Heliyon 10 (2024) e38662

Contents lists available at ScienceDirect

Heliyon

journal homepage: www.cell.com/heliyon

Research article

Integration of single-cell sequencing and bulk transcriptome data develops prognostic markers based on PCLAF^{+} stem-like tumor cells using artificial neural network in gastric cancer

Yong Shi [a,b], Ke An [a,b], ShaoX zhou [a,b], XuR. Zhang [a,b], QuanC. Kan [a,b,**], Xin Tian [a,b,*]

[a] Department of Pharmacy, The First Affiliated Hospital of Zhengzhou University, Zhengzhou, 450052, China
[b] Henan Key Laboratory of Precision Clinical Pharmacy, Zhengzhou University, Zhengzhou, Henan, 450052, China

ARTICLE INFO

Keywords:
Gastric cancer
Tumour stem cell
Prognosis
Single-cell sequencing
Artificial neural network

ABSTRACT

Background: Gastric cancer stem cells (GCSCs) are important tumour cells involved in tumourigenesis and gastric cancer development. However, their clinical value remains unclear due to the limitations of the available technologies. This study aims to explore the clinical significance of GCSCs, their connection to the tumour microenvironment, and their underlying molecular mechanisms.

Methods: Stem-like tumour cells were identified by mining single-cell transcriptomic data from multiple samples. Integrated analysis of single-cell and bulk transcriptome data was performed to analyse the role of stem-like tumour cells in predicting clinical outcomes by introducing the intermediate variable mRNA stemness degree (SD). Consensus clustering analysis was performed to develop an SD-related molecular classification strategy to assess the clinical characteristics in gastric cancer. A prognostic model was constructed using a customized approach that comprehensively considered SD-related gene signatures based on an artificial neural network.

Results: By analysing single-cell data and validating immunofluorescence results, we identified a PCLAF^{+} stem-like tumour cell population in GC. By calculating SD, we observed that PCLAF^{+} stem-like tumour cells were associated with poor prognosis and certain clinical features. The SD was negatively correlated with the abundance of most immune cell types. Furthermore, we proposed an SD-related classification method and prognostic model. In addition, the customised prognostic model can be used to predict whether a patient respond to PD-1/PD-L1 immunotherapy.

Conclusion: We identified a cluster of stem-like cells and elucidated their clinical significance, highlighting the possibility of their use as immunotherapeutic targets.

1. Introduction

Gastric cancer (GC) is one of the most common malignancies globally and the fourth most common cause of cancer-related death

* Corresponding author. Department of Pharmacy, The First Affiliated Hospital of Zhengzhou University, Zhengzhou 450052, China.
** Corresponding author. . Department of Pharmacy, The First Affiliated Hospital of Zhengzhou University, Zhengzhou 450052, China.
E-mail addresses: kanqc@zzu.edu.cn (QuanC. Kan), tianx@zzu.edu.cn (X. Tian).

https://doi.org/10.1016/j.heliyon.2024.e38662
Received 22 April 2024; Received in revised form 10 August 2024; Accepted 26 September 2024
Available online 29 September 2024

Abbreviations

TME	tumor microenvironment
GCSCs	gastric cancer stem cells
SD	stemness degree
GC	gastric cancer
CSCs	cancer stem cells
GEO	gene expression omnibus
TCGA	the cancer genome atlas
CNV	copy number variation
GSVA	gene set variation analysis
PCA	principal component analysis
DEGs	differentially expressed genes
CDF	cumulative distribution function
TMB	tumor mutation burden
AJCC	American Joint Committee on Cancer

[1]. Traditional treatments for gastric cancer include surgery, chemotherapy, radiotherapy, and targeted therapy [2]. However, owing to the heterogeneity of tumours and lack of effective tumour biomarkers to guide clinical decision-making, clinical outcomes have been unsatisfactory. Tumour immunotherapy, a novel treatment modality blocking immune checkpoints, is widely used to treat solid tumours, including GC [3]. Previous studies have shown that anti-PD1/PDL1 immunotherapy significantly improves the survival of patients with advanced gastric cancer [4]. However, only a subset of patients responds to immunotherapy, and many do not benefit from it [5]. Thus, there is an urgent need to discover new tumour biomarkers to guide clinical decisions and predict immunotherapy responses in the era of individualised GC treatment.

The main feature of cancer stem cells (CSCs) includes self-renewal, high tumorigenicity, differentiation potential, and drug resistance [6,7]. These dedifferentiated malignant cells can diffuse to distant sites, resulting in tumour progression [8,9]. Furthermore, both clinical studies and experimental models have shown that CSCs remain active after the use of cancer therapeutics, suggesting that they may be responsible for treatment failure [10]. Therefore, CSCs may play an important role in tumour progression. In the case of GC, CSCs have been identified by analysing various gastric cell lines [11], and further studies have confirmed the existence of gastric cancer stem cells (GCSCs) or stem-like cells in GC tissues [12]. Recently, Li et al. isolated GCSCs from GC cell lines. Further experiments to assess their tumourigenicity *in vivo* demonstrated that GCSCs facilitated the formation of xenografts in nude mice [13], indicating that they play a key role in maintaining the tumourigenicity of GC. In addition to tumourigenicity, GCSCs are associated with clinical prognosis and treatment failure [14]. For example, patients with high expression of GCSC markers such as LGR5, CD44, and CD24 have a short median survival [15-17]. Moreover, $CD44^{+}$ GCSCs exhibit stronger resistance to 5-FU than $CD44^{-}$ GC cells [18]. However, owing to technological limitations, previous studies have only indirectly investigated the clinical relevance of GCSCs using *in vitro* experiments or molecular markers. Moreover, no study has explained the association between GCSCs and the occurrence or development of GC in the clinical setting. Therefore, it is necessary to understand in detail the clinical significance of GCSCs.

Single-cell RNA sequencing (scRNA-seq) can precisely identify various cell types in complex tissues [19,20]. which enables researchers to understand the microenvironmental components of the tumour ecosystem [21]. Wu et al. systematically identified 11 cell clusters with distinct distribution patterns across different lung cancers [22]. Kumar et al. were the first to construct a single-cell atlas of GC across clinical stages and histological subtypes [23]. However, all the above-mentioned reports focused on resolving the cell components in tumour tissues. The clinical significance of certain cell clusters in tumour tissues has not been studied in detail, owing to the small sample capacity of single-cell sequencing and insufficient matching of clinical annotations. Bulk RNA sequencing (RNA-seq) requires a large sample size and provides complete clinical information. Thus, it is now possible to analyse the clinical relevance of GCSCs by integrating single-cell and bulk transcriptomes.

Here, using scRNA-seq data analysis obtained from the published literature, we identified stem-like tumour cells by referring to the expression of a panel of gene signatures. We used stemness degree (SD) to link stem-like tumour cells to bulk RNA-seq data to assess the clinical significance of stem-like tumour cells. We also demonstrated that stem-like tumour cells may participate in the remodelling of the immune microenvironment. Moreover, we proposed a novel classification method based on SD-specific gene signatures and performed a comprehensive comparison between the classification and conventional clinical traits. Finally, by mining valuable information on these SD-specific markers, we established a prognostic scoring system to evaluate clinical outcomes in patients with GC and observed that this model could predict the efficacy of anti-PD-1/L1 immunotherapy.

2. Materials and methods

2.1. Acquisition and preprocessing of gastric cancer data

Single-cell and bulk transcriptome datasets with clinical information were obtained from Gene Expression Omnibus (GEO) and the Cancer Genome Atlas (TCGA) database. Briefly, a single-cell sequencing dataset (dbGAP: phs001818.v1.p1) and six qualified bulk

transcriptomic cohorts [GSE57303, GSE15459, GSE62254/ACRG, GSE34942, GSE84437, and TCGA-STAD (The Cancer Genome Atlas-Stomach Adenocarcinoma)] were harvested in this study. For the single-cell sequencing dataset, normal samples were excluded, and only tumor samples were enrolled for further evaluation. For bulk transcriptome data, tumour samples without clinical information were deleted from downstream analysis. Regarding microarray data, the files of the normalized gene expression matrix were downloaded from the GEO database. For GC datasets in TCGA, the gene expression data (FPKM) were downloaded using the TCGA biolinks tool in the R software [24]. Due to few normal samples in TCGA-STAD datasets, these samples were removed from downstream analysis. Batch effects were corrected using the sva package with the "ComBat" algorithm [25]. The basic information of all GC datasets in this study is summarized in Table S1. The somatic mutation (MAF format) and Copy Number Variation (CNV) data were downloaded from the TCGA database. All data analyses were carried out using the R platform (version 4.0.1).

2.2. Single-cell sequencing data analysis

Low-quality cells and genes (min.cells <3, min.genes <200) were filtered out from further analysis. The cell cycle scores were normalized using the "var.regress.out" parameter in Seurat. For each pool, normalization was conducted by the "SCTransform" module from the Seurat [26] R package. Highly variable genes were selected based on the mean-variance curve, while principal-component analysis (PCA) was used to reduce data dimensionality. The Louvain algorithm was used to cluster cells. Finally, UMAP algorithm was used to visualize these cell populations.

2.3. Construction of cell trajectory

Monocle2 (version 2.10.1) [27], an unsupervised algorithm, was used for cell trajectory construction. Briefly, the new monocle object was created by inputting gene count matrix. Then, the high-variable genes were identified with default parameters for downstream analysis. Finally, we used "DDR-tree" algorithm to reduce the dimension of cells and ordered cells based on pseudo-time.

2.4. Derivation of stemness degree using OCLR

The mRNA expression-based stemness degree (SD) is a robust indicator to evaluate the degree of tumour dedifferentiation [28]. To calculate the SD of GC samples, a predictive method was established by one-class logistic regression (OCLR). First, the gene matrix of 906 $PCLAF^{+}$ stem-like tumour cells was extracted from single-cell sequencing data, and the genes expression values were preprocessed by mean-subtraction. Then, the predictive model was established using the OCLR algorithm from the R package gelnet (version 1.2.1) [29]. Next, we computed the Spearman's correlation between the model's weights and the individual GC sample's gene expression profile. Finally, Spearman's correlation coefficient was standardized to the [0, 1] range by a linear transformation, which was regarded as the stemness degree.

2.5. Gene set variation analysis (GSVA)

GSVA is generally used to evaluate the enrichment of pathways and biological processes for individual sample based on the gene expression profile [30]. GSVA enrichment analysis was performed by GSVA R package. The gene sets of "c2.cp.kegg.v6.2.symbols" from the MSigDB database were served as the baseline for subsequent GSVA analysis. An adjusted P-value of <0.05 was considered as statistically significant. Functional enrichment analysis for SD-related genes was performed using the ClusterProfiler software, with a criterion of adjusted $P < 0.05$.

2.6. Evaluation of immune cell abundance and tumour microenvironment

CIBERSORT is commonly employed to compute the enrichment of immune cell infiltration in the solid tumour microenvironment (TME) [31]. First, gene signature expression profiles of 28 immune cell types were obtained from previous studies [32]. Then, taking the reference expression profile of immune cell and the mix gene expression profiles of gastric cancer as input, CIBERSORT software was used for deconvolution with default parameters to calculate the abundance of immune cells for each sample. To exclude the influence of epithelial cells on the correlation between SD and immune cell abundance, a linear regression model was used to correct the immune cell abundance. The R package ESTIMATE was applied to evaluate the TME [33].

