

Research article

Development and validation of a risk prognostic model based on the *H. pylori* infection phenotype for stomach adenocarcinomaJing Zhou^{a,1}, Li Guo^{b,1}, Yuzhen Wang^a, Lina Li^a, Yahuan Guo^a, Lian Duan^a, Mi Jiao^a, Pan Xi^c, Pei Wang^{d,*}^a Department of Oncology, Shaanxi Province Tumor Hospital, Xi'an, 710061, China^b Department of Geriatrics, South District, 986th Hospital of the People's Liberation Army Air Force, Xi'an, 710054, China^c Department of Radiotherapy, Shaanxi Province Tumor Hospital, Xi'an, 710061, China^d Department of Anesthesiology, Shaanxi Province Tumor Hospital, Xi'an, 710061, China

ARTICLE INFO

Keywords:

Stomach adenocarcinoma (STAD)

Single cell RNA-Seq profile

RiskScore

H. pylori infection phenotype

ssGSEA

WGCNA

ABSTRACT

Background: Stomach adenocarcinoma (STAD) is one of the most common malignancies. Infection of helicobacter pylori (*H. pylori*) is a major risk factor that leads to the development of STAD. This study constructed a risk model based on the *H. pylori*-related macrophages for predicting STAD prognosis.

Methods: The single-cell RNA sequencing (scRNA-seq) dataset and the clinic information and RNA-seq datasets of STAD patients were collected for establishing a prognostic model and for validation. The “Seurat” and “harmony” packages were used to process the scRNA-seq data. Key gene modules were sectioned using the “limma” package and the “WGCNA” package. Kaplan-Meier (KM) and Receiver Operating Characteristic Curve (ROC) analyses were performed with “survminer” package. The “GSVA” package was employed for single sample gene set enrichment analysis (ssGSEA). Cell migration and invasion were measured by carrying out wound healing and trans-well assays.

Results: A total of 17397 were screened and classified into 8 cell type clusters, among which the macrophage cluster was closely associated with the *H. pylori* infection. Macrophages were further categorized into four subtypes (including C1, C2, C3, and C4), and highly variable genes of macrophage subtype C4 could serve as an indicator of the prognosis of STAD. Subsequently, we developed a RiskScore model based on six *H. pylori*-associated genes (*TNFRSF1B*, *CTLA4*, *ABCA1*, *IKBIP*, *AKAP5*, and *NPC2*) and observed that the high-risk patients exhibited poor prognosis, higher suppressive immune infiltration, and were closely associated with cancer activation-related pathways. Furthermore, a nomogram combining the RiskScore was developed to accurately predict the survival of STAD patients. *ABCA1* in the RiskScore model significantly affected the migration and invasion of tumor cells.

Conclusion: The gene expression profile served as an indicator of the survival for patients with STAD and addressed the clinical significance of using *H. pylori*-associated genes to treat STAD. The current findings provided novel understandings for the clinical evaluation and management of STAD.

* Corresponding author.

E-mail address: wangpeisean@163.com (P. Wang).¹ Equal Contribution.

1. Introduction

Stomach adenocarcinoma (STAD) is a gastrointestinal cancer with the fifth highest incidence and the third highest mortality rate worldwide. Early diagnosis and treatment can increase 5-year survival rate to 90–97 % but this figure drops below 30 % for patients with advanced or metastatic STAD [1,2]. Previous study has classified gastric carcinomas (GC) into intestinal and diffuse types, which all belong to adenocarcinomas [3] but with different pathological and epidemiological features [4]. The World Health Organization (WHO) also provides a histopathology-based classification guideline that divides the GC into mucinous, papillary, tubular and poorly cohesive carcinoma [5]. Although these classification methods are simple to use, pathological results may vary differently due to complex background (e.g. different pathogen infections) and subjective discrimination factors [6]. Currently, molecular subtypes classified based on the biological behaviors of STAD are poorly studied.

Studies have shown that the *H. pylori* is a gram-negative bacterium that colonizes the gastric mucous environment of 60.3 % of the world's population, especially in underdeveloped countries [7]. This is partly due to poor sanitation and high population density in these countries, which can easily lead to oral and fecal-oral transmission of *H. pylori* [8]. *H. pylori* is largely associated with increased risk of gastroduodenal disturbances (such as the gastroesophageal reflux disease, peptic ulcerations disease, and inflammatory bowel disease), particularly in mucosa-associated lymphoid tissue lymphoma, STAD and peptic ulcer [8,9]. The *H. pylori* infects and induces the normal mucosa to transfer chronic superficial gastritis, which further evolves to invasive GC from chronic gastritis, metaplasia and dysplasia [10]. In addition, several factors such as high intake of dietary salt, drinking and smoking [11], genetically high gastric acid secretion will all increase the risk of developing GC [12,13], but the inflammatory response to the persistent infection of *H. pylori* is a major etiological cause leading to carcinogenesis [14]. Patients with polymorphisms of encoding tumor necrosis factor- α , interleukin-1 β and IL-1 β receptor antagonist [15,16] are at higher risk of gastric atrophy expansion, hypochlorhydria and GC after *H. pylori* infection [17,18]. Due to a high mutational rate, *H. pylori* populations are highly genetically diverse, including large number of deletions, insertions and chromosomal rearrangements that affect the function of housekeeping genes [19]. *H. pylori* strains vary in their pathogenic properties, which could result in genetic variation in virulence factors, explaining why some infected individuals are asymptomatic and do not have severe pathological changes, while others exhibit peptic ulcer disease or cancers [20,21].

As STAD is often diagnosed at late stage, it is essential to identify factors that lead to the occurrence of the malignancy [22,23]. In addition, the next generation sequencing (NGS) technique has been widely used in molecular profiling research of many tumors. Fabio et al. revealed the genetic and epigenetic heterogeneity of STAD by conducting large-scale genomic and transcriptomic studies [24]. The single-cell RNA sequencing (scRNA-seq) characterizes the transcriptional states of a single cell [25], allowing unbiased analysis of cellular characteristics in tumor tissues for exploring tumor ecosystem [26] and cancer-immune heterogeneity [27]. This study performed an unbiased and systemic transcriptomic profile analysis using the data from The Cancer Genome Atlas (TCGA) and Gene Expression Omnibus (GEO) databases. The scRNA-seq landscape showed that macrophages were closely associated with the phenotype of *H. pylori* infection. We further used the differentially expressed genes (DEGs) and macrophage-related module to develop a reliable prognosis model [28] applying WGCNA, un/multivariate and LASSO cox regression analysis. The current findings provided novel insights into the cell-molecular landscape, clinical diagnosis and prognosis of STAD.

2. Material and methods

2.1. Data acquisition and preprocessing

The RNA-seq data and clinical follow-up information of STAD patients were downloaded from TCGA (<https://portal.gdc.cancer.gov/>) through the TCGA GDC API tool [29]. After excluding patients without survival time or state from the TCGA-STAD expression profile, a total of 353 STAD samples and 53 para-cancer control samples were obtained. The expression matrix were converted to TPM format and log₂-transformed to select protein-encoding genes. Subsequently, GSE66229 containing the expression of 300 STAD samples was retrieved from the GEO (<https://www.ncbi.nlm.nih.gov/geo/>) [30]. The probes were mapped to the genes based on the corresponding annotation information, and the mean value of expression was taken when multiple probes matched to one gene. In addition, through searching the keywords of KEGG_EPITHELIAL_CELL_SIGNALING_IN_HELICOBACTER_PYLORI_INFECTION and HP_HELICOBACTER_PYLORI_INFECTION, 73 *H. pylori* infection-related genes and the cancer-related Hallmark gene sets were collected from The Molecular Signatures Database (MSigDB, <https://www.gsea-msigdb.org/gsea/msigdb>) [31].

2.2. Single-cell RNA profile analysis and enrichment analysis

The scRNA-seq expression dataset (GSE167297) of STAD containing 10 tumor samples were collected from the GEO [30]. Genes expressed in fewer than three individual cells and containing less than 200 or more than 2000 genes were removed, while those with over 15 % of mitochondrial gene expression were retained. Subsequently, the scRNA-seq data were normalized using the “Seurat” R package [32] and FindVariableFeatures function was used for searching highly variable genes (log₂fc.threshold = 0.5 and min.pct = 0.25) [32]. Then, the ScaleData function was used to scale all the genes and principal component analysis (PCA) was performed to find the clustering anchor (dim = 10). Batch effect among samples were removed using the “harmony” R package, followed by using the RunTSNE function to further reduce dimension [33]. Finally, the cells were clustered by the FindNeighbors and FindClusters functions at resolution = 0.2, and different cell types were annotated according to marker genes.

Each cell associated with *H. pylori* infection was assigned with an enrichment score using the “AUCCell” package [34], with a higher

AUCell score indicating a closer biological relevance between the cell cluster and *H. pylori* infection. The “CellChat” R package was used for inferring the cell-to-cell communication. Finally, to assess the prognostic differences among patients with low and high scores, ssGSEA was performed using “GSVA” package [35] to compute the scores for macrophage subpopulations.

