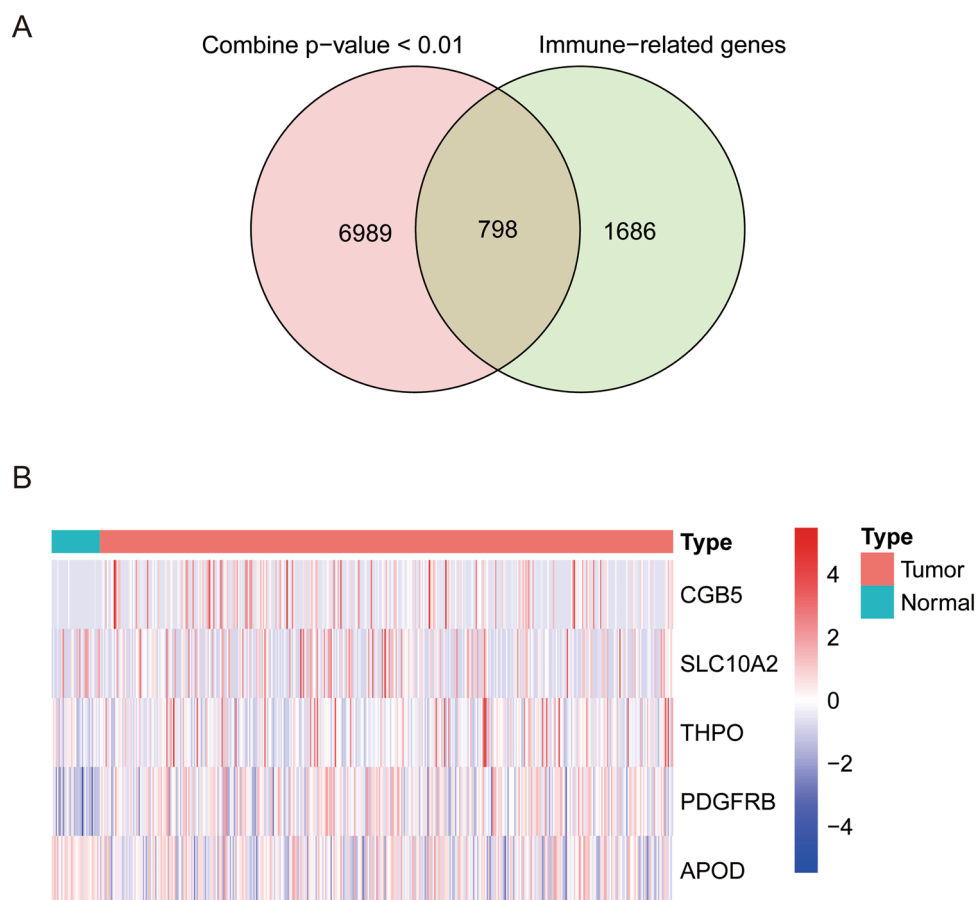


Figure 2. Identification of 5 prognostic genes. (A) Venn diagram showing 798 overlapping genes between genes screened out by multi-omics data and immune-related genes. (B) Heatmap showing differences in expression of the 5 prognostic genes between tumor and normal tissues.

Gene symbol	Coefficient	Hazard ratio	p-value
CGB5	0.157	1.170	<0.001
SLC10A2	0.077	1.080	0.025
THPO	0.112	1.119	0.002
PDGFRB	0.199	1.220	0.005
APOD	0.129	1.138	<0.001

Table 1. Univariate Cox analysis of associations between five hub genes and OS in the TCGA dataset.

Construction of an RS signature using the TCGA dataset

Through Cox and LASSO regression analyses, the 5 hub genes (Table 1) that most contributed to the OS of GC patients were screened out and used to construct an RS signature with the following formula (Formula 2): Risk score = $(0.157 \times \text{Exp}_{\text{CGB5}}) + (0.077 \times \text{Exp}_{\text{SLC10A2}}) + (0.112 \times \text{Exp}_{\text{THPO}}) + (0.199 \times \text{Exp}_{\text{PDGFRB}}) + (0.129 \times \text{Exp}_{\text{APOD}})$. The risk score of each patient was calculated, and patients in the TCGA training dataset were divided into two subgroups: high risk (n = 211) and low risk (n = 211), using the median score as the cutoff value. As shown in Fig. 3A–C, patients with high risk scores had higher mortality rates and expression of the 5 immune-related genes. KM survival analysis was subsequently performed to evaluate the effect of the RS signature on the OS of patients with GC in the training dataset (Fig. 3D). The results indicated that patients in the high risk subgroup had significantly poorer prognosis than did those in the low risk subgroup ($p < 0.001$). Time-dependent ROC analysis was further performed to assess the predictive performance of the RS signature. As presented in Fig. 3E,

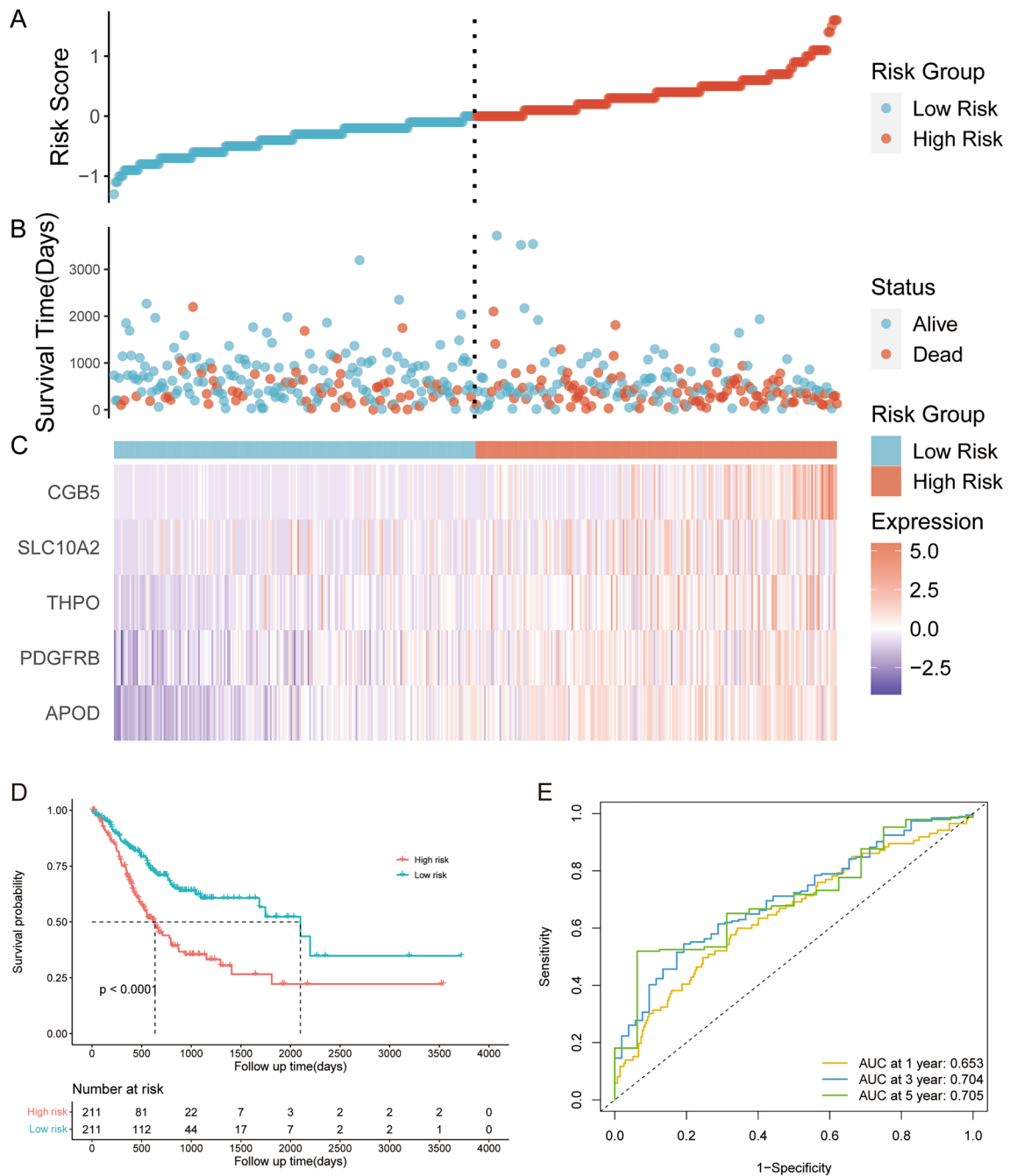


Figure 3. The prognostic value of the 5-genes RS signature in the training dataset. (A) The distribution of risk scores in the training dataset. (B) The scatter plot of the 5-genes RS signature distribution for patient survival status in the training dataset. (C) Expression of the 5 prognostic genes in patients with different RSs in the training dataset. (D) Survival analysis of OS in patients with different RSs in the training dataset.

the AUC reached 0.653 at 1 year, 0.704 at 3 years, and 0.704 at 5 years, demonstrating the prognostic value of the RS signature.