2.7. Scanning differentially expressed genes (DEGs)

To identify SD-related genes, all samples of GC were divided into two subsets based on the median SD value. The R package DESeq2 was employed to identify DEGs between high- and low-SD subsets. The cutoff value for statistical significance was set as adjusted $P < 0.001$.

2.8. Identification of the SD-related molecular subtypes in GC patients

Novel molecular classification of GC patients according to the DEGs was identified by the k-means machine learning algorithm in the R package "ConsensusClusterPlus" [34]. The entire clustering process consisted of 1000 iterations, each iteration sampling 80 % of

the clustering. The optimal number of clusters was determined by a comprehensive evaluation of the proportion of ambiguous clustering (PAC) algorithm, the area under the cumulative distribution function (CDF) curve, and the consensus heatmap.

2.9. Transformed features using an autoencoder neural network

Five preprocessed GC datasets comprising 1051 samples were used as the input for the autoencoders framework. The matrices were unit-norm scaled by sample. The autoencoder framework is widely used for data compression and dimensionality reduction [35]. For a given input layer x = (x1, …, xn) of dimension n, an autoencoder could use output x' to reconstruct x (x and x' have the same dimension), via transforming x in successive hidden layers.

The input layer is a two-dimension array:

$$X = \begin{pmatrix} x_1^1 & \cdots & x_i^1 \\ \vdots & \ddots & \vdots \\ x_1^j & \cdots & x_i^j \end{pmatrix}$$

The activation function for the hidden layers is tanh:

$$\tanh(z) = \frac{e^{2z} - 1}{e^{2z} + 1}$$

The output of a hidden layer is:

$$X' = \tanh\left(W_h * X + b\right)$$

Where X', W_h, and b are the output, weights, and bias of hidden layer respectively.

The loss function is:

$$Loss(X, X') = \frac{1}{2j}\sum_{j=1}^{j}\left(x'_j - x_j\right)^2 + \lambda\left(||W_h, b||_2\right)$$

Where Wh, b, are the weights and bias of all hidden layers respectively. $\lambda(||W_h, b||_2)$ is the L2 regulation term.

We constructed an autoencoder with five hidden layers consisting of 1000, 500, 50, 500, and 1000 nodes, respectively using the PyTorch (version 1.13.0) framework. We used the gradient descent algorithm with 100 epochs to train the autoencoder. Finally, the features from the bottleneck layer of the autoencoder were extracted for downstream analysis.

2.10. Generation of robust SD-related prognosis model based on autoencoder neural network

First, the preprocessed expression profile of 289 SD-related DEGs from 1051 samples were classified into training and testing sets in a proportion of 4:1. Then, the training set was used to optimize parameters for the autoencoder neural network model. The autoencoder neural network was used for dimensionality reduction, resulting in 50 feature values. Next, we selected prognostic-related features from the 50 feature values via a univariate Cox regression analysis. Subsequently, a prognostic model was constructed using multivariate Cox regression based on determined prognostic-related features. Finally, we used testing set to evaluate the robustness of the model.

2.11. Collection of immunotherapeutic datasets

Two PD-1/PD-L1 checkpoint immunotherapeutic datasets, namely, the GSE78220 dataset (advanced melanoma) and the IMvigor210 dataset (metastatic urothelial cancer), were enrolled in this study [36]. For the IMvigor210 cohort, the gene expression matrix and clinical annotations were obtained from the IMvigor210 R package. For the GSE78220 cohort, the gene expression matrix and detailed clinical annotations were obtained from the GEO database. The raw count data of both two datasets were normalized using the DEseq2 R package for downstream analysis. The basic information of two immunotherapeutic cohorts is compiled in Table S1.

2.12. Immunofluorescence

The thickness of the paraffin sections was set as 5 μm. For immunofluorescence staining, sections were deparaffinized using xylene (2 × 15 min) and subsequently rehydrated with a gradient of alcohol (100 %, 95 %, 80 %, 70 %), followed by distilled water, all at room temperature. Antigen retrieval of sections was carried out by high-temperature treatment at 95 °C for 30 min, with sodium citrate buffer (0.01 M, pH 6.0). After cooling, tissue sections were sequentially blocked for 25 min and 30 min at 25 °C in 3 % hydrogen peroxide solution and BSA buffer (0.3 % Triton X-100, 1 % BSA), followed by washing three times with PBS (3 × 5 min). Then, sections were incubated overnight at 4 °C with primary antibodies. Next day, after rinsing three times with phosphate-buffered saline (PBS, 3 × 15 min), sections were incubated for 50 min with HRP-labeled secondary antibody, followed by washing three times with PBS (3 × 5 min). Sections were incubated for 10 min with tyramide, followed by washing three times with TBST (3 × 5 min). Finally, sections were treated with DAPI (G1012, Servicebio) for nuclear visualization and imaged using fluorescence microscope (Nikon). The primary

antibodies included mouse anti-EPCAM (1:1000, MA5-11344, Servicebio), mouse anti-KI67 (1:500, sc-23900, SANTA), mouse anti-PCLAF (1:500, sc-390515, SANTA). Secondary antibody included goat HRP-labeled anti-IgG (1:500, GB23301 and GB23303, Servicebio). Goat anti-Rabbit IgG (1:2000, A-11008, ThermoFisher). Tyramide included iF488-Tyramide (1:500, G1231, Servicebio), iF555-Tyramide (1:500, G1233, Servicebio), iF647-Tyramide (1:500, G1232, Servicebio).

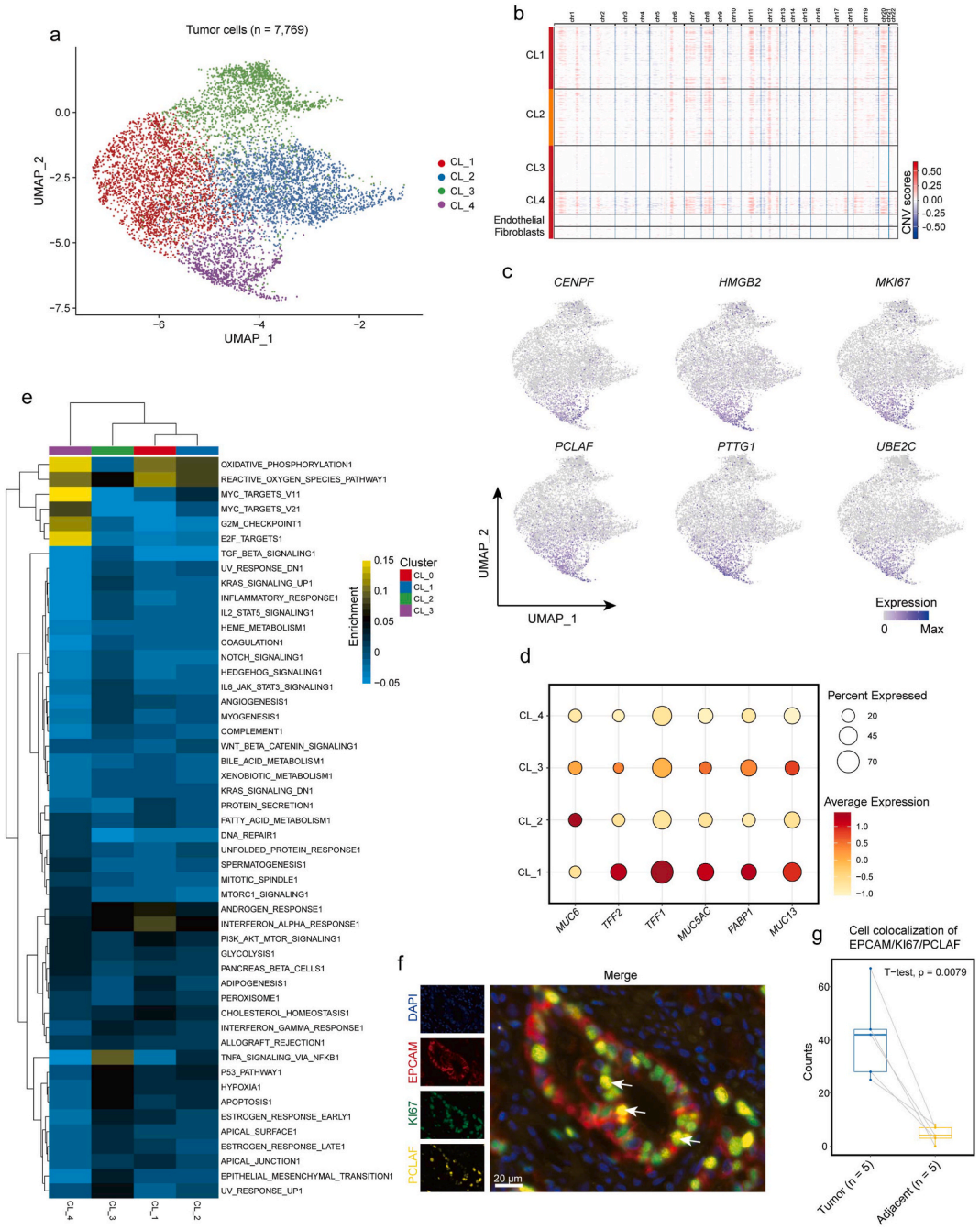


Fig. 1. Identification of stem-like tumor cells in gastric cancer. (a) UMAP visualization of 7769 cells from 10 primary gastric tumours labeled by four tumor cell types. (b) Heatmap showing the CNV scores across 4 epithelial cell subclusters. (c) UMAP plots of expression of cancer stem cell-related gene set (*CENPF*, *HMGB2*, *MKI67*, *PCLAF*, *PTTG1*, *UBE2C*) in tumor epithelial cells. (d) Dot plots showing the expression a panel of gastric glandular epithelium mature-related markers including *MUC6*, *TFF2*, *TFF1*, *MUC5AC*, *FABP1*, *MUC13*. (e) Heatmap showing the enrichment score of 50 tumor hallmark gene signatures in four tumor epithelial cell types. (f) Immunofluorescence staining for DAPI, EPCAM, KI67, PCLAF in GC tissues. EPCAM represented the marker of tumor cells; KI67 and PCLAF represented the markers of $PCLAF^{+}$ stem-like tumor cells. Scale bar = 20 μm. (g) boxplots showing the counts of $EPCAM^{+}KI67^{+}PCLAF^{+}$ tumor cells in five pairs of matched paracancerous and cancerous GC tissues.

2.13. Acquisition of tumour samples from other parts of the digestive system

Besides GC, we also included other digestive system tumour samples to evaluate the SD-related prognostic model. Four datasets, including colon adenocarcinoma, hepatocellular carcinoma, esophageal carcinoma, and pancreatic adenocarcinoma were obtained from the TCGA public database. Only samples with fully overall survival information were retained for downstream analysis. The basic information of four datasets is compiled in Table S1.