2.3. Weighted gene co-expression network analysis (WGCNA)

The DEGs analysis between tumor and para-cancer control samples in the TCGA cohort was used using the “limma” R package [32] (setting $p < 0.05$ and $|\log_2\text{Fold Change}| > \log_2(1.5)$). WGCNA was performed applying the “WGCNA” R package to determine the gene module associated with the phenotype of *H. pylori* infection. The “pickSoftThreshold” function in the “WGCNA” R package was applied for determining the soft threshold β (setting the sensibility to 2 and the module merge threshold to 0.3). Gene modules with at least 50 genes were sectioned [32]. The correlation analysis between the modules and the phenotype was performed using the Pearson method. The “clusterProfiler” R package [36] was used for the Gene Ontology (GO) and Kyoto Encyclopedia of Genes and Genomes (KEGG) analysis.

2.4. Identifying risk genes for developing a risk model

Univariate cox regression analysis was used to identify significant prognostic genes ($p < 0.05$). Then LASSO cox regression analysis was performed applying the “glmnet” R package to primarily shrink the number of candidate genes [32]. Multivariable cox with regression analysis and stepwise algorithm were employed to further reduce the candidate genes and calculate the regression coefficient, respectively. The risk model was established according to the formula of RiskScore = \sum (hazard cox coefficient * expression of risk gene).

2.5. Verification of model prognostic value

The RiskScore was calculated for all the patients and for classifying low- and high-risk groups according to the median RiskScore value. KM survival analysis and the ROC analysis with the Area Under Curve (AUC) were performed using the “survminer” R package [32].

2.6. Analysis of immune infiltration

CIBERSORT is an algorithm estimates the abundance of different cell types in mixed tissues based on suppressed gene expression profiles [37]. The TIMER algorithm is mainly used to quantify various types of immune cells in tumors [38]. This study used both CIBERSORT and TIMER algorithms to assess the differences in immune cell infiltration in patients from different risk groups.

2.7. Cell culture and transfection

From the American Type Culture Collection (ATCC), we obtained the GC cell line HGC27 and the gastric epithelial cell line GES-1. DEME medium (Invitrogen) containing 10 % fetal bovine serum (Gibco, Thermo Fisher Scientific), 50 $\mu\text{g}/\text{ml}$ streptomycin and 100 U/ml penicillin was used for cell culture in 5 % CO_2 at 37 °C [39]. The siRNA was used to silence ABCA1 applying the siRNA reagent (Sangon, shanghai, china) with forward sequence of 5'-GCGACTCCACATAGAAGAC-3' and reversed sequence of 5'-GACGTATGTG-CAGATCATA-3'. The Lipofectamine 3000 (Invitrogen) was used for the cell transfection. After incubation for 12 h, the cell samples were harvested for qPCR detection. Briefly, total RNA was extracted by using the TRizol Reagent (Invitrogen) and the cDNA was synthesized by the ReverTra Qpcr RT Master Mix kit (TOYOBO). Then the SYBR Green PCR Master Mix (Biosystems) was applied for qRT-PCR on LightCycler 96 (Roche) according to the manufacturer's specification. Target gene expression [40] from three times sample and technique repetition was calculated by the $2^{-\Delta\Delta\text{CT}}$ method, with gene β -actin as a reference. The specific primers were listed in Table S1.

2.8. Cell migration and invasion assays

Cell migration was measured by wound healing assay. 4×10^6 cells were seeded into a 6-well plate (Corning) and incubated until confluent, and then a rectilinear scratch was produced with a 100- μL pipette tip. After 24 h, the cells were fixed by 4 % paraformaldehyde for 15 min (min) and stained with 0.1 % crystal violet (Servicebio) for another 15 min. Next, wound closure was photographed with an inverted microscope (Leica) [40]. For invasion assays, a total of 4×10^4 cells were plated into the upper chamber well of 24-well plates (Corning, 8- μm pore) containing 200 μL serum-free DMEM, while the lower chamber was supplemented with 800 μL of DMEM containing 20 % FBS (Thermo Fisher Scientific). After 48-h incubation, the migrating cells were fixed by 4 % paraformaldehyde and stained by 0.1 % crystal violet for 15 min and then imaged with an inverted microscope [40].

2.9. Statistical analysis

All statistical analysis and data visualization was performed in the R software (version 4.3.1). The Pearson method was used for the correlation analysis. A p-value < 0.05 was defined as statistically significant. SangerBox (<http://sangerbox.com/home.html>) provided

certain data analysis.

3. Results

3.1. The landscape of single-cell RNA of STAD samples

After cell filtering (Figs. S1A and B), normalization and dimensionality reduction clustering (Fig. S1C), a total of 17397 cells from the scRNA-seq expression dataset were clustered into 8 clusters (Fig. 1A) including B cells, epithelial cells, T cells, dendritic cells, fibroblasts, macrophages, endothelial cells, and mast cells (Fig. 1B) according to the expression of marker genes (Fig. 1C–Table S2). The T cells and dendritic cells had a higher proportion in each sample (Fig. 1D). Analysis on the *H. pylori* infection-related genes showed that macrophages had the highest AUCell score (Fig. 1E), suggesting that macrophages were closely associated with the *H. pylori* infection in STAD.

3.2. Identifying *H. pylori* infection-associated macrophage subtypes

The macrophage population was selected to perform the t-distributed stochastic neighbor embedding (TSNE) clustering (resolution = 0.4). The macrophages were divided into 6 subclusters (cluster1-6, Fig. 2A), among which the C4 cluster exhibited the highest AUCell score of *H. pylori* infection (Fig. 2B). Subsequently, we identified the highly variable genes between these macrophage populations and found that several chemokines genes including the *CCL3L3*, *CXCL5* and *CCL3* were high-expressed in the C4 subpopulation (Fig. 2C). These protein-encoding genes promoted inflammatory response through regulating the migration and activation of leukocytes. KEGG pathway enrichment analysis further showed that these genes were enriched in inflammatory response, cytokine-cytokine receptor interaction, TNF and IL-17 signaling pathways (Fig. 2D).

3.3. C4 subcluster of macrophage-mediated cell communication

In multicellular organisms, cell communication plays an important part in cell life activity. According to the results of enrichment analysis, it was found that the C4 macrophages functioned crucially in regulating STAD progression. Further cell communication analysis revealed that macrophages exhibited obvious interaction relationship with other cell clusters (Fig. 3A), with the C4 macrophages having greater interaction intensity with other cell clusters (Fig. 3B). Further analysis of ligand-receptor information between different cell clusters showed that the C4 macrophages affected other cell clusters, especially some immune cell clusters (T cells and mast cells) through the *SPP1-CD44* and *MIF(CD74+CD44)* interaction (Fig. 3C), while epithelial cells and fibroblasts clusters affected the C4 macrophages through the *MDK-SDC2* interaction (Fig. 3D).

