Discover Oncology

Analysis

Multi-omics joint screening of biomarkers related to M2 macrophages in gastric cancer

Xilong Wang[1] · **Ying Zhang**[2]

Received: 9 September 2024 / Accepted: 25 November 2024
Published online: 02 December 2024

Abstract

Background Due to high mortality rate and limited treatments in gastric cancer (GC), call for deeper exploration of M2 macrophages as biomarkers is needed.

Methods The data for this study were obtained from the Gene Expression Omnibus (GEO) and Genomic Data Commons (GDC). The Seurat package was utilized for single-cell RNA sequencing (scRNA-seq) analysis. FindAllMarkers was used to identify genes highly expressed among different cell subsets. DESeq2 package was leveraged to screen differentially expressed genes (DEGs), while limma package was utilized for identifying differentially expressed proteins (DEPs). Enrichment analyses of the genes were conducted using KOBAS-i database. MultipleROC was applied to evaluate the diagnostic potential of biomarkers, and rms package was utilized to construct diagnostic models. hTFtarget database was utilized to predict potential transcription factors (TFs). Finally, cell-based assays were performed to validate the expression and potential biological functions of the screened key markers.

Results This study found that M2 macrophages were enriched in protein, endoplasmic reticulum, and virus-related pathways. A total of 4146 DEGs and 1946 DEPs were obtained through screening, with 254 common DEGs/DEPs. The results of gene function enrichment analysis suggested that it may affect the occurrence and development of GC through DNA replication and cell cycle. This study identified three biomarkers, *HSPH1*, *HSPD1*, and *IFI30*, and constructed a diagnostic model based on these three genes. The AUC value greater than 0.8 proved the reliability of the model. Through screening TFs, SPI1 and KLF5 were found to be the common TFs for the three biomarkers. The expression of the three genes *IFI30*, *HSPD1* and *HSPH1* was up-regulated in GC cells, and *IFI30* may play a facilitating role in the migration and invasion of GC cells.

Conclusion This study identified three biomarkers and constructed a diagnostic model, providing a new perspective for the research and treatment of GC.

Keywords Single-cell RNA sequencing · M2 macrophage · Gastric cancer · Diagnostic model · Transcription factor

Abbreviations

GC	Gastric cancer
TNF	Tumor necrosis factor
Th1	Type 1 helper T cells
Th2	Type 2 helper T cells
scRNA-seq	Single-cell transcriptome sequencing
GEO	Gene expression omnibus

✉ Ying Zhang, zhangyingdoctor@163.com | [1]Tumor Hematology Department, Liaoyang Central Hospital, Liaoyang 111000, China. [2]General Surgery Department, Liaoyang Central Hospital, Liaoyang 111000, China.

GDC	Genomic data commons
TCGA	The Cancer Genome Atlas
PCA	Principal component analysis
UMAP	Uniform manifold approximation and projection
KNN	K-nearest neighbor
DEG	Differentially expressed gene
DEP	Differentially expressed proteins
GO	Gene ontology
KEGG	Kyoto Encyclopedia of Genes and Genomes
BP	Biological process
CC	Cellular component
MF	Molecular function
ROC	Receiver operating characteristic
DCA	Decision curve analysis
TF	Transcriptional factors
TME	Tumor microenvironment
NF-κB	Nuclear factor-κB
HSP	Heat shock protein
mtDNA	Mitochondrial DNA

1 Introduction

Gastric cancer (GC) is a common malignancy, ranking fifth in both global cancer incidence and related mortality causes [1, 2]. East Asia and South America are hotspots for the incidence and mortality of GC, and *Helicobacter pylori* is the primary cause of GC [3]. Due to the fact that GC is usually diagnosed at an advanced stage, its mortality rate is high [4]. Systemic chemotherapy, radiotherapy, and surgery are common treatment methods for GC; howbeit, the heterogeneity of the tumor has led to the poor prognosis [5]. In recent years, the development of immunotherapy and targeted therapy has provided new directions for the treatment of GC, but its research is still limited [6]. In summary, GC remains a fatal disease, and current treatment options or early detection strategies cannot control it in a meaningful way, requiring further research.

Macrophages, originating from bone marrow cells, are vital members of the innate immune response [7]. They are a type of white blood cell distributed in tissues and originate from monocytes [8]. Macrophages are characterized by plasticity and versatility, playing crucial roles in tissue development and homeostasis, clearing cellular debris, eliminating pathogens, and regulating inflammatory responses [9, 10]. Macrophages existing in different tissues polarize according to their environmental changes, forming distinct macrophage subtypes [11]. Based on their activation status and functions, macrophages can be categorized into M1 (classically activated macrophages) and M2 (alternatively activated macrophages) [12, 13]. M1 macrophages initiate pro-inflammatory responses and produce pro-inflammatory cytokines such as IL-6, IL-12, and tumor necrosis factor (TNF) to induce Th1 (Type 1 helper T cells) immune responses [14]. Although tumor-associated macrophages do not strictly adhere to the M1 and M2 classifications, they predominantly exhibit M2-like characteristics and promote tumor proliferation by fostering immune suppression [15]. M2 macrophages predominantly secrete anti-inflammatory cytokines like Arginase-I, IL-10, and TGF-β, suppressing Th1 immune responses, promoting Th2 (Type 2 helper T cells) immunity, and participating in wound healing and tissue remodeling [16]. In the research of Chen et al., macrophage secreted proteins were characterised and their results showed that M2 macrophages were recruited by tumour cells. It was also found that CHI3L1 secreted by M2 can promote breast cancer cell metastasis by activating the MAPK signalling pathway, which in turn promotes breast cancer cell metastasis [17]. However, there is limited research on M2 macrophages in relation to cancer, and their potential as biomarkers in cancer remains to be further explored.

The development of single-cell RNA sequencing (scRNA-seq) technology has significantly enhanced our understanding of biological systems [18, 19]. By isolating individual cells, capturing their transcripts, and generating sequencing libraries at the single-cell level, scRNA-seq can reveal the states and functions of single cells [20]. In recent years, scRNA-seq technology has been widely applied in cancer research [21–23]. Based on this, the current study aims to identify M2 macrophage-related genes through single-cell sequencing, and jointly employ transcriptome sequencing and proteomics

sequencing to search for M2 macrophage-associated biomarkers that play a role in GC, ultimately establishing a diagnostic model to provide new insights into the subsequent research and treatment of GC.

2 Material and methods

2.1 Data acquisition

The GC single-cell sequencing dataset GSE163558 was downloaded from Gene Expression Omnibus (GEO) https://www.ncbi.nlm.nih.gov/geo/). This dataset comprises single-cell transcriptomic data from three primary GC cases. The The Cancer Genome Atlas (TCGA) GC dataset was retrieved from the Genomic Data Commons (GDC) database (https://portal.gdc.cancer.gov/). Through rigorous screening, only high-quality (class A) samples were retained, ultimately yielding 377 cancer samples and 34 matched adjacent normal tissue control samples. Furthermore, we integrated proteomic expression data from previous research [24], encompassing 84 GC tumor samples and 84 adjacent normal tissue control samples. Additionally, data on 15 pro-inflammatory factor genes were obtained from prior studies [25].

2.2 Filtering, dimension reduction, and clustering of scRNA-seq data

The Seurat package [26] was utilized to process scRNA-seq data. Initially, the data were first read and subjected to a series of filtering steps aimed at excluding cells that did not meet specific quality criteria. These criteria included the number of genes expressed in the cells, which was required to be between 200 and 10,000, as well as a limitation on the proportion of mitochondrial gene expression, which was not allowed to exceed 10% [18]. Through this rigorous screening process, 9446 high-quality cells were successfully retained from the original dataset for subsequent in-depth analysis. Subsequently, the SCTransform function was applied to normalize the data, effectively reducing technical variability while preserving important biological differences. Following normalization, principal component analysis (PCA) was performed to reduce the dimensionality of the data, making it more manageable for subsequent analyses. To address potential batch effects across samples, the PCA results were corrected using the harmony package [27], ensuring the accuracy and reliability of our analytical outcomes. Next, the uniform manifold approximation and projection (UMAP) algorithm [28] was employed for further dimension reduction, enabling the visualization of cellular similarities and differences in a two-dimensional space. Based on the top 20 principal components, a K-nearest neighbor (KNN) graph was constructed using the FindNeighbors function [26], which leverages Euclidean distance. With the KNN graph established, clustering analysis was conducted on the cells using the FindClusters function [26]. The overall cell population, a resolution parameter of 0.05 was set for clustering. For the clustering of macrophages, the resolution parameter was adjusted to 0.1. Finally, the clustered cells were annotated by leveraging the Cellmarker2.0 database [29].

