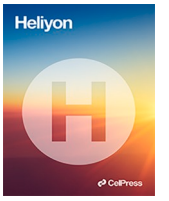
Heliyon 10 (2024) e35549

Contents lists available at ScienceDirect

Heliyon

journal homepage: www.cell.com/heliyon



Research article

Single-cell sequencing analysis revealed that WDR72 was a novel cancer stem cells related gene in gastric cancer

Lei Zheng [a,1], Jia Lu [b,1], Dalu Kong [a,*], Yang Zhan [a,**]

[a] *Department of Colorectal Oncology, Tianjin Medical University Cancer Institute & Hospital, National Clinical Research Center for Cancer, Tianjin's Clinical Research Center for Cancer, Key Laboratory of Cancer Prevention and Therapy, Tianjin Key Laboratory of Digestive Cancer, Tianjin, 300060, China*

[b] *Department of Infection Management, Tianjin Medical University Cancer Institute & Hospital, National Clinical Research Center for Cancer, Tianjin's Clinical Research Center for Cancer, Key Laboratory of Cancer Prevention and Therapy, Tianjin Key Laboratory of Digestive Cancer, Tianjin, 300060, China*

ARTICLE INFO

Keywords:
Cancer stem cells
Gastric cancer
Single-cell sequencing
WDR72
Prognosis

ABSTRACT

Background: Cancer stem cells (CSCs) are pivotal in tumor resistance to chemotherapy and gastric cancer's rapid proliferation and metastasis. We aimed to explore the CSCs-related genes in gastric cancer epithelial cells.

Methods: The mRNA expression profile and single-cell sequencing data of gastric cancer were downloaded from the public database.

Results: We identified WDR72 as a CSCs-related gene in gastric cancer epithelial cells. WDR72 was highly expressed in gastric cancer tissues, and high expression of WDR72 was associated with inferior prognosis of patients. WDR72 expression had a significant negative correlation with the infiltration of CD8 + T cells and activated memory CD4 + T cells. PD-L1 expression was significantly reduced in gastric cancer patients with high WDR72 expression. WDR72 was correlated with IC50 of multiple small-molecule drugs.

Conclusion: We identified a novel CSCs-related gene in gastric cancer epithelial cells, WDR72, which was highly expressed in patients with high stemness scores.

1. Introduction

Gastric cancer is the fifth most prevalent cancer and it is also the third biggest cause of cancer-related deaths worldwide each year [1]. It generally begins in the cells that produce mucus in the stomach's inner lining and can spread to other body parts if left untreated [2]. Factors that increase the risk of developing gastric cancer include a family history of the disease, *Helicobacter pylori* infection, smoking, a diet high in salt, and certain genetic mutations [3,4]. Symptoms of gastric cancer include indigestion, heartburn, bloating,

* Corresponding author. Department of Colorectal Oncology, Tianjin Medical University Cancer Institute and Hospital, National Clinical Research Center for Cancer, Tianjin's Clinical Research Center for Cancer, Key Laboratory of Cancer Prevention and Therapy, Tianjin Key Laboratory of Digestive Cancer, Huanhuxi Road, Hexi District, Tianjin, 300060, China.

** Corresponding author. Department of Colorectal Oncology, Tianjin Medical University Cancer Institute and Hospital, National Clinical Research Center for Cancer, Tianjin's Clinical Research Center for Cancer, Key Laboratory of Cancer Prevention and Therapy, Tianjin Key Laboratory of Digestive Cancer, Huanhuxi Road, Hexi District, Tianjin, 300060, China.

E-mail addresses: 18622626115@163.com (D. Kong), zhanyang@tmu.edu.cn (Y. Zhan).

[1] These authors contributed equally this work.

https://doi.org/10.1016/j.heliyon.2024.e35549
Received 22 January 2024; Received in revised form 29 July 2024; Accepted 31 July 2024
Available online 31 July 2024

Table 1
Clinicopathological characteristics of STAD patients from database.

Characteristics		Patients(N = 348)	
		NO.	%
Gender	**Female**	124	35.63 %
	Male	224	64.37 %
Age	**≤67(Median)**	181	52.01 %
	>67(Median)	167	47.99 %
Stage	**I**	46	13.22 %
	II	110	31.61 %
	III	144	41.38 %
	IV	35	10.06 %
	Unknown	13	3.74 %
T	**T0**	4	1.15 %
	T1	16	4.60 %
	T2	74	21.26 %
	T3	160	45.98 %
	T4	94	27.01 %
M	**M0**	325	93.39 %
	M1	23	6.61 %
N	**N0**	112	32.18 %
	N1	93	26.72 %
	N2	72	20.69 %
	N3	71	20.40 %
Survival Time	**Long(>3 years)**	47	13.51 %
	Short(<3 years)	301	86.49 %
OS status	**Dead**	146	41.95 %
	Alive	202	58.05 %

loss of appetite, unintentional weight loss, and stomach pain [5]. Early detection is crucial for successful treatment, as advanced-stage gastric cancer can be challenging to cure [6]. Treatment options for gastric cancer include surgery [7], chemotherapy [8], radiation therapy [9], targeted therapy [10], and immunotherapy [11]. Regular screenings and maintaining a healthy lifestyle can help reduce the risk of developing gastric cancer.

Cancer stem cells (CSCs) constitute a subset of cancer cells capable of initiating tumor development owing to their possession of stem cell traits, such as the ability to self-renew and differentiate into multiple lineages [12]. It has been learned via the progressive advancement of CSCs research that CSCs are the precursor cells of malignant tumors [13,14]. They play an important role in promoting the spread of cancer, its recurrence, and even resistance to chemotherapy [15,16]. Previous studies have shown that gastric cancer stem cells (gCSCs) possess distinct characteristics and play a crucial role in treatment resistance [17]. Stem cell markers CD44, CD133, and Musashi-1 are highly expressed in precancerous lesions, contributing to the malignant transformation of gastric cancer tissues [18]. The expression of CD44 and CD133 correlates with the pathological stage of the disease [19]. Additionally, Zhao et al. have identified AQP5 as a specific surface marker of CSCs in gastric cancer, and its cooperation with LGR5 to promote tumorigenesis and activate autophagy in gastric cancer stem cells [20]. Guo et al. discovered six CSC-related genes (BUB1, KIF18A, MAD2L1, NCAPG, RAD54L, and PLK4) in different types of gastric cancer, all associated with the cell cycle [21]. Furthermore, Zhao et al. found high expression of the CSC-related gene CXCR4 in gastric cancer, which is linked to poor patient survival [22]. Knockdown of CXCR4 inhibited the aggressive behavior of CSCs *in vitro* and reduced tumor growth and liver metastasis in mice [22]. Zhang et al. have demonstrated that CSCs-related gene CYB5R1 is correlated with drug resistance and M2 macrophage polarization in gastric cancer [23]. Accordingly, further exploration of CSCs-related genes is crucial for devising effective therapeutic strategies to combat gastric cancer.

Thus, in this study, we collected the single-cell sequencing data of gastric cancer from public databases and identified *WDR72* as a key CSCs-related gene in gastric cancer. Moreover, we analyzed the correlation of *WDR72* with prognosis and immune cell infiltration of gastric cancer patients. Our findings highlighted the role of *WDR72* in gastric cancer and are expected to provide more insight into the development of therapeutic strategies for gastric cancer.

2. Materials and methods

2.1. Data collection

The single-cell sequencing data of gastric cancer tissue were collected from the GSE163558 dataset for single-cell analysis, including 3 primary gastric cancer samples, 1 para-cancerous sample, and 6 metastatic samples. The raw data of each sample was processed by the software Cell Ranger (version 6.0.2) from 10x Genomics to obtain three files: barcodes.tsv, genes.tsv, and matrix.tsv. These files were downloaded and used directly for subsequent analysis without any preprocessing. Furthermore, the sequencing transcriptome data of a total of 1010 tumor samples and 69 para-cancerous samples were collected from The Cancer Genome Atlas (TCGA) database and the Gene Expression Omnibus (GEO, https://www.ncbi.nlm.nih.gov/geo/) database for subsequent analysis.

The samples included in subsequent survival-related analysis must meet the following criteria: 1. Gastric cancer tissue samples; 2. The patients from whom the cancer samples come have recorded corresponding survival time and survival status information. specific criteria. The distribution of the final number of samples was as follows: TCGA-STAD (tumor = 348, para-cancerous = 32); GSE56807 (tumor = 5, paired para-cancerous tissue = 5); GSE65801 (tumor = 32, para-cancerous = 32); GSE84437 (tumor = 433); GSE15459 (tumor = 192). TCGA-STAD, GSE56807, and GSE65801 were utilized for screening and verifying differentially expressed genes (DDEGs), while TCGA-STAD, GSE84437, and GSE15459 were used for screening and verifying prognosis-related genes. The clinicopathological characteristics of STAD patients from TCGA database are presented in Table 1.

The data in the GSE163558, GSE56807, GSE65801, GSE84437 and GSE15459 were obtained using Illumina NovaSeq 6000, Affymetrix Human Exon 1.0 ST Array, Agilent-028004 SurePrint G3 Human GE 8 × 60K Microarray, Illumina HumanHT-12 V3.0 expression beadchip and Affymetrix Human Genome U133 Plus 2.0 Array platforms, respectively. The probe names of GSE56807, GSE65801, GSE84437 and GSE15459 datasets were converted to Genesymbol using their respective platform annotation files.

2.2. Single-cell dimensionality reduction and cell clustering, mutation analysis, and pseudotime analysis

The fundamental concept of clustering was to analyze gene expression data to identify variations in expression patterns among cells, subsequently categorizing cells into subgroups based on these differences. It is important to note that the cell grouping generated at this stage is solely based on mathematical calculations and may evolve as the threshold for classification is modified. In the GSE163558 dataset, the quality control, statistical analysis, and exploration of single-cell sequencing data were completed using the R package Seurat. t-distributed stochastic neighbor embedding (tSNE) was used for dimensionality reduction and cluster classification analysis. The singleR package was used to annotate cell clusters with signature marker genes.