3.4. Identifying gene module related to C4 macrophages

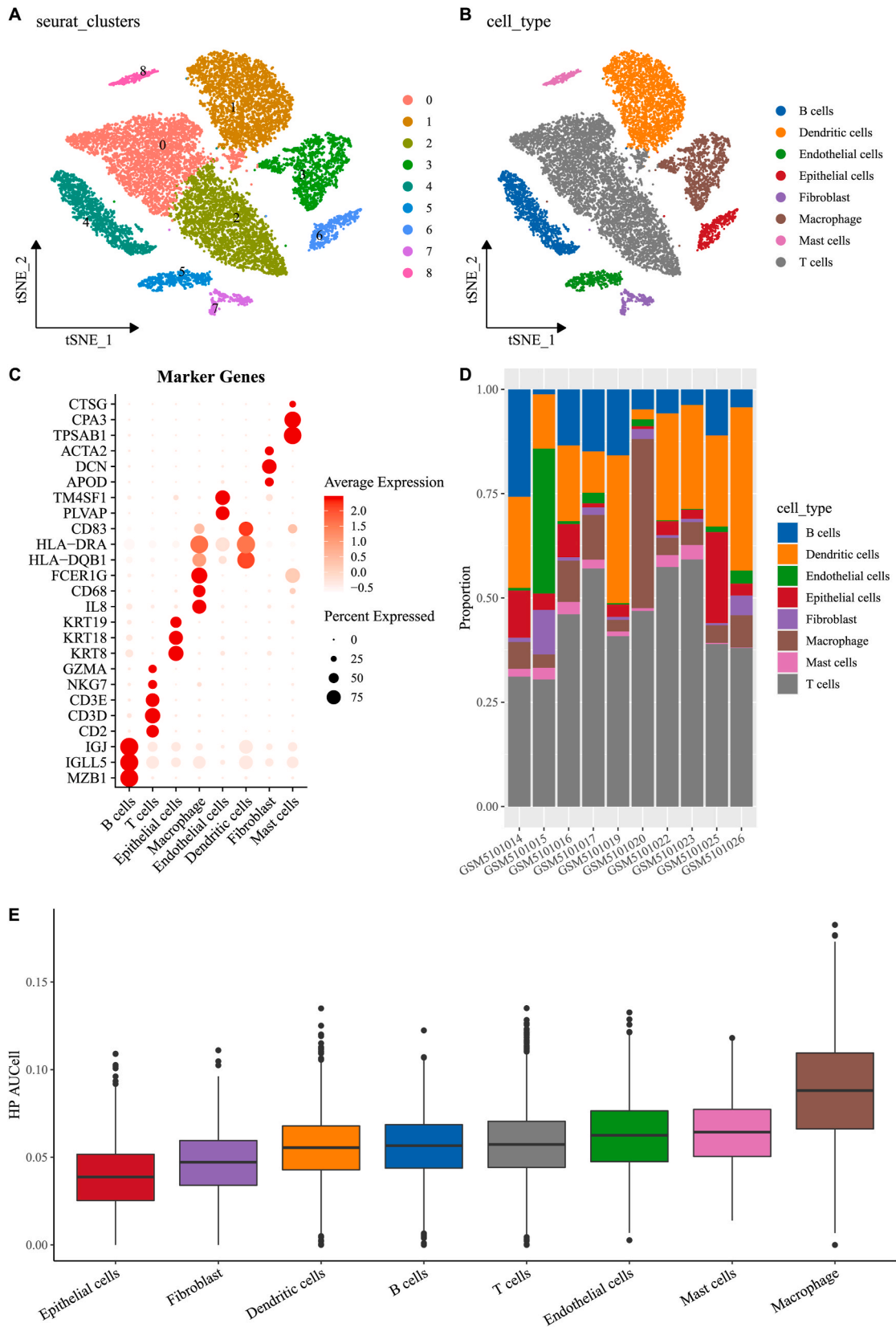
We calculated the ssGSEA score in the TCGA cohort according to the expression of the highly variable genes of C4 macrophages, and found that STAD patients with a higher score tended to have a worse prognosis (Fig. 4A). Next, WGCNA was used to identify gene module related to C4 macrophages based on the DEGs. The soft threshold β was set at 12 to ensure a scale-free network (Fig. 4B). After hierarchical clustering and module merging, a total of 5 co-expression modules were obtained (Fig. 4C). As the grey module cannot be merged with other modules, it was considered as an invalid module. Further correlation analysis showed that the green module had a strong correlation with the C4 subcluster (Fig. 4D). KEGG enrichment analysis revealed that the genes in the green module were enriched in cytokine receptor interaction, inflammatory response pathways, and cell adhesion molecules (Fig. S2A). GO analysis showed that these genes were closely related to immune cell activation and immune response regulation pathways in biological process (Fig. S2B), protein and immune complex pathways in cell component (Fig. S2C) and receptor bind and activity regulation pathways in molecular function (Fig. S2D).

3.5. Establishment of a risk classification model

To establish risk model, the samples in the TCGA-STAD cohort were divided into the training set and test set at the ratio of 7:3, with the GSE66229 cohort as an independent validation set. The clinical information of training set and test set was listed in Table 1. Chi-square test showed no significant difference between varying clinical groups, indicating that our grouping was random and reliable. Univariate, LASSO and multivariate cox regression analysis filtered 6 key prognostic genes and used them to establish a risk model: $\text{RiskScore} = (-0.25 * TNFRSF1B) + (-0.2 * CTLA4) + (0.384 * ABCA1) + (0.343 * IKBIP) + (-0.564 * AKAP5) + (0.4 * NPC2)$.

3.6. Validation of the effectiveness of the model classification

Based on the optimal cutoff point, the patients in training set were divided into high-risk and low-risk groups. It was found that high-risk patients had a poor prognosis (Fig. 5A), with an AUC value of 0.67, 0.72, 0.71 and 0.65 for 1-, 2-, 3- and 4-year survival, respectively, which indicated that a high accuracy of the RiskScore in long- and short-term prediction and classification (Fig. 5A). Similar survival results were also observed in the test set (Fig. 5B) and the TCGA cohort (Fig. 5C), and ABCA1, IKBIP and NPC2 were high-expressed in high-risk group (Fig. 5C). We further evaluated the accuracy and robustness of the model in the GSE66229 validation



(caption on next page)

Fig. 1. Single cell atlas of stomach adenocarcinoma. (A) TSEN plot of single cell clustering. (B) TSEN plot of the annotated cell clusters. (C) The bubble plot of the expression of marker genes in each cell cluster. (D) The proportion of cell cluster in different samples. (E) The AUCell score of *H. pylori* infection in each cell cluster.

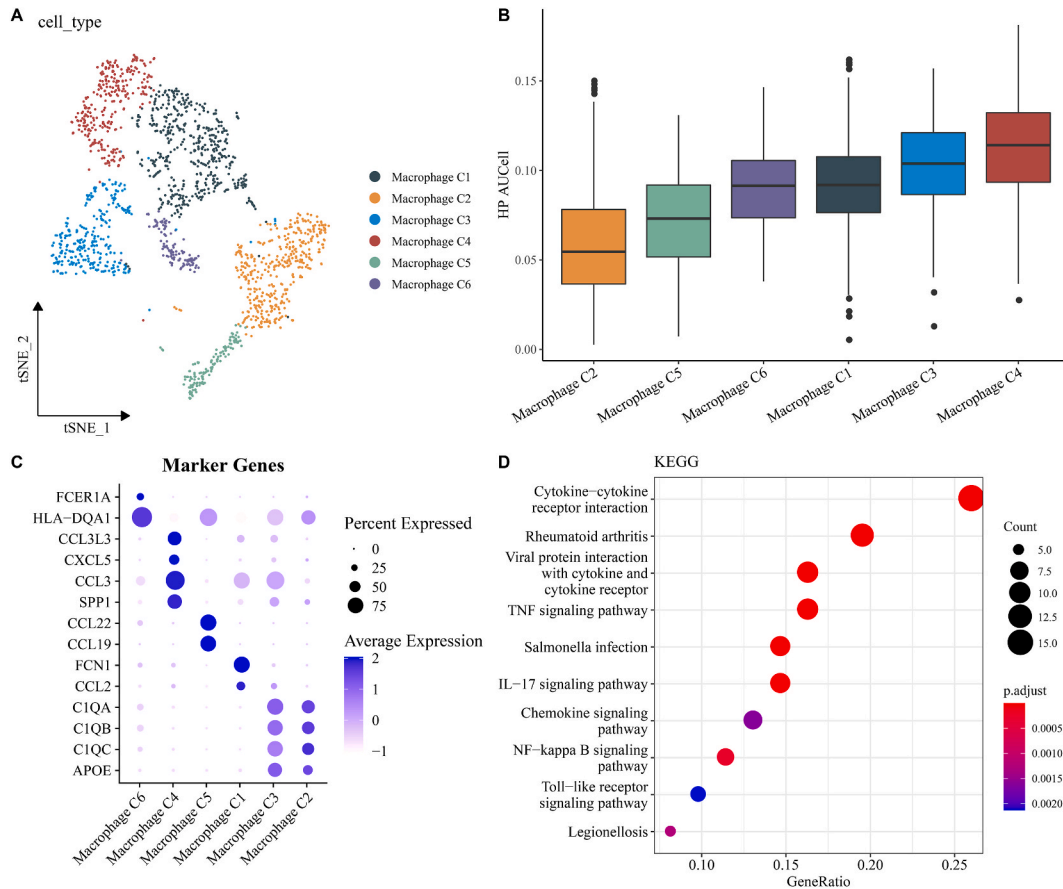


Fig. 2. Macrophage clustering. (A) The t-NSE plot of macrophage clustering. (B) The AUCell score of *H. pylori* infection in different macrophage cluster. (C) The bubble plot of the expression of marker genes in different macrophage cluster. (D) The KEGG analysis of highly variable gene of macrophage (C4).