2.3 Differential expression analysis among cell subpopulations

During the differential expression analysis among cell subpopulations, the FindAllMarkers function [26] was utilized to identify specifically highly expressed genes between different subpopulations. For the comparison among general subpopulations, the parameters were set as follows: logfc.threshold = 0.10, min.pct = 0.25, and only.pos = T. For the identification of highly expressed genes in macrophages, the parameters were adjusted slightly to accommodate the characteristics of macrophages: logfc.threshold remains at 0.10, min.pct was lowered to 0.20, and only.pos was set to T [30].

2.4 Differentially expressed genes (DEGs) analysis

The 411 samples, consisting of 377 cancer samples and 34 adjacent tissue control samples obtained previously, were subjected to differential gene expression analysis between the cancer group and the control group using the DESeq2 package [31]. Significant DEGs were selected based on the criteria of $|\log2FC| \geq 1$ and $p.adj < 0.01$.

2.5 Differentially expressed proteins (DEPs) analysis

A total of 168 samples, including 84 GC tumor samples and 84 adjacent non-tumor control samples, were distinguished into tumor and control groups. The limma package was employed to calculate the differences in protein expression levels between the cancer group and the control group [30]. Subsequently, significantly differentially expressed proteins (DEPs) were screened based on the criteria of $|\log_2 FC| \geq 0.585$ (Foldchange = 1.5) and $p.adj < 0.01$.

2.6 Functional enrichment analysis

The gene set was uploaded to the KOBAS-i database (http://bioinfo.org/kobas) for Gene Ontology (GO) and Kyoto Encyclopedia of Genes and Genomes (KEGG) functional enrichment analyses, with the objective of identifying significantly enriched ($p < 0.05$) functional pathways. GO enrichment analysis included Biological Process (BP), Cellular Component (CC) and Molecular Function (MF) analysis.

2.7 Construction of diagnostic model

Firstly, the diagnostic potential of individual biomarkers was evaluated using the multipleROC package [32]. Subsequently, the rms package [33] was employed to construct the diagnostic model, and a visual nomogram was generated to present the model's predictive outcomes. To further validate the model's performance, the corresponding receiver operating characteristic (ROC) curve of the nomogram, as well as the calibration curve and decision curve analysis (DCA), were plotted to evaluate the predictive capability and reliability of the model.

2.8 Construction of the transcriptional factor (TF) regulatory network for biomarkers

To explore the potential regulatory mechanisms of biomarkers within gastric tissue, the hTFtarget database (https://guolab.wchscu.cn/hTFtarget/#!/) was utilized to predict the potential TFs associated with these biomarkers in gastric tissue. Subsequently, to gain a deeper understanding of the molecular regulatory network surrounding these biomarkers, the ENCORI database (https://rnasysu.com/encori/) was employed to predict the miRNAs targeted by these biomarkers. A rigorous selection process was implemented to retain those predictions with experimental validation counts of no less than 5 only, thereby ensuring the reliability and accuracy of the predictions. Ultimately, to visually represent these intricate regulatory relationships, the aforementioned predictions were imported into Cytoscape 3.8.0 [34], constructing of a comprehensive TF regulatory network.

2.9 Cell acquisition, culture and transfection

Human gastric epithelial cells (GES-1) and the gastric cancer cell line (AGS) were obtained from Cobioer Company in Nanjing, China. These cells were cultured in DMEM medium (GIBCO) which was supplemented with 10% fetal bovine serum (FBS) and 1% L-glutamic acid and were maintained in a 5% CO_2 atmosphere at 37℃. The IFI30-specific small interfering RNA (sense strand: 5′-GGAGAGAGGACAGACGAGAGTT-3′, antisense strand: 5′-CUCUCGUCUGUCCUCUCUCCTT-3′) was purchased from Sangon company (Shanghai, China) and used for the transfection assay based on the manufacturer's instructions, the Lipofectamine 3000 (ThermoFisher) was used for the lentivirus infection based on the previous transfection protocols. Total RNA extraction was performed utilizing the TRIzol Reagent (Invitrogen), and cDNA synthesis was achieved with the ReverTra qPCR RT Master Mix kit (TOYOBO). Subsequently, qPCR was carried out on the LightCycler 96 (Roche) using the SYBR Green PCR Master Mix (Invitrogen), following the provided protocols along with specific primers

Table 1 The sequence of specific primers

Name	forward (5′—3′)	Reversed (5′—3′)
HSPH1	GCTGGTCAACTTGGTGGTG	CTTGGGTCTCCTTGGGTTG
HSPD1	GGCTGATGCCGAGATCAAG	CGAGTTGATGGTGTTGGAGG
IFI30	GAGGACTACAGGCAGGACG	TGGTACAGGGTGAGGTGTTG
GAPDH	GGTGGTCTCCTCTGACTTCAACA	GTTGCTGTAGCCAAATTCGTTGT

(Table 1). The expression levels were calculated using the $2^{-\Delta\Delta CT}$ method, with GAPDH serving as a reference. Each group comprised three samples and technical replicates.

2.10 Wound healing and transwell assay

In this research, we investigated cell migration capacity through a wound healing assay. 4×10^5 resuspended AGS cells in 10 ml of medium and transfer them into a 10 cm dish. When cell confluence reaches 95%, utilize the tip of a 100 µl pipette to create scratches in the cell layer, ensuring all wounds are of uniform width. Subsequently, rinse the scratches with PBS, and place the samples in a complete medium supplemented with 1% fetal bovine serum, incubating them at 37 °C with 5% CO_2. The width of the cell wounds was assessed using an inverted microscope at both 0 and 48 h post-scratching.

For conducting invasion assays, AGS cells were resuspended to a concentration of 2×10^4 cells/ml. A 200 µl aliquot of this cell suspension was then placed in the upper chamber of a Matrigel-coated transwell, which was serum-free. Meanwhile, the lower chamber was filled with 500 µl of complete medium supplemented with 10% fetal bovine serum. The Transwell apparatus was incubated at 37 °C in a 5% CO_2 environment for 48 h. Subsequently, the cells that migrated to the lower chamber were stained with a 0.1% crystal violet solution for a duration of 15 min and were then counted using an optical microscope [35].

2.11 Statistical analysis

All statistical evaluations were executed utilizing the R programming language (version 3.6.0). For comparing differences between two groups of continuous variables, the Wilcoxon rank-sum test was applied. And statistical significance set at $p<0.05$. Sangerbox (http://sangerbox.com/) was leveraged as a supportive resource, offering valuable assistance in conducting the analysis [36].

3 Results

3.1 Single-cell atlas of GC

After cellular filtering, normalization, batch effect removal across samples, dimensionality reduction, and clustering, we identified eight major cell subsets (Fig. 1A–B), namely, T cell, neutrophil, epithelial cell, fibroblast, B cell, plasma cell, mast cell, and the focus of this study, macrophage. We marked T cells with *CCL5*, *CD2*, and *NKG7*, neutrophils with *S100A9* and *S100A8*, epithelial cells with *KRT8*, *KRT18*, and *KRT19*, macrophages with *LYZ*, *C1QB*, and *C1QA*, fibroblasts with *DCN* and *LUM*, B cells with *CD79B* and *CD79A*, plasma cells with *IGHG1*, *IGHG3*, and *IGHG4*, as well as mast cells with *KIT* and *TPSB2*. We further plotted a bubble chart of biomarker expression levels (Fig. 1C), and the results revealed that these marker genes are predominantly expressed in their respective major cell subsets.

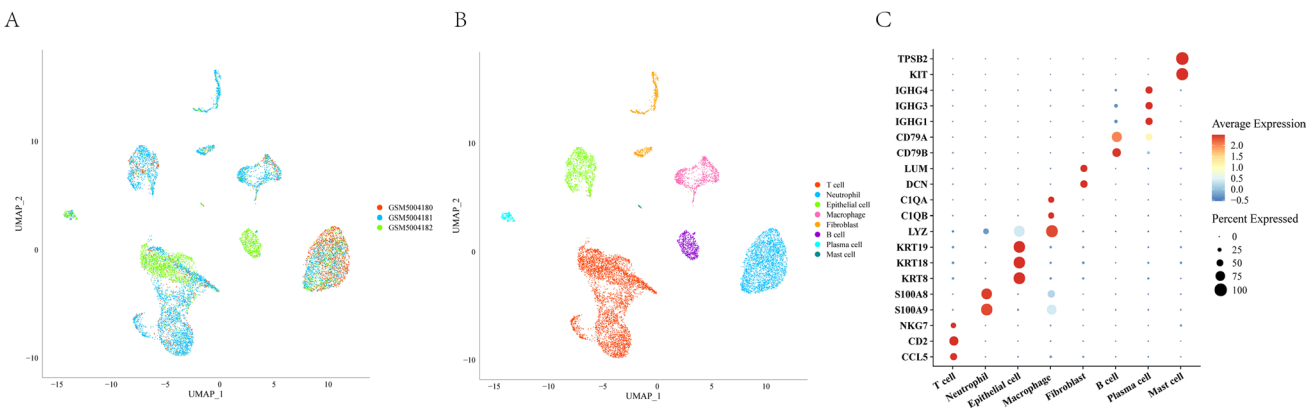


Fig. 1 Single-cell atlas of GC and expression of marker genes **A** Distribution of different samples after batch effect removal. **B** UMAP visualization of the distribution of 8 major cell subpopulations. **C** Expression levels of different cell marker genes.