The copy number variant (CNV) analysis was applied to screen tumor cells via R package infercnv. Selected tumor cells were subjected to stemness analysis using the R language "synapser" package. finally, tumor cells were subjected to pseudotime analysis using Monocle2, determining the correlation between stemness score and differentiation status.

2.3. Differential gene analysis

The differential gene analysis was performed using the "limma" package of R language, based on the principle of "t.test". The differentially expressed genes (DEGs) between the high and low stemness score groups (using GSE163558), gastric cancer samples and para-cancerous samples (employing TCGA, GSE56807, and GSE65801) were identified using $|\text{Log2Fold change (FC)}| > 1$ and $p < 0.05$.

2.4. Gene set enrichment analysis (GSEA)

In the TCGA cohort, the gastric cancer samples were divided into *WDR72*^{high} and *WDR72*^{low} groups according to the cutoff value (−1.260) of *WDR72* expression. Next, the DEGs between the *WDR72*^{high} and *WDR72*^{low} groups were identified. GSEA for DEGs was performed using the "ClusterProfiler" package in the R language [24]. Significantly enriched Kyoto Encyclopedia of Genes and Genomes (KEGG) signaling pathways were screened via $|\text{NES}| > 1$ and $p < 0.05$.

2.5. Survival analysis

WDR27 was standardized within the TCGA dataset to determine the most suitable cutoff value (−1.260). Subsequently, gastric cancer patients were segregated into *WDR72*^{high} and *WDR72*^{low} groups according to this optimal cutoff value. The overall survival rate of gastric cancer patients in the *WDR72*^{high} and *WDR72*^{low} groups were determined using the R language "survival" and "survminer" package (https://CRAN.R-project.org/package=survival), based on the Kaplan-Meier (KM) method. The log-rank test was used to assess the statistical significance of variations in survival between the two groups. $p < 0.05$ was deemed statistically significant. The target genes were tested using multivariate Cox regression analysis to see if they could predict the survival of gastric cancer patients independently of other variables.

2.6. Immune cell infiltration

The infiltration proportions of 22 human immune cells were calculated using CIBERSORT software [25]. The CIBERSORT developers initially created and validated a leukocyte gene signature set called LM22, consisting of 574 genes. This set is capable of distinguishing 22 hematopoietic cell phenotypes, including 7 types of T cells, initial and memory B cells, plasma cells, natural killer (NK) cells, and bone marrow subtypes. By utilizing LM22, researchers can calculate the relative proportions of these 22 types of immune cells by integrating tissue RNA sequencing data. The total of all estimated proportions of immune cell types in each sample equals 1. Immune scores of gastric cancer patients were calculated using the "estimate" package (https://R-Forge.R-project.org/projects/estimate/).

2.7. Drug sensitivity analysis

The GDSC database is the largest public resource for tumor cellular drug sensitivity and tumor therapeutic genomic data, containing tumor cell line anticancer drug sensitivity data and cell line genomics data, and is dedicated to the discovery of tumor

Table 2
Primer sequences used for quantitative real-time PCR (qRT-PCR) analysis.

Genes	Forward Primer (5′-3′)	Reverse Primer (5′-3′)
WDR72	AGAGCATGCCACTGGAAACA	GTACTAGGACAGGCCTCCCA
GAPDH	GAAGGTGAAGGTCGGAGTC	GAAGATGGTGATGGGATTTC

therapeutic targets to improve tumor therapy. CTRP covers the link between compound sensitivity and genetic or genealogical characteristics in 70,000 cancer cell lines, a dataset that researchers can use to find therapeutic targets available for different cancer types.

The IC860 of 265 small compounds and the IC1001 of 481 small molecules in 50 cell lines and their associated gene mRNA expression were extracted from the genomics of drug sensitivity in cancer (GDSC) and the response portal for therapeutic genomics (CTRP). The relevant gene mRNA expression data and drug sensitivity data from IC860 and IC101 were merged. Pearson correlation analysis was used to calculate the relationship between WDR72 expression and medication IC50 (half-maximal drug inhibitory concentration). The *p*-value was adjusted using the Benjamini-Hochberg method, false discovery rate (FDR) value < 0.05 was the threshold for screening significantly related drugs.

2.8. Patient tissue samples

A total of 20 gastric cancer tissues and 20 normal tissues were obtained from the Tianjin Medical University Cancer Institute & Hospital. All experiments were approved by the ethics committee of the Tianjin Medical University Cancer Institute & Hospital (ethic code: No. bc2023174), conformed to the declaration of Helsinki, and informed consent was obtained from all subjects. The information on all subjects is shown in Table S1.

2.9. qRT-PCR assay

Total RNA from tissues was extracted using the RNAprep Pure Tissue Kit (DP431, Tiangen Biotechnology Co., Ltd., Beijing, China). Subsequently, the RNA was reverse transcribed with PrimeScript™ RT reagent Kit with gDNA Eraser (RR047A, Takara) and amplified using the All-in-One™ qPCR Mix (QP001, GeneCopoeia) on an iQ5 Real-Time PCR amplicon (Applied Biosystems). The primer sequences are shown in Table 2. The PCR conditions were as follows: pre-denaturation at 95 °C for 30 s, followed by 40 cycles of denaturation at 95 °C for 5 s, and annealing at 60 °C for 30 s. GADPH was used as the internal reference, and each sample was run in triplicate. The mRNA expression levels were calculated using the $2^{-\Delta\Delta CT}$ formula.

2.10. Western blot assay

The protein was extracted using RIPA lysis solution (R0010, Solarbio) and PMSF solution. The Western blot was consistent with previous methods [26]. The first antibodies included WDR72 Polyclonal Antibody (PA5-63780, Thermo), and Beta Actin Monoclonal antibody (66009-1-lg, Proteintech). The second antibody was Horseradish enzyme-labeled Goat anti-rabbit IgG (H + L) (ZB-2301, ZSGB-BIO, USA) and Horseradish enzyme-labeled Goat anti-mouse IgG (ZB-2305, ZSGB-BIO, USA). The gray values of the bands were analyzed using Image J software.

2.11. Statistical analysis

In this study, the comparison between two groups in the box plot was analyzed using the Wilcoxon rank sum test method, while the comparison between multiple groups utilized the “anova” method. A statistically significant difference was determined when the *p*-value was less than 0.05. A multivariate Cox regression proportional hazards model was used to determine the effects of mRNA expression of *WDR72* and clinicopathological characteristics on the overall survival rate of patients. All the above statistical analyses used R software. All experimental data were analyzed with GraphPad 8.0.2 and expressed as the mean ± SD. Statistical analyses were performed using Student’s t-test for two group comparisons. Differences were considered statistically significant when $p < 0.05$.

3. Results

3.1. A single-cell transcriptome atlas in diverse gastric cancer types

In the GSE163558 dataset, the single-cell sequencing data of 3 primary gastric cancer tissues, 6 metastatic gastric cancer tissues, and 1 para-cancerous tissue were first subjected to principal component analysis (PCA) dimensionality reduction and then clustered. Following PCA, an ’Elbowplot’ (Fig. S1) revealed that the curve no longer experiences a significant drop after reaching a PC value of 10. This suggested that the first 10 principal components encapsulate the majority (over 90 %) of the data information. Thus, the PC value was selected as 10, and the UMAP method was used to divide the cells into 17 cell types (Fig. 1A). The distribution of cells in different samples among these 17 cell types was shown in Fig. 1B. Next, we used the Human Primary Cell Atlas Data in the SingleR function package to annotate these 17 cell types, and discovered that epithelial cells were ubiquitously present in cluster 10 (Fig. 1C).