2.14. Statistical analysis

The independent student's t-test for continuous variables was employed for pairwise comparisons between groups. The Mann-Whitney *U* test was applied to compare categorical data between two groups. The Kruskal-Wallis test was used to compare three or

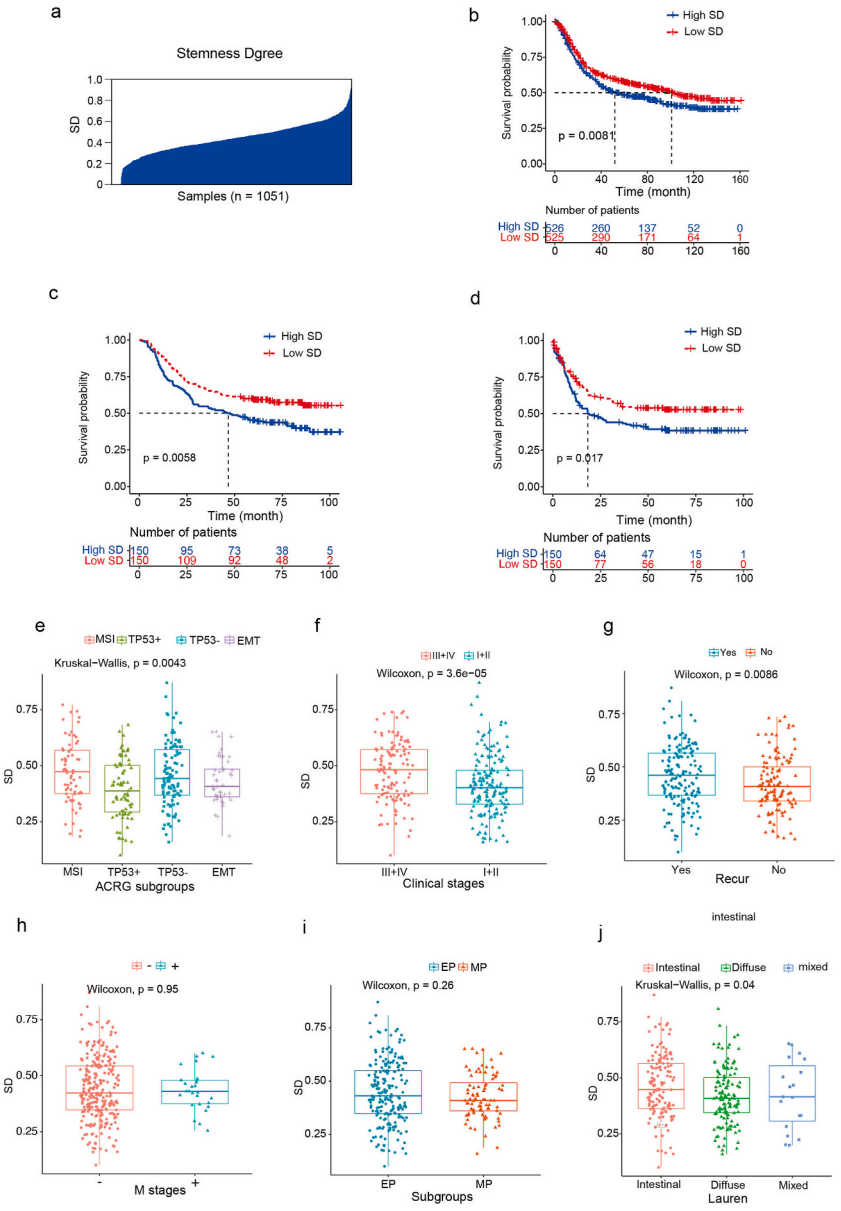


Fig. 2. Comparison of the clinical features and the SD in GC patients. (a) An overview of the SD level in GC patients. the SD values were ranked from low to high among 1051 samples. (b) The Kaplan–Meier curves showed patients with higher SD are associated with worse OS in integrated data including 1051 GC patients ($p < 0.05$). (c, d) The Kaplan–Meier curves showed patients with higher SD were correlated with worse OS (c) and FPS (d) in ACRG dataset ($p < 0.05$). (e–j) Boxplots showing the level of SD in GC patients, stratified by ACRG subgroups (e), clinical stages (f), Recur states (g), M stages (h), molecular subtypes (i), lauren classification (j).

more groups. Correlation analyses between Gaussian distribution variables were evaluated with Pearson's correlation test, while those between non-normal distribution variables were evaluated with Spearman's correlation test. All statistical P-values were two-sided and $p < 0.05$ denoted statistical significance. The Statistical analyses were conducted using the R platform (4.0.1).

3. Results

3.1. Identification of PCLAF$^+$ stem-like tumour cells in gastric tumours

An overview of this study is presented in Fig. S1a. Single-cell transcriptomic data for GC cells derived from ten tumour samples were obtained from a public database (Table S1). A total of 21,461 genes and 21,920 cells were assessed for downstream analysis after quality control and filtering. To differentiate the cell types among all samples, a correction method (Harmonypy) was first used to minimise patient-specific effects. We grouped these cells into seven clusters: epithelial cells, T cells, macrophages, fibroblasts, endothelial cells, B cells, and mast cells (Fig. S1b). Signature genes for each cell type were determined, and their expression scores are presented as dot plots (Fig. S1c). The distribution of these cell clusters varied considerably among all samples, demonstrating widespread tumour heterogeneity (Fig. S1d). To explore the transcriptomic heterogeneity of tumour cells, we extracted total tumour epithelial cells from all samples and clustered them into subpopulations. Four subsets with distinct transcriptional characteristics were detected (Fig. 1a). CNV analysis revealed that CL3 represented a non-malignant cell subpopulation, whereas CL1, CL2, and CL4 represented malignant cell subpopulations (Fig. 1b). It is noteworthy that a cluster of cells specifically expressed tumour stemness-related gene sets, such as *CENPF* [37], *HMGB2* [38], *MKI67* [39], *PTTG1* [40], *UBE2C* [41]. We also found that gastric cancer stem cell gene set 1 (*OLFM4*, *EPHB2*, *AXIN2*, *SOX4*, *SOX9*) [42] and set 2 (*REG1A*, *REG3A*, *CTGF*, *LYZ*) [43] had higher scores in PCLAF + tumor cells (Fig. S1e). Therefore, we considered them to be PCLAF$^+$ stem-like tumour cells owing to the top high expression of *PCLAF* (Fig. 1c). Further cross-referencing of conventional gastric glandular epithelium differentiation-related markers indicated that the PCLAF$^+$ stem-like tumour population represented an undifferentiated or dedifferentiated phenotype (Fig. 1d). Moreover, the tumour hallmark pathway-related cell cycles were significantly enriched in PCLAF$^+$ stem-like tumour cells, reflecting their increased proliferative ability (Fig. 1e). In addition, pseudo-time analysis showed that the PCLAF$^+$ population was presented at the beginning of cell differentiation (Figs. S1f and S1g). To further validate this cell type, we performed multiplex immunofluorescence assays on paired cancerous and adjacent tissues from five GC cases. Indeed, we observed the presence of PCLAF$^+$ stem-like tumour cells in GC tissues, with a significantly higher number in cancer tissues than in adjacent tissues (Fig. 1f–g, Fig. S1h; $p < 0.01$). In summary, we identified a cluster of stem-like tumour cells with high *PCLAF* expression and validated their presence in GC.

3.2. Associations between SD and clinical features in gastric cancer

Tumour stem cells are essential factors that drive cancer development. To investigate the role of PCLAF$^+$ stem-like tumour cells in the development of GC, we introduced the SD variable to evaluate the similarity of the transcriptome between tumour cells and PCLAF$^+$ stem-like tumour cells. Subsequently, SD was used to evaluate the association between PCLAF$^+$ stem-like tumour cells and clinical parameters in GC. Five GC datasets with clinical annotations (GSE15459, GSE34942, GSE57303, GSE62254/ACRG, and GSE84437; Table S1) were integrated to generate a multicentre dataset for further analysis. By applying the one-class logistic regression (OCLR) method, SD was calculated according to the gene expression matrix of PCLAF$^+$ stem-like tumour cells defined in the GC scRNA-seq data (Fig. 2a). Consistent with expectations, the SD value positively correlated with several key characteristic genes of PCLAF$^+$ stem-like cells (Fig. S2a). To determine whether SD is associated with GC prognosis, Kaplan–Meier survival curves were constructed. As shown in Fig. 2b, the high-SD group exhibited poorer overall survival (OS) than the low-SD group in the aggregated dataset, suggesting that SD may be a prognostic factor ($p < 0.05$). To investigate the clinical association between SD and GC, we focused on the ACRG dataset due to its comprehensive clinical information. In this dataset, patients with high SD also demonstrated poorer OS and progression-free survival (PFS) compared to those with low SD ($p < 0.05$, Fig. 2c and d). Moreover, the SD value was higher in the Microsatellite Instability (MSI) subtype than in other subtypes ($p < 0.05$, Fig. 2e) and was not related to the TP53 mutation ($p > 0.05$, Fig. 2e). In addition, SD was related to clinical stage and disease recurrence, but not to M stage. The III + IV and recurring groups demonstrated higher SD than the I + II and non-recurring groups, respectively ($p < 0.05$), while the level of SD was not significantly different between distinct M stages ($p > 0.05$, Fig. 2f–h). We observed no significant differences in SD values between the epithelial phenotype (EP) and mesenchymal phenotype (MP) ($p > 0.05$, Fig. 2i). Regarding the Lauren classification, the intestinal subtype showed a higher SD than the diffuse and mixed subtypes ($p < 0.05$; Fig. 2j). Collectively, these results revealed that low SD may benefit clinical outcomes in patients with GC, and that SD was correlated with several clinicopathological parameters, including MSI status, clinical stage, disease recurrence, and histological subtype.

3.3. Associations between the SD and tumour immune microenvironment (TIME) characteristics in GC

To determine whether CSCs modulate the TIME in GC, the enrichment of 28 immune gene sets, denoting the most common immune cell types in the TIME, was calculated using the CIBERSORT algorithm (Fig. 3a). Then, 1051 patients with GC were clustered into two subtypes according to the abundance of immune cells using the K-means clustering algorithm (Figs. S2b–h). To further investigate the TIME features between the two immune subtypes, we evaluated immune activity using several indicators, including the stromal score (immunosuppressive factor), immune score (immune activation factor), ESTIMATE score (immune score minus stromal score), and tumour purity (Fig. 3a). The results showed that immune cluster A had lower tumour purity and higher immune, stromal, and