Table 2 shows the distributions of clinical characteristics among patients in the high risk and low risk subgroups. The distributions of patients according to American Joint Committee on Cancer (AJCC) TNM stage and tumor status were significantly different between the high and low risk subgroups.

Validation of the RS signature in GEO datasets

We used two independent validation datasets from the GEO database to assess the prognostic significance of this novel RS signature in patients with GC. With the risk score calculated by the Formula (2) mentioned above, the patients with GC in GSE62254 (validation dataset 1; $n = 300$) were divided into high risk ($n = 150$) and low risk ($n = 150$) subgroups according to the median risk score. Due to the limited sample size, we combined the GSE13861 and GSE26942 datasets as validation dataset 2 ($n = 191$), and the patients were also divided into high risk ($n = 95$) and low risk ($n = 96$) subgroups using the same methods mentioned before. Similar to the results found for the training datasets, the patients in the high risk subgroup tended to die earlier and have a significantly shorter survival time than did those in low risk subgroup in validation datasets 1 ($p = 0.003$, Fig. 4A) and 2 ($p = 0.001$, Fig. 4B). As shown in Fig. 4C,D, the AUC for validation datasets 1 and 2 reached 0.609 and 0.605 at 1 year, 0.642 and 0.652 at 3 years, and 0.630 and 0.695 at 5 years, respectively.

Prognostic prediction in patients with different tumor stages

To further investigate the ability of the RS signature to predict OS, we applied KM survival analysis to OS in the training dataset based on patients with AJCC TNM stages I and II or III and IV. The RS signature showed an excellent predictive value for OS in patients with stage I or II disease ($p = 0.022$, Fig. 5A) or stage III and IV disease ($p < 0.001$, Fig. 5B). The AUC for the patients in stages I and II reached 0.679 at 1 year, 0.676 at 3 years,

Clinical indicators ^a	High risk (n = 211)	Low risk (n = 211)	p value for chi-square tests ^b
Age			0.254
≥ 65 years	113 (53.6)	125 (59.2)	
< 65 years	97 (46.0)	84 (39.8)	
Missing	1 (0.5)	2 (0.9)	
Sex			1.000
Male	138 (65.4)	137 (64.9)	
Female	73 (34.6)	74 (35.1)	
AJCC TNM stage			0.011
Stage I	18 (8.5)	39 (18.5)	
Stage II	67 (31.8)	70 (33.2)	
Stage III	91 (43.1)	82 (38.9)	
Stage IV	26 (12.3)	15 (7.1)	
Stage x or missing	9 (4.3)	5 (2.4)	
T			0.011
T1	5 (2.4)	17 (8.1)	
T2	42 (19.9)	52 (24.6)	
T3	95 (45.0)	95 (45.0)	
T4	66 (31.3)	46 (21.8)	
Tx or missing	3 (1.4)	1 (0.5)	
N			0.132
N0	52 (24.6)	73 (34.6)	
N1	62 (29.4)	60 (28.4)	
N2	43 (20.4)	43 (20.4)	
N3	46 (21.8)	33 (15.6)	
Nx or missing	8 (3.8)	2 (0.9)	
M			0.138 ^c
M0	182 (86.3)	195 (92.4)	
M1	17 (8.1)	9 (4.3)	
Mx or missing	12 (5.7)	7 (3.3)	

Table 2. Clinical characteristics of the high risk and low risk groups. ^aData are presented as the number (%), AJCC: American Joint Committee on Cancer, T: tumor status, N: regional lymph node status, M: metastasis status. ^bData containing missing values were not included in the statistical analysis. ^cChi-square test with Yates' continuity correction.

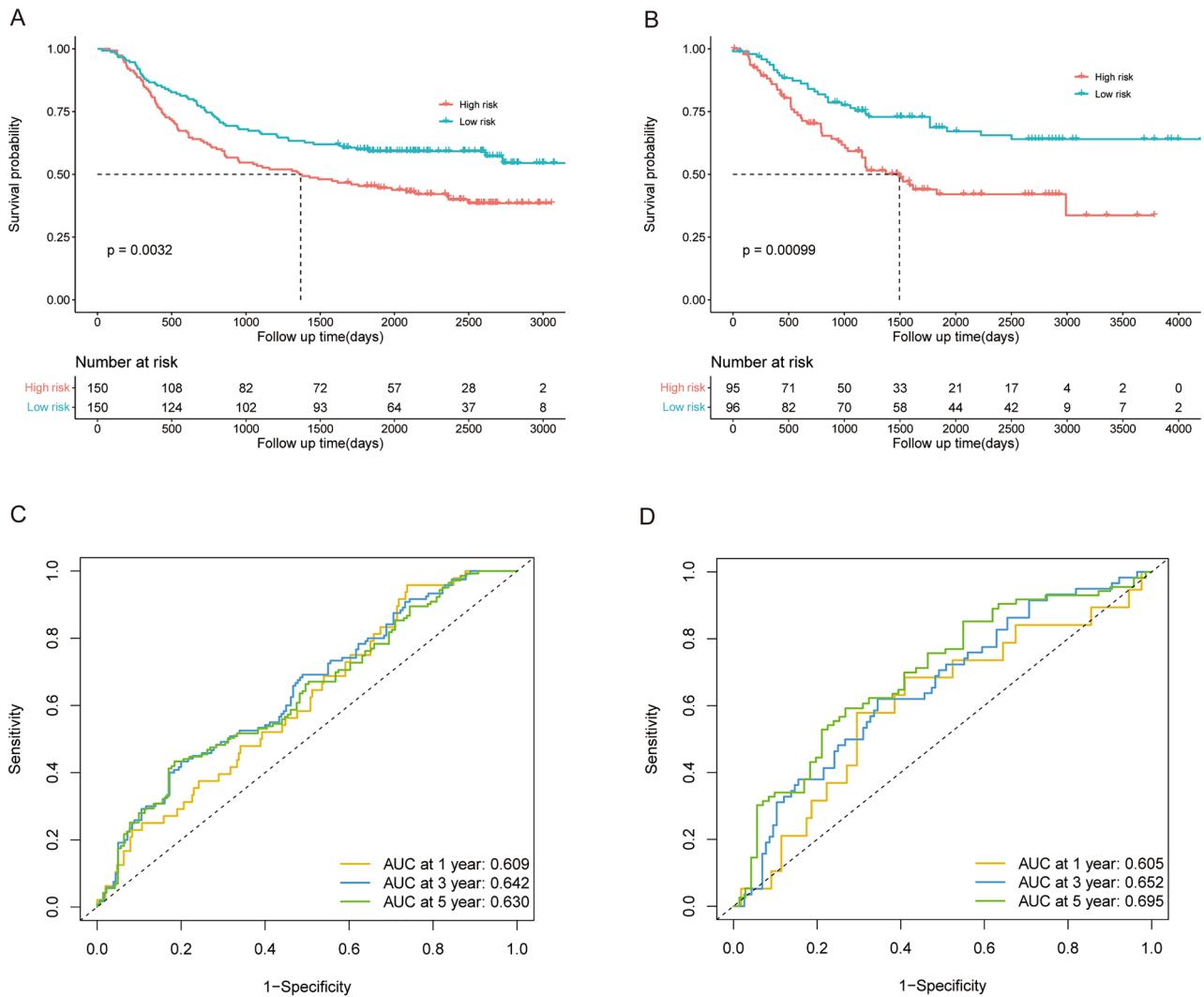


Figure 4. The prognostic value of the 5-genes RS signature in GEO validation datasets 1 (GSE62254) and 2 (GSE13861 and GSE26942). (A,B) Survival analysis of OS in patients with different RSs in the validation datasets 1 and 2. (C,D) ROC analysis of patients in validation datasets 1 and 2.

and 0.618 at 5 years (Fig. 5C), and it performed better in stage III and IV patients, with AUCs reaching 0.642, 0.696, and 0.733 at 1, 3, and 5 years, respectively (Fig. 5D).

Independent prognostic value of the risk score

We explored whether the risk score is an independent prognostic factor. In the training dataset, univariate Cox regression analyses showed that the risk score had a significant relationship with OS (hazard ratio (HR) = 2.114, 95% CI 1.672–2.673, $p < 0.001$; Fig. 6A) and a stronger predictive ability than other classical prognostic predictors, including age and the American Joint Committee on Cancer (AJCC) TNM stage. In multivariate Cox regression, risk score, age, and the AJCC TNM stage were evaluated for independent predictive capacity. The findings are shown in Fig. 6B. In terms of predictive ability, the risk score (HR = 2.084, 95% CI = 1.626–2.672, $p < 0.001$) was superior than age (HR = 1.033, 95% CI = 1.016–1.050, $p < 0.001$) and American Joint Committee on Cancer (AJCC) TNM stage (HR = 1.676, 95% CI = 1.356–2.073, $p < 0.001$). A nomogram containing the AJCC TNM stage, sex, age, and RS is presented in Fig. 6C.