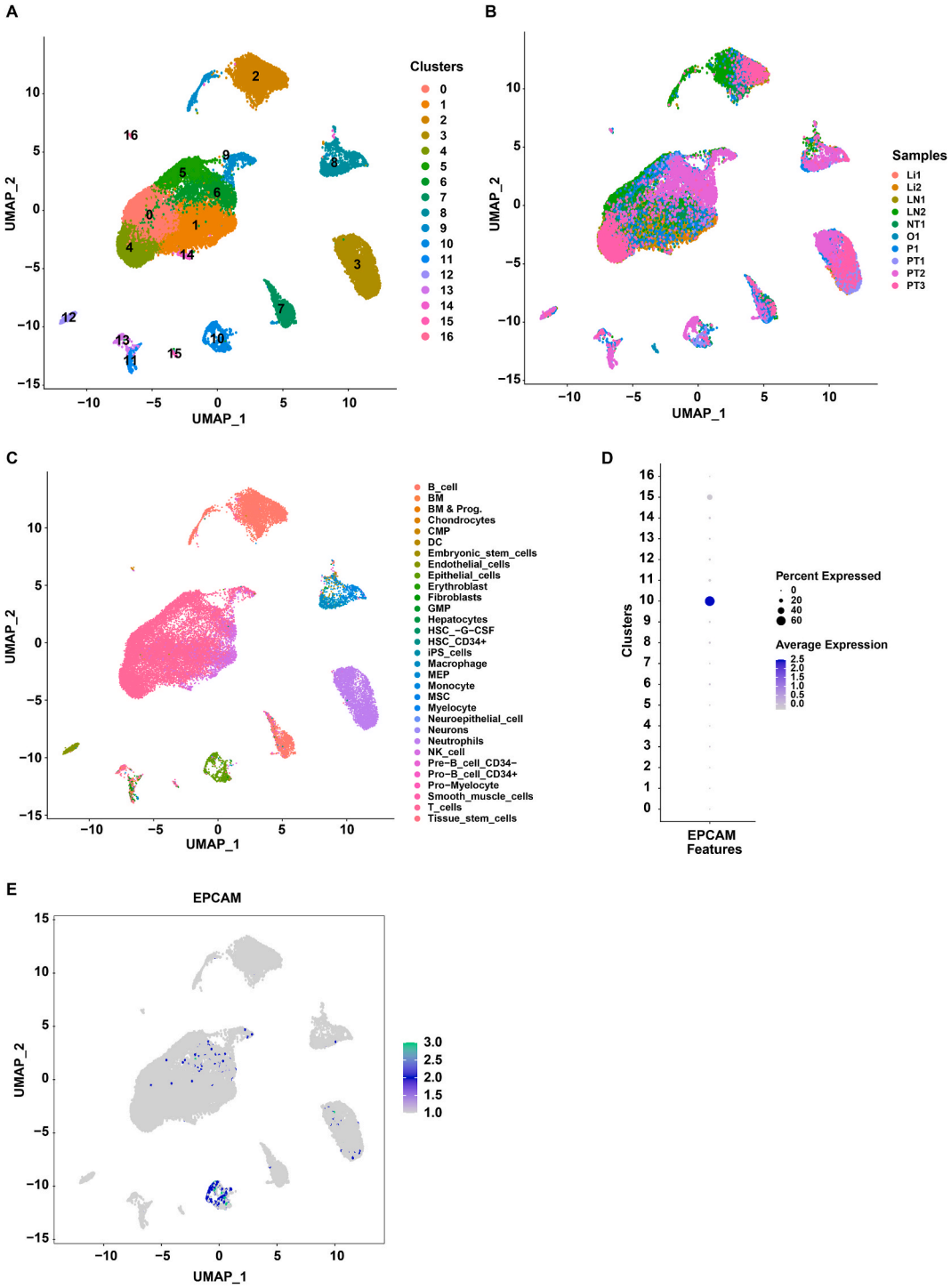


Fig. 1. A single-cell transcriptome atlas in diverse gastric cancer types. A, Clustered Uniform Manifold Approximation and Projection (UMAP) plot of single-cell sequencing data in the GSE163558 dataset, the PC value was selected as 10. B, Distribution of primary gastric cancer, metastatic gastric cancer, and para-cancerous samples in cell clusters. C, The main cell types annotated by known cell lineages in cell clusters were illustrated by UMAP plots using the "Human Primary Cell Atlas Data" database in the SingleR function package. D-E, The expression of epithelial cell marker gene (*EPCAM*) in different cell clusters.

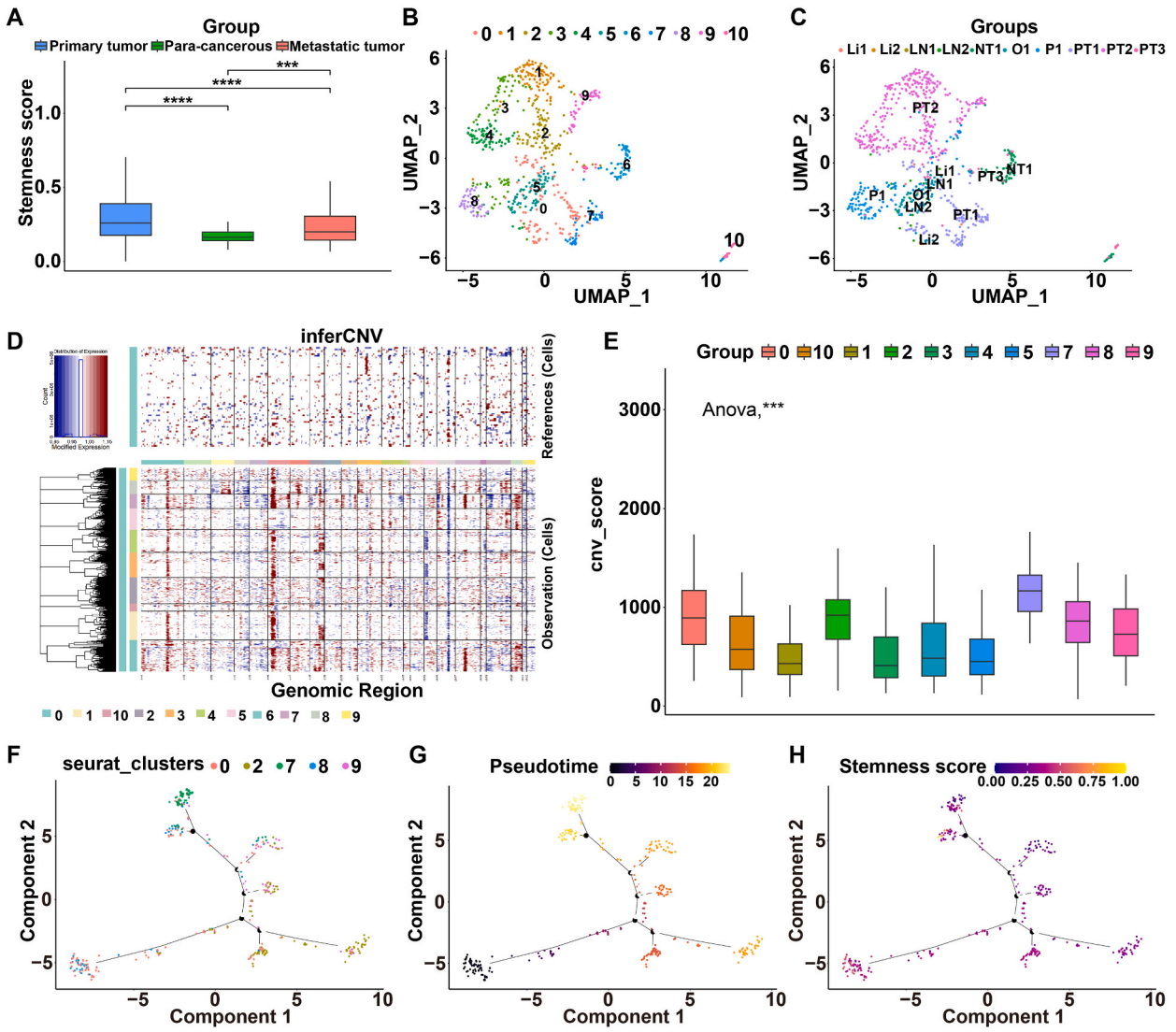


Fig. 2. Identified gastric cancer-related epithelium. A, The score of stemness in primary gastric cancer samples, normal samples, and metastatic samples. B–C, Re-clustered UMAP plot of epithelial cells. Heatmap of specific gene expression in different epithelial cell clusters. Li1: Gastric cancer liver metastasis; Li2: Gastric cancer liver metastasis; LN1: Gastric cancer lymph node metastasis; LN2: Gastric cancer lymph node metastasis; NT1: Adjacent non-tumoral; O1: Gastric cancer ovary metastasis; P1: Gastric cancer peritoneum metastasis; PT1: Primary gastric cancer; PT2: Primary gastric cancer; PT3: Primary gastric cancer. D, Specifically expressed genes in different epithelial cell clusters via inferCNV. E, The score of copy number variant (CNV) in different epithelial cell types. F, High CNV score cell clustering proposed time series diagram, the horizontal and vertical coordinates are the two principal components, each dot in the diagram represents a cell, different colors indicate different cell clusters, black dots are nodes of different cell states, the diagram shows that there are five different cell nodes. G, Plot of the pseudotime time trajectory, darker to lighter colors indicate the proposed chronological order. H, Pseudotime plot of stemness scores, lighter colors indicate higher dryness scores. ***$p < 0.001$.

To further verify the annotation conclusions, we analyzed the expression of epithelial cell marker gene EPCAM in clusters, and found that EPCAM expression was predominantly observed in cluster 10 (Fig. 1D and E).

3.2. *Identified gastric cancer-related epithelial cell clusters*

In cluster 10, the stemness score was calculated in primary, metastatic gastric cancer tissues, and para-cancerous tissues. The stemness score was found to be highest in primary tissues and lowest in para-cancerous tissues (Fig. 2A, primary > metastatic > para-cancerous). Subsequently, the epithelial cells in cluster 10 were re-clustered into 11 cell types, and the para-cancerous cells were presented in cell type 6 (Fig. 2B and C).