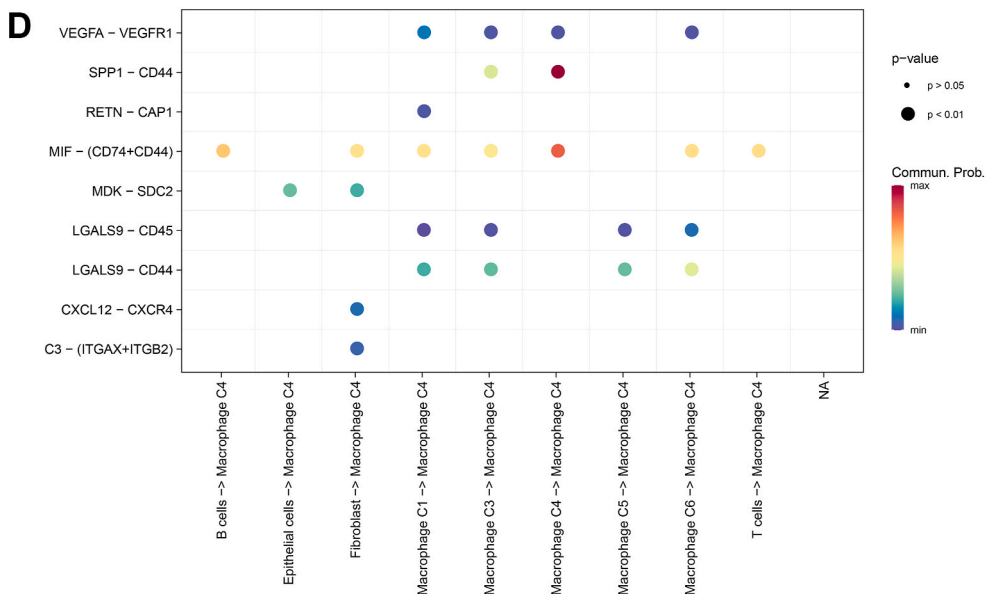
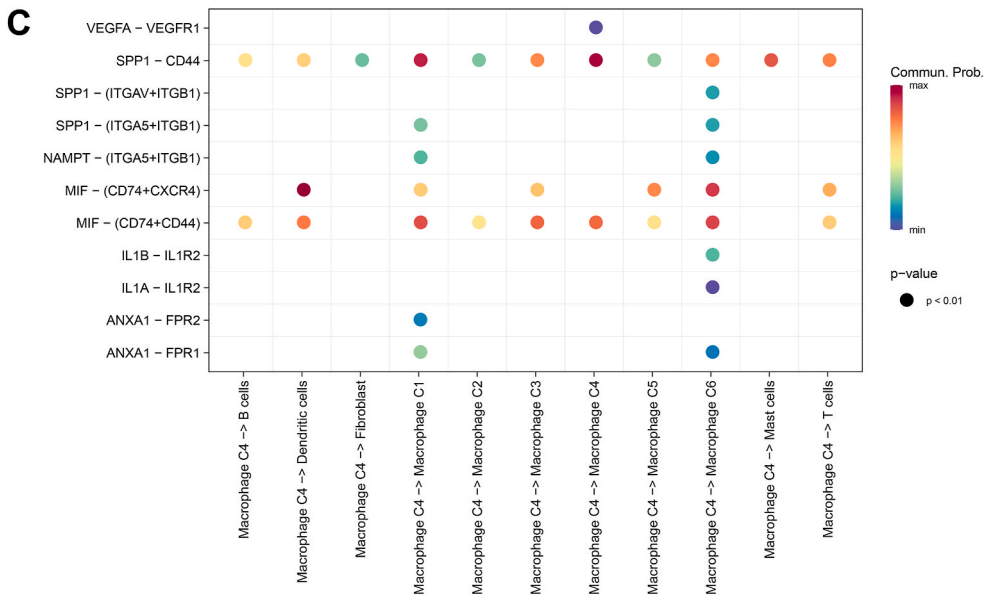
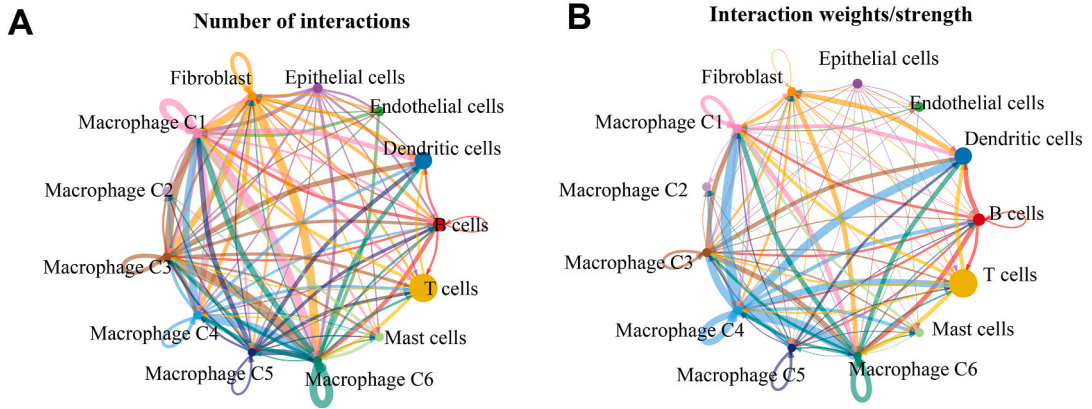
set. The results showed that the patients with a higher RiskScore had the worst prognosis and shorter survival time, with an AUC of 1- to 5-year survival rate higher than 0.65 (Fig. 5D), which demonstrated that the RiskScore was highly effective in predicting long- and short-term prognosis of STAD (Fig. 5D).

3.7. Identifying independent prognostic factors and establishing a nomogram

The RiskScore and other clinical factors were incorporated into univariate cox regression analysis, which showed that the Age, American Joint Committee on Cancer (AJCC) stage and RiskScore were significant influencing factors for STAD prognosis ($p < 0.05$, Fig. 6A). Multivariate cox regression analysis also proved that these three factors were independent prognostic factors ($p < 0.05$, Fig. 6B). To further improve the risk assessment and survival prediction for STAD, the Age, AJCC stage and RiskScore were combined to develop a nomogram model (Fig. 6C). The results showed that the RiskScore had the greatest influence on predicting patients' survival (Fig. 6C). The calibration curve of nomogram presented that the 1-, 3- and 5- year calibration curve was close to the standard curve (Fig. 6D), suggesting that the nomogram model had an excellent prognostic prediction performance. In the decision curve analysis (DCA), the net benefit of the nomogram and RiskScore was obviously higher than the extreme curve (Fig. 6E), indicating that the current prognostic model had the strongest survival prediction ability.

3.8. Immune infiltration and pathway activation difference

Differences in tumor microenvironment (TME) were compared based on the immune infiltration score. CIBERSOER analysis



(caption on next page)

Fig. 3. Cell communication analysis. (A) The interaction relationship between the macrophage and other cell cluster. (B) The interaction strength analysis between the macrophage and other cell cluster. (C) The receptor-ligand interaction ways of macrophage to others cell cluster. (D) The receptor-ligand interaction ways of immune cell clusters to macrophage.

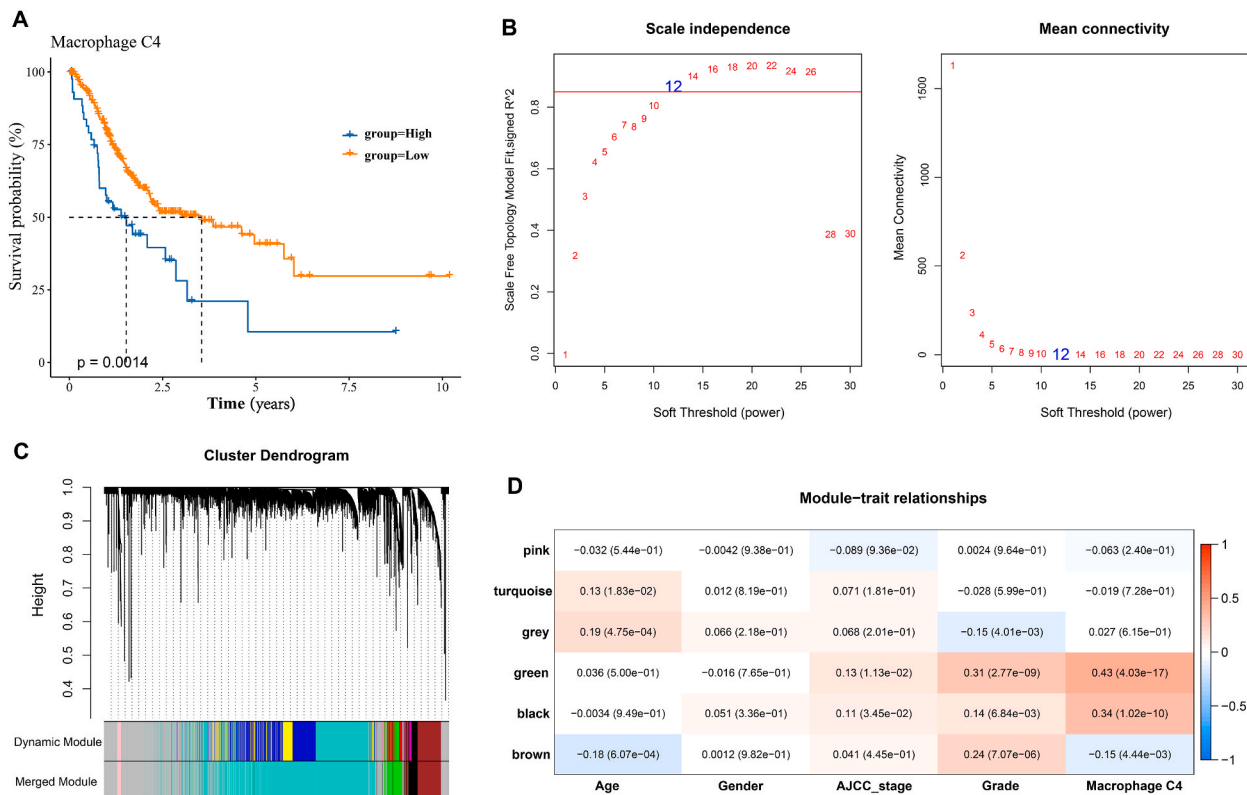


Fig. 4. WGCNA for Macrophage(C4)-related gene module searching. (A) The KM survival analysis of patients with different AUCell score of highly variable gene of macrophage (C4). (B) Analysis of the mean connectivity for various soft-thresholding powers for WGCNA. (C) Dendrogram of genes clustered based on a dissimilarity measure (1-TOM). (D) The correlation between module and feature.