Favorable prognostic value of the risk score in different GC subtypes

GC can be divided into 4 different molecular subtypes, CIN, EBV, GS, and MSI²⁶. Figure 7A shows the risk score distribution in patients with different GC subtypes, which revealed higher RSs for GS and CIN, which are considered to have poorer prognoses than EBV and MSI^{28,29}. Since each molecular subtype involves a different mutation, methylation, and immune signature³⁰, we applied KM survival analysis of OS in the 4 different subtypes of patients in training datasets to further evaluate the prognostic value of the RS signature in GC patients with different subtypes. Figure 7B,C shows that the RS signature had good prognostic value for CIN ($p < 0.001$, $n = 127$) with an AUC reaching 0.690 at 1 year, 0.755 at 3 years, and 0.774 at 5 years. As the CIN subtype is considered to

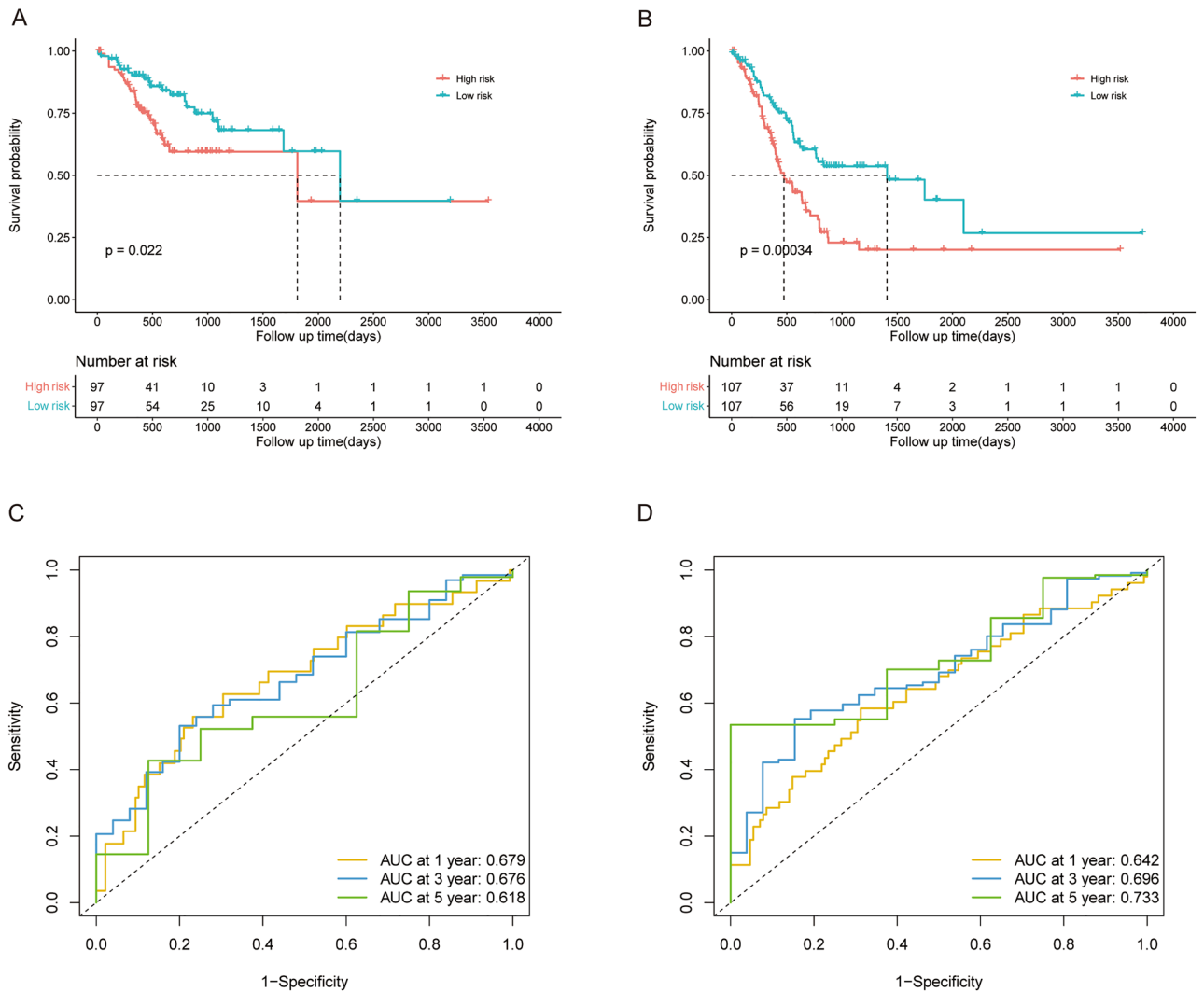


Figure 5. The predictive performance of the 5-genes RS signature in patients with different disease stages in the training cohort. **(A)** Survival analysis of OS in patients with different RSs and TNM stage I or II disease in the training dataset. **(B)** Survival analysis of OS in patients with different RSs and TNM stage III or IV disease in the training dataset. **(C)** ROC analysis of the RS in patients with TNM stage I or II disease in the training dataset. **(D)** ROC analysis of the RS in patients with TNM stage III or IV disease in the training dataset.

be related to poor prognosis in GC²⁹, these results revealed the favorable prognostic value of the RS signature in patients with different GC subtypes. As shown in Fig. 7D–F, the RS signature had limited prognostic value for EBV ($p = 0.180$, $n = 25$), MSI ($p = 0.560$, $n = 52$), and GS ($p = 0.082$, $n = 51$). While the maximum sample size was only 52, the possibility could not be excluded that the sample size restricted the predictive ability of the RS signature.

Functional enrichment analysis of genes screened out by multi-omics data

To clarify biological process (BP), cellular compartment (CC), molecular function (MF) terms and pathways correlating with the genes screened out by multi-omics data in the training dataset, enrichment analysis of GO terms and KEGG pathways was performed. According to GO enrichment analysis (Fig. 8A), the most enriched (sorted by p values) BP was muscle contraction, the most common CC was receptor ligand activity, and the most common MF was collagen-containing extracellular matrix. The top KEGG pathways (sorted by p value) related to the genes screened out by multi-omics data were the cAMP signaling pathway and calcium signaling pathway (Fig. 8B). These findings may indicate molecular changes in GC patients according to multi-omics data.

Immune characteristics of patients with different risk scores

We also compared immune characteristics between high risk subgroup and low risk subgroup, and the results are shown in Fig. 9. As shown in Fig. 9A, there were multiple immune checkpoint differences between the two high and low risk patient groups in the training cohort, but only the number of resting dendritic cells was significantly different between the two groups in immune infiltration analysis (Fig. 9C). However, in patients in validation dataset 1 from the GEO database, expression of several immune checkpoint genes and the proportions of several