Additionally, CNV variant analysis was conducted using “infercnv” to distinguish tumor cells from non-malignant cells within

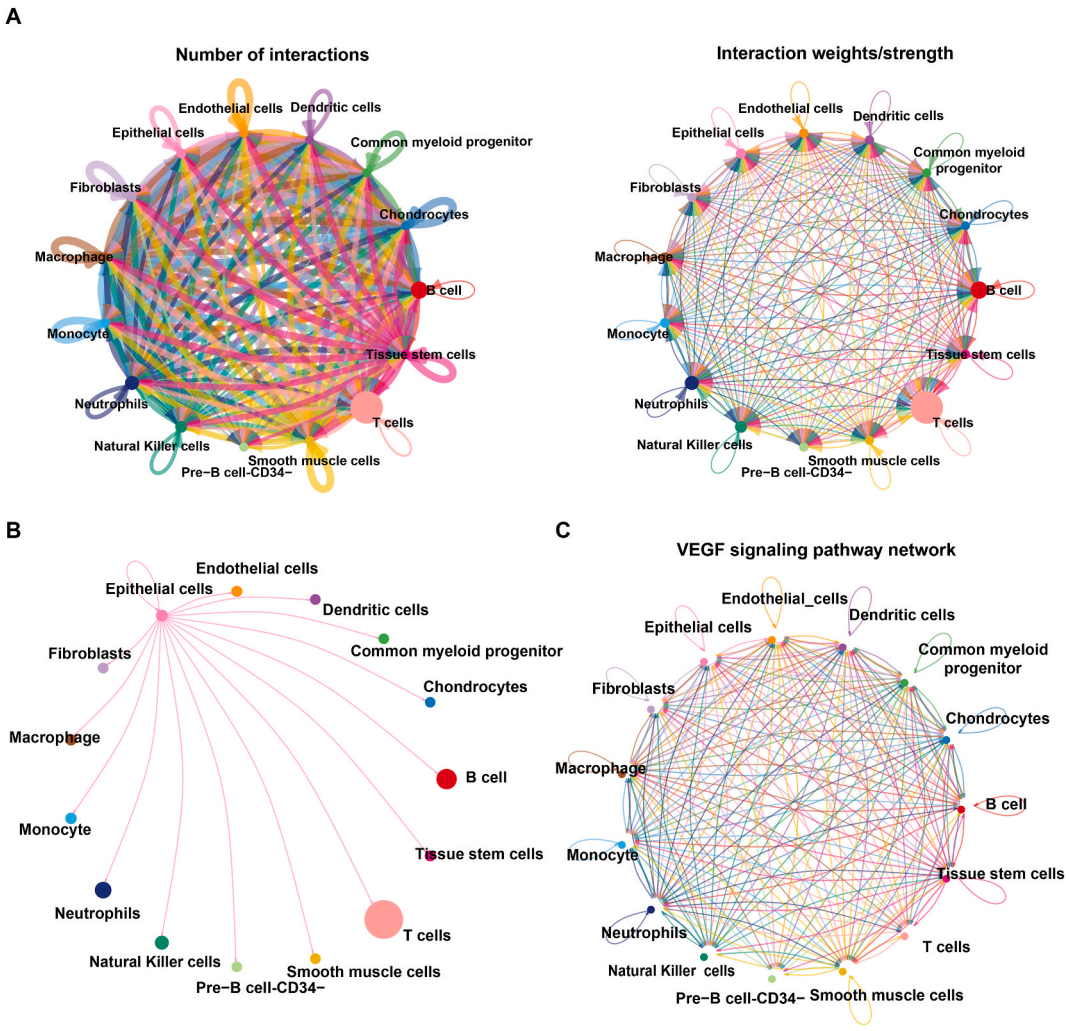


Fig. 3. Epithelial cells were correlated with immune cells in gastric cancer. A, The result of cell-cell communication of all cell types. The left side is the number of communication, the wider the line means more communication, the right side of the graph is the strength of communication, the bigger the point means the stronger the communication between the cell and other cells. B, The communication of epithelial cells with other immune cells. C, The communication of epithelial cells with other cells in the vascular endothelial growth factor (VEGF) signaling pathways.

gastric epithelial cell types. The specifically expressed genes in different epithelial cell types are presented in Fig. 2D. We selected epithelial cell types (0: 945.8, 2: 892.5, 7: 1160.7, 8: 829.5, 9: 726.7) with high-CNV scores (an average CNV score greater than 720.5) as gastric cancer-related epithelial cells (Fig. 2E).

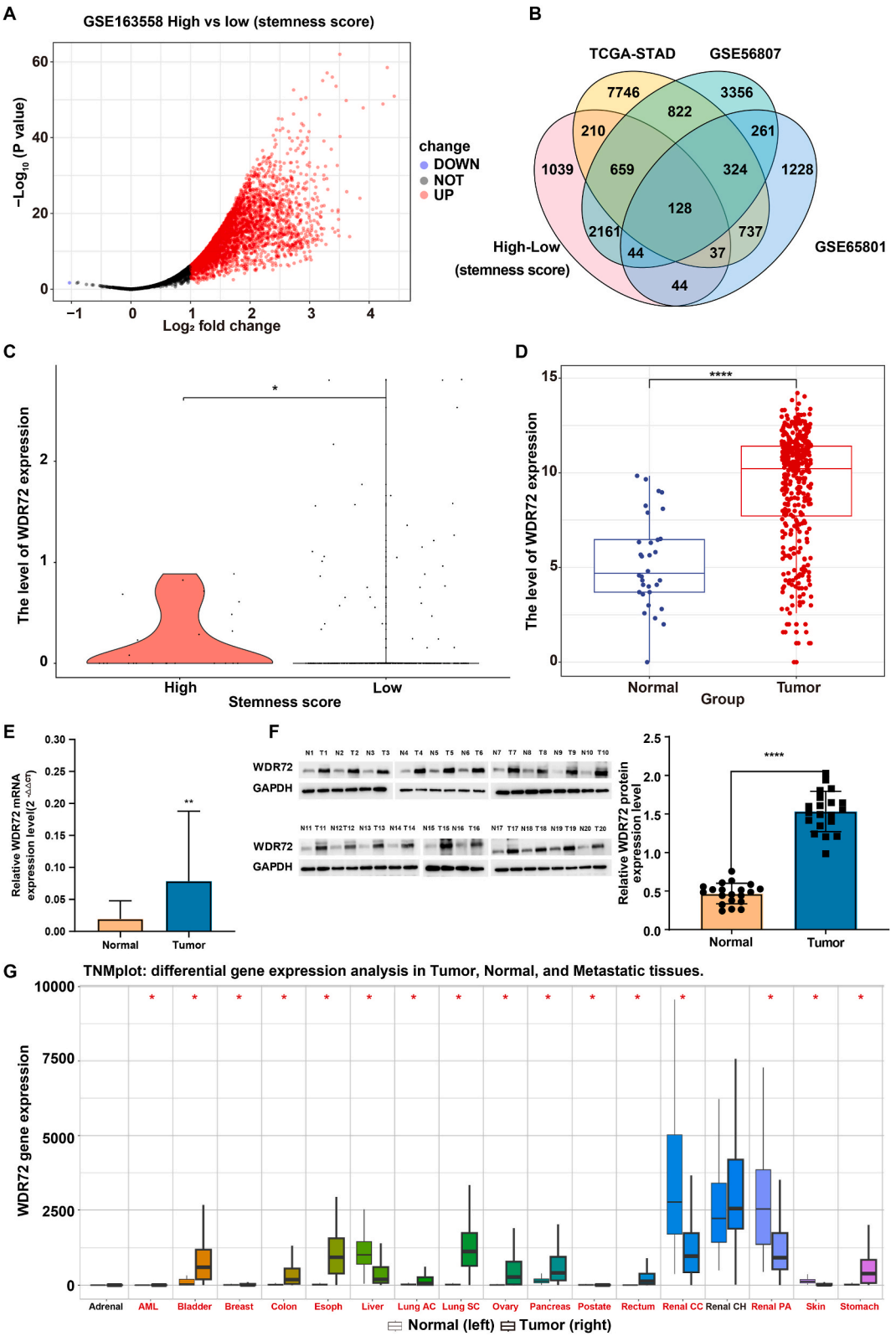
Pseudotime analysis was conducted on cell types with high-CNV scores (0, 2, 7, 8, 9), resulting in trajectories of cells across various cell types (Fig. 2F). It was observed that cells in different types followed distinct differentiation trajectories. Specifically, the differentiation times of cell types 2 and 7 were notably longer compared to cell types 0 and 8 (Fig. 2G). Moreover, the stemness scores of cell types 0, 2, 7, and 8 were higher (Fig. 2H).

3.3. Gastric cancer-related epithelial cells were correlated with immune cells

Cell communication analysis showed that each cell type had interaction with each other (Fig. 3A). Gastric cancer-related epithelial cells exhibited strong interaction with T cells, B cells, and neutrophils (Fig. 3B). In addition, in the vascular endothelial growth factor (VEGF) signaling pathway, the correlations between gastric cancer-related epithelial cells and T cells, B cells, macrophage and neutrophils were remarkably higher intensity (Fig. 3C).

3.4. WDR72 as a key gene associated with stemness in gastric cancer

Based on the stemness score, the cell types 0, 2, 7, 8, and 9 were categorized into low stemness score group (score <0.5) and high



(caption on next page)

Fig. 4. *WDR72* as a key gene associated with stemness in gastric cancer. A, Differentially expressed genes between high and low stemness score groups in the GSE163558 dataset. Red dots represent up-regulated genes, blue dots represent down-regulated genes, and p value < 0.05, $|Log2FC| > 1$ were used as thresholds to screen differentially expressed genes. B, The cross-over analysis between four gene sets. C, The expression of *WDR72* in the high and low stemness score groups in the GSE163558 dataset. D, The expression of *WDR72* in gastric cancer samples and normal samples in the TCGA dataset. E, The levels of WDR72 mRNA expression in gastric cancer tissues and normal tissues were determined by qRT-PCR assay. F, the level of WDR72 protein expressions in gastric cancer tissues and normal tissues was determined via Western blot assay. G, The expression of *WDR72* in pan-cancer. $^{**}p < 0.01$, $^{****}p < 0.0001$.