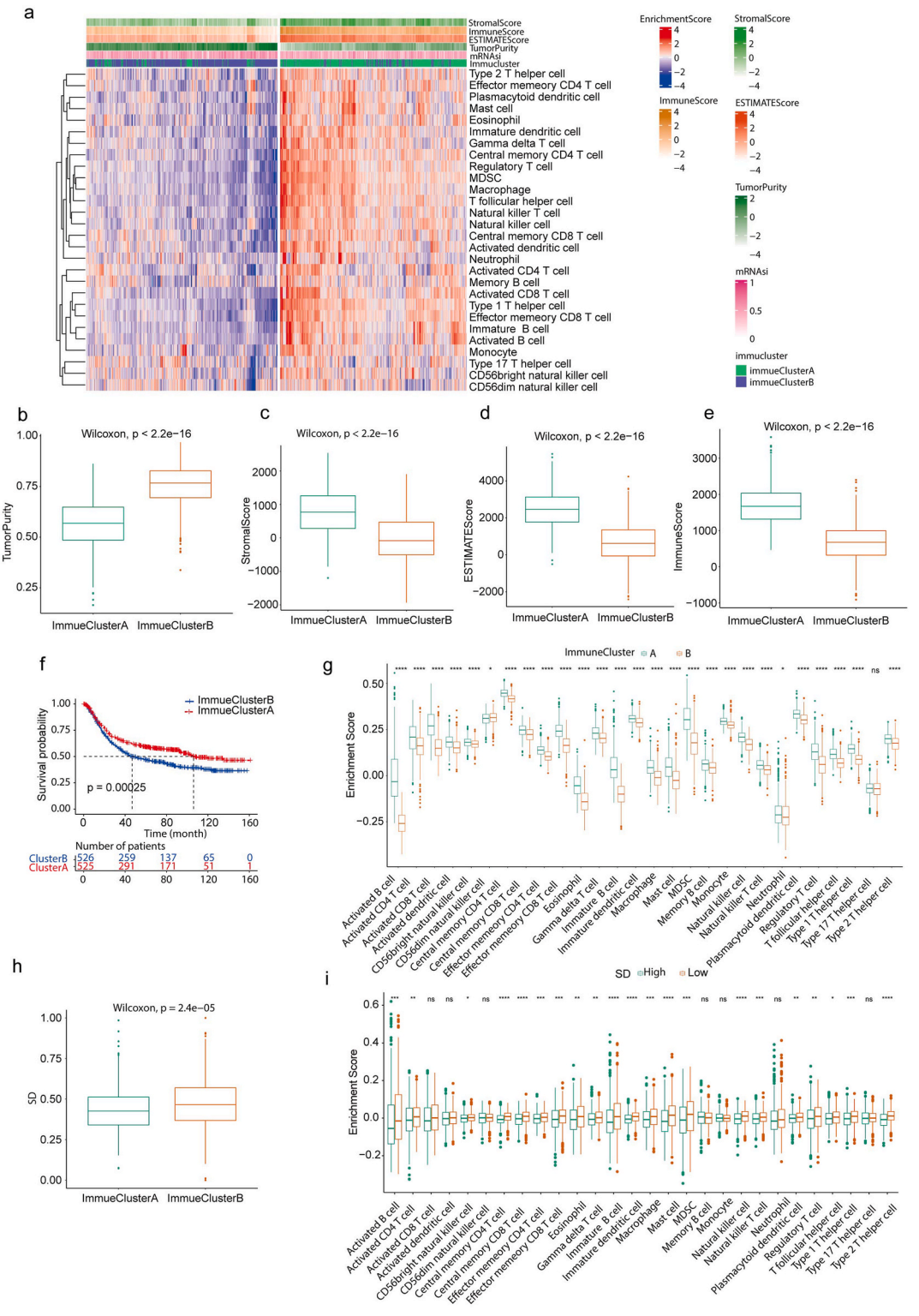
ESTIMATE scores than immune cluster B, indicating that patients in the former group were more likely to develop an immune response to the tumour (Fig. 3b–e). Hence, we considered cluster A as the 'immune response' type and cluster B as the 'immune deficiency' type. Coincidentally, immune cells including $CD8^+$ T cells, $CD4^+$ T cells, and B cells were enriched in cluster A (Fig. 3a–g). Furthermore, patients in immune cluster A experienced longer survival times, indirectly confirming that an anti-tumour effect tended to occur in this cluster ($p < 0.05$, Fig. 3f). We determined the SD level between distinct immune clusters to compare the association between immune status and SD. We observed that immune cluster A presented a lower SD ($p < 0.05$, Fig. 3h), which was negatively correlated with the abundance of most immune cell types (Fig. 3i). Therefore, we concluded that SD was closely correlated with the immune status in GC, suggesting that $PCLAF^+$ stem-like tumour cells may be involved in shaping different TIME landscapes in GC.

3.4. Generation of stemness degree gene signatures

To explore the potential molecular traits of stemness, 289 SD-related DEGs were identified using the DEseq2 package (Fig. S3a and Table S2). Gene ontology (GO) enrichment analyses of the DEGs were performed using ClusterProfiler software. Significant GO terms were listed in Table S3. These genes were consistently predominantly involved in certain cell differentiation- and immunity-related biological processes, confirming that tumour stem-like cells may participate in immune regulation (Fig. S3b). Subsequently, K-means clustering analysis, based on 289 SD-related DEGs, was performed to classify the samples into genetically distinct subtypes. The clustering algorithm identified two distinct SD molecular types, named as gene clusters A and B (Figs. S3c–i). These two gene clusters were characterised by distinct gene signatures (Fig. 4a). Moreover, we observed distinct SD levels between the two subtypes, with a lower SD in cluster B and a higher SD in cluster A (Fig. 4a and b). Consistent with the Kaplan–Meier survival analyses based on SD, gene cluster B (low-SD subtype) was associated with a markedly better prognosis, whereas cluster A (high-SD subtype) was associated with poorer clinical outcomes ($p < 0.001$, Fig. 4c). In addition, patients in cluster B had higher immune scores, indicating that this cluster may be more sensitive to immunotherapy (Fig. 4d). To further investigate the biological behaviour of distinct gene clusters, GSVA enrichment analysis was employed to annotate the molecular functions of the two transcriptomic subtypes. Surprisingly, both tumour stemness-related GO terms (cell proliferation and differentiation) and immunity-related GO terms were significantly enriched in gene cluster A (Fig. 4e–Table S4), further confirming that CSCs may play an important role in shaping the GC immune microenvironment. We focused on the ACRG cohort because of its explicit and relatively complete clinical information (Fig. 4f). As shown in Fig. 4g–i, no significant differences were observed in age, sex, or Lauren subtype between the two transcriptomic subtypes. Importantly, we found that cluster A was preferentially presented in epithelial-mesenchymal transition (EMT) subtypes (Fig. 4k; 84.8 % vs. 15.2 %), suggesting that cluster A is characterised by EMT activation. Gene cluster A was mainly concentrated in stage I (Fig. 4j; 63.3 % vs. 36.7 %), whereas gene cluster B was widely distributed in stages II and III (Fig. 4j; 41.2 % vs. 58.8 % and 40.7 % vs. 59.3 %, respectively). We also observed that gene cluster B presented significantly longer OS and FPS based on Kaplan–Meier survival analysis (Fig. 4l and m). Based on the above results, we suggested that SD signatures, as novel molecular subtypes, can be used to evaluate the clinical characteristics of gastric tumours.

3.5. Construction of prognostic model derived from the stemness degree gene signatures

SD-related phenotypes were relevant to the prognosis and progression of GC. However, we could not accurately quantify the effect of stemness transcriptome subtypes on the prognosis of individual patients because our conclusions were based on patient populations. To address this, we developed a scoring system, termed the stemness score, by integrating SD-related genes using a customized autoencoder artificial neural network (Fig. 5a). First, all SD-related gene signatures were inputted into the autoencoder model for dimensionality reduction. Univariate Cox regression analysis was performed for 50 features derived from the bottleneck layer of the autoencoder model (Table S5). Nine significant prognosis-related features were selected to construct the prognostic model (Table S6). The equation of the prognostic model is as follows: stemness score = exp (1.12 × features_8) + exp (−0.66 × features_18) + exp (0.7 × features_32) + exp (−0.93 × features_42) + exp (0.62 × features_50). Further Kaplan–Meier survival analyses demonstrated that patients with GC exhibited poorer clinical outcomes, with increasing stemness scores in both the training and testing groups (Fig. 5b and c, $p < 0.05$). Moreover, the multivariate Cox regression analyses, which enrolled the factors of age, gender, TNM status, MHL1 status, and clinical stage, indicated stemness score as a robust and independent prognostic biomarker for the evaluation of patient outcomes (Fig. 5g, HR 2.0 (1.47–2.7), $p < 0.001$). To comprehensively illustrate the characteristics of the stemness score, we investigated whether it correlated with SD-related genomic phenotypes. Statistically significant differences were observed in stemness scores between the SD gene clusters. Gene cluster A had a higher stemness score (Fig. S4a, $p < 0.05$). We also observed that the median score for immune cluster B was greater than that for immune cluster A, indicating that a lower stemness score may represent an immune activation state, whereas a higher stemness score denotes an immune inhibition state (Fig. S4b, $p < 0.05$). Moreover, we compared various components of immune cells between the groups with distinct stemness scores using the CIBERSORT method. The group with a low stemness score was prominently enriched in adaptive immune cells, including B-cells, $CD4^+$, and $CD8^+$ T cells ($p < 0.05$; Fig. S4c). This confirmed that a low stemness score was related to immune activation. Furthermore, enrichment analyses showed that low stemness scores were significantly enriched in immune pathways, whereas high scores were correlated with stromal pathway activation (Fig. 5e). Therefore, the stemness score not only predicted the prognosis of individual patients but also contributed to the assessment TIME cell-infiltration characterisation of tumours. To further compare the association between stemness score and known gastric signatures, we employed Spearman's correlation analysis among these signatures (Fig. 5f). In addition, patients with the EMT subtype showed the highest stemness score compared to the other three ACRG molecular subtypes, whereas those with the MSI subtype had the lowest scores (Fig. 5d, $p < 0.05$). These findings highlight the stemness score as a valuable predictor for assessing clinical



(caption on next page)

Fig. 3. The association of SD and tumor immune microenvironment subtypes. (a) Heatmap showing the enrichment of 28 immune signatures between immune cluster A and B. SD, tumor purity, ESTIMATEScore, ImmuneScore, StromalScore also were mapped. (b–e) Boxplots showing the level of immune signatures including tumor purity (b), ESTIMATEScore (c), ImmuneScore (d), StromalScore (e) between immune cluster A and B. (f) The Kaplan–Meier curves of patients with immune cluster A and B in integrated dataset including 1051 GC patients ($p < 0.05$). (g) Boxplots showing the enrichment score of 28 immune signatures between immune cluster A and B in GC patients. * represents $p < 0.05$; ** represents $p < 0.01$; *** represents $p < 0.001$; ns, no significance. (h) Boxplots demonstrating SD level between immune cluster A and B ($p < 0.0001$). (i) Boxplots showing the enrichment score of 28 immune signatures between high and low SD groups in GC patients. The enrichment scores were corrected to remove the effect of epithelial cells using a linear regression method. * represents $p < 0.05$; ** represents $p < 0.01$; *** represents $p < 0.001$; ns, no significance.

outcomes in GC.

3.6. Validation of prognosis model in the independent dataset from TCGA

To assess the robustness of the prognostic model, we enrolled another independent STAD cohort from the TCGA database. By applying the algorithm, the stemness score for each patient was calculated and then the patients were divided into two groups according to the median stemness score (Fig. 6a). As anticipated, the low-stemness score group demonstrated significantly longer overall survival (OS) compared to the high-stemness score group (Fig. 6a, $p < 0.05$). Further validation using multivariate Cox regression analysis, which included factors such as patient age, gender, and TNM status, affirmed that the stemness score was an independent prognostic biomarker (HR = 1.2 (1.12–1.2), p < 0.05, Fig. 6b). Subsequently, to evaluate the model's applicability across various digestive system cancers, this algorithm was applied to colon adenocarcinoma, hepatocellular carcinoma, esophageal carcinoma, and pancreatic adenocarcinoma. We observed significant differences in clinical outcomes between the high and low stemness scores in the PAAD cohort. However, no significant correlation was observed among the other cohorts (Fig. 6c). These findings demonstrated that the stemness score, a prominent predictor of GC, also facilitated the assessment of the prognosis of PAAD, but not of all digestive system tumours.