3.2 Differentiation of macrophage subtypes

We extracted macrophages from the results of single-cell analyses and classified macrophages into two major cell subpopulations, cluster_0 and cluster_1, based on the Seurat package (Fig. 2A). Due to the dynamic conversion between M1 and M2 macrophages, it is difficult to distinguish them through markers alone. Therefore, we scored pro-inflammatory factors in the two groups using AddModuleScore (Fig. 2B). The results showed that the pro-inflammatory score of cluster_1 was significantly higher than that of cluster_0. Based on this, we labeled cluster_0 as M2 macrophages and cluster_1 as M1 macrophages (Fig. 2C). Subsequently, we conducted GO enrichment analysis on the specifically highly expressed genes in these two subtypes of macrophages. The results showed that M2 macrophages were mainly involved in entries such as protein targeting to endoplasmic reticulum (ER), establishment of protein localization to endoplasmic reticulum, andviral transcription, most of which are related to proteins, endoplasmic reticulum, and viruses (Fig. 2D). On the other hand, M1 macrophages were primarily involved in entries like neutrophil mediated immunity, neutrophil activation, and antigen processing and presentation, all of which are related to immunity (Fig. 2E). The enrichment results of M1 and M2 macrophages demonstrated their functional differences.

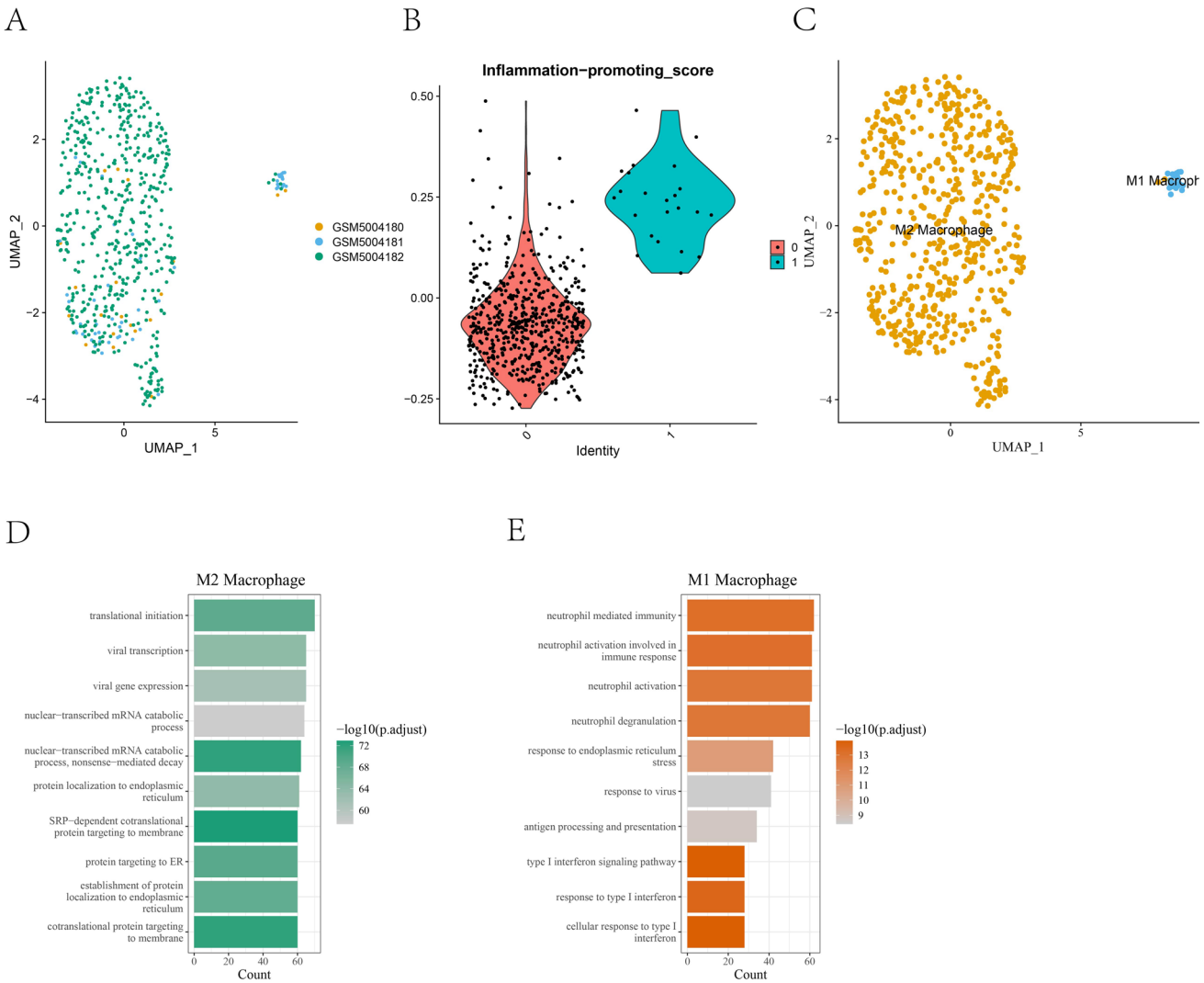


Fig. 2 Differentiation of macrophage subtypes and enrichment analysis of specifically highly expressed genes **A** UMAP visualization of the distribution of macrophage subtypes across different samples. **B** Violin plot of pro-inflammatory factor scores for cluster_0 and cluster_1. **C** UMAP visualization of the distribution of M2 and M1 macrophages. **D** Bar chart showing the top 10 enriched terms from the enrichment analysis of specifically highly expressed genes in M2 macrophages. **E** Bar chart showing the top 10 enriched terms from the enrichment analysis of specifically highly expressed genes in M1 macrophages.

3.3 Screening of DEGs and DEPs

Through screening, a total of 4146 DEGs were obtained, including 1963 up-regulated genes and 2183 down-regulated genes. The volcano plot of gene differential expression is shown in Fig. 3A. We selected the top 20 genes with the most significant up- and down-regulation to plot a heatmap of expression levels (Fig. 3B). Similarly, 1946 DEPs were identified through screening, comprising 1,306 up-regulated proteins and 640 down-regulated proteins. The volcano plot of protein differential expression is presented in Fig. 3C. We chose the top 20 proteins with the most significant up- and down-regulation to illustrate the heatmap of expression levels (Fig. 3D). Subsequently, the results from transcriptomics were integrated with those from proteomics to generate a nine-quadrant plot (Fig. 3E). The results revealed that there were 118 commonly up-regulated DEGs/DEPs and 136 commonly down-regulated DEGs/DEPs shared between the two datasets.

3.4 Gene function enrichment analysis

We conducted a functional enrichment analysis on the aforementioned total of 254 DEGs. The results indicated significant enrichment in GO terms such as DNA replication, DNA strand elongation involved in DNA replication, and DNA replication origin binding (Fig. 4A–C). Meanwhile, in KEGG enrichment analysis, the genes were primarily implicated in pathways like DNA replication and cell cycle (Fig. 4D). These findings suggested that the genes may influence the initiation and progression of tumors through mechanisms involving DNA replication and the cell cycle.

3.5 Construction of a diagnostic model based on M2 macrophage-related biomarkers

To establish a diagnostic model, we intersected the genes specifically overexpressed in M2 macrophages with those differentially upregulated in both transcriptomic and proteomic analyses, yielding three biomarkers associated with M2 macrophages in GC: *HSPH1*, *HSPD1*, and *IFI30* (Fig. 5A). To evaluate the diagnostic efficacy of these three biomarkers in GC, ROC curves were plotted for each biomarker, as shown in Fig. 5B. The AUC values for all three biomarkers exceeded 0.8, indicating excellent diagnostic performance. Furthermore, to quantify the risk assessment for GC patients, we developed a nomogram incorporating the expression levels of these three biomarkers (Fig. 5C). Initially, the ROC curve was utilized to assess the model's performance, revealing an AUC of 0.982 (Fig. 5D), suggesting a highly reliable model. Subsequently, the calibration curve was employed to evaluate the predictive accuracy of the model, demonstrating close alignment with the ideal curve, indicating good predictive performance of the nomogram (Fig. 5E). Additionally, the DCA was utilized to assess the model's reliability, revealing that the benefit of the nomogram was significantly higher than that of any individual gene, underscoring the robust predictive capability of our model (Fig. 5F).