showed that naïve B cells, CD4⁺ memory activated T cells, CD8⁺ T cells were infiltrated in low-risk group, while neutrophils and macrophage M2 were infiltrated in high-risk group (Fig. S3A). Higher macrophage M2 infiltration suggests the presence of immunosuppressive TME [41]. Similar results were observed in TIMER analysis, as CD4⁺ T cells, B cells, and CD8⁺ T cells were enriched in the low-risk group (Fig. S3B). Further analysis showed that the RiskScore was positively related to tumorigenesis and development pathways such as angiogenic hypoxia, Notch signaling, epithelial mesenchymal transition (EMT) and negatively correlated with immune and cell cycle pathways (Fig. S3C). This suggested that STAD patients with a higher RiskScore were more prone to activate these typical cancer activation-related pathways.

3.9. ABCA1 mediated the migration and invasion of tumor cells

Finally, we examined the expression of model genes and the role of ABCA1 in cell migration and invasion. The results of qPCR showed that these genes including *ABCA1*, *CTLA4*, *IKBIP*, *NPC2* and *TNFRSF1B* were significantly overexpressed in the HGC27 cancer cells ($p < 0.05$, Fig. 7A–E), while *AKAP5* was significantly downregulated in the HGC27 cells in comparison to that in the epithelial cell line GES-1 (Fig. 7F). In addition, the invasion ability of cancer was greatly inhibited after *ABCA1* silencing and the number of blue migrated cells in the si-*ABCA1* was significantly lower than that in the si-NC groups ($p < 0.05$, Fig. 7G). Wound healing assay revealed that silencing *ABCA1* significantly affected the rate of wound closure, which was significantly reduced in the si-*ABCA1* groups (Fig. 7H).

4. Discussion

Infection of *H. pylori* could easily cause corpus-predominant gastritis and has been considered a high-risk factor that significantly promotes the occurrence of STAD [42]. Some infected patients develop gastritis, while others with *H. pylori* infection may have gastric

Table 1
The clinical information of training set and test set.

Characteristics	Train cohort(N = 247)	Test cohort(N = 106)	Total(N = 353)	pvalue	FDR
Age					
Mean ± SD	65.52 ± 10.53	65.50 ± 10.88	65.51 ± 10.62		
Median[<i>min-max</i>]	67.00[35.00,90.00]	68.00[41.00,90.00]	67.00[35.00,90.00]		
Gender				0.63	1
FEMALE	85(24.08 %)	40(11.33 %)	125(35.41 %)		
MALE	162(45.89 %)	66(18.70 %)	228(64.59 %)		
AJCC stage				1	
I	29(8.22 %)	19(5.38 %)	48(13.60 %)		
II	74(20.96 %)	35(9.92 %)	109(30.88 %)		
III	107(30.31 %)	39(11.05 %)	146(41.36 %)		
IV	25(7.08 %)	10(2.83 %)	35(9.92 %)		
unknown	12(3.40 %)	3(0.85 %)	15(4.25 %)		
Grade				0.54	1
G1	8(2.27 %)	1(0.28 %)	9(2.55 %)		
G2	86(24.36 %)	42(11.90 %)	128(36.26 %)		
G3	147(41.64 %)	60(17.00 %)	207(58.64 %)		
unknown	6(1.70 %)	3(0.85 %)	9(2.55 %)		
Status				0.92	1
Alive	146(41.36 %)	64(18.13 %)	210(59.49 %)		
Death	101(28.61 %)	42(11.90 %)	143(40.51 %)		
OS.time					
Mean ± SD	608.00 ± 527.65	625.69 ± 594.02	613.31 ± 547.63		
Median[<i>min-max</i>]	476.00[3.00,3540.00]	406.00[20.00,3720.00]	468.00[3.00,3720.00]		

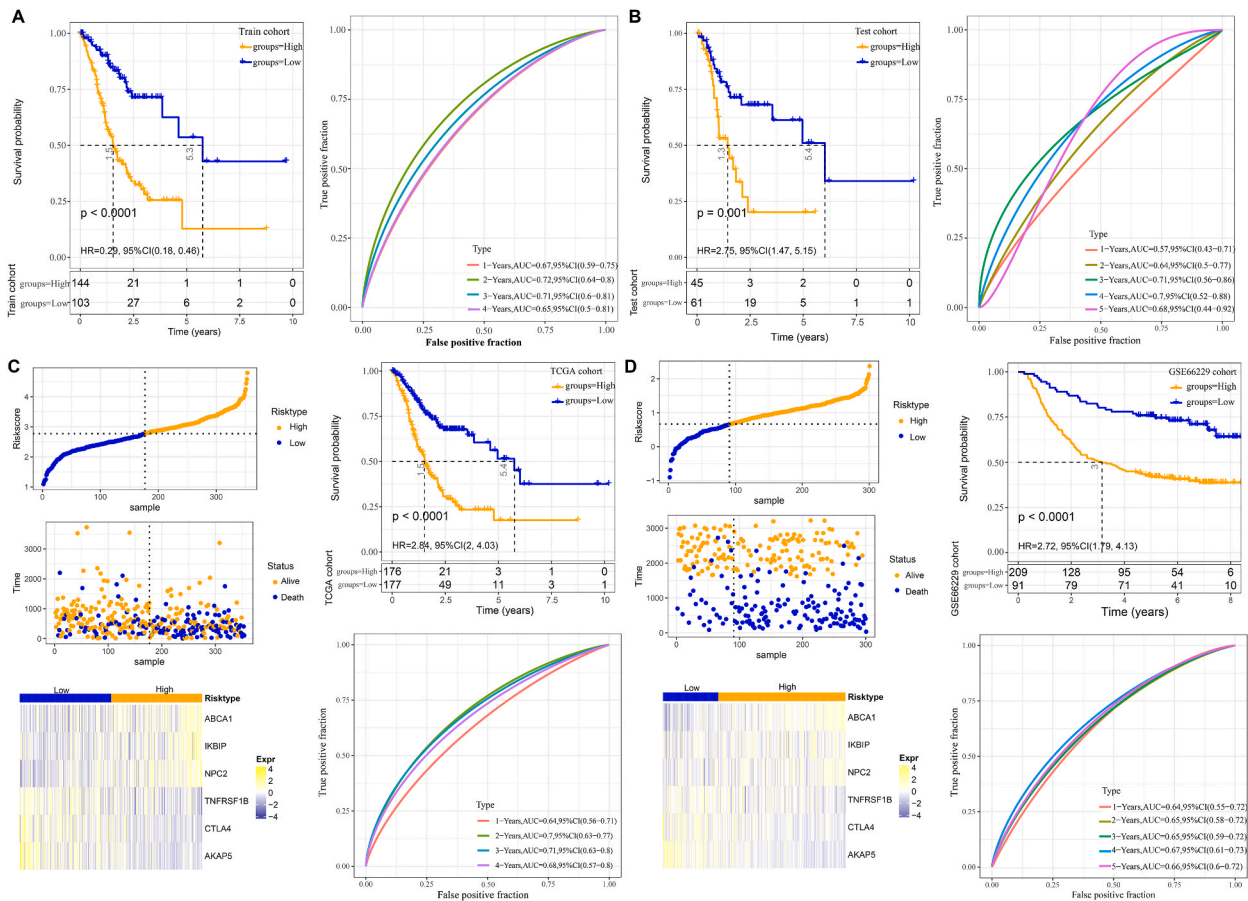


Fig. 5. Validation of model prognostic performance. (A) KM survival and ROC analysis of varying patients in training set. (B) KM survival and ROC analysis of varying patients in test set. (C) KM survival, ROC and living time analysis of varying patients in TCGA cohort. (D) KM survival, ROC and living time analysis of varying patients in validation cohort.