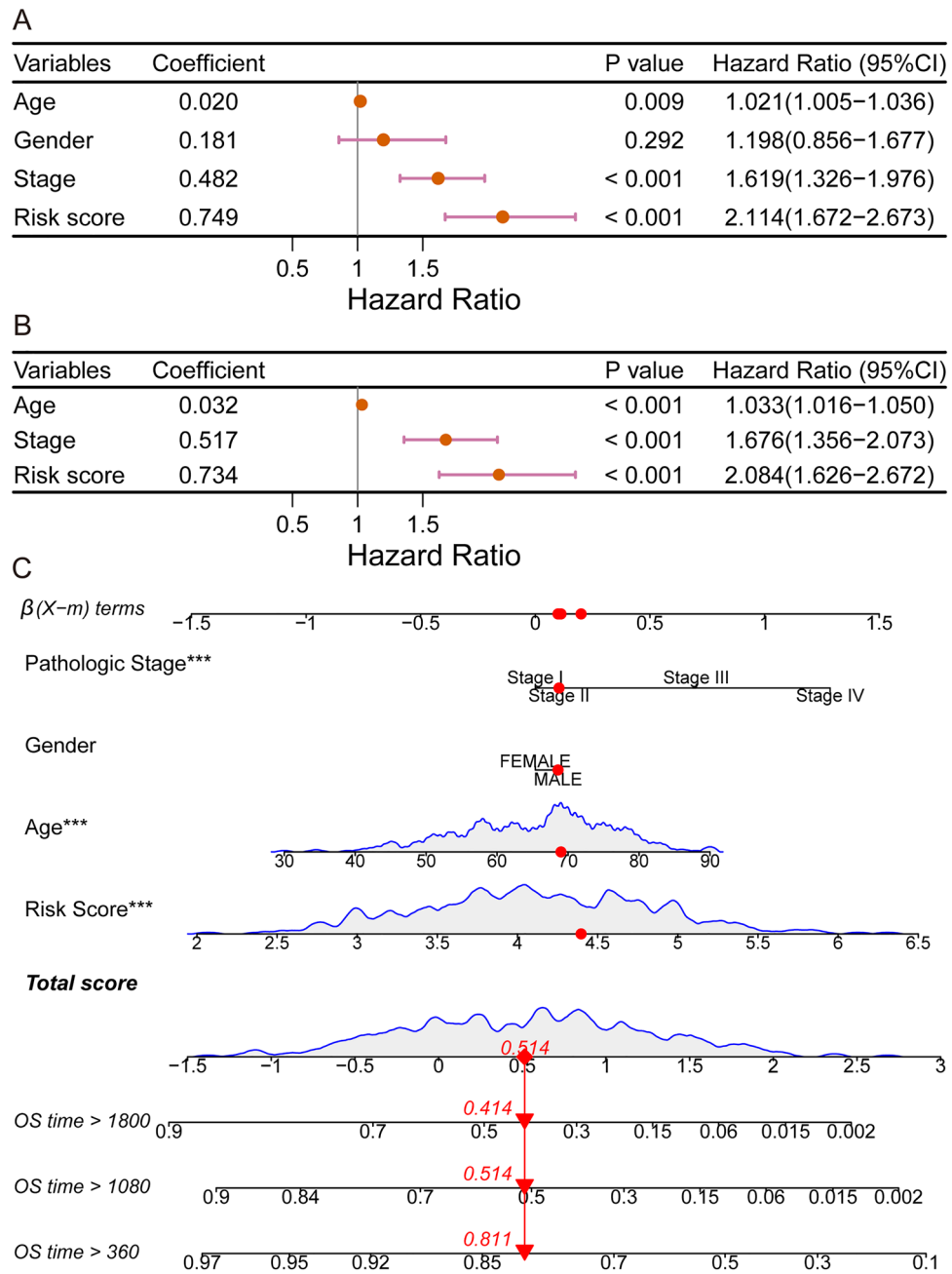


Figure 6. Comparison of the ability of the RS with other clinical indicators to predict the prognosis of GC patients (A,B) Results of univariate and multivariate Cox regression analyses in the training dataset. (C) Nomogram showing the performance of the AJCC TNM stage, sex, age, and RS in predicting the prognosis of GC patients according to multivariate Cox regression analysis.

immune cells were altered (Fig. 9B and D). In both the training dataset and the validation dataset 1, expression of BTLA, CD200, CD28, CD86, HAVCR2, LAIR1, TNFRSF4, and TNFSF4 was upregulated in high risk patients, which indicated the association between the risk score and tumor immunity.

Molecular docking

Figure 10 shows that for the five constituent model genes, binding of the drug docetaxel differed significantly between patients in the high- and low-risk groups in the training dataset. CGB5 forms a side chain with Thr-C269 and Arg-B94. SLC10A2 forms a backbone with Ser108, Ala107, and Thr106. THPOs form backbones with Phe-128, Leu-129 and Arg-136. PDGFRB forms a backbone with Glu-A664 and Ser-A660. APODs form backbones with Gln-98 and solvent residues with Phe-96.

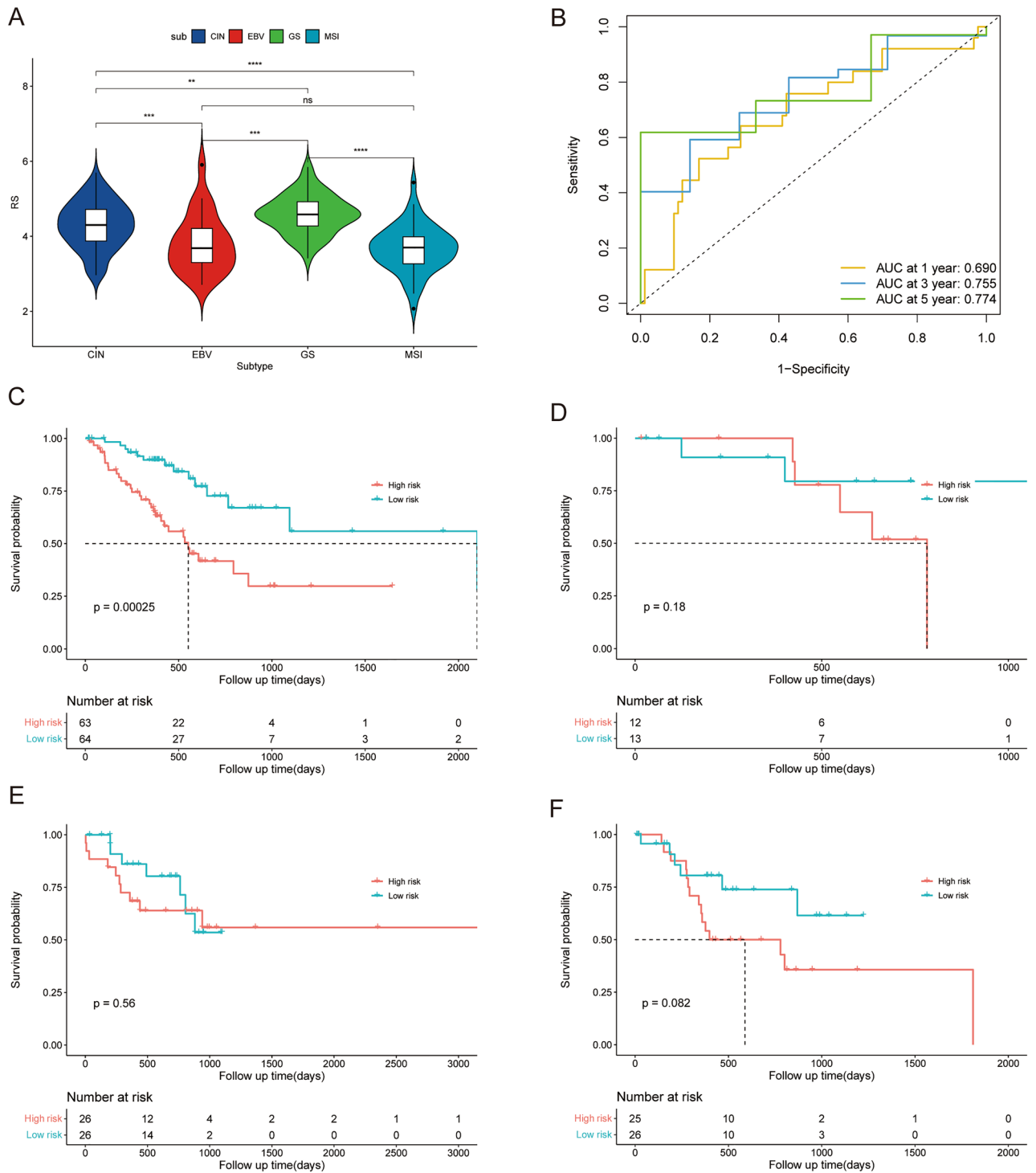


Figure 7. The predictive performance of the 5-gene signature in GC patients with different subtypes. **(A)** The distribution of the RS in patients with different GC subtypes in the training dataset. **(B)** ROC analysis of the RS in patients with the CIN subtype of GC in the training dataset. **(C–F)** Survival analysis of OS between patients with different RSs and those with four subtypes of GC in the training dataset. ns: $p > 0.05$, *: $p \leq 0.05$, **: $p \leq 0.01$, ***: $p \leq 0.001$, ****: $p \leq 0.0001$.