3.6 TF regulatory network

We predicted and screened for TFs that regulate the expression of the three biomarkers in gastric tissues. For *HSPH1*, four TFs were identified; for *HSPD1*, three TFs were selected; and for *IFI30*, two TFs were found. Notably, SPI1 and KLF5 were common TFs for all three genes, GATA4 was shared by *HSPD1* and *HSPH1*, and GATA6 was unique to *HSPH1*. Subsequently, we predicted the miRNAs that target these biomarkers. The results showed that *HSPH1* had 19 experimentally validated target miRNAs with more than five occurrences, *HSPD1* had 16, and *IFI30* had 7. The detailed TF regulatory network is presented in Fig. 6.

3.7 Validation of in vitro cellular assays for key markers

To further validate our screened key markers, we first validated the mRNA expression levels of the three genes based on qRT-PCR. As shown in Fig. 7A, the mRNA expression levels of *IFI30*, *HSPD1* and *HSPH1* were significantly up-regulated in AGS cells relative to controls. Due to the significantly high expression level of *IFI30* in AGS cells, we evaluated its possible role in affecting cancer cells ($P < 0.0001$). As shown in Fig. 7B, we observed that silencing of *IFI30* significantly inhibited the migratory ability of AGS cells. In addition, *IFI30* silencing also significantly reduced the invasive ability of AGS cells

Fig. 3 Screening and expression analysis of DEGs and DEPs **A** Volcano plot of DEGs, where red represents up-regulated DEGs, blue represents down-regulated DEGs, and gray represents non-differentially expressed genes. **B** Heatmap of gene expression levels, with blue indicating low expression and red indicating high expression. **C** Volcano plot of DEPs. **D** Heatmap of protein expression levels. **E** Nine-quadrant plot, where red indicates genes and their corresponding proteins that are differentially up-regulated in both transcriptomics and proteomics, and blue indicates genes and their corresponding proteins that are differentially down-regulated in both datasets. ▶

relative to the control group (Fig. 7C). These results *IFI30* may play a promotional role in the invasion and migration of GC cells, suggesting its potential oncogenic role in GC progression.

4 Discussion

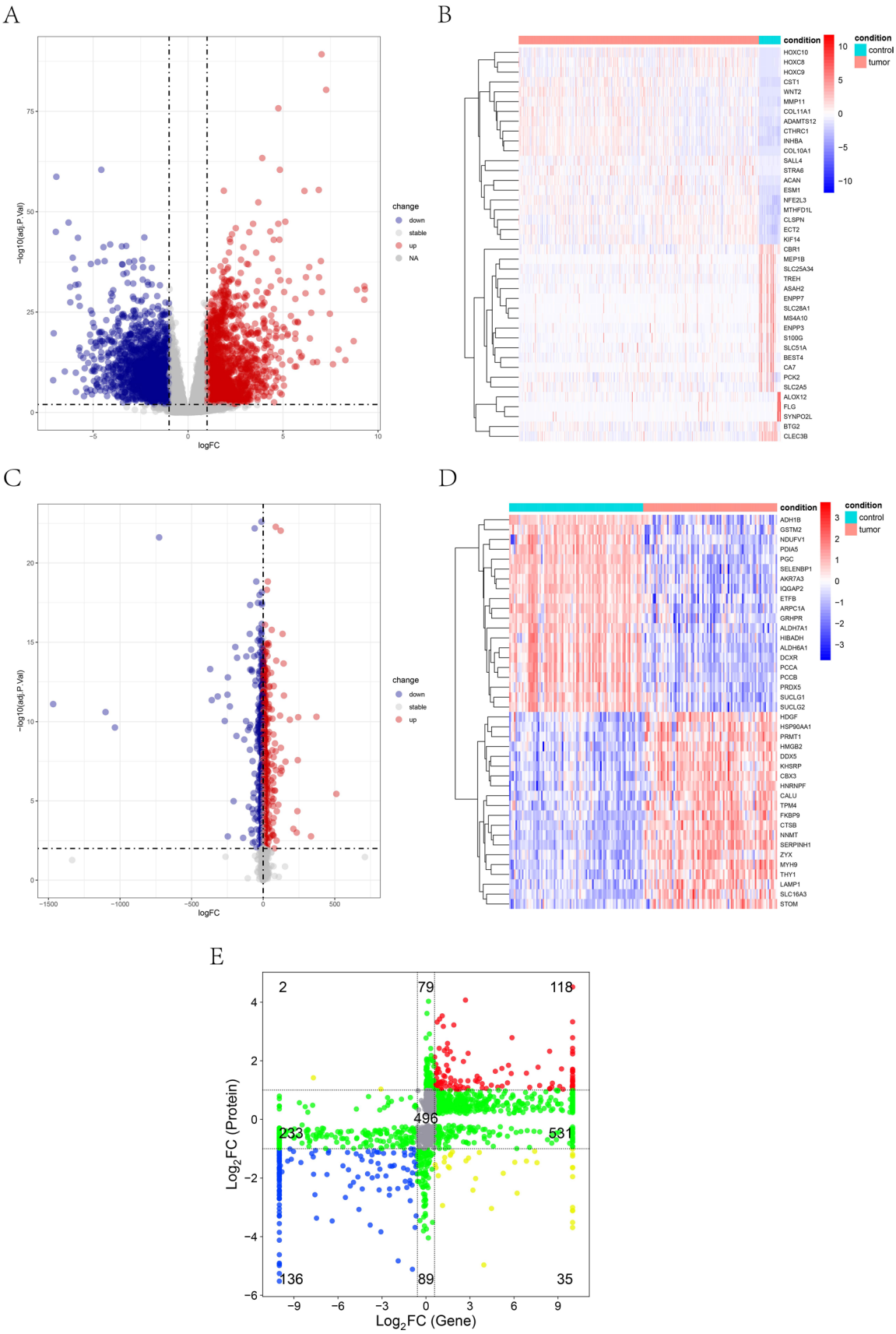
As a highly prevalent malignant tumor, GC has an annual increase of over one million new cases worldwide, with a notably high mortality rate [37]. M2 macrophages are capable of conducting anti-inflammatory responses [38]. Recent studies have indicated that M2 macrophages can promote cancer progression by inhibiting anti-tumor immune responses [39]. This research has identified three biomarkers related to M2 macrophages, namely *HSPH1*, *HSPD1*, and *IFI30*, and has established a diagnostic model with good predictive performance. It provides potential intervention targets and therapeutic strategies, offering new possibilities for the treatment of GC.

The enrichment analysis of this study revealed that M2 macrophages are primarily involved in pathways related to proteins, endoplasmic reticulum, and viruses, while M1 macrophages mainly participate in immune-related pathways. Generally, M1 macrophages exert immune protection by secreting proinflammatory cytokines, whereas M2 macrophages, characterized by their anti-inflammatory properties, contribute to tissue remodeling and tumor progression [40]. During the early stages of repair, approximately 85% of macrophages in the wound exhibit the M1 proinflammatory phenotype, which shifts to approximately 80–85% of anti-inflammatory M2 macrophages by day 5–7 [41]. In TME, macrophage polarization is a continuous process spanning the two extremes from M1 to M2 macrophages [42]. The transition between M1 and M2 is closely related to the timing of inflammation, and the M2 macrophage-associated biomarkers discovered in this study offer a novel perspective for diagnosis of GC.

This study has identified three biomarkers: *HSPH1*, *HSPD1*, and *IFI30*. *HSPH1* is a pivotal member of the *HSP70* superfamily, not only serves as a crucial component of the β-catenin degradation complex but also exhibits unique functions as a molecular chaperone in cancer contexts [43, 44]. It possesses anti-aggregation capabilities and collaborates with *HSP70*'s protein refolding activity, jointly safeguarding intracellular protein homeostasis [45]. Prior studies have demonstrated that *HSPH1* exhibits a notable upregulation trend in various malignancies, including head and neck cancer [46] and breast cancer [47]. In the tumor microenvironment (TME), *HSPH1* plays intricate regulatory roles, implicating multiple signaling pathways intimately linked to inflammation [48]. Specifically, *HSPH1* directly interacts with STAT3, facilitating its phosphorylation and activation, thereby influencing the expression of downstream genes [49]. Additionally, *HSPH1* participates in regulating the hyperactivation of Wnt signaling and maintaining the persistent activation of the nuclear factor-κB (NF-κB) signaling pathway [50]. Notably, *HSPH1* can be actively secreted by cancer cells into the extracellular space, further promoting the polarization of macrophages towards a precancerous phenotype conducive to tumor growth [51].