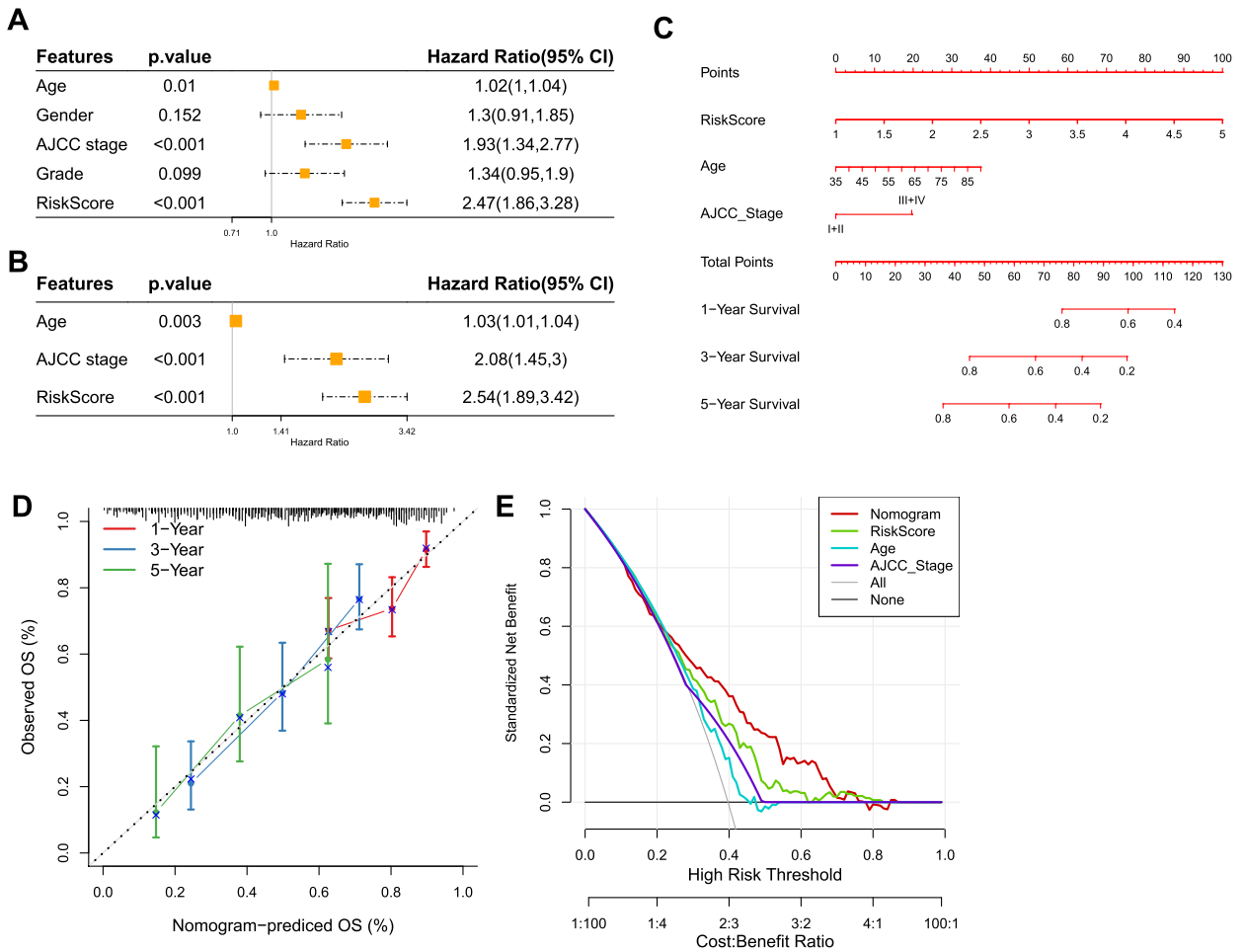


Fig. 6. Independent factor and nomogram developing. (A) Univariate cox regression analysis for significant prognostic factors. (B) Multivariate cox regression analysis for significant independent prognostic factors. (C) A nomogram developing. (D) Calibration curve of nomogram model. (E) Decision curve analysis (DCA) of nomogram model.

cancer [43]. Many patients with gastric inflammation are asymptomatic, and gastritis symptoms in certain types of patients are more persistent or will recur after eradication treatment [44]. Varied clinical outcomes could be explained by multiple factors [28] such as virulence factors, but studies demonstrated that the pathogenic mechanisms of *H. pylori* infection are more complex than generally accepted [45]. This study established a risk prognosis model according to the *H. pylori* infection phenotype to explore the potential pathogenic mechanism of *H. pylori* risk factor at molecular level and the model was able to achieve precise risk stratification for STAD patients. The nomogram model developed based on the RiskScore, AJCC stage and age can be used to accurately assess the survival probability of patients with STAD, meanwhile, the RiskScore also acted as an indicator of cancer activation-related pathways. Our finding could assist the clinical diagnosis and treatment of STAD patients.

Using the AUCell algorithm, the current results demonstrated that *H. pylori* was closely associated with macrophages. The interaction between *H. pylori* and macrophages plays a significant role in the progression, pathogenesis, and suppression of the infection [46]. In a mouse model, Zhuang et al. found that the NF-kappa B pathway is involved in the macrophage response to *H. pylori* and produces IL-6, IL-23, and CCL20, which in turn inhibit the NF-kappa-B pathway in macrophages and ultimately lead to reduced differentiation of Th17 cells [47]. Wen et al. discovered that in a co-culture model of macrophages and *H. pylori*, the use of γ -secretase to inhibit Notch signaling causes a downregulation of the expression of inducible nitric oxide synthase (iNOS) and its product, nitric oxide (NO). Such an intervention reduces the secretion of pro-inflammatory cytokines and suppresses the phagocytic and bactericidal functions of macrophages against *H. pylori* [48]. These results further strengthened a strong association between *H. pylori* and macrophages in STAD, providing new insights into the clinical management of STAD patients. Based on the coefficient of genes in the RiskScore model, *TNFRSF1B* and *CTLA4* were regarded as the protective factors (coefficient <0), while the *ABCA1*, *IKBIP*, *AKAP5* and *NPC2* were regarded as the risk factors (coefficient >0). *TNFRSF1B* is a tumor necrosis factor receptor that recognizes their cognate ligands (TNF) and promotes the differentiation, clonal expansion and survival of antigen-primed CD8 and CD4 T cells, mediating adaptive immunity to kill cancer cells [49]. Cytotoxic T lymphocyte antigen 4 (*CTLA4*) is an immune checkpoint molecule. Targeting

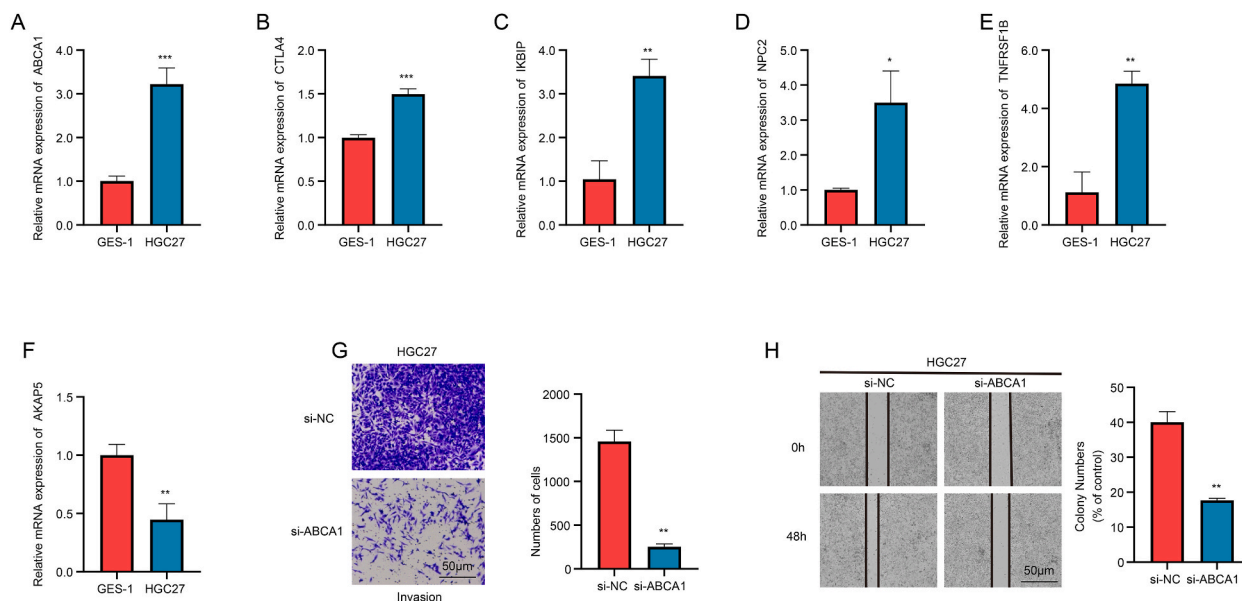


Fig. 7. qPCR, wound healing and invasion assay. (A–F) The expression of model genes in cancer cells and epithelial cells. (G) Wound healing assay for cell migration. (H) Trans-well assay for cell invasion. * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$.