Discussion

As one of the most prevalent cancers in the world, early detection of GC is problematic, and most patients are diagnosed at an advanced stage; even if they receive treatment, most patients experience recurrence or metastasis, resulting in poor prognosis and a 5-year survival rate of less than 30%³¹. Therefore, a signature that can accurately predict the prognosis of GC patient needs to be developed. Extensive study of multiple levels of biomolecules utilizing multi-omics is advantageous for exploring relationships among biological processes and is beneficial

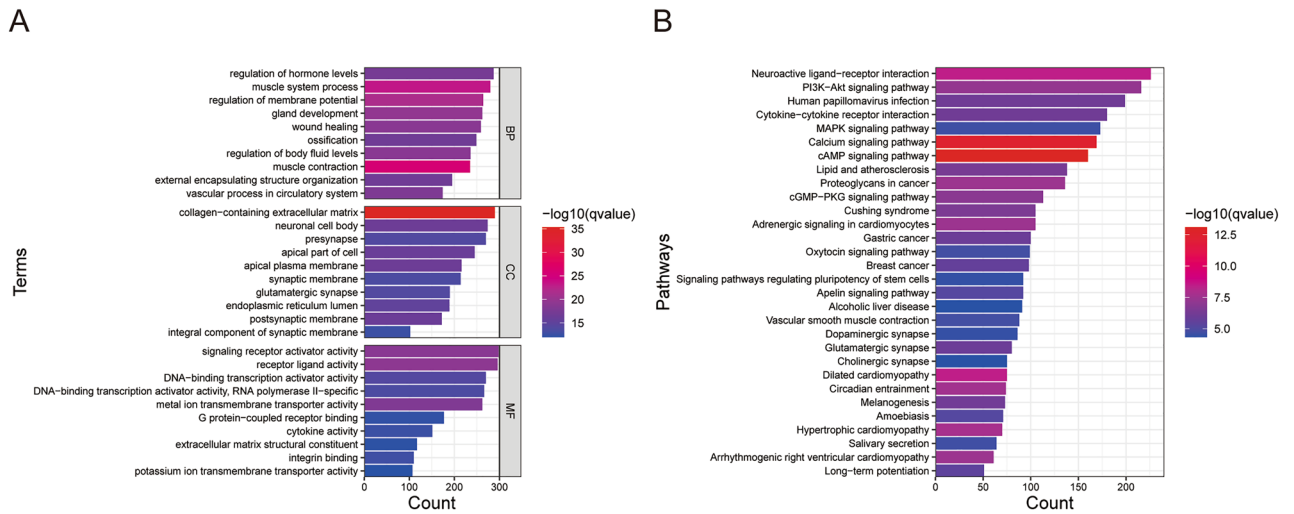


Figure 8. Functional enrichment analysis of genes screened out from multi-omics data from the training dataset in GC patients. (A) The results of GO analysis. (B) The results of KEGG enrichment analysis.

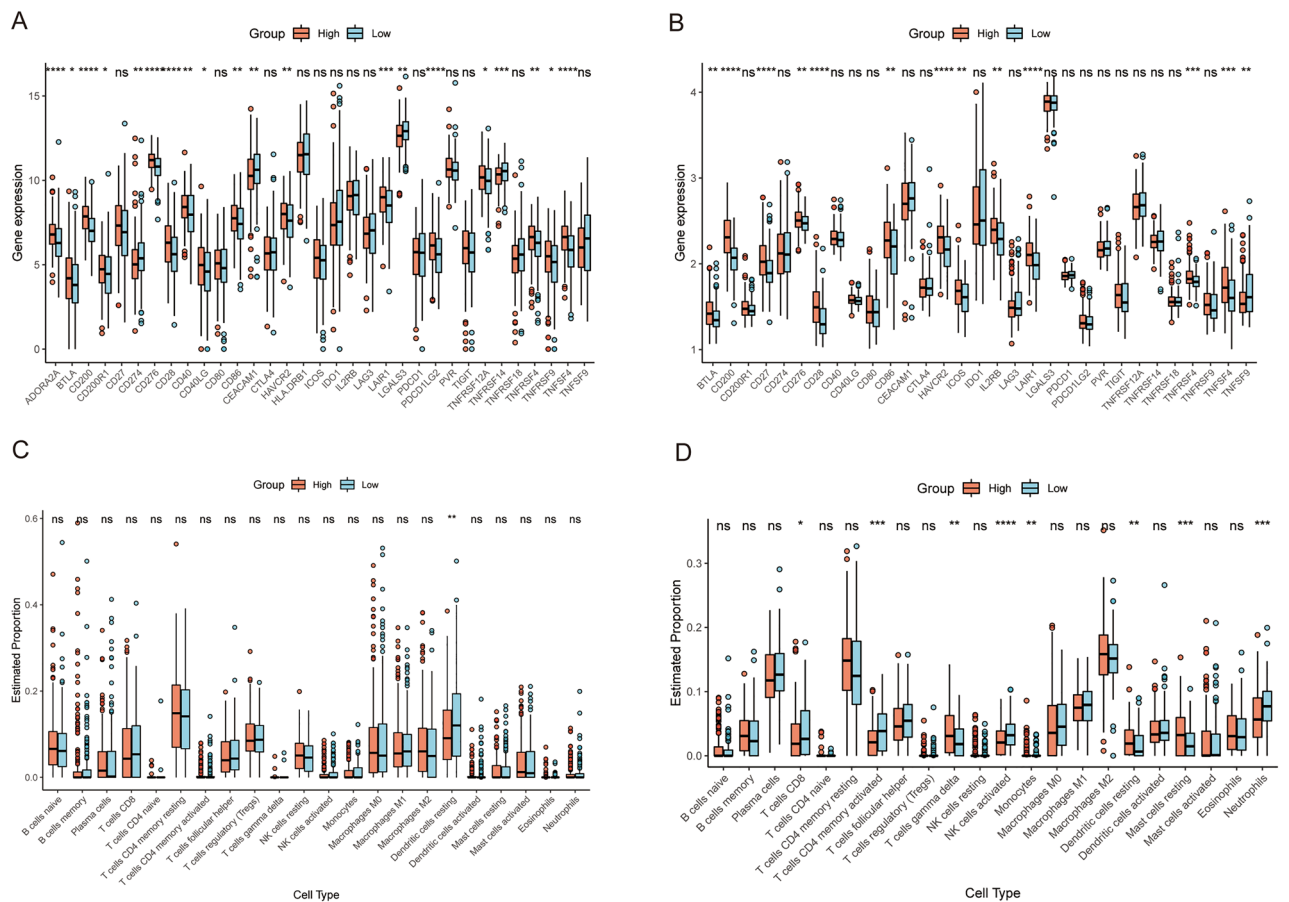


Figure 9. The immune characteristics of the 5-gene RS signature in the training dataset and GEO validation dataset 1 (GSE62254). (A,B) Expression of immune checkpoint genes among patients with different RSs in the training dataset and validation dataset 1. (C,D) Immune cell infiltration among patients with different RSs in the training dataset and validation dataset 1. ns: $p > 0.05$, *: $p \leq 0.05$, **: $p \leq 0.01$, ***: $p \leq 0.001$, ****: $p \leq 0.0001$.

for determining the underlying mechanism in GC. The characteristics of single histology are insufficient for describing complex signaling pathways in organisms because the nature of life activities involves interaction of complex signaling pathways involving multiple molecules. Indeed, analysis of single-level molecules often omits essential information on physiological processes. In addition, molecules interact with each other at multiple levels in terms of the pathways and processes occurring in GC, which can increase the accuracy of data mining³².