HSPD1 also known as *HSP60*, belongs to the heat shock protein (HSP) family along with *HSPH1* [52]. This family of proteins plays a pivotal role in maintaining the cellular internal environment homeostasis. Notably, cancer cells cleverly harness the natural protective mechanisms of HSPs during malignant transformation to promote their survival and proliferation [53]. *HSPD1*, primarily localized in the mitochondrial matrix, functions as a crucial molecular chaperone that is indispensable for the integrity of mitochondrial function. It is responsible for ensuring the folding and refolding of nuclear-encoded proteins as well as peptides encoded by mitochondrial DNA (mtDNA) [54]. Additionally, *HSPD1* demonstrates a unique role in immune system activation by actively driving the activation processes of T cells and B cells and regulating the production of various interferons and interleukins, thereby participating in immune responses [55]. Research has shown that downregulation of *HSPD1* can trigger apoptosis in cancer cells across multiple cancer types, including oral cancer [56] and non-small cell lung cancer [57]. Furthermore, studies have also revealed that the HSP family, including *HSP60*, is involved in the process which macrophages impact cancer [58].

IFI30 is a reductase located in endosomes, lysosomes, and phagosomes [59]. In the exploration of the biological mechanisms of malignant tumors, IFI30 plays an important role in malignant diseases such as melanoma [60] and prostate cancer [61]. Particularly noteworthy is the finding that IFI30 exhibits highly expressed characteristics in malignant glioma subtypes,


A
−log10(adj.P.Val)
logFC
change
down
stable
up
NA
B
condition
control
tumor
HOXC10
HOXC8
HOXC9
CST1
WNT2
MMP11
COL11A1
ADAMTS12
CTHRC1
INHBA
COL10A1
SALL4
STRA6
ACAN
ESM1
NFE2L3
MTHFD1L
CLSPN
ECT2
KIF14
CBR1
MEP1B
SLC25A34
TREH
ASAH2
ENPP7
SLC28A1
MS4A10
ENPP3
S100G
SLC51A
BEST4
CA7
PCK2
SLC2A5
ALOX12
FLG
SYNPO2L
BTG2
CLEC3B
C
−log10(adj.P.Val)
logFC
change
down
stable
up
D
condition
control
tumor
ADH1B
GSTM2
NDUFV1
PDIA5
PGC
SELENBP1
AKR7A3
IQGAP2
ETFB
ARPC1A
GRHPR
ALDH7A1
HIBADH
ALDH6A1
DCXR
PCCA
PCCB
PRDX5
SUCLG1
SUCLG2
HDGF
HSP90AA1
PRMT1
HMGB2
DDX5
KHSRP
CBX3
HNRNPF
CALU
TPM4
FKBP9
CTSB
NNMT
SERPINH1
ZYX
MYH9
THY1
LAMP1
SLC16A3
STOM
E
Log2FC (Protein)
Log2FC (Gene)
2
79
118
496
233
581
136
89
35

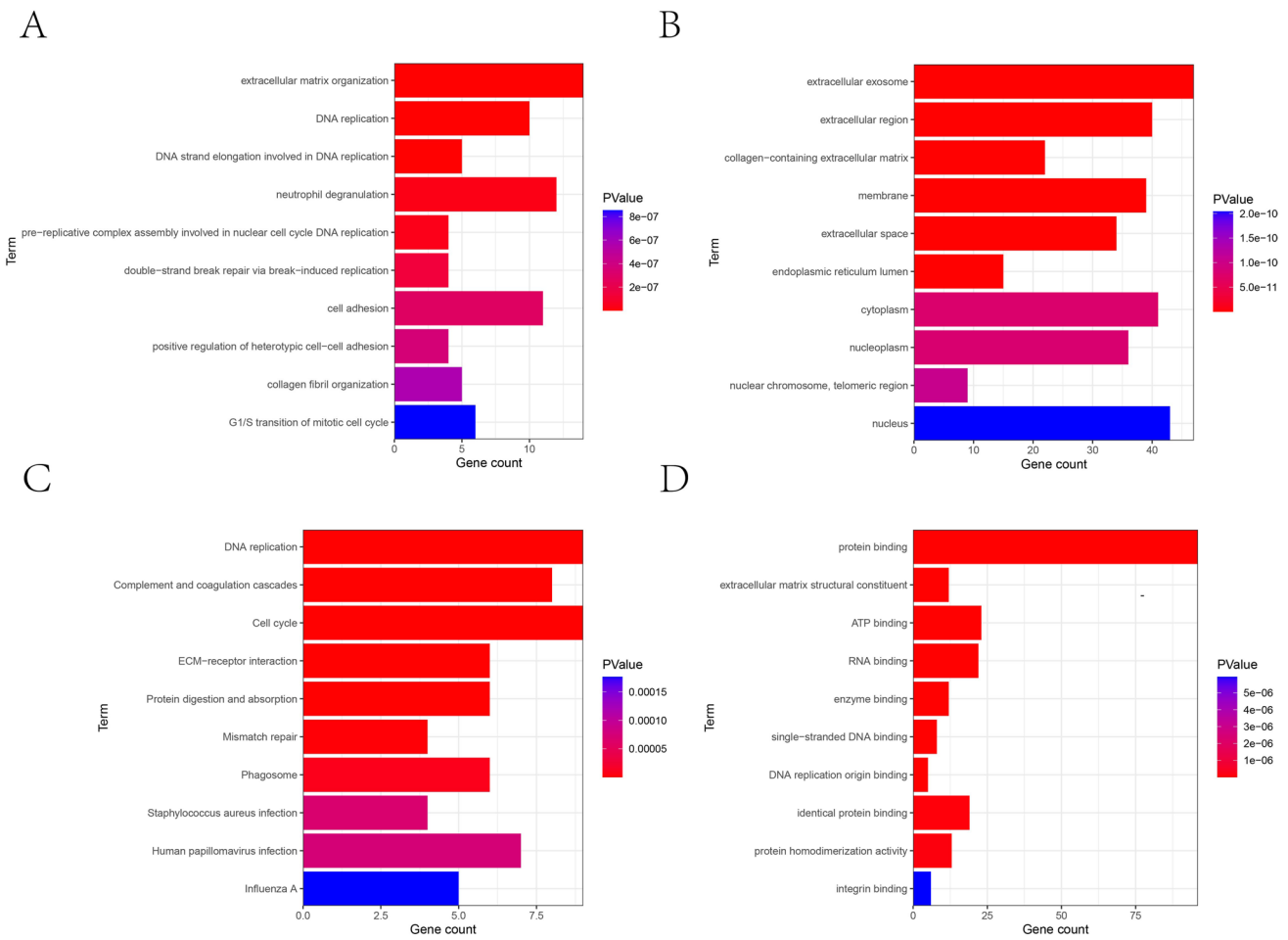


Fig. 4 Gene function enrichment analysis **A** Bar plot of GO enrichment analysis for BP. The x-axis represents the number of genes included in each term, and the color represents the significance p-value, ranging from blue (least significant) to red (most significant). **B** Bar plot of GO enrichment analysis for CC. **C** Bar plot of GO enrichment analysis for MF. **D** Bar plot of KEGG enrichment analysis.

and this phenomenon of high expression is closely related to the significant infiltration of M2 macrophages [62]. In summary, the expression of these three genes, namely *HSPH1*, *HSPD1*, and *IFI30*, may exert an influence on the development of cancers such as GC, potentially through modulating the infiltration of M2 macrophages.

This study successfully identified SPI1 and KLF5 as two TFs that serve as core regulatory elements for three key biomarkers. SPI1, also known as PU.1, emerged from groundbreaking research on Friend's hemoglobinopathy [63]. In macrophages, SPI1 demonstrates its potent regulatory capacity by engaging with closed chromatin to activate the transcription of its target genes [64, 65]. Notably, SPI1 plays a pivotal oncogenic role in various cancers, including breast, lung, colon, and cervical cancers [66–68], and its elevated expression has been firmly linked to poor prognosis and disease progression in GC [69]. On the other hand, KLF5, a prominent member of the KLF transcription factor family, extensively participates in regulating critical biological processes such as cell proliferation, cell cycle control, apoptosis, and cell differentiation [70]. In the field of oncology, KLF5 has garnered significant attention due to its potential roles in multiple cancer types, particularly in gastrointestinal tumors like GC and colon cancers [71, 72]. In conclusion, the revelation of SPI1 and KLF5 in this study not only strengthens the tight correlation between the selected biomarkers and GC but also sheds light on their potential impact on GC progression through mechanisms involving the regulation of M2 immune cell infiltration.