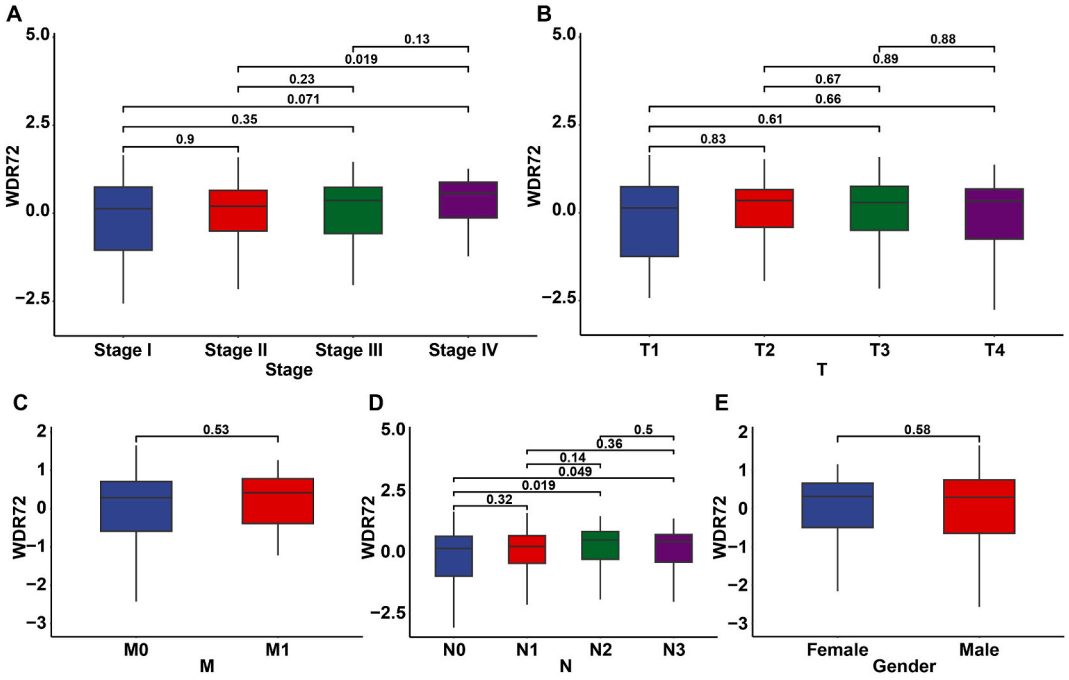


Fig. 5. *WDR72* expression varies in different pathological stages of gastric cancer. A, The expression of *WDR72* in stageI~stageIV of gastric cancer. B, The expression of *WDR72* in T1~T4 of gastric cancer. C, The expression of *WDR72* in M0 and M1 of gastric cancer. D, The expression of *WDR72* in N0~N3 of gastric cancer. E, The expression of *WDR72* in male and female gastric cancer patients. $^{*}p < 0.05$.

stemness score group (score >0.5, with a score closer to 1 indicating closer proximity to cancer stem cells). A total of 4322 DEGs were identified between the high and low stemness score groups (Fig. 4A). Furthermore, DEGs between gastric cancer samples and para-cancerous samples in TCGA, GSE56807, and GSE65801 cohorts were analyzed. Subsequently, a cross-analysis was conducted among these four DEG groups, resulting in the identification of 128 overlapping genes (Fig. 4B). Among these 128 genes, *WDR72* was selected for further investigation based on existing literature reports.

Additionally, we analyzed the expression of *WDR72* in gastric cancer samples. In the GSE163558 dataset, the *WDR72* expression was increased in the high stemness score group than in the low stemness score group (Fig. 4C). In the TCGA cohort, the *WDR72* was highly expressed in gastric cancer samples (Fig. 4D, tumor vs. para-cancerous). The levels of WDR72 mRNA and protein expressions were significantly increased in gastric cancer tissues compared to normal tissues (Fig. 4E and F, Fig. S2). The pan-cancer analysis showed that compared to the normal group, the *WDR72* was highly expressed in multiple tumor types (Fig. 4G).

3.5. *WDR72 expression varies in different pathological stages of gastric cancer*

In the TCGA cohort, we conducted an analysis of *WDR72* expression in gastric cancer patients at various pathological stages. Our findings revealed that WDR72 expression was significantly higher in patients with Stage IV compared to those with Stage II (Fig. 5A). Interestingly, there was no significant difference in *WDR72* expression among patients with different T stages (T1~T4), nor between those with M0 and M1 stages (Fig. 5B and C). The expression of *WDR72* was markedly increased in patients with N2 and N3 compared to those with N0 (Fig. 5D). The gender of the patient did not exert a significant influence on the expression of the gene (Fig. 5E).

3.6. *WDR72 might be an independent predictor of the prognosis of gastric cancer patients*

In the TCGA, GSE84437, and GSE15459 cohorts, gastric cancer samples were classified into $WDR72^{high}$ and $WDR72^{low}$ groups based on the cutoff value of *WDR72* expression, it was discovered that patients with gastric cancer displaying high *WDR72* expression

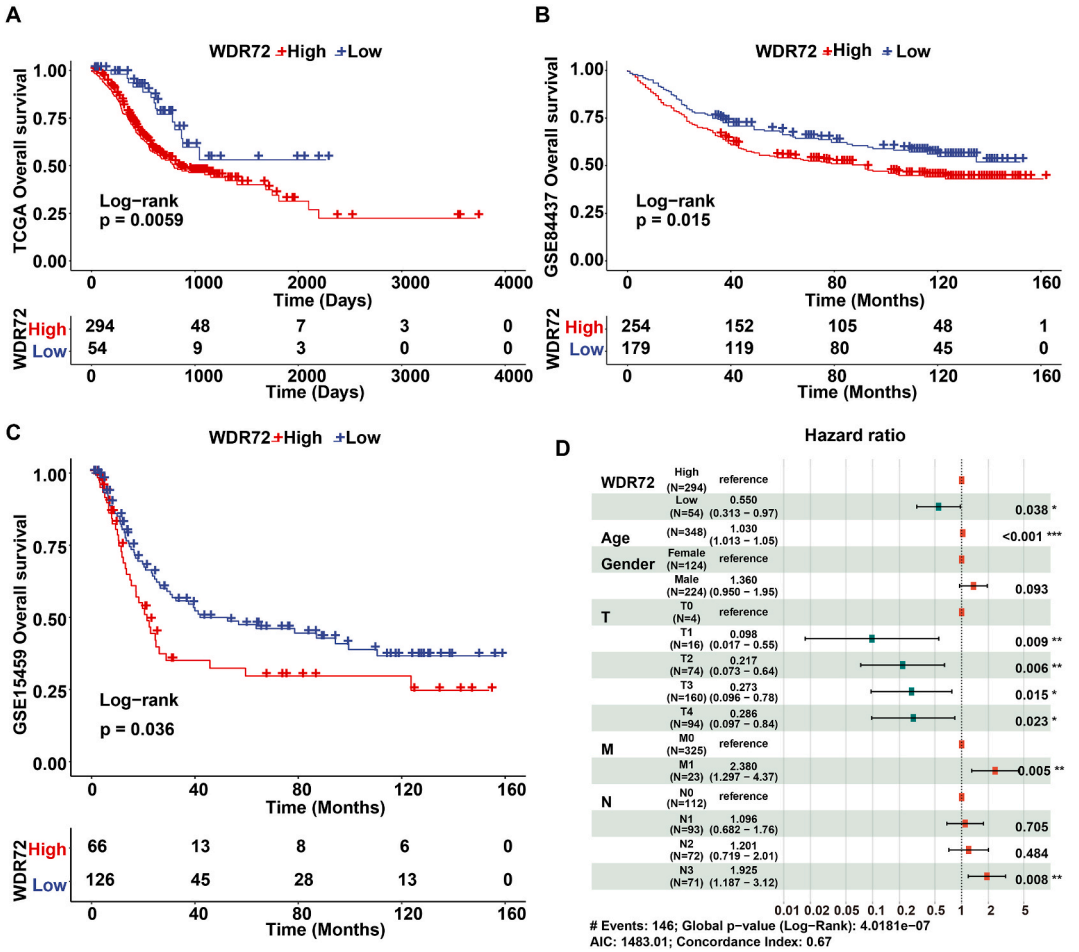


Fig. 6. *WDR72* might be an independent predictor of prognosis of gastric cancer patients. A, The overall survival of gastric cancer patients with high or low *WDR72* expression in the TCGA cohorts. B, The overall survival of gastric cancer patients with high or low *WDR72* expression in the GSE84437 cohorts. C, The overall survival of gastric cancer patients with high or low *WDR72* expression in the GSE15459 cohorts. D, Multivariate Cox regression analysis including age, gender, T, M, N, *WDR72* high and low expression. Compared to the reference sample, samples with a Hazard ratio greater than 1 had a higher risk of death, and samples with a Hazard ratio less than 1 had a lower risk of death.

had a poorer prognosis (Fig. 6A-C, high vs. low). In the TCGA cohort, multivariate Cox regression analysis considered six factors: age, gender, T, M, N, *WDR72*^{high}, and *WDR72*^{low} expression. The analysis suggested that *WDR72* could potentially serve as an independent predictor of prognosis in patients with gastric cancer (Fig. 6D).

3.7. *Potential pathways between gastric cancer patients with high and low WDR72 expression*

DEGs between *WDR72*^{high} and *WDR72*^{low} groups were identified and subjected to GSEA for functional analysis. The analysis revealed a total of 50 activated and 2 suppressed signaling pathways in the *WDR72*^{high} group compared to the *WDR72*^{low} group (Table S2, Fig. 7A), such as Wnt signaling pathway, Gastric cancer and Signaling pathways regulating pluripotency of stem cells (Fig. 7B).

3.8. *The correlation between WDR72 and immune cell infiltration in gastric cancer*

The infiltration of 22 immune cells in gastric cancer samples in the TCGA cohort was calculated (Fig. 8A). Patients with high *WDR72* expression showed reduced infiltration of CD8 + T cells, activated memory CD4 + T cells, and macrophages M1, while increased infiltration of resting memory CD4 + T cells (Fig. 8B, high vs. low). Additionally, a significant negative correlation was observed between *WDR72* expression and the infiltration of CD8 + T cells and activated memory CD4 + T cells (Fig. 8C and D).

Furthermore, we analyzed the expression of 8 immune checkpoints *(PD-1* (*PDCD1*), *CTLA4*, *PD-L1* (*CD274*), *PDL-2* (*PDCD1LG2*), *CD80*, *CD86*, *LAG3*, *TIGIT*) in *WDR72*^{high} and *WDR72*^{low} groups. As shown in Fig. 8E, the *PD-L1 (CD274)* expression was significantly reduced in the *WDR72*^{high} group (high vs. low).