CTLA4 is widely used to activate anti-cancer immune response through stimulating T cell activation [50]. Guan et al. indicated that *CTLA4* enhances macrophage recruitment and increases macrophage proportion in glioblastoma [51]. This indicates that *CTLA4* may influence tumor immunosuppression and progression of STAD by affecting macrophage polarization and, in turn, tumor immunosuppression. ATP binding cassette protein A1 (*ABCA1*) is a crucial molecule in cholesterol homeostasis, and the expression of *ABCA1* is upregulated during the EMT of breast cancer to promote the metastatic capacity of tumor [52]. We also found that the *ABCA1* was overexpressed in the gastric cancer cells, and that its silencing affected the migration and invasion of tumors. Notably, a significantly positive correlation between the RiskScore and EMT indicated that EMT increased tumor spread and metastasis of STAD, and that *H. pylori* infection further exacerbated the EMT process through the activating EMT-related signaling pathways with *ABCA1* as a crucial contributor during the process. I kappa B kinase interacting protein (*IKIP*) is a biomarker that maintains abnormal proliferation of the glioblastoma cells through suppressing the ubiquitination and degradation of *CDK4* [53]. Also, *IKIP* is an immunosuppressive microenvironment biomarker of digestive system malignancies [54]. Zhong et al. revealed that the expression of protein kinase A-anchoring protein 5 (*AKAP5*) is upregulated and is closely associated with the clinical stages in the STAD, whereas low-expressed *AKAP5* can act as a protective factor [55]. In addition, some researchers found that the recruitment of immature macrophages to the TME in lung cancer is inhibited by *NPC2*, and confirmed that *NPC2* is secreted by tumor cells and absorbed by immature macrophages [56]. Macrophages play a dual role in gastritis induced by *H. pylori*, ulcers and gastric cancer. On one hand, *H. pylori* induces macrophage polarization to promote inflammatory responses and eliminate *H. pylori*, on the other hand, macrophages are key cells in the primary immune response against *H. pylori* infection [46]. Analytical method validation is the process used to prove that a test method consistently yields what it is expected to do, and its purpose is to establish that an accurate, precise, and rugged method has been developed [57]. Further experimental results showed that *ABCA1* promoted cancer cell migration and invasion. Taken together, the six prognostic genes identified in this study could influence the process by promoting macrophage-tumor cell interactions and their interactions with other immune cells.

Previous studies have developed long non-coding RNA (lncRNA) signatures related to *H. pylori* infection for predicting the prognosis of gastric cancer patients [58]. Zheng et al. integrated multiple gastric cancer cohorts and identified 28 key prognostic genes for evaluating the risk for patients [59]. However, our model only contained 6 prognostic genes with a high classification effectiveness, indicating that our model had greater potential for clinical application. Although the current model had certain prognostic value, there were also some limitations. Firstly, all the datasets were public databases and more sample data from different populations are required to achieve greater generalization. Secondly, the mechanisms through which these model genes affected macrophages in the STAD microenvironment and their role in *H. pylori* infection have not been validated. Therefore, in the future, *in vivo* experiments should be carried out to explore the specific biological functions of these genes. In addition, bias was inevitable due to the heterogeneity of individual tumor, therefore large-scale clinical trials are needed to provide individualized treatment for STAD patients.

5. Conclusion

This study analyzed the single-cell profile of gastric cancer and annotated a total of 8 cell clusters, among which T cells and dendritic cells accounted for the highest proportion in tumor tissues and the C4 macrophages were closely associated with the *H. pylori*

infection phenotype. Furthermore, a risk prognostic model related to *H. pylori* infection was developed based on the DEGs and the module genes related to C4 subtype. Our model exhibited short- and long-term prognostic value and patients with higher RiskScore were prone to activate typical cancer activation-related pathways. A nomogram model was developed to accurately predict the survival probability of STAD patients. The present discoveries provided new insights for the diagnosis and management of patients with STAD.

Ethics approval and consent to participate

Not applicable.

Consent for publication

Not applicable.

Availability of data and material

The datasets generated during and/or analyzed during the current study are available in the GSE repository [GSE66229] (<https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE66229>) and repository [GSE16279] (<https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE16279>).

Funding

This work was supported by Shaanxi Provincial Health Research Fund (2021D036).

CRediT authorship contribution statement

Jing Zhou: Writing – review & editing, Writing – original draft, Software, Resources, Project administration, Formal analysis, Data curation, Conceptualization. **Li Guo:** Writing – review & editing, Writing – original draft, Validation, Software, Resources, Methodology, Investigation. **Yuzhen Wang:** Supervision, Resources, Project administration, Investigation, Formal analysis. **Lina Li:** Validation, Resources, Project administration, Methodology, Formal analysis. **Yahuan Guo:** Visualization, Validation, Resources, Methodology, Investigation. **Lian Duan:** Visualization, Software, Project administration, Methodology, Investigation. **Mi Jiao:** Visualization, Supervision, Project administration, Investigation, Formal analysis, Data curation. **Pan Xi:** Visualization, Supervision, Resources, Methodology, Formal analysis, Data curation. **Pei Wang:** Writing – review & editing, Visualization, Validation, Supervision, Software, Funding acquisition, Formal analysis, Conceptualization.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgements

Not applicable.

Appendix A. Supplementary data

Supplementary data to this article can be found online at <https://doi.org/10.1016/j.heliyon.2024.e36882>.

Abbreviations

STAD	Stomach adenocarcinoma
DEGs	differential expression gene
KM	Kaplan-Meier
ROC,	Receiver Operating Characteristic Curve
ssGSEA	single sample gene set enrichment analysis
GI	gastrointestinal
CG	chronic gastritis
WHO	World Health Organization
<i>H. pylori</i>	HP, helicobacter pylori
GC	gastric carcinoma
NGS	next generation sequencing

TCGA	the cancer genome atlas
GEO	Gene Expression Omnibus
WGCNA	Weighted gene co-expression network analysis
lasso	Least Absolute Shrinkage and Selection Operator
MSigDB	Molecular Signatures Database
PCA	Principal Component Analysis
GO	Gene Ontology
KEGG	Kyoto Encyclopedia of Genes and Genomes
AUC	Area Under Curve
TSNE	t-distributed stochastic neighbor embedding
DCA	decision curve analysis
TME	tumor microenvironment
EMT	epithelial mesenchymal transition
LncRNA	long non-coding RNA