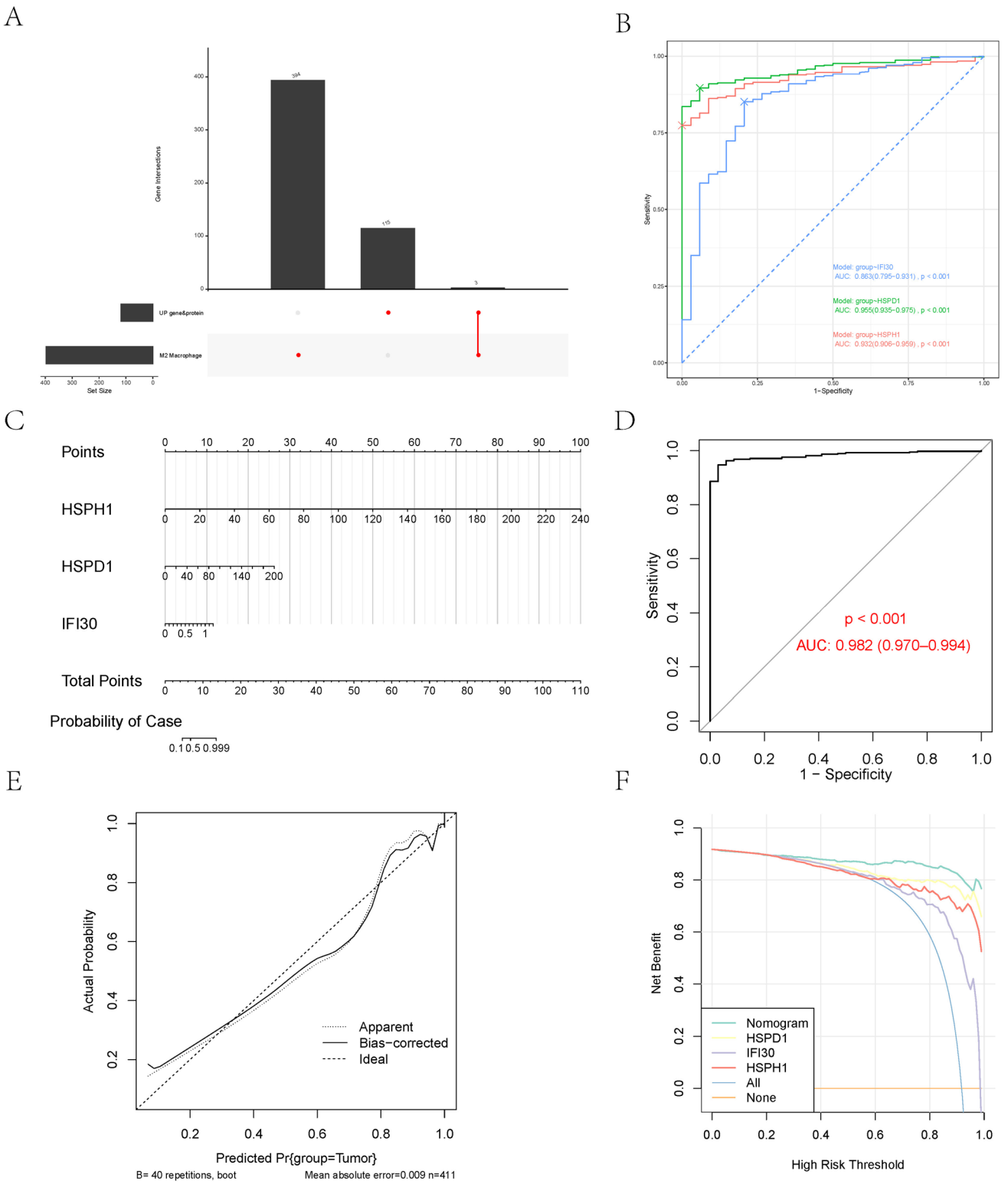


Fig. 5 Construction and reliability validation of the diagnostic model for M2 macrophage-related biomarkers **A** Upset plot of M2 macrophage-specific highly expressed genes and common significantly highly expressed genes/proteins in transcriptomics and proteomics. The left bar chart shows the number of genes in each subset, and the upper bar chart indicates the number of genes in each intersection. **B** ROC curves for *HSPH1*, *HSPD1*, and *IFI30*. **C** Nomogram for biomarkers, with the expression levels of *HSPH1*, *HSPD1*, and *IFI30* in the middle, the corresponding scores at the top, and the risk probability compared with the total score at the bottom. **D** ROC curve of the nomogram. **E** Calibration curve of the nomogram. **F** DCA of the nomogram.

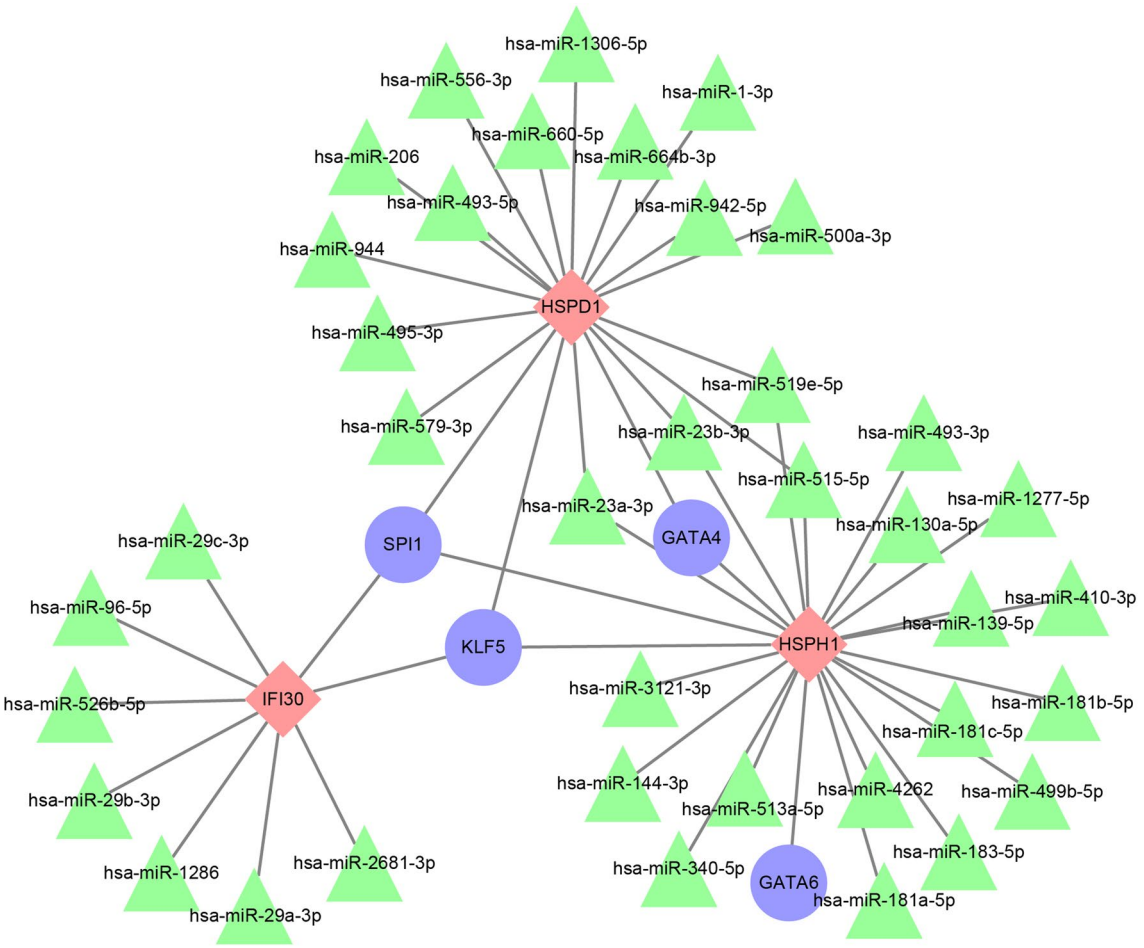


Fig. 6 TF regulatory network diagram Red diamonds represent biomarkers, blue circles represent TFs, and green triangles represent miRNAs

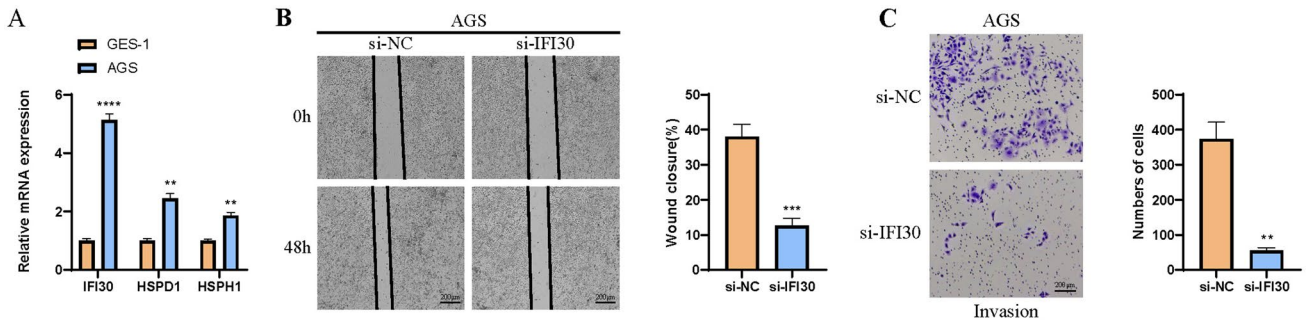


Fig. 7 In-vitro validation assay. **A** The mRNA level of the screened three markers (*IFI30*, *HSPD1*, and *HSPH1*) in GES-1 and AGS cells via qRT-PCR. **B** Wound healing assay to assess the effect of silencing *IFI30* expression on the migratory capacity of AGS cells. **C** Transwell assay to assess the effect of silencing *IFI30* expression on the invasive capacity of AGS cells.