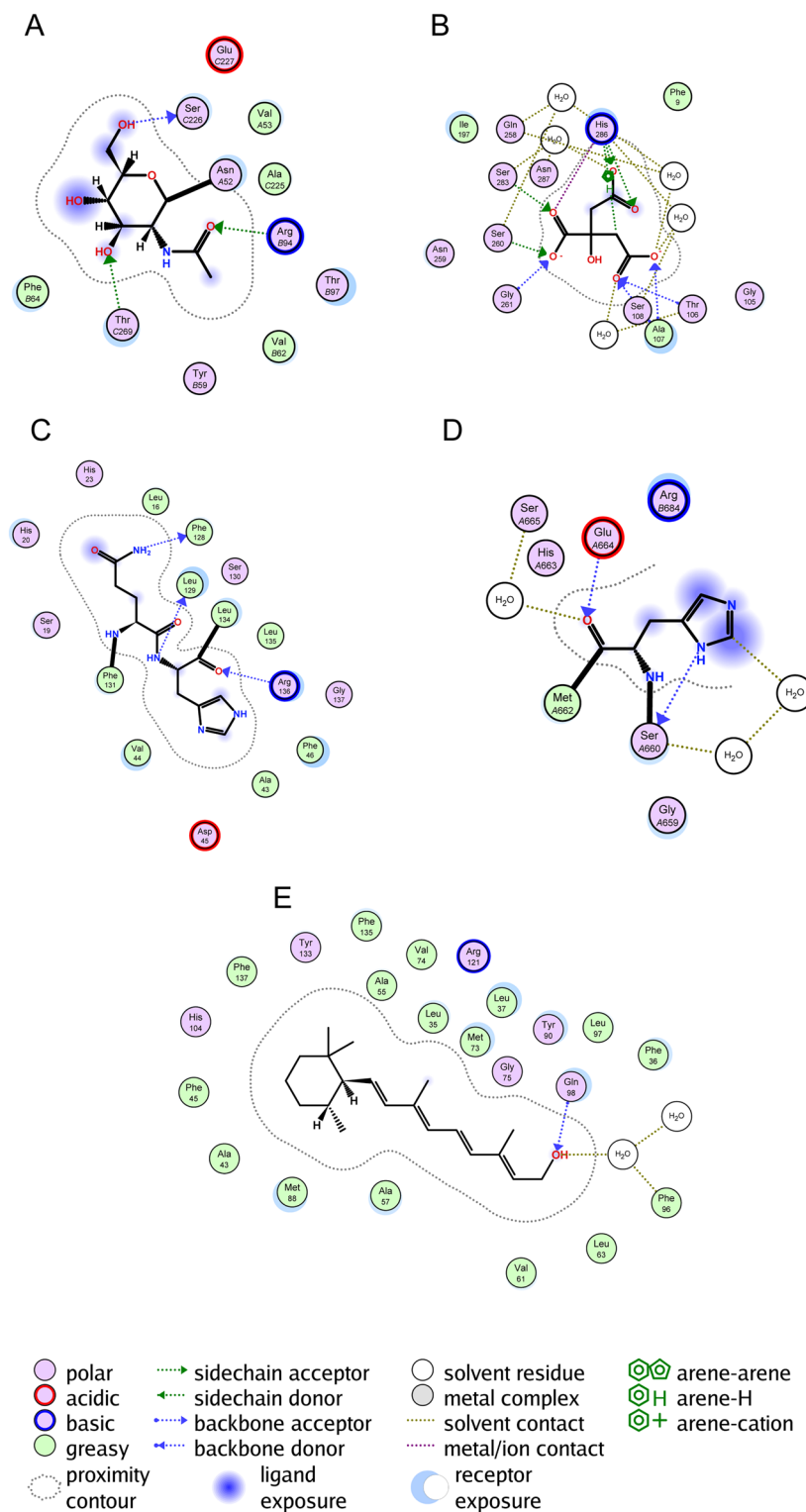


Figure 10. Molecular docking of docetaxel with (A) CGB5, (B) SLC10A2, (C) THPO, (D) PDGFRB, and (E) APOD.

In this study, we used TCGA-STAD gene expression RNA-Seq data, DNA methylation 27k array data, and gene-level CNV data, and integrated the significance of each gene using Fisher's method. Five genes strongly associated with the prognosis of GC patients were screened using univariate Cox regression analysis and LASSO regression analysis to construct the RS signature. The predictive power of the RS signature was subsequently

validated using survival analysis and ROC curves analysis in a training dataset and two independent validation datasets. The results showed that the RS signature was effective at predicting prognosis of patients with GC. Patients classified by the RS signature into high risk subgroups in all three datasets had significantly worse survival probabilities than did those in the low risk subgroup (Figs. 3D,E, and 4). The following univariate and multivariate Cox regression analyses also showed that the RS signature correlated independently and significantly with GC patient prognosis (Fig. 6A,B). Among the four molecular subtypes of GC, patients with the EBV subtype had the lowest risk score, while patients with the GS subtype had the highest risk score (Fig. 7A). This conclusion is also consistent with previous research²⁸, demonstrating that patients with the EBV subtype had the best prognosis and patients with the GS subtype the worst prognosis among the four molecular subtypes of GC. We then performed survival analysis and plotted ROC curves in the training dataset for patients with different disease stages and four MSs to verify the broad applicability of the RS signature. The results showed that the RS signature still had good predictive power for patients with different disease stages (Fig. 5A–D) and CIN subtypes (Fig. 7B,C) (with non-excessive sample size), demonstrating that the RS signature can effectively predict prognosis in a wide range of GC patient populations.

According to the results of functional enrichment analysis, the most enriched MF was the collagen-containing extracellular matrix (Fig. 8A). Collagens in the extracellular matrix act as ligands for immune inhibitory receptors³³. One such receptor is LAIR1, which was more highly expressed in the high-risk subgroup than in the low-risk subgroup according to immune checkpoint analysis. LAIR1 signaling results in T-cell exhaustion and suppression and inhibition of natural killer, monocyte, and dendritic cell activation and function^{34–36}, which reflects the intense immunosuppression in the high-risk subgroup and the predictive power of the RS signature for immune characteristics from another aspect. GO enrichment analysis also revealed enrichment of the regulation of membrane potential (Fig. 8A). Membrane potential can modulate critical cellular activities, which may impact tumor cell proliferation, migration, and differentiation^{37,38}. Changes in membrane potential promote cell cycle checkpoint transition and are likely to trigger intracellular signaling messengers such as Ca²⁺ to drive sustained proliferation³⁷. Moreover, the calcium signaling pathway was enriched according to KEGG enrichment analysis (Fig. 8B), which revealed some of the characteristics of the TME in GC patients. Furthermore, several terms or pathways related to signaling pathways, including receptor–ligand activity, signaling receptor activator activity, the cAMP signaling pathway, the PI3K–Akt signaling pathway, and the MAPK signaling pathway, were enriched according to GO and KEGG enrichment analyses. These findings reflect the complex signal transduction and immune regulation in the TME of GC. In summary, the enrichment landscape revealed by multi-omics data reflected several critical features of GC, providing clues for improving the treatment and prognosis of GC patients.

The immune characteristics of patients with different RSs were further examined. Immune checkpoints are several suppressive immune receptors/ligands that act as gatekeepers for the immune response^{39,40}. In this study, we found that expression of 8 immune checkpoint genes, namely, BTLA, CD200, CD28, CD86, HAVCR2, LAIR1, TNFRSF4, and TNFSF4, was significantly increased in the high-risk subgroup in both TCGA and GEO cohorts (Fig. 9A,B).

BTLA is an inhibitory receptor belonging to the CD28 superfamily and a ligand of HVEM¹⁷. By preventing B and T-cell activation and proliferation, BTLA can cause immunosuppression. An increase in expression of BTLA and HVEM is considered to be associated with poor prognosis in GC patients^{17,41}. A crucial costimulatory protein on the surface of T lymphocytes is CD28, which competes with other CD28 family members, such as CTLA-4, for binding to ligands of the B7 family, including CD80 and CD86⁴². In this study, we observed an increase in expression of CD86, a ligand for CD28. However, expression of CTLA-4, a competitive receptor of CD28, was not significantly different between the high- and low-risk subgroups, which may indicate a stronger CD28 costimulatory signal in the high-risk subgroup. CD28 costimulation is thought to enhance metabolic adaptation of tumor-infiltrating lymphocytes to restore metabolism and function in the TME^{43–45}. However, the high-risk subgroup had poorer prognosis, which may reveal that other immune regulatory pathways inhibit the effect of CD28 co-stimulation. Successful checkpoint blockade treatment requires positive CD28 expression and co-stimulation^{46–49}; a stronger co-stimulatory signal in patients with high risk scores may predict the effectiveness of immunotherapy.

There are several other immune checkpoint genes with altered expression. HAVCR2, often called TIM3, is highly expressed within the TME and correlates with suppression of T-cell responses and T-cell exhaustion, suggesting its role in tumor immunity^{17,50,51}. The signal transduction generated by CD200 and its ligand CD200R is thought to regulate T-cell function, but its function in tumors is complex, and there is no consistent conclusion yet. LAIR1 is a kind of collagen domain-binding receptor³⁵, that suppresses lymphocytic activity when binding to collagen, resulting in CD8⁺ T cell exhaustion and tumor immune suppression^{52–54}. TNFRSF4 (OX40) and its ligand TNFSF4 (OX40L) are members of the TNFR/TNF superfamily⁵⁵. Research has shown that there is increased expression of OX40 in GC patients while metastatic GC patients have higher soluble OX40 levels^{56,57}; moreover, upregulated expression of OX40 is associated with better prognosis in such tumors^{58,59}. Therefore, evaluating the relationship between GC prognosis and OX40 or OX40L is difficult. These findings of increased expression of immune checkpoint genes in the TCGA and GEO datasets demonstrated the high performance of the RS signature for risk-based grouping of GC patients in this study; the immune characteristics of the patients were well distinguished, providing information for treatment to achieve better prognosis.

Interestingly, the immune cell infiltration patterns of GC patients in the training and GEO (GSE62254) datasets were quite different (Fig. 9C,D). Among the American population in the TCGA cohort, only the proportion of resting dendritic cells was significantly greater in the low risk subgroup than in the high risk subgroup (Fig. 9C). As in the population from Korea in GSE62254, proportions of CD8⁺ T cells, activated CD4⁺ memory T cells, activated NK cells, and neutrophils were significantly greater in the low risk subgroup while those of gamma delta T cells, monocytes, resting dendritic cells, and resting mast cells were significantly lower in the low risk subgroup (Fig. 9D), revealing a stronger immune response in the low risk subgroup. We examined patient

age in the two datasets to understand this difference. The results showed that the median (lower quartile, upper quartile) age of patients in the TCGA dataset was 67 (58, 74) years and that of patients in the GSE62254 was 64 years (55, 70). A rank sum test showed that the patient age in the TCGA dataset was greater than that in the GSE62254 dataset. We acknowledge that younger individuals usually have stronger immunity, which may partially explain the difference in immune cell infiltration. Studies have reported racial and ethnic differences in the incidence of GC worldwide and in America⁶⁰, which suggests the influence of genetic background on GC and may also be the reason for the different results of immune infiltration analysis in populations from different regions. These results indicated that in patients from Korea, the different risk subgroups distinguished by our RS signature had distinct immune cell infiltration signatures.