We then explored the associations between genomic variations and stemness scores in the GC dataset from TCGA using maftools software. The high stemness score group had a higher tumour mutation burden (TMB) frequency than the low stemness group, with the frequencies for the first most significantly mutated gene, *TTN,* being 59 % and 44 %, respectively (Fig. 6d and e). Moreover, patients with high stemness scores presented a higher level of CNV in both amplifications and deletions (Fig. 6f). TMB quantification analysis also showed a positive correlation between the stemness score and TMB level (Fig. 6g and h). These findings suggest that the stemness score may reflect the degree of genomic variation in GC.

3.7. Association between stemness score and anti-PD-1/L1 immunotherapy

In recent years, PD-L1- and PD-1-blocking immunotherapy has been a breakthrough in cancer treatment. However, owing to the heterogeneity of tumours, the therapeutic effects of immunotherapy in some patients are unsatisfactory. To investigate the predictive value of the stemness score for anti-PD-1/L1 immunotherapy, two immunotherapy datasets from the GEO database were included in the subsequent analysis. In both datasets, lower stemness scores were significantly associated with higher PD-L1 levels, suggesting that patients with low stemness scores were more likely to respond to immunotherapy (Fig. 7a and b). Consistently, the low stemness score group exhibited a higher proportion of patients responding to anti-PD-1/L1 checkpoint therapy (Fig. 7c and d 44.6 % versus 21.8 %, $p < 0.001$), and the patients responding to immunotherapy demonstrated lower stemness scores (Fig. 7e and f). Prognostic analyses further supported these findings, showing that patients with low stemness scores had significantly prolonged survival and better clinical outcomes (Fig. 7g and h). Further analysis indicated that Tregs and stromal signatures were enriched in patients with high stemness scores, which suppressed the immune response to tumours (Fig. 7i). Given that tumor mutation burden (TMB) is associated with immunotherapy efficacy, we combined stemness scores and TMB for a comprehensive evaluation. Survival analyses revealed that the combination group with low stemness scores and high TMB values had better clinical outcomes (Fig. 7j). These results suggested that the stemness score is an effective biomarker for predicting both prognosis and clinical response to immunotherapy (Fig. 7k). In summary, our study provides concrete evidence that the stemness score is significantly correlated with the response to immunotherapy. Thus, consideration of stemness score and other clinical traits would be useful for predicting the therapeutic effects of anti-PD-1/L1 immunotherapy.

4. Discussion

Increasing evidence has demonstrated that GCSCs play an indispensable role in tumour resistance to chemoradiotherapy [15, 44-46]. While many studies have focused on the molecular features and mechanisms of cancer stem cells (CSCs), their clinical significance remains underexplored. Identifying the link between CSCs and clinical outcomes of patients with GC will enhance our understanding of the role of CSCs in the development of GC and guide more effective and precise treatment strategies.

Due to limitations in material availability and technology, few studies have addressed the clinical significance of tumor stem cells, especially in GC. Chen et al. [12] isolated CSCs from human GC tissues using surface markers, and validated their capacity to generate tumours both *in vivo* and *in vitro*. Owing to the limited sample size and clinical information, their study was restricted to CSC identification. The main advantage of our study is the combination of single-cell sequencing and bulk transcriptome data with detailed clinical information, which enabled us to identify CSC subtypes and assess their clinical relevance. Our study covered several critical

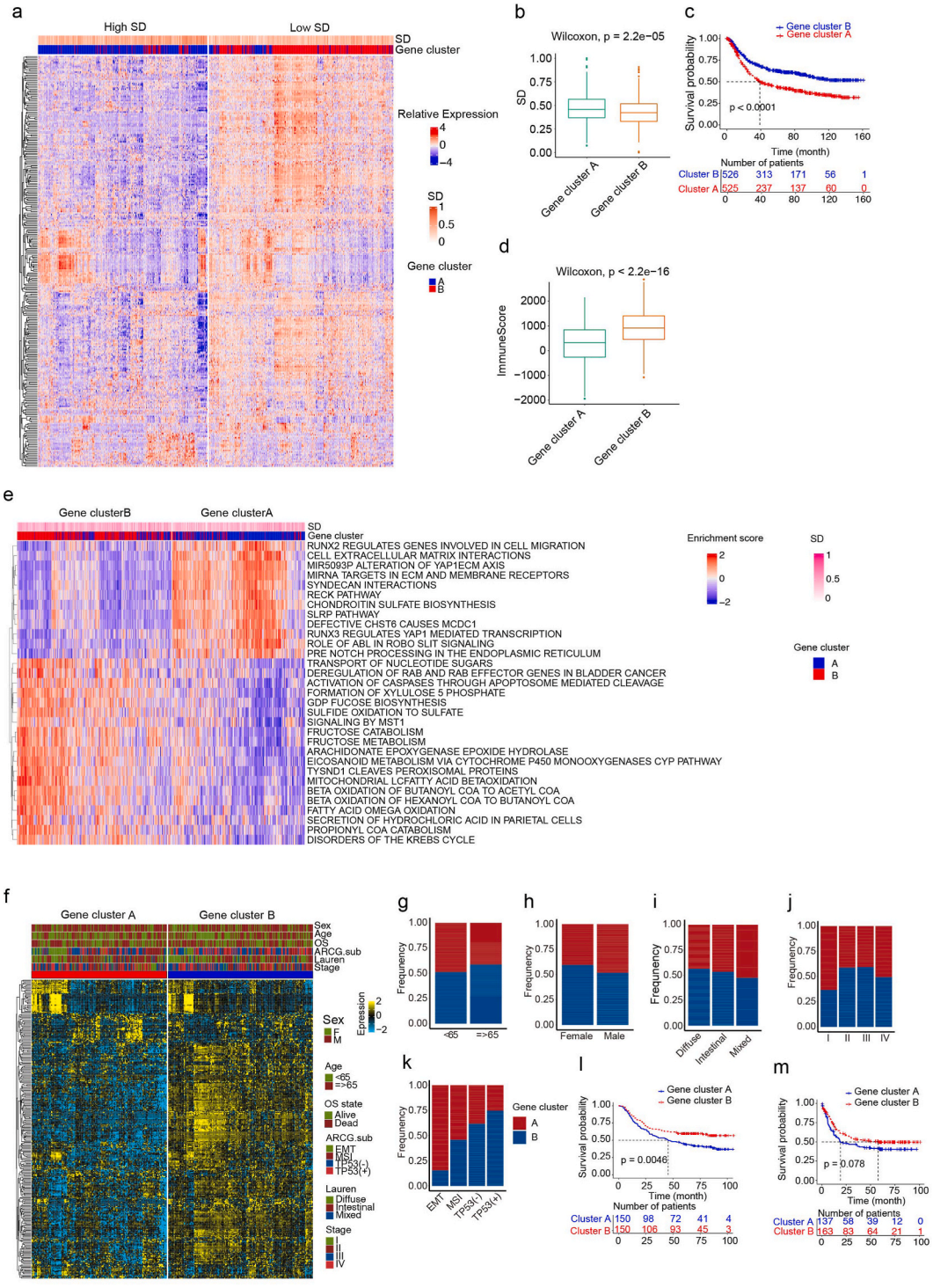


Fig. 4. The clinical significance of the two stemness subtypes. (a) Heatmap showing the signature genes expression between two gene clusters in 1051 GC samples. SD level also was mapped. (b) Boxplots showing the SD level of two gene clusters among 1051 GC patients ($p < 0.0001$). (c) The Kaplan–Meier curves analysis of gene cluster A and B among GC patients ($p < 0.0001$). (d) Boxplots showing the ImmuneScore level across distinct gene clusters in GC patients ($p < 0.0001$). (e) Heatmap illustrating the enrichment of 31 molecular pathways identified by GSVA analyses between gene cluster A and B. (f) Heatmap showing the signature genes expression between two gene clusters in ACRG dataset. Corresponding clinical and molecular characteristics also were mapped. (g–k) Comparing the frequencies of two gene clusters between different clinical parameters including age (g), sex (h), lauren classification (i), clinical stages (j), ACRG subgroups (k) in ACRG dataset. (l, m) The Kaplan–Meier curves of OS (l) and FPS (m) between two gene clusters in ACRG dataset.