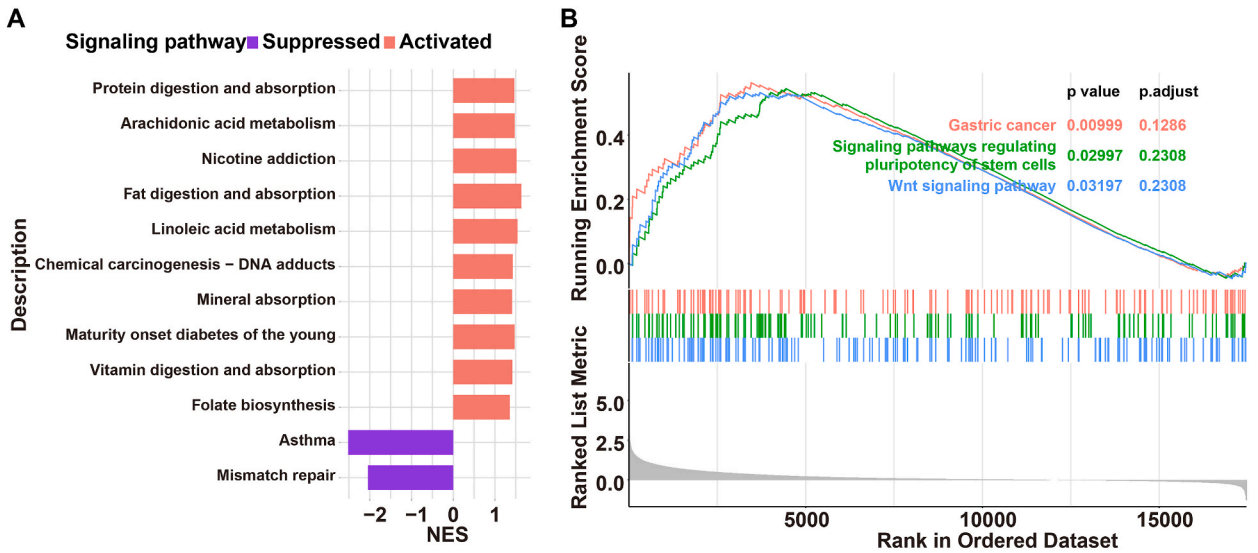


Fig. 7. Potential pathways between gastric cancer patients with high and low *WDR72* expression. A, The top 10 and 2 significantly activated and suppressed signaling pathways in *WDR72*^{high} group. B, Wnt signaling pathway, Gastric cancer and Signaling pathways regulating pluripotency of stem cells were significantly activated in *WDR72*^{high} group compared to *WDR72*^{low} group.

3.9. WDR72 might be associated with stem cells in gastric cancer, and it was a target gene of multiple small molecule drugs

Finally, we analyzed the correlation between marker genes (Table S3) of stem cells and *WDR72* expression and found that *WDR72* expression was significantly positively correlated with *TDG*, *SLCO1A2*, and *WFDC2* expression (Fig. 9A-C, $p < 0.05$ and $|correlation| > 0.3$). In addition, drug sensitivity analysis in the GDSC database indicated that the IC50 of Vorinostat was positively correlated with *WFDC2* expression and negatively correlated with *TDG* expression. Moreover, *SCLO1A2* expression showed a positive correlation with the IC50 of JW-7-52-1, and a negative correlation with the IC50 of SB 505124 and Vorinostat (Fig. 9D–Table S4).

In the CTRP database, it was found that *WFDC2* and *WDR72* expression showed a significant positive correlation with the IC50 of 27 and 15 drugs, respectively. Conversely, they exhibited a negative association with the IC50 of PD 153035, afatinib, and lapatinib. *SLCO1A2* expression was remarkably negatively related to IC 50 of 27 drugs (Fig. 9E–Table S5). These findings suggested a potential association between *WDR72* and stem cells in gastric cancer, highlighting it as a target gene for multiple small-molecule drugs.

4. Discussion

Patients with gastric cancer die mostly as a result of metastasis, recurrence, and treatment resistance of GC cells. CSCs are defined as a crucial element in cancer metastasis, recurrence, and treatment resistance by the cancer stem cell theory [27]. Previous studies discovered that some drugs could target the stemness of gastric cancer by modulating gene expression, potentially opening up new avenues for therapeutic therapy of gastric cancer [28]. In this study, we identified a novel CSCs-related gene *WDR72* in gastric cancer epithelial cells. High expression of *WDR72* in gastric cancer patients was associated with a high stemness score, as well as poor prognosis and lower PD-L1 expression. Additionally, *WDR72* expression was correlated with immune cell infiltration and drug sensitivity of gastric cancer.

Focused on malignant cells, we identified five gastric cancer-related epithelial cell types (0, 2, 7, 8, 9) using single-cell sequencing data. These five epithelial cell types were further categorized into high and low stemness score groups. A total of 4322 DEGs were identified between high and low stemness score groups. Among these DEGs, 128 genes exhibited significant differences between gastric cancer and normal tissues. Of these 128 genes, *WDR72* showed higher expression in patients with high stemness scores compared to those with low stemness scores. *WDR72*, a member of the WD40-repeat domain superfamily, is known for its vesicle-related functions in enamel-producing ameloblasts [29]. Previous studies have highlighted the role of *WDR72* in various cancers. In renal cell carcinoma (RCC), increased *WDR72* expression could inhibit the survival and invasion of cancer cells *in vitro* [30]. *WDR72* demonstrated high expression in non-small cell lung cancer (NSCLC) and colon adenocarcinoma and exhibited a positive correlation with poor prognosis [31,32]. In addition, *WDR72* might be a candidate biomarker for early diagnosis of bladder cancer [33]. In the present study, *WDR72* was also up-regulated in gastric cancer tissues and its higher expression was associated with poor outcomes of patients. These results confirmed the clinical and prognostic significance of WDR72 in gastric cancer and suggested that targeting *WDR72* expression could be a promising strategy for developing therapeutic interventions for cancers like gastric cancer. the results of this study.

Gastric cancer patients with high *WDR72* expression exhibited elevated stemness scores and showed a significant positive correlation with stem cell marker genes, *TDG*, *SLCO1A2,* and *WFDC2*. In lung cancer tissues, *WDR72* was found to be highly expressed in

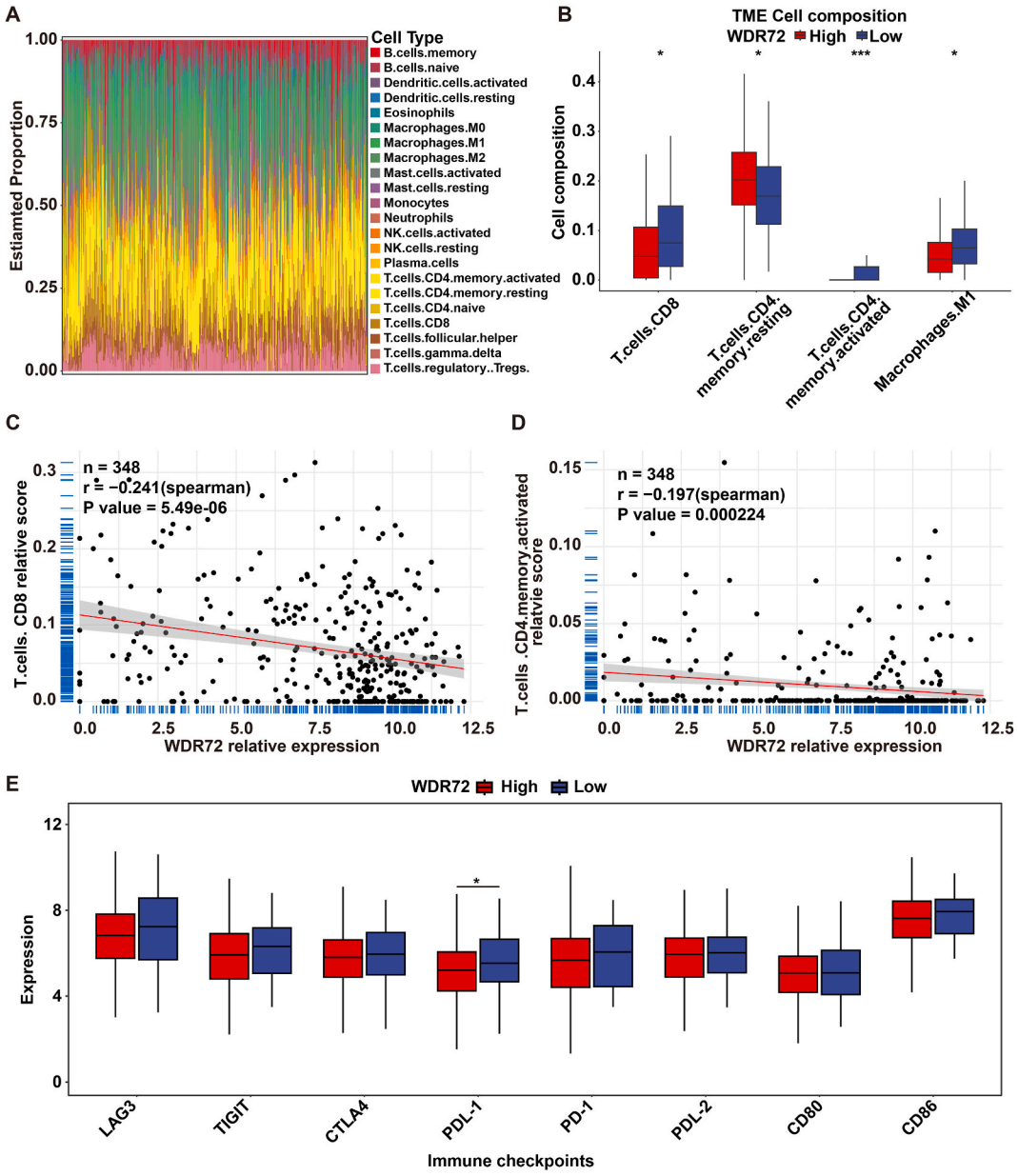


Fig. 8. The correlation between *WDR72* and immune cell infiltration in gastric cancer. A, The infiltration of 22 immune cells in gastric cancer samples in the TCGA cohort was determined by CIBERSORT. B, The infiltration of immune cells in gastric cancer patients with high and low *WDR72* expression. C-D, The correlation of *WDR72* with infiltration of T.cells.CD8 and T.cells.CD4.memory activated. The correlation coefficient is calculated by Pearson correlation coefficients. E. The expression of 8 immune checkpoints (PD-1 (PDCD1), CTLA4, PD-L1 (CD274), PDL-2 (PDCD1LG2), CD80, CD86, LAG3, TIGIT) in $WDR72^{high}$ and $WDR72^{low}$ groups. $^{*}p < 0.05$, $^{***}p < 0.001$.