The five hub genes that comprise the RS signature have been demonstrated in earlier research to be connected to the development of gastric or other cancers or to significantly impact patient prognosis. Overexpression of CGB5 in ovarian cancer cells results in increased receptor expression, and interaction between the two accelerates tumor growth and the development of ovarian cancer⁶¹. Sequence variants in SLC10A2 were observed to correlate with the risk of colorectal cancer⁶². Overexpression of THPO in gastric adenocarcinoma tumor tissues has been reported, and its high expression leads to poor prognosis⁶³. PDGFRB affects GS metastasis and prognosis, and its co-expression with other genes is associated with reduced patient survival^{64–66}. The prognosis of breast cancer patients is significantly impacted by APOD, which can be utilized as a biomarker^{67–69}. These findings establish the relationship between the five genes that constitute the RS signature and cancer prognosis, and validate the RS signature in this study, which can be used to predict GC patient prognosis effectively. Molecular docking analysis reveals a strong binding affinity between docetaxel and the amino acid residues of PDGFRB and SLC10A2 proteins. The results of prior research findings also indicate that the products of these genes influence the action of the drugs. Inhibition of PDGFRB transcription has been found to be an important factor in docetaxel's effect on breast cancer⁷⁰. The study by Deeken et al. found a correlation between SLC10A2 and docetaxel toxicity, which suggests the possibility that there is an association of this gene with docetaxel therapy, with potential implications for its efficacy⁷¹. Therefore, it suggests a potential therapeutic efficacy of docetaxel against GC. While the remaining genes also exhibit a binding affinity with docetaxel, the underlying mechanisms and precise impact remain contentious, warranting further research.

Conclusion

In conclusion, this study used gene expression RNA-Seq, DNA methylation, and CNV data for gastric cancer patients in the TCGA cohort and Fisher's test in combination with multi-omics data to screen for five immune-related genes with high prognostic relevance for GC patients and to construct an RS signature. The results illustrated that the RS can be used to predict the prognosis of GC patients effectively and is independent of other clinical indicators. The RS signature provides a new diagnostic approach and therapeutic target for GC, which might improve the prognosis of GC patients if validated by further experiments.

Data availability

The data used to support the findings of this study are available at UCSC (<https://xenabrowser.net/datapages/>) and GEO (<http://www.ncbi.nlm.nih.gov/geo/>) databases, accession numbers: GSE62254, GSE26942, and GSE13861.

Received: 4 September 2023; Accepted: 13 January 2024

Published online: 16 January 2024

References

- Smyth, E. C. *et al.* Gastric cancer. *Lancet* **396**(10251), 635–648 (2020).
- Liu, N. *et al.* Identification of novel prognostic biomarkers by integrating multi-omics data in gastric cancer. *BMC Cancer* **21**(1), 460 (2021).
- Wei, J., Wu, N. D. & Liu, B. R. Regional but fatal: Intraperitoneal metastasis in gastric cancer. *World J. Gastroenterol.* **22**(33), 7478–7485 (2016).
- Tsai, M. M. *et al.* Potential prognostic, diagnostic and therapeutic markers for human gastric cancer. *World J. Gastroenterol.* **20**(38), 13791–13803 (2014).
- Shimada, H. *et al.* Clinical significance of serum tumor markers for gastric cancer: A systematic review of literature by the Task Force of the Japanese Gastric Cancer Association. *Gastric Cancer* **17**(1), 26–33 (2014).
- Necula, L. *et al.* Recent advances in gastric cancer early diagnosis. *World J. Gastroenterol.* **25**(17), 2029–2044 (2019).
- Rappoport, N. & Shamir, R. Multi-omic and multi-view clustering algorithms: Review and cancer benchmark. *Nucleic Acids Res.* **46**(20), 10546–10562 (2018).
- Luo, N. *et al.* Prognostic role of M6A-associated immune genes and cluster-related tumor microenvironment analysis: A multi-omics practice in stomach adenocarcinoma. *Front. Cell Dev. Biol.* **10**, 935135 (2022).
- Zhao, J. *et al.* A multi-omics deep learning model for hypoxia phenotype to predict tumor aggressiveness and prognosis in uveal melanoma for rationalized hypoxia-targeted therapy. *Comput. Struct. Biotechnol. J.* **20**, 3182–3194 (2022).
- Shrestha, R. *et al.* Multiomics characterization of low-grade serous ovarian carcinoma identifies potential biomarkers of MEK inhibitor sensitivity and therapeutic vulnerability. *Cancer Res.* **81**(7), 1681–1694 (2021).
- Lv, S. Q. *et al.* Comprehensive omics analyses profile genesets related with tumor heterogeneity of multifocal glioblastomas and reveal LIF/CCL2 as biomarkers for mesenchymal subtype. *Theranostics* **12**(1), 459–473 (2022).
- Zhang, H. & Wu, Z. The generalized Fisher's combination and accurate p-value calculation under dependence. *Biometrics* **79**, 1159–1172 (2022).
- Fisher, R. A. Statistical methods for research workers. In *Breakthroughs in Statistics* 66–70 (Springer, 1992).
- Won, S. *et al.* Choosing an optimal method to combine P-values. *Stat. Med.* **28**(11), 1537–1553 (2009).
- Li, L. *et al.* The landscape and prognostic value of tumor-infiltrating immune cells in gastric cancer. *PeerJ* **7**, e7993 (2019).
- Liu, Z. *et al.* Tumor stroma-infiltrating mast cells predict prognosis and adjuvant chemotherapeutic benefits in patients with muscle invasive bladder cancer. *Oncoimmunology* **7**(9), e1474317 (2018).
- Toor, S. M. *et al.* Immune checkpoints in the tumor microenvironment. *Semin. Cancer Biol.* **65**, 1–12 (2020).