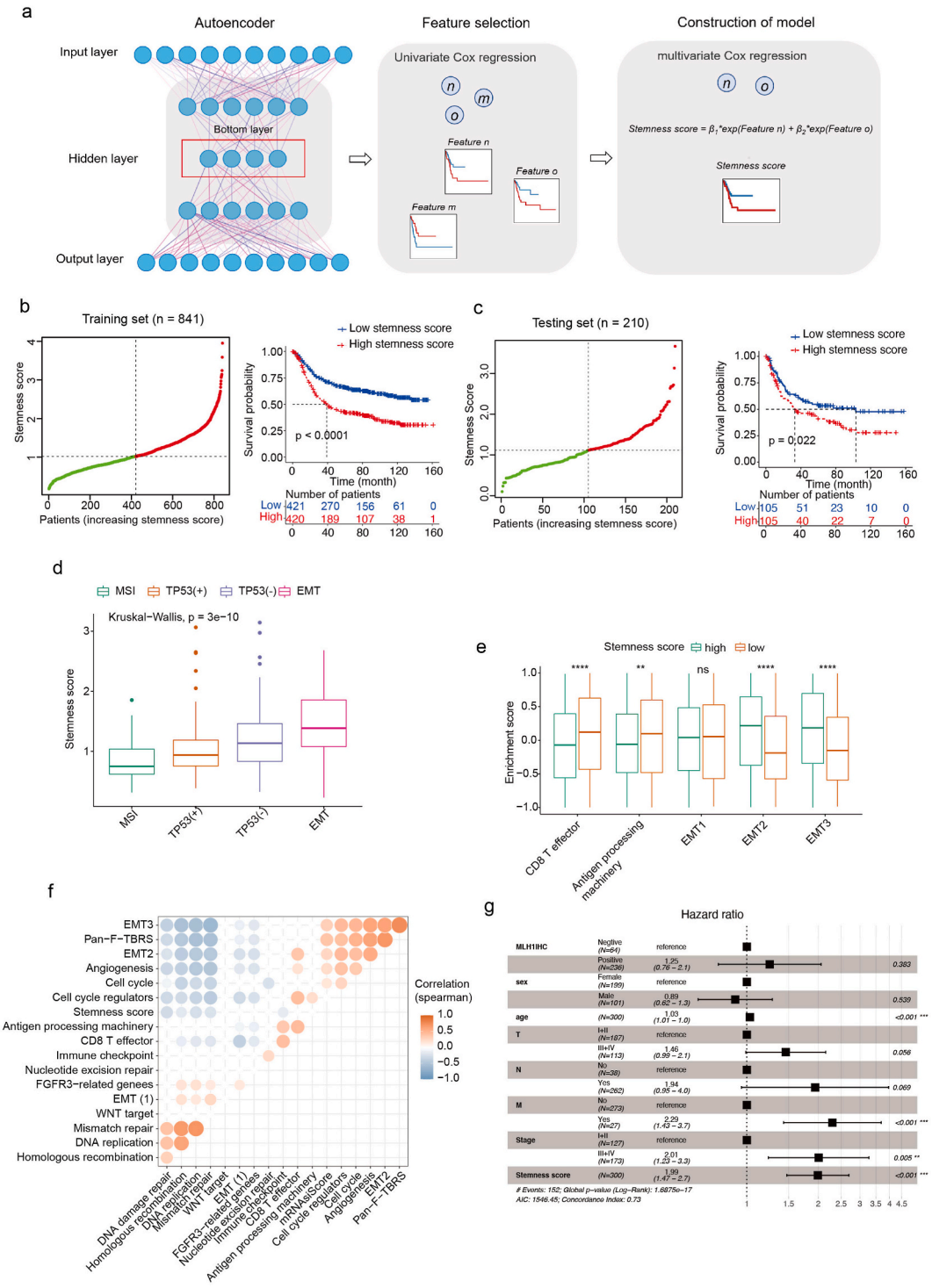
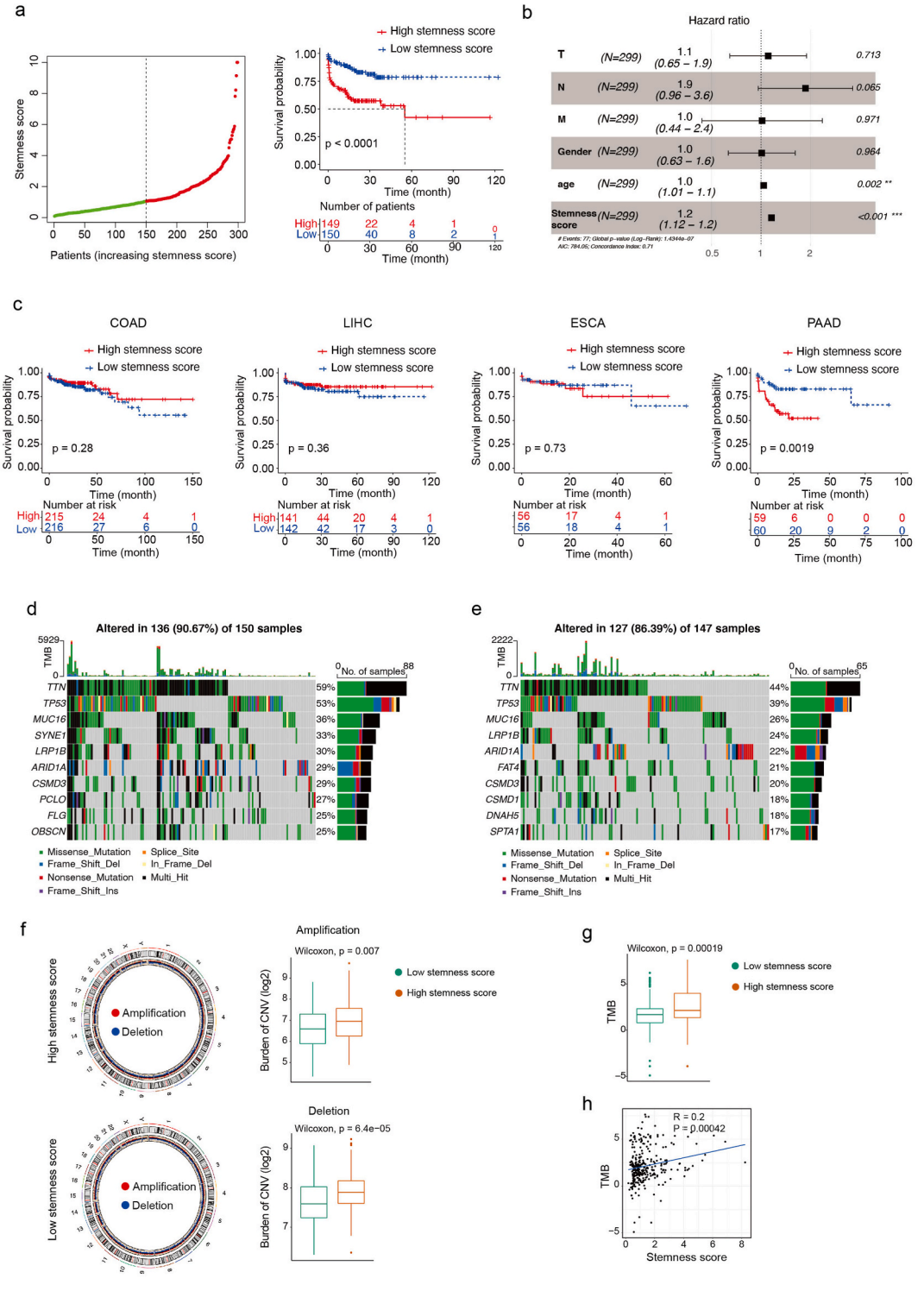


Fig. 5. Construction of the predictor derived from stemness signatures. (a) An overview of the workflow for constructing prognostic model. (b) Scatter plots showing stemness score in training dataset containing 841 GC patients (left panel). Kaplan-Meier curve analyses of low and high stemness score groups in in training dataset ($p < 0.0001$; right panel). (c) Scatter plot illustrating stemness score in testing dataset including 210 GC patients (left panel). Kaplan-Meier curve analyses of low and high stemness score groups in testing dataset ($p = 0.022$; right panel). (d) Differences in stemness score among distinct ACRG subtypes. (e) Boxplots showing stroma-activated pathways activity between high and low stemness score groups. * represents $p < 0.05$; ** represents $p < 0.01$; *** represents $p < 0.001$; ns, no significance. (f) Spearman analyses showing correlations between stemness score and the gene signatures derived from previous studies in ACRG dataset. (g) Forest plots show Multivariate Cox regression analyses for stemness score, age, gender, TNM stage, MHL1, and clinical stage in ACRG dataset.



(caption on next page)

Fig. 6. Validation of Stemness score in TCGA dataset. (a) An overview of stemness score in 299 GC patients from TCGA database (left) and survival analysis for GC patients stratified by stemness score in TCGA dataset (right). (b) Forest plots show Multivariate Cox regression analyses for stemness score, TNM status, age, and gender in TCGA cohort. (c) Survival analysis for tumor samples among distinct cancers including COAD, LIHC, ESCA, PAAD in TCGA dataset, which stratified by stemness score. (d, e) The waterfall plots showing top 10 gene mutation frequencies grouped by high (d) and low stemness score (e) in TCGA dataset. (f) Left panel: Circos plots showing the CNVs between high and low stemness score groups in TCGA dataset. Red denoting amplification of chromosomes, blue denoting deletion of chromosomes. Right panel: Boxplots showing burdens of CNVs in distinct stemness score groups. (g) Boxplots showing the level of tumor mutation burden (TMB) between GC patients classified by stemness score in TCGA. (h) Pearson correlation analysis for stemness score and TMB shown by scatter plots.

aspects, namely the identification of stem-like tumour cells, relevance of SD and clinical features, involvement of stem-like tumour cells in shaping TIME, classification based on SD, construction of a prognostic model based on SD, and the association between stemness score and immune therapy. Thus, we provide comprehensive insights into the role of CSCs in the occurrence and progression of tumours.

PCLAF is upregulated in various tumours and associated with patient prognosis [47]. It can promote tumor progression through multiple pathways, such as enhancing tumor proliferation [48], metastasis [49], and drug resistance [50]. Previous studies have shown that PCLAF maintains the stemness of breast cancer stem cells through the Wnt signaling pathway [51]. However, the correlation between PCLAF and tumor stem cells is still unclear. Here, based on single-cell sequencing data of GC from a public database, we identified $PCLAF^{+}$ stem-like tumour cells in tumour tissue. We introduced SD to integrate single-cell and bulk transcriptomes to evaluate the association between $PCLAF^{+}$ stem-like tumour cells and clinical parameters. We observed both associations and differences between the SD and clinical parameters. Importantly, a high SD was associated with worse clinical outcomes, suggesting that $PCLAF^{+}$ stem-like tumour cells may be a crucial factor in promoting the development of GC. Previous studies have shown that GCSC markers are positively correlated with the degree of tumour malignancy [52,53]. Moreover, MSI and the intestinal subtype showed a higher SD than other molecular subtypes. Therefore, we inferred that stem-like tumour cells contribute to the development of MSI or intestinal GC. However, the mechanism by which stem-like tumour cells are involved in the formation of these subtypes, as well as the reasons involved, require further investigation.

With an increasing understanding of the diversity and complexity of the TME, research has increasingly highlighted that it is closely related to tumour progression [54-56]. However, the crosstalk between CSCs and the TME remains unclear because most animal models used for CSC studies lack the major immune cell components. Recent studies have demonstrated that CSCs may regulate biological functions of immune cells [57-59]. In this study, two immune microenvironment phenotypes were identified in GC, including immunoreactive subtype (enriched immune cells) and immunodeficient subtype (shortage of immune cells). Moreover, the immunoreactive subtype is associated with better clinical outcomes. Recent studies have consistently revealed that the abundance of immune effector cells in tumours is positively correlated with patient prognosis [60,61]. Importantly, we observed that stem-like tumour cells may be involved in shaping the TME, as evidenced by that the high SD observed in the immunodeficient subtype. Therefore, our study further validated the crosstalk between CSCs and the TIME.

TNM stags remains the gold standard for guiding traditional treatment [62]. However, it is not applicable to emerging treatments such as molecular therapy and immunotherapy. Therefore, we developed a novel gastric cancer classifier based on tumor stemness, offering a theoretical foundation for patient classification in the context of immunotherapy. The two clusters exhibit different TME patterns. Cluster B was characterised by a lack of immune cells, consistent with the immune-deficient phenotype, whereas cluster A was characterised by the enrichment of immune cells, consistent with the immune-enriched phenotype. Thus, we concluded that the patients in cluster A were more likely to benefit from immunotherapy. Moreover, differences in mRNA transcriptomes between distinct GC classifications were significantly associated with the biological pathways involved in differentiation and immunity. We also compared the new classification with traditional clinical stratification and observed that gene cluster A was significantly associated with the EMT subtype, but not with other clinical features. Notably, patients in gene cluster A had a worse clinical prognosis for GC. We concluded that, in addition to the traditional clinical markers, transcriptomic subtypes based on SD should also be used as indicators to evaluate tumour progression in GC. Hence, treatment strategies developed by combining novel classifications with traditional clinical parameters may benefit patients with GC.

Given the association between stemness signatures and prognosis, there is an urgent need for a robust and reliable prognostic model to assess clinical outcomes. Therefore, we established a scoring model (stemness score) to assess the prognosis of patients with GC. As expected, high stemness scores indicated poor outcomes, whereas low stemness scores indicated better outcomes. In addition, we obtained consistent results from two additional datasets: IMvigor210 and TCGA cohorts [36]. These findings suggest that the stemness score is a robust prognostic biomarker for GC. In addition to prognosis, our study also compared the association between the stemness score and clinical parameters. High stemness scores demonstrated a strong link between the EMT subtype and the extent of genomic variation. Consistent with this observation, previous studies have demonstrated that GC with the EMT subtype is more malignant [63], highlighting that the stemness score also predicts certain clinical features. Interestingly, comparative analysis showed that *TTN* mutation rates were higher in the high stemness score group. Previous studies have shown that *TTN* is an important oncogene involved in the development of various cancers, including gastric cancer [64-66]. However, there is a lack of effective methods for predicting the mutation status of *TTN*, and our study suggests that the stemness score may predict it. Although our model has demonstrated robust performance in assessing patient prognosis and certain molecular characteristics, it also has some limitations. For example, in this study, we used 50 features for the dimensionality reduction of the autoencoder neural network and then constructed models based on these 50 features. However, there is currently no suitable method for selecting the optimal number of dimensions. In the future, using a more scientific method to select the optimal number of dimensions would further improve the performance of the model.