both lung cancer tissues and lung CSCs, and repression of *WDR72* expression inhibits the stemness protumorigenic effects of lung CSCs [34]. Additionally, our analysis revealed a significant activation of the Wnt signaling pathway in gastric cancer patients with high *WDR72* expression. Previous studies have highlighted the critical role of the Wnt signaling pathway in CSCs [35]. Ji et al. demonstrated that capillary morphogenesis gene 2 (*CMG2*) could activate the Wnt/β-catenin pathway by interacting with LRP6 in GCSLCs to maintain gastric cancer stem-like cell phenotype [36]. Tumor necrosis factor-α-inducing protein (Tipα) was shown to enhance CSC markers (CD44, Oct4, and Nanog) by activating the Wnt/β-catenin pathway, thereby promoting gastric cancer progression [37]. Accordingly, we hypothesized that *WDR72* might regulate the Wnt signaling pathway to enhance gCSCs, ultimately contributing to the development of gastric cancer. Further investigations are warranted to validate this hypothesis in future studies.

In NSCLC, *WDR72* was associated with immune cell infiltration [31]. Our study revealed a negative correlation between *WDR72* and the infiltration of CD8 + T cells and activated memory CD4 + T cells in gastric cancer. You et al. found that the higher infiltrations

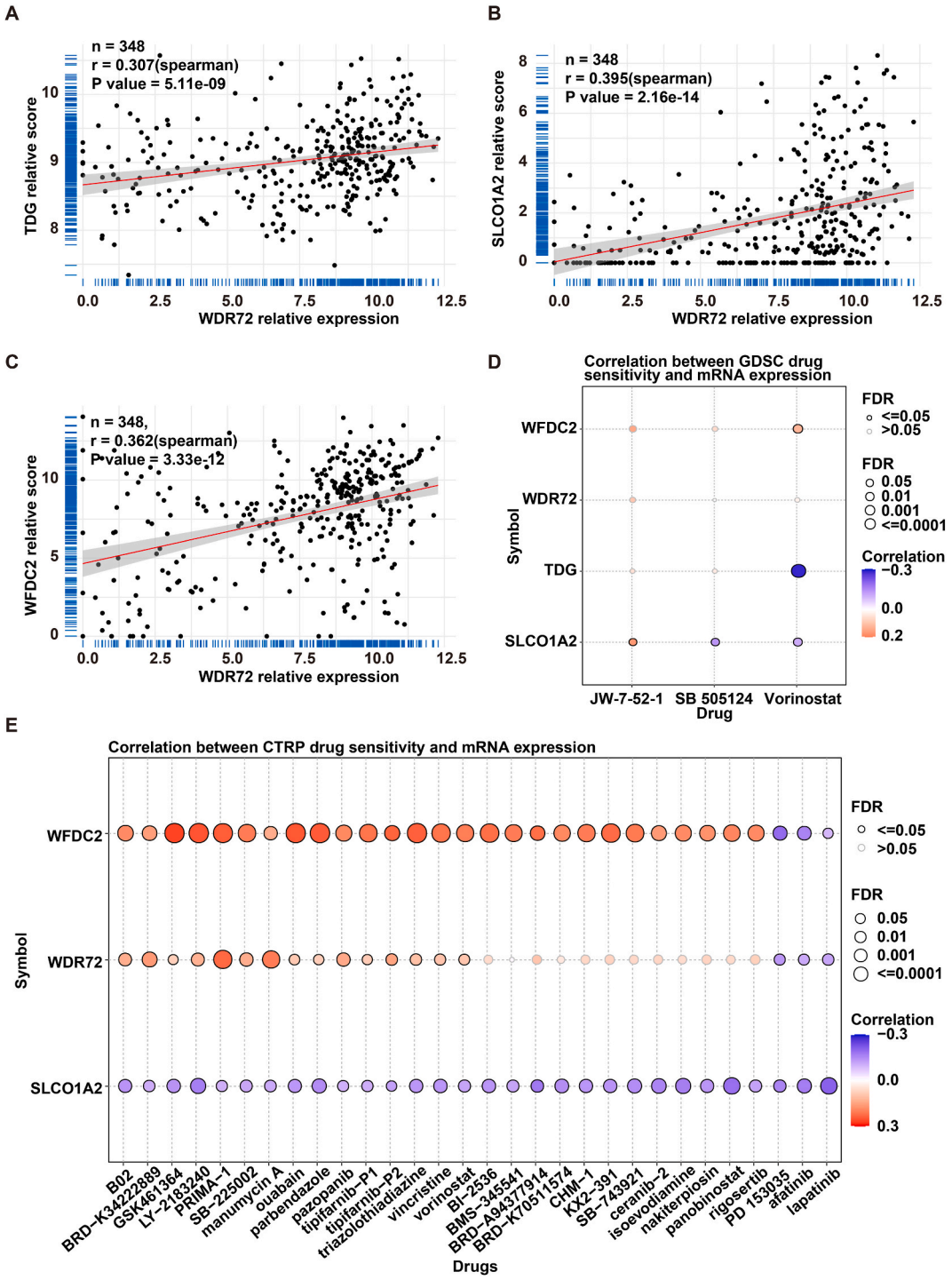


Fig. 9. *WDR72* might be associated with stem cell in STAD, and it was a target gene of multiple small molecule drugs. A-C, The correlation of *WDR72* expression with *TDG*, *SLCO1A2* and *WFDC2* expression. The correlation coefficient is calculated by Pearson correlation coefficients. D, The correlation of *WDR72* mRNA expression with the IC50 of drugs in GDSC. FDR was corrected by the Benjamini-Hochberg method. E, The correlation of *WDR72* mRNA expression with the IC50 of drugs in CTRP. FDR was corrected by the Benjamini-Hochberg method.

of CD4 + T cells and CD8 + T cells were associated with a shorter overall survival rate of gastric cancer patients [38]. Additionally, well-differentiated gastric cancer exhibited increased infiltration of CD4 + T cell populations [39]. This suggested that in gastric cancer, *WDR72* might regulate the infiltration of CD4 + and CD8 + T cells to promote the progression of the tumor. Furthermore, we

observed that patients with low *WDR72* expression had higher levels of *PD-L1* expression. Notably, gastric cancer patients with high *PD-L1* expression had better overall survival compared to those with low *PD-L1* expression [40]. Treatment with pembrolizumab resulted in tumor regression in 53 % of PD-L1-positive advanced gastric cancer patients, with 22 % achieving partial remission on imaging [41]. Therefore, anti-PD-L1 antibody therapy could be a promising approach for gastric cancer patients with low *WDR72* expression.

Although the present study revealed the role of the stemness-related gene *WDR72* in gastric cancer prognosis and immunity using multiple public databases, some limitations should be acknowledged. First, public data sets might not fully represent the overall situation of gastric cancer patients, and there might be biases in the data collection process and data standardization issues might also impact the quality of the data. Future research should involve collecting tumor specimens from gastric cancer patients and conducting single-cell RNA sequencing to confirm the findings. Second, the mechanism of *WDR72* in regulating the immune cell' infiltration and immunotherapy in gastric cancer should be further validated by prospective studies. Additionally, the relationship between *WDR72* and gastric cancer progression should be investigated by in vivo experiments or clinical trials in future studies.

5. Conclusion

We identified a novel CSCs-related gene *WDR72* in gastric cancer epithelial cells. Gastric cancer patients with high *WDR72* expression had high stemness scores and exhibited inferior prognosis and lower *PD-L1* expression. Furthermore, *WDR72* expression was linked to immune cell infiltration and drug sensitivity of gastric cancer. In conclusion, the results of this study confirmed the clinical and prognostic significance of *WDR72* in gastric cancer and suggested that inhibiting the expression of *WDR7*2 might be an important approach for developing therapeutic strategies for gastric cancer.

Funding

This study was supported by the Science and Technology Development Fund of Tianjin Education Commission for Higher Education (2020KJ133), and the Tianjin Key Medical Discipline (Specialty) Construction Project (TJYXZDXK-009A).

Ethical approval and consent to participate

This study was reviewed and approved by Tianjin Medical University Cancer Institute & Hospital, with the approval number: No. bc202317, dated November 24, 2023.

Consent for publication

All participants provided informed consent to participate in the study and for their data to be published.

CRediT authorship contribution statement

Lei Zheng: Writing – original draft, Validation, Software, Methodology, Funding acquisition, Formal analysis, Data curation, Conceptualization. **Jia Lu:** Writing – original draft, Software, Methodology, Formal analysis, Data curation, Conceptualization. **Dalu Kong:** Writing – review & editing, Visualization. **Yang Zhan:** Software, Formal analysis, Data curation.