18. Love, M. I., Huber, W. & Anders, S. Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. *Genome Biol.* **15**(12), 550 (2014).
19. Hogben, C. A. A practical and simple equivalent for Student's T test of statistical significance. *J. Lab. Clin. Med.* **64**, 815–819 (1964).
20. Kolde, R. & Kolde, M. R. J. R. P. *Package 'heatmap'*. **1** (2018).
21. Yu, G. *et al.* clusterProfiler: An R package for comparing biological themes among gene clusters. *OMICS* **16**(5), 284–287 (2012).
22. M., C. *org.Hs.eg.db: Genome wide annotation for Human.* (2015).
23. Therneau, T. & Grambsch, P. *Modeling Survival Data: Extending The Cox Model.* vol. 48 (2000).
24. Kassambara, A. *et al.* *Package 'survminer'* (2017).
25. Heagerty, P. J., Lumley, T. & Pepe, M. S. J. B. *Time-Dependent ROC Curves for Censored Survival Data and a Diagnostic Marker.* vol. 56, no. 2, 337–344 (2000).
26. Network, T.C.G.A.R. Comprehensive molecular characterization of gastric adenocarcinoma. *Nature* **513**(7517), 202–209 (2014).
27. Chen, B. *et al.* Profiling tumor infiltrating immune cells with CIBERSORT. In *Cancer Systems Biology* 243–259 (Springer, 2018).
28. Sohn, B. H. *et al.* Clinical significance of four molecular subtypes of gastric cancer identified by the cancer genome atlas project. *Clin. Cancer Res.* **23**(15), 4441–4449 (2017).
29. Silva, A. N. S. *et al.* Increasing frequency of gene copy number aberrations is associated with immunosuppression and predicts poor prognosis in gastric adenocarcinoma. *Br. J. Surg.* **109**(3), 291–297 (2022).
30. Comprehensive molecular characterization of gastric adenocarcinoma. *Nature*. **513**(7517), 202–209 (2014).
31. Siegel, R. L., Miller, K. D. & Jemal, A. Cancer statistics, 2016. *CA Cancer J. Clin.* **66**(1), 7–30 (2016).
32. Chakraborty, S. *et al.* Onco-multi-OMICS approach: A new frontier in cancer research. *Biomed. Res. Int.* **2018**, 9836256 (2018).
33. Horn, L. A. *et al.* Remodeling the tumor microenvironment via blockade of LAIR-1 and TGF-beta signaling enables PD-L1-mediated tumor eradication. *J. Clin. Investig.* <https://doi.org/10.1172/JCI155148> (2022).
34. Carvalheiro, T. *et al.* Leukocyte associated immunoglobulin like receptor 1 regulation and function on monocytes and dendritic cells during inflammation. *Front. Immunol.* **11**, 1793 (2020).
35. Keerthivasan, S. *et al.* Homeostatic functions of monocytes and interstitial lung macrophages are regulated via collagen domain-binding receptor LAIR1. *Immunity* **54**(7), 1511–1526.e8 (2021).
36. Sivori, S. *et al.* Inhibitory receptors and checkpoints in human NK cells, implications for the immunotherapy of cancer. *Front. Immunol.* **11**, 2156 (2020).
37. Yang, M. & Brackenbury, W. J. Membrane potential and cancer progression. *Front. Physiol.* **4**, 185 (2013).
38. Silver, B. B. & Nelson, C. M. The bioelectric code: Reprogramming cancer and aging from the interface of mechanical and chemical microenvironments. *Front. Cell Dev. Biol.* **6**, 21 (2018).
39. Wang, W. *et al.* Effector T cells abrogate stroma-mediated chemoresistance in ovarian cancer. *Cell* **165**(5), 1092–1105 (2016).
40. Gandhi, L. *et al.* Pembrolizumab plus chemotherapy in metastatic non-small-cell lung cancer. *N. Engl. J. Med.* **378**(22), 2078–2092 (2018).
41. Lan, X. *et al.* Increased BTLA and HVEM in gastric cancer are associated with progression and poor prognosis. *Onco Targets Ther.* **10**, 919–926 (2017).
42. Nagai, S. & Azuma, M. The CD28-B7 family of co-signaling molecules. *Adv. Exp. Med. Biol.* **1189**, 25–51 (2019).
43. Beckermann, K. E. *et al.* CD28 costimulation drives tumor-infiltrating T cell glycolysis to promote inflammation. *JCI Insight* <https://doi.org/10.1172/jci.insight.138729> (2020).
44. Teijeira, A. *et al.* Metabolic consequences of T-cell costimulation in anticancer immunity. *Cancer Immunol. Res.* **7**(10), 1564–1569 (2019).
45. Marangoni, F. *et al.* Tumor tolerance-promoting function of regulatory T cells is optimized by CD28, but strictly dependent on calcineurin. *J. Immunol.* **200**(10), 3647–3661 (2018).
46. Flemming, A. T cells: Successful checkpoint blockade requires positive co-stimulation. *Nat. Rev. Immunol.* **17**(4), 215 (2017).
47. Hui, E. *et al.* T cell costimulatory receptor CD28 is a primary target for PD-1-mediated inhibition. *Science* **355**(6332), 1428–1433 (2017).
48. Kamphorst, A. O. *et al.* Rescue of exhausted CD8 T cells by PD-1-targeted therapies is CD28-dependent. *Science* **355**(6332), 1423–1427 (2017).
49. Kim, K. H. *et al.* PD-1 blockade-unresponsive human tumor-infiltrating CD8(+) T cells are marked by loss of CD28 expression and rescued by IL-15. *Cell Mol. Immunol.* **18**(2), 385–397 (2021).
50. Wherry, E. J. & Kurachi, M. Molecular and cellular insights into T cell exhaustion. *Nat. Rev. Immunol.* **15**(8), 486–499 (2015).
51. Bucktrout, S. L., Bluestone, J. A. & Ramsdell, F. Recent advances in immunotherapies: From infection and autoimmunity, to cancer, and back again. *Genome Med.* **10**(1), 79 (2018).
52. Meynard, L. The inhibitory collagen receptor LAIR-1 (CD305). *J. Leukoc. Biol.* **83**(4), 799–803 (2008).
53. Lebbink, R. J. *et al.* Identification of multiple potent binding sites for human leukocyte associated Ig-like receptor LAIR on collagens II and III. *Matrix Biol.* **28**(4), 202–210 (2009).
54. Peng, D. H. *et al.* Collagen promotes anti-PD-1/PD-L1 resistance in cancer through LAIR1-dependent CD8(+) T cell exhaustion. *Nat. Commun.* **11**(1), 4520 (2020).
55. Croft, M. Control of immunity by the TNFR-related molecule OX40 (CD134). *Annu. Rev. Immunol.* **28**, 57–78 (2010).
56. Martins, M. R. *et al.* Could OX40 agonist antibody promote activation of the anti-tumor immune response in gastric cancer?. *J. Surg. Oncol.* **117**(5), 840–844 (2018).
57. Lima, C. A. C. *et al.* High soluble OX40 levels correlate with metastatic gastric cancer. *J. Surg. Oncol.* **126**(1), 139–143 (2022).
58. Li, Y. *et al.* Stress-induced upregulation of TNFSF4 in cancer-associated fibroblast facilitates chemoresistance of lung adenocarcinoma through inhibiting apoptosis of tumor cells. *Cancer Lett.* **497**, 212–220 (2021).
59. Roszik, J. *et al.* TNFSF4 (OX40L) expression and survival in locally advanced and metastatic melanoma. *Cancer Immunol. Immunother.* **68**(9), 1493–1500 (2019).
60. Shah, S. C. *et al.* Population-based analysis of differences in gastric cancer incidence among races and ethnicities in individuals age 50 years and older. *Gastroenterology* **159**(5), 1705–1714.e2 (2020).
61. Gao, S. *et al.* Effects of HCG on human epithelial ovarian cancer vasculogenic mimicry formation in vivo. *Oncol. Lett.* **12**(1), 459–466 (2016).
62. Wang, W. *et al.* An association between genetic polymorphisms in the ileal sodium-dependent bile acid transporter gene and the risk of colorectal adenomas. *Cancer Epidemiol. Biomark. Prev.* **10**(9), 931–936 (2001).
63. Zhou, C. L., Su, H. L. & Dai, H. W. Thrombopoietin is associated with a prognosis of gastric adenocarcinoma. *Rev. Assoc. Med. Bras.* (1992) **66**(5), 590–595 (2020).
64. Guo, Y. *et al.* Clinicopathological significance of platelet-derived growth factor B, platelet-derived growth factor receptor-β, and E-cadherin expression in gastric carcinoma. *Contemp. Oncol./Współczesna Onkologia* **17**(2), 150–155 (2013).
65. Gong, Y., Chen, L. & Chu, X. Expression of platelet-derived growth factor and PDGF receptors and its local invasiveness and metastasis in human pancreatic cancer. *J.-Nanjing Univ. Nat. Sci. Ed.* **34**, 564–568 (1998).
66. Wang, G. *et al.* Hypomethylated gene NRP1 is co-expressed with PDGFRB and associated with poor overall survival in gastric cancer patients. *Biomed. Pharmacother.* **111**, 1334–1341 (2019).
67. Wu, M., Li, Q. & Wang, H. Identification of novel biomarkers associated with the prognosis and potential pathogenesis of breast cancer via integrated bioinformatics analysis. *Technol. Cancer Res. Treat.* **20**, 1533033821992081 (2021).

68. Soiland, H. *et al.* Apolipoprotein D predicts adverse outcome in women ≥ 70 years with operable breast cancer. *Breast Cancer Res. Treat.* **113**(3), 519–528 (2009).
69. Mitchel, J. *et al.* A translational pipeline for overall survival prediction of breast cancer patients by decision-level integration of multi-omics data. In *Proceedings (IEEE Int Conf Bioinformatics Biomed)*, 2019, 1573–1580 (2019).
70. Zhang, J. *et al.* Regulation of docetaxel chemosensitivity by NR2F6 in breast cancer. *Endocr. Relat. Cancer* **27**(5), 309–323 (2020).
71. Deeken, J. F. *et al.* A pharmacogenetic study of docetaxel and thalidomide in patients with castration-resistant prostate cancer using the DMET genotyping platform. *Pharmacogenomics J.* **10**(3), 191–199 (2010).

Author contributions

Y.W., W.H., and X.P.: conception and design of study. Y.W., W.H., S.Z., and L.Z.: acquisition, analysis and of data. Y.W., W.H., and L.W.: drafting the manuscript. X.P. and L.W.: revising the manuscript critically for important intellectual content. Y.W., W.H., and X.P.: approval of the version of the manuscript to be published. All authors contributed to the article and approved the submitted version.

Funding

The present study was supported by the present study was supported by the National Natural Science Foundation of China (grant nos. 81772631; grant nos. 81974362), the Shenzhen Science and Technology Innovation Committee (Grant no. JCYJ20190814121001751, JCYJ20190814110203636 and 20220530154209021).

Competing interests

The authors declare no competing interests.

Additional information

Supplementary Information The online version contains supplementary material available at <https://doi.org/10.1038/s41598-024-52087-3>.

Correspondence and requests for materials should be addressed to Y.W. or X.P.

Reprints and permissions information is available at www.nature.com/reprints.

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.



Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.

© The Author(s) 2024