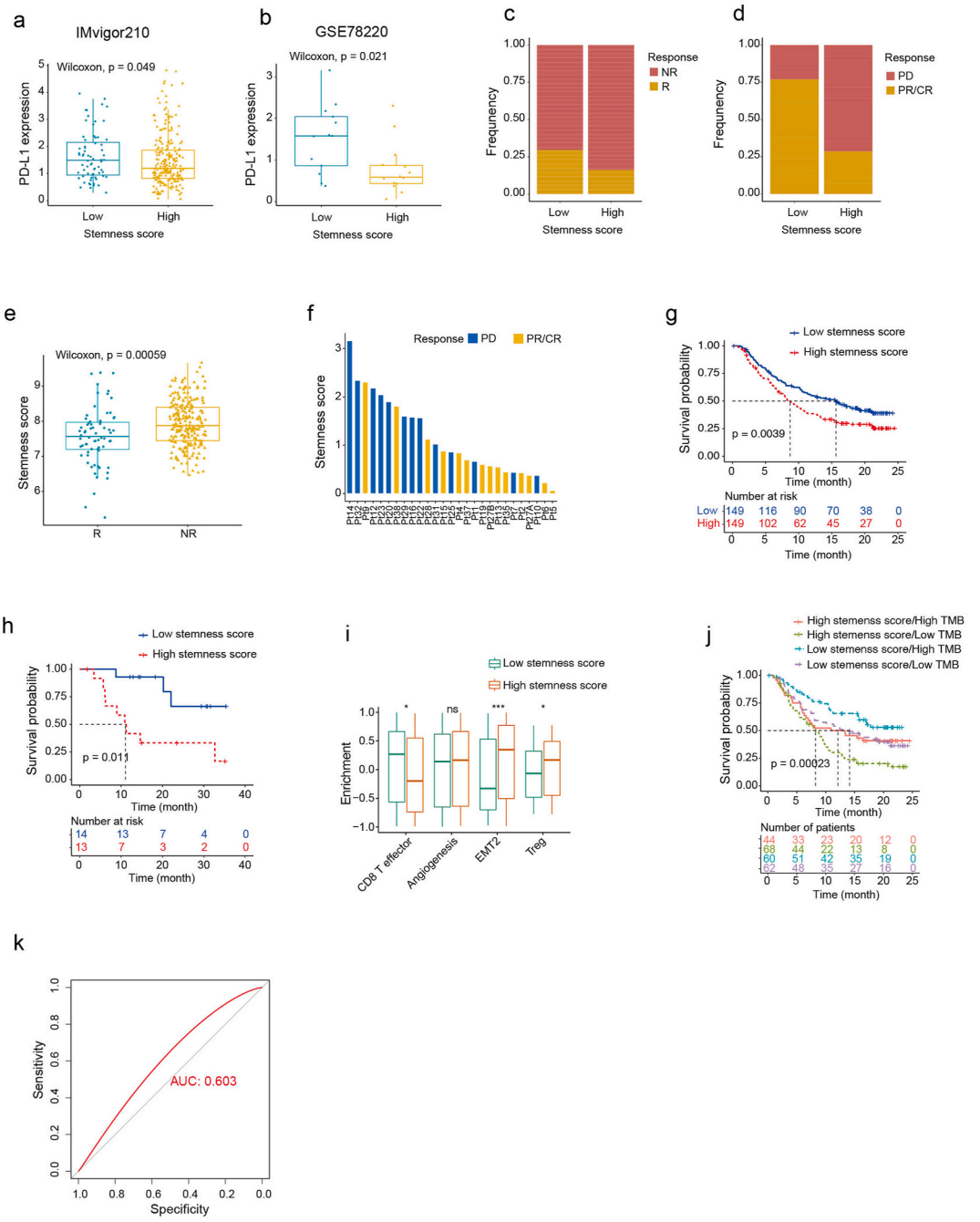


Fig. 7. The predictive values of stemness score in anti-PD-1/L1 immunotherapy. (a, b) mRNA expression of PD-L1 between low and high stemness score groups in IMvigor210 (a) and GSE78220 (b) datasets ($p < 0.05$). (c, d) The distribution of patients responded to anti-PD-L1 (c) or PD-1 (d) immunotherapy in low or high stemness score groups. CR representing complete response; PR representing partial response; PD representing progressive disease; NR representing no response; R representing response. (e) Boxplots showing stemness score between no response and response groups in IMvigor210 dataset ($p < 0.001$). (f) Bar plots showing stemness score among samples receiving anti-PD-1 immunotherapy in GSE78220 datasets. (g, h) Survival analyses of high and low stemness score groups in IMvigor210 (g) and GSE78220 (h) datasets ($P < 0.05$). (i) Boxplots showing the level of stemness score between different immune subtypes in IMvigor210 dataset. (j) Kaplan-Meier curve analysis for patients classified based on both stemness score and TMB in IMvigor210 dataset. ($p < 0.001$). (k) The AUC curve of the stemness score for predicting patients whether responding to anti-PD-1/L1 immunotherapy (AUC, 0.603).

Although anti-PD-1/L1immunotherapy benefits patients with GC, it is difficult to assess its efficacy in individual patients [67]. Therefore, it is important to evaluate the responses of patients to anti-PD-1/L1 immunotherapy. In this study, we observed that stemness score was a predictor of the efficacy of anti-PD-1/L1immunotherapy, as patients with low stemness scores were more responsive to anti-PD-1/L1immunotherapy and exhibited high PD-L1expression. Hence, the stemness score may have predictive capabilities for patient prognosis and the response to anti-PD-1/L1immunotherapy. This may also assist in determining whether immunotherapy should be administered.

5. Conclusions

In conclusion, our study uncovers significant associations between cancer stem cells (CSCs) and clinical traits in GC. CSCs are essential factors that shape the TME landscape of GC. A thorough assessment of individual tumor stemness scores will enhance our understanding of the molecular mechanisms underlying GC and inform more effective immunotherapeutic strategies.

Consent for publication

Not applicable.

Data availability statement

The GEO datasets (GSE15459, GSE34942, GSE57303, GSE62254, GSE78220, GSE84437) analyzed during this study are available in https://www.ncbi.nlm.nih.gov/gds/. The TCGA datasets (STAD, ESCA, COAD, LIHC) analyzed during this study are available in https://portal.gdc.cancer.gov/. The single-cell sequencing dataset (phs001818.v1.p1) could be download from https://www.ncbi.nlm.nih.gov/gap/.

Funding

This study was financially supported by grants from the National Natural Science Foundation of China (No.32170594 and No.31870809), the Province and Ministry Co-construction Major Program of Medical Science and Technique Foundation of Henan Province (No. SBGJ202001007).

CRediT authorship contribution statement

Yong Shi: Writing – review & editing, Writing – original draft, Methodology, Formal analysis, Data curation. **Ke An:** Writing – review & editing. **ShaoX zhou:** Resources. **XuR. Zhang:** Data curation. **QuanC. Kan:** Supervision. **Xin Tian:** Writing – review & editing, Supervision, Funding acquisition.

Declaration of competing interest

We declare that we have no known conflict of interest.

Appendix A. Supplementary data

Supplementary data to this article can be found online at https://doi.org/10.1016/j.heliyon.2024.e38662.

References

[1] 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 (2020), https://doi.org/10.3390/ijms21114012.
[2] S.S. Joshi, B.D. Badgwell, Current treatment and recent progress in gastric cancer, CA Cancer J Clin 71 (2021) 264–279, https://doi.org/10.3322/caac.21657.
[3] X. Jin, Z. Liu, D. Yang, K. Yin, X. Chang, Recent progress and future perspectives of immunotherapy in advanced gastric cancer, Front. Immunol. 13 (2022) 948647, https://doi.org/10.3389/fimmu.2022.948647.
[4] W.-L. Guan, Y. He, R.-H. Xu, Gastric cancer treatment: recent progress and future perspectives, J. Hematol. Oncol. 16 (2023) 57, https://doi.org/10.1186/s13045-023-01451-3.
[5] Y.Y. Janjigian, K. Shitara, M. Moehler, M. Garrido, P. Salman, L. Shen, L. Wyrwicz, K. Yamaguchi, T. Skoczylas, A. Campos Bragagnoli, T. Liu, M. Schenker, P. Yanez, M. Tehfe, R. Kowalyszyn, M.V. Karamouzis, R. Bruges, T. Zander, R. Pazo-Cid, E. Hitre, K. Feeney, J.M. Cleary, V. Poulart, D. Cullen, M. Lei, H. Xiao, K. Kondo, M. Li, J.A. Ajani, First-line nivolumab plus chemotherapy versus chemotherapy alone for advanced gastric, gastro-oesophageal junction, and oesophageal adenocarcinoma (CheckMate 649): a randomised, open-label, phase 3 trial, Lancet 398 (2021) 27–40, https://doi.org/10.1016/S0140-6736(21)00797-2.
[6] M.F. Clarke, J.E. Dick, P.B. Dirks, C.J. Eaves, C.H.M. Jamieson, D.L. Jones, J. Visvader, I.L. Weissman, G.M. Wahl, Cancer stem cells–perspectives on current status and future directions: AACR Workshop on cancer stem cells, Cancer Res. 66 (2006) 9339–9344.
[7] L.V. Nguyen, R. Vanner, P. Dirks, C.J. Eaves, Cancer stem cells: an evolving concept, Nat. Rev. Cancer 12 (2012) 133–143, https://doi.org/10.1038/nrc3184.

[8] T. Shibue, R.A. Weinberg, EMT, CSCs, and drug resistance: the mechanistic link and clinical implications, Nat. Rev. Clin. Oncol. 14 (2017) 611–629, https://doi.org/10.1038/nrclinonc.2017.44.

[9] Y. Ge, N.C. Gomez, R.C. Adam, M. Nikolova, H. Yang, A. Verma, C.P.-J. Lu, L. Polak, S. Yuan, O. Elemento, E. Fuchs, Stem cell lineage infidelity drives wound repair and cancer, Cell 169 (2017), https://doi.org/10.1016/j.cell.2017.03.042.

[10] A. Kreso, J.E. Dick, Evolution of the cancer stem cell model, Cell Stem Cell 14 (2014) 275–291, https://doi.org/10.1016/j.stem.2014.02.006.

[11] L. Vermeulen, M. Todaro, F. de Sousa Mello, M.R. Sprick, K. Kemper, M. Perez Alea, D.J. Richel, G. Stassi, J.P. Medema, Single-cell cloning of colon cancer stem cells reveals a multi-lineage differentiation capacity, Proc Natl Acad Sci U S A 105 (2008) 13427–13432, https://doi.org/10.1073/pnas.0805706105.

[12] T. Chen, K. Yang, J. Yu, W. Meng, D. Yuan, F. Bi, F. Liu, J. Liu, B. Dai, X. Chen, F. Wang, F. Zeng, H. Xu, J. Hu, X. Mo, Identification and expansion of cancer stem cells in tumor tissues and peripheral blood derived from gastric adenocarcinoma patients, Cell Res. 22 (2012) 248–258, https://doi.org/10.1038/cr.2011.109.

[13] R. Li, X. Wu, H. Wei, S. Tian, Characterization of side population cells isolated from the gastric cancer cell line SGC-7901, Oncol. Lett. 5 (2013) 877–883.