5 Conclusion

This study explored the role of M2 macrophages in the progression of GC and identified three biomarkers: *HSPH1*, *HSPD1*, and *IFI30*. Furthermore, a diagnostic model with good predictive capabilities was established based on these biomarkers. This discovery not only enriches our understanding of GC but also provides new potential targets for early diagnosis and disease assessment of GC.

6 Limitation

This study has achieved preliminary results in exploring the role of M2 macrophages in GC progression and its related biomarkers, yet it also possesses non-negligible limitations. Firstly, constrained by the limited sample size and scarcity of single-cell sequencing data, the broad applicability of the research findings requires further validation. In the future, we plan to integrate data from a wider range of public databases and studies to expand the sample size, thereby enhancing the credibility and universality of our conclusions. Furthermore, although this study has verified the diagnostic efficacy of the biomarkers through ROC analysis, direct evidence from in vivo experiments is crucial for comprehensively assessing their application potential. Therefore, we anticipate conducting in-depth in vivo and in vitro experiments in future research to further validate and optimize the practical application value of these biomarkers in GC diagnosis and treatment.

Acknowledgements None

Author contributions All authors contributed to this present work: [XLW] and [YZ] designed the study. [XLW] and [YZ] collected data and analyzed data. [XLW] and [YZ] drafted the manuscript. [XLW] and [YZ] reviewed and revised the manuscript. The manuscript has been approved by all authors for publication.

Funding The authors declare that they have receive no funding.

Data availability The datasets generated and/or analyzed during the current study are available in the [GSE163558] repository, [https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE163558].

Declarations

Ethics approval and consent to participate Not applicable.

Patient consent statement Not applicable.

Competing interests The authors declare no competing interests.

References

1. Bray F, Laversanne M, Sung H, Ferlay J, Siegel RL, Soerjomataram I, et al. Global cancer statistics 2022: GLOBOCAN estimates of incidence and mortality worldwide for 36 cancers in 185 countries. CA Cancer J Clin. 2024;74:229–63.
2. Alavi Dana SMM, Farkhondeh T, Aschner M, Darroudi M, Samini H, Samarghandian S. Chrysin Effect Against Gastric Cancer: Focus on its Molecular Mechanisms. Curr Mol Pharmacol. 2023;16:707–11.
3. Smyth EC, Nilsson M, Grabsch HI, van Grieken NC, Lordick F. Gastric cancer. The Lancet Diabetes. Endocrinology. 2020;396:635–48.
4. Sexton RE, Al Hallak MN, Diab M, Azmi AS. Gastric cancer: a comprehensive review of current and future treatment strategies. Cancer Metastasis Rev. 2020;39:1179–203.
5. Wagner AD, Syn NLX, Moehler M, Grothe W, Yong WP, Tai BC, et al. Chemotherapy for advanced gastric cancer. Cochrane Database Syst Rev. 2017. https://doi.org/10.1002/14651858.CD004064.pub4.
6. Guan W-L, He Y, Xu R-H. Gastric cancer treatment: recent progress and future perspectives. J Hematol Oncol. 2023;16:57.
7. Boutilier AJ, Elsawa SF. Macrophage polarization states in the tumor microenvironment. Int J Mol Sci. 2021;22:6995.
8. Dijkgraaf EM, Heusinkveld M, Tummers B, Vogelpoel LTC, Goedemans R, Jha V, et al. Chemotherapy alters monocyte differentiation to favor generation of cancer-supporting m2 macrophages in the tumor microenvironment. Can Res. 2013;73:2480–92.
9. Varol C, Mildner A, Jung S. Macrophages: development and tissue specialization. Annu Rev Immunol. 2015;33:643–75.
10. Anderson NR, Minutolo NG, Gill S, Klichinsky M. Macrophage-based approaches for cancer immunotherapy. Can Res. 2021;81:1201–8.
11. Martinez FO, Gordon S. The M1 and M2 paradigm of macrophage activation: time for reassessment. Prime Rep. 2014;6:13.

12. Chávez-Galán L, Olleros ML, Vesin D, Garcia I. Much more than M1 and M2 macrophages, there are also CD169+ and TCR+ macrophages. Front immunol. 2015. https://doi.org/10.3389/fimmu.2015.00263.
13. Abdollahi E, Johnston TP, Ghaneifar Z, Vahedi P, Goleij P, Azhdari S, Moghaddam AS. Immunomodulatory Therapeutic Effects of Curcumin on M1/M2 Macrophage Polarization in Inflammatory Diseases. Curr Mol Pharmacol. 2023;16:2–14.
14. Yunna C, Mengru H, Lei W, Weidong C. Macrophage M1/M2 polarization. Eur J Pharmacol. 2020;877: 173090.
15. Mehla K, Singh PK. Metabolic regulation of macrophage polarization in cancer. Trends Cancer. 2019;5:822–34.
16. Pritchard A, Tousif S, Wang Y, Hough K, Khan S, Strenkowski J, et al. Lung tumor cell-derived exosomes promote M2 macrophage polarization. Cells. 2020;9:1303.
17. Chen Y, Zhang S, Wang Q, Zhang X. Tumor-recruited M2 macrophages promote gastric and breast cancer metastasis via M2 macrophage-secreted CHI3L1 protein. J Hematol Oncol. 2017;10:36.
18. Zulibiya A, Wen J, Yu H, Chen X, Xu L, Ma X, et al. Single-cell RNA sequencing reveals potential for endothelial-to-mesenchymal transition in tetralogy of fallot. Congenit Heart Dis. 2023;18:611–25.
19. Luecken MD, Theis FJ. Current best practices in single-cell RNA-seq analysis: a tutorial. Mol Syst Biol. 2019;15: e8746.
20. Tang F, Barbacioru C, Wang Y, Nordman E, Lee C, Xu N, et al. mRNA-Seq whole-transcriptome analysis of a single cell. Nat Methods. 2009;6:377–82.
21. Ding S, Chen X, Shen K. Single-cell RNA sequencing in breast cancer: Understanding tumor heterogeneity and paving roads to individualized therapy. Cancer Commun. 2020;40:329–44.
22. Kodous AS, Balaiah M, Ramanathan P. Single cell RNA sequencing—a valuable tool for cancer immunotherapy: a mini review. Oncogene. 2023;25:635–9.
23. Zhang J, He J, Chen W, Chen G, Wang L, Liu Y, et al. Single-cell RNA-binding protein pattern-mediated molecular subtypes depict the hallmarks of the tumor microenvironment in bladder urothelial carcinoma. Oncogene. 2024. https://doi.org/10.1515/oncologie-2024-0071.
24. Ge S, Xia X, Ding C, Zhen B, Zhou Q, Feng J, et al. A proteomic landscape of diffuse-type gastric cancer. Nat Commun. 2018;9:1012.
25. He Y, Jiang Z, Chen C, Wang X. Classification of triple-negative breast cancers based on Immunogenomic profiling. J Exp Clin Cancer Res. 2018;37:327.
26. Hao YH, Stuart T, Kowalski MH, Choudhary S, Hoffman P, Hartman A, et al. Dictionary learning for integrative, multimodal and scalable single-cell analysis. Nat Biotechnol. 2024;42:293–304.
27. Korsunsky I, Millard N, Fan J, Slowikowski K, Zhang F, Wei K, et al. Fast, sensitive and accurate integration of single-cell data with Harmony. Nat Methods. 2019;16:1289–96.
28. McInnes L, Healy J. UMAP: Uniform Manifold Approximation and Projection for Dimension Reduction. ArXiv. 2018; abs/1802.03426.
29. Hu C, Li T, Xu Y, Zhang X, Li F, Bai J, et al. Cell marker 2.0: an updated database of manually curated cell markers in human/mouse and web tools based on scRNA-seq data. Nucl Acids Res. 2022;51:870–6.
30. Song Z, Yu J, Wang M, Shen W, Wang C, Lu T, et al. CHDTEPDB: transcriptome expression profile database and interactive analysis platform for congenital heart disease. Congenit Heart Dis. 2023;18:693–701.
31. Love M, Anders S, Huber W. Differential analysis of count data—the DESeq2 package. Genome Biol. 2014;15:10–1186.
32. Du Z, Hao Y. reportROC: an easy way to report ROC analysis. R package version. 2020. 3.
33. Harrell Jr FE, Harrell Jr MFE, Hmisc D. Package 'rms'. Vanderbilt University. 2017. 229.
34. Shannon P, Markiel A, Ozier O, Baliga NS, Wang JT, Ramage D, et al. Cytoscape: a software environment for integrated models of biomolecular interaction networks. Genome Res. 2003;13:2498–504.
35. Zeng Q, Feng K, Yu Y, Lv Y. Hsa_Circ_0000021 Sponges miR-3940-3p/KPNA2 Expression to Promote Cervical Cancer Progression. Curr Mol Pharmacol. 2024;17(1):e170223213775.
36. Shen W, Song Z, Zhong X, Huang M, Shen D, Gao P, et al. Sangerbox: a comprehensive, interaction-friendly clinical bioinformatics analysis platform. Imeta. 2022;1: e36.
37. Rawla P, Barsouk A. Epidemiology of gastric cancer: global trends, risk factors and prevention. Gastroenterol Review/Przegląd Gastroenterol. 2019;14:26–38.
38. Murray Peter J, Allen Judith E, Biswas Subhra K, Fisher Edward A, Gilroy Derek W, Goerdt S, et al. Macrophage activation and polarization: nomenclature and experimental guidelines. Immunity. 2014;41:14–20.
39. Erlandsson A, Carlsson J, Lundholm M, Fält A, Andersson S-O, Andrén O, et al. M2 macrophages and regulatory T cells in lethal prostate cancer. Prostate. 2019;79:363–9.
40. Cai H, Zhang Y, Wang J, Gu J. Defects in macrophage reprogramming in cancer therapy: the negative impact of PD-L1/PD-1. Front Immunol. 2021. https://doi.org/10.3389/fimmu.2021.690869.
41. Aitcheson SM, Frentiu FD, Hurn SE, Edwards K, Murray RZ. Skin wound healing: normal macrophage function and macrophage dysfunction in diabetic wounds. Molecules. 2021;26:4917.
42. Sousa S, Brion R, Lintunen M, Kronqvist P, Sandholm J, Mönkkönen J, et al. Human breast cancer cells educate macrophages toward the M2 activation status. Breast Cancer Res. 2015;17:101.
43. Finka A, Mattoo RUH, Goloubinoff P. Experimental milestones in the discovery of molecular chaperones as polypeptide unfolding enzymes. Annu Rev Biochem. 2016;85:715–42.
44. Mogk A, Bukau B, Kampinga HH. Cellular handling of protein aggregates by disaggregation machines. Mol Cell. 2018;69:214–26.
45. Mattoo RUH, Sharma SK, Priya S, Finka A, Goloubinoff P. Hsp110 is a bona fide chaperone using ATP to unfold stable misfolded polypeptides and reciprocally collaborate with Hsp70 to solubilize protein aggregates. J Biol Chem. 2013;288:21399–411.
46. Fan G, Tu Y, Wu N, Xiao H. The expression profiles and prognostic values of HSPs family members in Head and neck cancer. Cancer Cell Int. 2020;20:220.
47. Zhang K, Jiang K, Hong R, Xu F, Xia W, Qin G, et al. Identification and characterization of critical genes associated with tamoxifen resistance in breast cancer. PeerJ. 2020;8: e10468.
48. Li Y, Zhang N, Zhang L, Song Y, Liu J, Yu J, et al. Oncogene HSPH1 modulated by the rs2280059 genetic variant diminishes EGFR-TKIs efficiency in advanced lung adenocarcinoma. Carcinogenesis. 2020;41:1195–202.