References

- [1] H. Suzuki, et al., High rate of 5-year survival among patients with early gastric cancer undergoing curative endoscopic submucosal dissection, *Gastric Cancer* 19 (1) (2016) 198–205.
- [2] C. Allemani, et al., Global surveillance of cancer survival 1995–2009: analysis of individual data for 25,676,887 patients from 279 population-based registries in 67 countries (CONCORD-2), *Lancet* 385 (9972) (2015) 977–1010.
- [3] P. Lauren, The two histological main types of gastric carcinoma: diffuse and SO-called intestinal-type carcinoma. AN attempt at a histo-clinical classification, *Acta Pathol. Microbiol. Scand.* 64 (1965) 31–49.
- [4] F. Carneiro, M. Seixas, M. Sobrinho-Simões, New elements for an updated classification of the carcinomas of the stomach, *Pathol. Res. Pract.* 191 (6) (1995) 571–584.
- [5] Z.S. Li, Q. Li, [The latest 2010 WHO classification of tumors of digestive system], *Zhonghua Bing Li Xue Za Zhi* 40 (5) (2011) 351–354.
- [6] Q. Liang, et al., Integrative identification of Epstein-Barr virus-associated mutations and epigenetic alterations in gastric cancer, *Gastroenterology* 147 (6) (2014) 1350, 62.e4.
- [7] L.G.V. Coelho, et al., 14th Brazilian consensus conference on *Helicobacter pylori* infection, *Arq. Gastroenterol.* 55 (2) (2018) 97–121.
- [8] C. Burucoa, A. Axon, Epidemiology of *Helicobacter pylori* infection, *Helicobacter* 22 (Suppl 1) (2017).
- [9] Y. Urita, et al., Role of infected grandmothers in transmission of *Helicobacter pylori* to children in a Japanese rural town, *J. Paediatr. Child Health* 49 (5) (2013) 394–398.
- [10] P. Correa, Human gastric carcinogenesis: a multistep and multifactorial process—first American cancer society award lecture on cancer epidemiology and prevention, *Cancer Res.* 52 (24) (1992) 6735–6740.
- [11] N. Raei, et al., *Helicobacter pylori* infection and dietary factors act synergistically to promote gastric cancer, *Asian Pac. J. Cancer Prev. APJCP* 17 (3) (2016) 917–921.
- [12] C.A. González, et al., Smoking and the risk of gastric cancer in the European prospective investigation into cancer and nutrition (EPIC), *Int. J. Cancer* 107 (4) (2003) 629–634.
- [13] S. Tsugane, Salt, salted food intake, and risk of gastric cancer: epidemiologic evidence, *Cancer Sci.* 96 (1) (2005) 1–6.
- [14] J.C. Machado, et al., A proinflammatory genetic profile increases the risk for chronic atrophic gastritis and gastric carcinoma, *Gastroenterology* 125 (2) (2003) 364–371.
- [15] C.L. Shaffer, et al., *Helicobacter pylori* exploits a unique repertoire of type IV secretion system components for pilus assembly at the bacteria-host cell interface, *PLoS Pathog.* 7 (9) (2011) e1002237.
- [16] C. Xu, R. Ding, C. Xu, Analysis of the effects of different surgical procedures for the treatment of thyroid cancer on the expression levels of IL-17, IL-35, and SIL-2R and the prognostic factors, *Oncologie* 22 (1) (2020) 43–51.
- [17] H. Yang, B. Hu, Immunological perspective: *Helicobacter pylori* infection and gastritis, *Mediat. Inflamm.* 2022 (2022) 2944156.
- [18] Q. Cai, et al., Inflammation-associated senescence promotes *Helicobacter pylori*-induced atrophic gastritis, *Cell Mol Gastroenterol Hepatol* 11 (3) (2021) 857–880.
- [19] D. Brennan, et al., Molecular detection of antibiotic-resistant *Helicobacter pylori*, *Methods Mol. Biol.* 2283 (2021) 29–36.
- [20] A. Hanafiah, B.S. Lopes, Genetic diversity and virulence characteristics of *Helicobacter pylori* isolates in different human ethnic groups, *Infect. Genet. Evol.* 78 (2020) 104135.
- [21] P. Boquet, V. Ricci, Intoxication strategy of *Helicobacter pylori* VacA toxin, *Trends Microbiol.* 20 (4) (2012) 165–174.
- [22] T.L. Cover, R.M. Peek Jr., Diet, microbial virulence, and *Helicobacter pylori*-induced gastric cancer, *Gut Microb.* 4 (6) (2013) 482–493.
- [23] H. Alzoubi, et al., The integrated histopathologic and molecular approach to adult-type diffuse astrocytomas: status of the art, based on the 2021 WHO classification of central nervous system tumors, *Oncologie* 24 (1) (2022) 51–63.
- [24] F. Farinati, et al., *Helicobacter pylori*, inflammation, oxidative damage and gastric cancer: a morphological, biological and molecular pathway, *Eur. J. Cancer Prev.* 17 (3) (2008) 195–200.
- [25] T. Baslan, J. Hicks, Unravelling biology and shifting paradigms in cancer with single-cell sequencing, *Nat. Rev. Cancer* 17 (9) (2017) 557–569.
- [26] I. Pastushenko, et al., Identification of the tumour transition states occurring during EMT, *Nature* 556 (7702) (2018) 463–468.
- [27] P. Savas, et al., Single-cell profiling of breast cancer T cells reveals a tissue-resident memory subset associated with improved prognosis, *Nat Med* 24 (7) (2018) 986–993.
- [28] R.-T.-C. Il, et al., Peri-procedural risk stratification and management of patients with Williams syndrome, *Congenit. Heart Dis.* 12 (2) (2017) 133–142.
- [29] L.A. Cooper, et al., PanCancer insights from the Cancer Genome Atlas: the pathologist's perspective, *J. Pathol.* 244 (5) (2018) 512–524.
- [30] E. Clough, T. Barrett, The gene expression Omnibus database, *Methods Mol. Biol.* 1418 (2016) 93–110.
- [31] A. Liberzon, et al., The Molecular Signatures Database (MSigDB) hallmark gene set collection, *Cell Syst* 1 (6) (2015) 417–425.
- [32] Z. Tan, et al., Comprehensive analysis of scRNA-Seq and bulk RNA-Seq reveals dynamic changes in the tumor immune microenvironment of bladder cancer and establishes a prognostic model, *J. Transl. Med.* 21 (1) (2023) 223.
- [33] N. Chen, et al., Identification of HBEGF+ fibroblasts in the remission of rheumatoid arthritis by integrating single-cell RNA sequencing datasets and bulk RNA sequencing datasets, *Arthritis Res. Ther.* 24 (1) (2022) 215.
- [34] S. Aibar, et al., SCENIC: single-cell regulatory network inference and clustering, *Nat. Methods* 14 (11) (2017) 1083–1086.
- [35] S. Hänzelmann, R. Castelo, J. Guinney, GSEA: gene set variation analysis for microarray and RNA-seq data, *BMC Bioinf.* 14 (2013) 7.

- [36] Y. Geng, et al., Single-cell transcriptome analysis of tumor immune microenvironment characteristics in colorectal cancer liver metastasis, *Ann. Transl. Med.* 10 (21) (2022) 1170.
- [37] B. Chen, et al., Profiling tumor infiltrating immune cells with CIBERSORT, *Methods Mol. Biol.* 1711 (2018) 243–259.
- [38] T. Li, et al., TIMER2.0 for analysis of tumor-infiltrating immune cells, *Nucleic Acids Res.* 48 (W1) (2020) W509–w514.
- [39] C. Liu, et al., The interaction between RACK1 and WEE1 regulates the growth of gastric cancer cell line HGC27, *Oncol. Lett.* 14 (4) (2017) 4784–4792.
- [40] W. Liang, et al., KRT17 promotes T-lymphocyte infiltration through the YTHDF2-CXCL10 Axis in colorectal cancer, *Cancer Immunol. Res.* 11 (7) (2023) 875–894.
- [41] D. Aran, et al., Reference-based analysis of lung single-cell sequencing reveals a transitional profibrotic macrophage, *Nat. Immunol.* 20 (2) (2019) 163–172.
- [42] E. Lahner, et al., Helicobacter pylori immunoproteomics in gastric cancer and gastritis of the carcinoma phenotype, *Expert Rev. Proteomics* 7 (2) (2010) 239–248.
- [43] K.A. Bockerstett, R.J. DiPaolo, Regulation of gastric carcinogenesis by inflammatory cytokines, *Cell Mol Gastroenterol Hepatol* 4 (1) (2017) 47–53.
- [44] P. Malfertheiner, et al., Management of Helicobacter pylori infection—the maastricht IV/florence consensus report, *Gut* 61 (5) (2012) 646–664.
- [45] J. Budzyński, M. Kłopotka, Brain-gut axis in the pathogenesis of Helicobacter pylori infection, *World J. Gastroenterol.* 20 (18) (2014) 5212–5225.
- [46] Y.F. Wei, S.A. Xie, S.T. Zhang, Current research on the interaction between Helicobacter pylori and macrophages, *Mol. Biol. Rep.* 51 (1) (2024) 497.
- [47] Y. Zhuang, et al., Helicobacter pylori-infected macrophages induce Th17 cell differentiation, *Immunobiology* 216 (1–2) (2011) 200–207.
- [48] J. Wen, et al., Notch signaling ligand Jagged1 enhances macrophage-mediated response to Helicobacter pylori, *Front. Microbiol.* 12 (2021) 692832.
- [49] T. So, N. Ishii, The TNF-TNFR family of Co-signal molecules, *Adv. Exp. Med. Biol.* 1189 (2019) 53–84.
- [50] M.M. Halpert, et al., Dendritic cell-secreted cytotoxic T-lymphocyte-associated protein-4 regulates the T-cell response by downmodulating bystander surface B7, *Stem Cells Dev* 25 (10) (2016) 774–787.
- [51] X. Guan, et al., CTLA4-Mediated immunosuppression in glioblastoma is associated with the infiltration of macrophages in the tumor microenvironment, *J. Inflamm. Res.* 14 (2021) 7315–7329.
- [52] S. Prijic, J.T. Chang, ABCA1 expression is upregulated in an EMT in breast cancer cell lines via MYC-mediated de-repression of its proximal ebox element, *Biomedicines* 10 (3) (2022).
- [53] K. Li, et al., IKBIP, a novel glioblastoma biomarker, maintains abnormal proliferation of tumor cells by inhibiting the ubiquitination and degradation of CDK4, *Biochim. Biophys. Acta, Mol. Basis Dis.* 1869 (1) (2023) 166571.
- [54] Q. Li, et al., IKBIP is a predictive biomarker related to immunosuppressive microenvironment in digestive system malignancies, *Discov. Med.* 35 (174) (2023) 57–72.
- [55] Z. Zhong, et al., Low expression of A-kinase anchor protein 5 predicts poor prognosis in non-mucin producing stomach adenocarcinoma based on TCGA data, *Ann. Transl. Med.* 8 (4) (2020) 115.
- [56] T. Kamata, et al., The cholesterol-binding protein NPC2 restrains recruitment of stromal macrophage-lineage cells to early-stage lung tumours, *EMBO Mol. Med.* 7 (9) (2015) 1119–1137.
- [57] B. Seyfinejad, et al., Importance of Method Validation in the Analysis of Biomarker, *Curr. Pharm. Anal.* 18 (6) (2022) 567–569.
- [58] Z. Xin, et al., Helicobacter pylori infection-related long non-coding RNA signatures predict the prognostic status for gastric cancer patients, *Front. Oncol.* 11 (2021) 709796.
- [59] K. Zheng, et al., Integrated analysis of Helicobacter pylori-related prognostic gene modification patterns in the tumour microenvironment of gastric cancer, *Front Surg* 9 (2022) 964203.