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.

Data availability

The data utilized and analyzed during the current study are openly available in The Cancer Genome Atlas (TCGA, https://tcga-data.nci.nih.gov/tcga/) database and the Gene Expression Omnibus (GEO, https://www.ncbi.nlm.nih.gov/geo/) database.

Acknowledgement

Not applicable.

Appendix A. Supplementary data

Supplementary data to this article can be found online at https://doi.org/10.1016/j.heliyon.2024.e35549.

References

[1] E.C. Smyth, M. Nilsson, H.I. Grabsch, N.C. Van Grieken, F. Lordick, Gastric cancer, Lancet 396 (10251) (2020) 635–648.
[2] M. Riihimaki, A. Hemminki, K. Sundquist, J. Sundquist, K. Hemminki, The epidemiology of metastases in neuroendocrine tumors, Int. J. Cancer 139 (12) (2016) 2679–2686.
[3] J. Machlowska, J. Baj, M. Sitarz, R. Maciejewski, R. Sitarz, Gastric cancer: epidemiology, risk factors, classification, genomic characteristics and treatment strategies, Int. J. Mol. Sci. 21 (11) (2020).
[4] A.R. Yusefi, K. Bagheri Lankarani, P. Bastani, M. Radinmanesh, Z. Kavosi, Risk factors for gastric cancer: a systematic review, Asian Pac. J. Cancer Prev. APJCP : APJCP. 19 (3) (2018) 591–603.
[5] J. Hippisley-Cox, C. Coupland, Symptoms and risk factors to identify women with suspected cancer in primary care: derivation and validation of an algorithm, Br. J. Gen. Pract. : J. Roy. Coll. Gen. Pract. 63 (606) (2013) e11–e21.
[6] R.E. Sexton, M.N.A. Hallak, M.H. Uddin, M. Diab, A.S. Azmi, Gastric cancer heterogeneity and clinical outcomes, Technol. Cancer Res. Treat. 19 (2020) 1533033820935477.
[7] Z. Tan, Recent advances in the surgical treatment of advanced gastric cancer: a review, Med. Sci. Mon. Int. Med. J. Exp. Clin. Res. : international medical journal of experimental and clinical research 25 (2019) 3537–3541.
[8] A.H. Choi, J. Kim, J. Chao, Perioperative chemotherapy for resectable gastric cancer: MAGIC and beyond, World J. Gastroenterol. 21 (24) (2015) 7343–7348.
[9] S.P. Ng, T. Leong, Role of radiation therapy in gastric cancer, Ann. Surg Oncol. 28 (8) (2021) 4151–4157.
[10] F. Meric-Bernstam, A.M. Johnson, E.E.I. Dumbrava, et al., Advances in HER2-targeted therapy: novel agents and opportunities beyond breast and gastric cancer, Clin. Cancer Res. 25 (7) (2019) 2033–2041.
[11] A. Hogner, M. Moehler, Immunotherapy in gastric cancer, Curr. Oncol. 29 (3) (2022) 1559–1574.
[12] R. Huang, E.K. Rofstad, Cancer stem cells (CSCs), cervical CSCs and targeted therapies, Oncotarget 8 (21) (2017) 35351–35367.
[13] J.P. Medema, Targeting the colorectal cancer stem cell, N. Engl. J. Med. 377 (9) (2017) 888–890.
[14] S.W. Shin, Clinical and therapeutic implications of cancer stem cells, N. Engl. J. Med. 381 (10) (2019) e19.
[15] M. Najafi, B. Farhood, K. Mortezaee, Cancer stem cells (CSCs) in cancer progression and therapy, J. Cell. Physiol. 234 (6) (2019) 8381–8395.
[16] J.C. Chang, Cancer stem cells: role in tumor growth, recurrence, metastasis, and treatment resistance, Medicine (Baltim.) 95 (1 Suppl 1) (2016) S20–S25.
[17] M. Otaegi-Ugartemendia, A. Matheu, E. Carrasco-Garcia, Impact of cancer stem cells on therapy resistance in gastric cancer, Cancers 14 (6) (2022).
[18] T. Wang, C.W. Ong, J. Shi, et al., Sequential expression of putative stem cell markers in gastric carcinogenesis, Br. J. Cancer 105 (5) (2011) 658–665.
[19] S.E. Duzcu, H.M. Astarci, N. Tunc, C. Boran, Expressions of putative cancer stem cell markers, CD44 and CD133, are correlated with pathological tumour stage in gastric adenocarcinoma, Journal of the College of Physicians and Surgeons–Pakistan : JCPSP 30 (5) (2021) 553–558.
[20] R. Zhao, B. He, Q. Bie, et al., AQP5 complements LGR5 to determine the fates of gastric cancer stem cells through regulating ULK1 ubiquitination, J. Exp. Clin. Cancer Res. 41 (1) (2022) 322.
[21] R. Guo, A. Chu, Y. Gong, Identification of cancer stem cell-related biomarkers in intestinal-type and diffuse-type gastric cancer by stemness index and weighted correlation network analysis, J. Transl. Med. 18 (1) (2020) 418.
[22] H. Zhao, R. Jiang, C. Zhang, Z. Feng, X. Wang, The regulatory role of cancer stem cell marker gene CXCR4 in the growth and metastasis of gastric cancer, npj Precis. Oncol. 7 (1) (2023) 86.
[23] Q. Zhang, Y. Ma, Y. Yan, L. Zhang, Y. Zhang, CYB5R1 is a potential biomarker that correlates with stemness and drug resistance in gastric cancer, Transl Oncol 39 (2024) 101766.
[24] G. Yu, L.G. Wang, Y. Han, Q.Y. He, clusterProfiler: an R package for comparing biological themes among gene clusters, OMICS 16 (5) (2012) 284–287.
[25] A.M. Newman, C.L. Liu, M.R. Green, et al., Robust enumeration of cell subsets from tissue expression profiles, Nat. Methods 12 (5) (2015) 453–457.
[26] P. Li, S.N.A. Bukhari, T. Khan, et al., Apigenin-loaded solid lipid nanoparticle attenuates diabetic nephropathy induced by streptozotocin nicotinamide through Nrf2/HO-1/NF-kB signalling pathway, Int. J. Nanomed. 15 (2020) 9115–9124.
[27] X. Rao, C. Zhang, H. Luo, et al., Targeting gastric cancer stem cells to enhance treatment response, Cells 11 (18) (2022).
[28] C.H. Lee, H.Y. Tsai, C.L. Chen, et al., Isoliquiritigenin inhibits gastric cancer stemness, modulates tumor microenvironment, and suppresses tumor growth through glucose-regulated protein 78 downregulation, Biomedicines 10 (6) (2022).
[29] K.A. Katsura, J.A. Horst, D. Chandra, et al., WDR72 models of structure and function: a stage-specific regulator of enamel mineralization, Matrix Biol. 38 (2014) 48–58.
[30] Y. Zou, Q. Lu, Q. Yao, D. Dong, B. Chen, Identification of novel prognostic biomarkers in renal cell carcinoma, Aging (Albany NY) 12 (24) (2020) 25304–25318.
[31] G. Shi, Q. Ni, Y. Miao, et al., Identification of WD-repeat protein 72 as a novel prognostic biomarker in non-small-cell lung cancer, Mediat. Inflamm. 2023 (2023) 2763168.
[32] Q. Zhao, H. Li, W. Li, et al., Identification and verification of a prognostic signature based on a miRNA-mRNA interaction pattern in colon adenocarcinoma, Front. Cell Dev. Biol. 11 (2023) 1161667.
[33] J. Mares, M. Szakacsova, V. Soukup, J. Duskova, A. Horinek, M. Babjuk, Prediction of recurrence in low and intermediate risk non-muscle invasive bladder cancer by real-time quantitative PCR analysis: cDNA microarray results, Neoplasma 60 (3) (2013) 295–301.
[34] X. Ouyang, X. Shi, N. Huang, et al., WDR72 enhances the stemness of lung cancer cells by activating the AKT/HIF-1alpha signaling pathway, Journal of oncology 2022 (2022) 5059588.
[35] Y. Zhang, X. Wang, Targeting the Wnt/beta-catenin signaling pathway in cancer, J. Hematol. Oncol. 13 (1) (2020) 165.
[36] C. Ji, L. Yang, W. Yi, et al., Capillary morphogenesis gene 2 maintains gastric cancer stem-like cell phenotype by activating a Wnt/beta-catenin pathway, Oncogene 37 (29) (2018) 3953–3966.
[37] K. Guo, J. Duan, J. Lu, et al., Tumor necrosis factor-alpha-inducing protein of Helicobacter pylori promotes epithelial-mesenchymal transition and cancer stem-like cells properties via activation of Wnt/beta-catenin signaling pathway in gastric cancer cells, Pathogens and disease 80 (1) (2022).
[38] Q. You, T. Fang, X. Yin, et al., Serum CD4 is associated with the infiltration of CD4(+)T cells in the tumor microenvironment of gastric cancer, J Immunol Res 2021 (2021) 6539702.
[39] F. Li, Y. Sun, J. Huang, W. Xu, J. Liu, Z. Yuan, CD4/CD8 + T cells, DC subsets, Foxp3, and Ido expression are predictive indictors of gastric cancer prognosis, Cancer Med. 8 (17) (2019) 7330–7344.
[40] S.T. Kim, R. Cristescu, A.J. Bass, et al., Comprehensive molecular characterization of clinical responses to PD-1 inhibition in metastatic gastric cancer, Nat. Med. 24 (9) (2018) 1449–1458.
[41] E.C. Smyth, D. Cunningham, Encouraging results for PD-1 inhibition in gastric cancer, Lancet Oncol. 17 (6) (2016) 682–683.