49. Berthenet K, Bokhari Ad, Lagrange A, Marcion G, Boudesco C, Causse S, et al. HSP110 promotes colorectal cancer growth through STAT3 activation. Oncogene. 2017;36:2328–36.
50. Boudesco C, Verhoeyen E, Martin L, Chassagne-Clement C, Salmi L, Mhaidly R, et al. HSP110 sustains chronic NF-κB signaling in activated B-cell diffuse large B-cell lymphoma through MyD88 stabilization. Blood. 2018;132:510–20.
51. Liang Y, Luo J, Yang N, Wang S, Ye M, Pan G. Activation of the IL-1β/KLF2/HSPH1 pathway promotes STAT3 phosphorylation in alveolar macrophages during LPS-induced acute lung injury. 2020. Biosci Rep. https://doi.org/10.1042/BSR20193572.
52. Yun CW, Kim HJ, Lim JH, Lee SH. Heat shock proteins: agents of cancer development and therapeutic targets in anti-cancer therapy. Cells. 2020;9:60.
53. Ghane Shahrbaf F, Assadi F. Drug-induced renal disorders. J Renal Inj Prev. 2015;4:57–60.
54. Klebl DP, Feasey MC, Hesketh EL, Ranson NA, Wurdak H, Sobott F, et al. Cryo-EM structure of human mitochondrial HSPD1. iScience. 2021. https://doi.org/10.1016/j.isci.2020.102022.
55. Kim SK, Kim K, Ryu JW, Ryu TY, Lim JH, Oh JH, et al. The novel prognostic marker, EHMT2, is involved in cell proliferation via HSPD1 regulation in breast cancer. Int J Oncol. 2019;54:65–76.
56. Kang B-H, Shu C-W, Chao J-K, Lee C-H, Fu T-Y, Liou H-H, et al. HSPD1 repressed E-cadherin expression to promote cell invasion and migration for poor prognosis in oral squamous cell carcinoma. Sci Rep. 2019;9:8932.
57. Parma B, Ramesh V, Gollavilli PN, Siddiqui A, Pinna L, Schwab A, et al. Metabolic impairment of non-small cell lung cancers by mitochondrial HSPD1 targeting. J Exp Clin Cancer Res. 2021;40:248.
58. Madore A-M, Perron S, Turmel V, Laviolette M, Bissonnette ÉY, Laprise C. Alveolar macrophages in allergic asthma: an expression signature characterized by heat shock protein pathways. Hum Immunol. 2010;71:144–50.
59. West LC, Cresswell P. Expanding roles for GILT in immunity. Curr Opin Immunol. 2013;25:103–8.
60. Buetow KH, Meador LR, Menon H, Lu Y-K, Brill J, Cui H, et al. High GILT Expression and an active and intact MHC class II antigen presentation pathway are associated with improved survival in melanoma. J Immunol. 2019;203:2577–87.
61. Bao B-Y, Pao J-B, Huang C-N, Pu Y-S, Chang T-Y, Lan Y-H, et al. Polymorphisms inside MicroRNAs and MicroRNA target sites predict clinical outcomes in prostate cancer patients receiving androgen-deprivation therapy. Clin Cancer Res. 2011;17:928–36.
62. Zhu C, Chen X, Guan G, Zou C, Guo Q, Cheng P, et al. IFI30 Is a novel immune-related target with predicting value of prognosis and treatment response in glioblastoma. Onco Targets Ther. 2020;13:1129–43.
63. Moreau-Gachelin F, Tavitian A, Tambourin P. Spi-1 is a putative oncogene in virally induced murine erythroleukaemias. Nature. 1988;331:277–80.
64. Ghisletti S, Barozzi I, Mietton F, Polletti S, De Santa F, Venturini E, et al. Identification and characterization of enhancers controlling the inflammatory gene expression program in macrophages. Immunity. 2010;32:317–28.
65. Gregoricchio S, Polit L, Esposito M, Berthelet J, Delestré L, Evanno E, et al. HDAC1 and PRC2 mediate combinatorial control in SPI1/PU.1-dependent gene repression in murine erythroleukaemia. Nucl Acids Res. 2022;50:7938–58.
66. Du B, Gao W, Qin Y, Zhong J, Zhang Z. Study on the role of transcription factor SPI1 in the development of glioma. Chinese Neurosurg J. 2022;8:7.
67. Wang J, Wang X, Guo Y, Ye L, Li D, Hu A, et al. Therapeutic targeting of SPIB/SPI1-facilitated interplay of cancer cells and neutrophils inhibits aerobic glycolysis and cancer progression. Clin Transl Med. 2021;11: e588.
68. Elton TS, Selemon H, Elton SM, Parinandi NL. Regulation of the MIR155 host gene in physiological and pathological processes. Gene. 2013;532:1–12.
69. Huang J, Chen W, Jie Z, Jiang M. Comprehensive analysis of immune implications and prognostic value of SPI1 in gastric cancer. Front Oncol. 2022. https://doi.org/10.3389/fonc.2022.820568.
70. Luo Y, Chen C. The roles and regulation of the KLF5 transcription factor in cancers. Cancer Sci. 2021;112:2097–117.
71. Gao Y, Ding Y, Chen H, Chen H, Zhou J. Targeting Kruppel-like factor 5 (KLF5) for cancer therapy. Curr Top Med Chem. 2015;15:699–713.
72. Li J-C, Chen Q-H, Jian R, Zhou J-R, Xu Y, Lu F, et al. The partial role of KLF4 and KLF5 in gastrointestinal tumors. Gastroenterol Res Pract. 2021;2021:2425356.