[14] H. Taniguchi, C. Moriya, H. Igarashi, A. Saitoh, H. Yamamoto, Y. Adachi, K. Imai, Cancer stem cells in human gastrointestinal cancer, Cancer Sci. 107 (2016) 1556–1562, https://doi.org/10.1111/cas.13069.

[15] N. Barker, M. Huch, P. Kujala, M. van de Wetering, H.J. Snippert, J.H. van Es, T. Sato, D.E. Stange, H. Begthel, M. van den Born, E. Danenberg, S. van den Brink, J. Korving, A. Abo, P.J. Peters, N. Wright, R. Poulsom, H. Clevers, Lgr5(+ve) stem cells drive self-renewal in the stomach and build long-lived gastric units in vitro, Cell Stem Cell 6 (2010) 25–36, https://doi.org/10.1016/j.stem.2009.11.013.

[16] Y. Chen, Z. Fu, S. Xu, Y. Xu, P. Xu, The prognostic value of CD44 expression in gastric cancer: a meta-analysis, Biomed. Pharmacother. 68 (2014) 693–697, https://doi.org/10.1016/j.biopha.2014.08.001.

[17] J.-X. Wu, Y.-Y. Zhao, X. Wu, H.-X. An, Clinicopathological and prognostic significance of CD24 overexpression in patients with gastric cancer: a meta-analysis, PLoS One 9 (2014) e114746, https://doi.org/10.1371/journal.pone.0114746.

[18] W. Chen, X. Zhang, C. Chu, W.L. Cheung, L. Ng, S. Lam, A. Chow, T. Lau, M. Chen, Y. Li, Y. Nie, B.C. Wong, R. Pang, Identification of CD44+ cancer stem cells in human gastric cancer, Hepato-Gastroenterology 60 (2013) 949–954, https://doi.org/10.5754/hge12881.

[19] S. Wang, Y. Zheng, J. Li, Y. Yu, W. Zhang, M. Song, Z. Liu, Z. Min, H. Hu, Y. Jing, X. He, L. Sun, L. Ma, C.R. Esteban, P. Chan, J. Qiao, Q. Zhou, J.C. Izpisua Belmonte, J. Qu, F. Tang, G.-H. Liu, Single-cell transcriptomic atlas of primate ovarian aging, Cell 180 (2020), https://doi.org/10.1016/j.cell.2020.01.009.

[20] J.-J. Wang, W. Ge, Q.-Y. Zhai, J.-C. Liu, X.-W. Sun, W.-X. Liu, L. Li, C.-Z. Lei, P.W. Dyce, M. De Felici, W. Shen, Single-cell transcriptome landscape of ovarian cells during primordial follicle assembly in mice, PLoS Biol. 18 (2020) e3001025, https://doi.org/10.1371/journal.pbio.3001025.

[21] T. Baslan, J. Hicks, Unravelling biology and shifting paradigms in cancer with single-cell sequencing, Nat. Rev. Cancer 17 (2017) 557–569, https://doi.org/10.1038/nrc.2017.58.

[22] F. Wu, J. Fan, Y. He, A. Xiong, J. Yu, Y. Li, Y. Zhang, W. Zhao, F. Zhou, W. Li, J. Zhang, X. Zhang, M. Qiao, G. Gao, S. Chen, X. Chen, X. Li, L. Hou, C. Wu, C. Su, S. Ren, M. Odenthal, R. Buettner, N. Fang, C. Zhou, Single-cell profiling of tumor heterogeneity and the microenvironment in advanced non-small cell lung cancer, Nat. Commun. 12 (2021) 2540, https://doi.org/10.1038/s41467-021-22801-0.

[23] V. Kumar, K. Ramnarayanan, R. Sundar, N. Padmanabhan, S. Srivastava, M. Koiwa, T. Yasuda, V. Koh, K.K. Huang, S.T. Tay, S.W.T. Ho, A.L.K. Tan, T. Ishimoto, G. Kim, A. Shabbir, Q. Chen, B. Zhang, S. Xu, K.-P. Lam, H.Y.J. Lum, M. Teh, W.P. Yong, J.B.Y. So, P. Tan, Single-cell atlas of lineage states, tumor microenvironment, and subtype-specific expression programs in gastric cancer, Cancer Discov. 12 (2022) 670–691, https://doi.org/10.1158/2159-8290.CD-21-0683.

[24] A. Colaprico, T.C. Silva, C. Olsen, L. Garofano, C. Cava, D. Garolini, T.S. Sabedot, T.M. Malta, S.M. Pagnotta, I. Castiglioni, M. Ceccarelli, G. Bontempi, H. Noushmehr, TCGAbiolinks: an R/Bioconductor package for integrative analysis of TCGA data, Nucleic Acids Res. 44 (2016) e71, https://doi.org/10.1093/nar/gkv1507.

[25] J.T. Leek, W.E. Johnson, H.S. Parker, A.E. Jaffe, J.D. Storey, The sva package for removing batch effects and other unwanted variation in high-throughput experiments, Bioinformatics 28 (2012) 882–883, https://doi.org/10.1093/bioinformatics/bts034.

[26] A. Butler, P. Hoffman, P. Smibert, E. Papalexi, R. Satija, Integrating single-cell transcriptomic data across different conditions, technologies, and species, Nat. Biotechnol. 36 (2018) 411–420, https://doi.org/10.1038/nbt.4096.

[27] X. Qiu, Q. Mao, Y. Tang, L. Wang, R. Chawla, H.A. Pliner, C. Trapnell, Reversed graph embedding resolves complex single-cell trajectories, Nat. Methods 14 (2017) 979–982, https://doi.org/10.1038/nmeth.4402.

[28] T.M. Malta, A. Sokolov, A.J. Gentles, T. Burzykowski, L. Poisson, J.N. Weinstein, B. Kamińska, J. Huelsken, L. Omberg, O. Gevaert, A. Colaprico, P. Czerwińska, S. Mazurek, L. Mishra, H. Heyn, A. Krasnitz, A.K. Godwin, A.J. Lazar, J.M. Stuart, K.A. Hoadley, P.W. Laird, H. Noushmehr, M. Wiznerowicz, Machine learning identifies stemness features associated with oncogenic dedifferentiation, Cell 173 (2018), https://doi.org/10.1016/j.cell.2018.03.034.

[29] A. Sokolov, D.E. Carlin, E.O. Paull, R. Baertsch, J.M. Stuart, Pathway-based genomics prediction using generalized elastic net, PLoS Comput. Biol. 12 (2016) e1004790, https://doi.org/10.1371/journal.pcbi.1004790.

[30] S. Hänzelmann, R. Castelo, J. Guinney, GSVA: gene set variation analysis for microarray and RNA-seq data, BMC Bioinf. 14 (2013) 7, https://doi.org/10.1186/1471-2105-14-7.

[31] A.M. Newman, C.L. Liu, M.R. Green, A.J. Gentles, W. Feng, Y. Xu, C.D. Hoang, M. Diehn, A.A. Alizadeh, Robust enumeration of cell subsets from tissue expression profiles, Nat. Methods 12 (2015) 453–457, https://doi.org/10.1038/nmeth.3337.

[32] P. Charoentong, F. Finotello, M. Angelova, C. Mayer, M. Efremova, D. Rieder, H. Hackl, Z. Trajanoski, Pan-cancer immunogenomic analyses reveal genotype-immunophenotype relationships and predictors of response to checkpoint blockade, Cell Rep. 18 (2017) 248–262, https://doi.org/10.1016/j.celrep.2016.12.019.

[33] K. Yoshihara, M. Shahmoradgoli, E. Martínez, R. Vegesna, H. Kim, W. Torres-Garcia, V. Treviño, H. Shen, P.W. Laird, D.A. Levine, S.L. Carter, G. Getz, K. Stemke-Hale, G.B. Mills, R.G.W. Verhaak, Inferring tumour purity and stromal and immune cell admixture from expression data, Nat. Commun. 4 (2013) 2612, https://doi.org/10.1038/ncomms3612.

[34] M.D. Wilkerson, D.N. Hayes, ConsensusClusterPlus: a class discovery tool with confidence assessments and item tracking, Bioinformatics 26 (2010) 1572–1573, https://doi.org/10.1093/bioinformatics/btq170.

[35] C. Zhang, Y. Geng, Z. Han, Y. Liu, H. Fu, Q. Hu, Autoencoder in autoencoder networks, IEEE Trans Neural Netw Learn Syst (2022), https://doi.org/10.1109/tnnls.2022.3189239.

[36] S. Mariathasan, S.J. Turley, D. Nickles, A. Castiglioni, K. Yuen, Y. Wang, E.E. Kadel, H. Koeppen, J.L. Astarita, R. Cubas, S. Jhunjhunwala, R. Banchereau, Y. Yang, Y. Guan, C. Chalouni, J. Ziai, Y. Şenbabaoğlu, S. Santoro, D. Sheinson, J. Hung, J.M. Giltnane, A.A. Pierce, K. Mesh, S. Lianoglou, J. Riegler, R.A. D. Carano, P. Eriksson, M. Höglund, L. Somarriba, D.L. Halligan, M.S. van der Heijden, Y. Loriot, J.E. Rosenberg, L. Fong, I. Mellman, D.S. Chen, M. Green, C. Derleth, G.D. Fine, P.S. Hegde, R. Bourgon, T. Powles, TGFβ attenuates tumour response to PD-L1 blockade by contributing to exclusion of T cells, Nature 554 (2018) 544–548, https://doi.org/10.1038/nature25501.

[37] L. Guo, H. Li, W. Li, J. Tang, Construction and investigation of a combined hypoxia and stemness index lncRNA-associated ceRNA regulatory network in lung adenocarcinoma, BMC Med Genomics 13 (2020) 166, https://doi.org/10.1186/s12920-020-00816-8.

[38] C. Zhang, Y.N. Fondufe-Mittendorf, C. Wang, J. Chen, Q. Cheng, D. Zhou, Y. Zheng, H. Geiger, Y. Liang, Latexin regulation by HMGB2 is required for hematopoietic stem cell maintenance, Haematologica 105 (2020) 573–584, https://doi.org/10.3324/haematol.2018.207092.

[39] X. Zhao, J. Hu, Y. Li, M. Guo, Volumetric compression develops noise-driven single-cell heterogeneity, Proc Natl Acad Sci U S A 118 (2021), https://doi.org/10.1073/pnas.2110550118.

[40] S. Parte, I. Virant-Klun, M. Patankar, S.K. Batra, A. Straughn, S.S. Kakar, PTTG1: a unique regulator of stem/cancer stem cells in the ovary and ovarian cancer, Stem Cell Rev Rep 15 (2019) 866–879, https://doi.org/10.1007/s12015-019-09911-5.

[41] Y. Hu, H. Liu, X. Xiao, Q. Yu, R. Deng, L. Hua, J. Wang, X. Wang, Bone marrow mesenchymal stem cell-derived exosomes inhibit triple-negative breast cancer cell stemness and metastasis via an ALKBH5-dependent mechanism, Cancers 14 (2022), https://doi.org/10.3390/cancers14246059.

[42] P. Zhang, M. Yang, Y. Zhang, S. Xiao, X. Lai, A. Tan, S. Du, S. Li, Dissecting the single-cell transcriptome network underlying gastric premalignant lesions and early gastric cancer, Cell Rep. 27 (2019), https://doi.org/10.1016/j.celrep.2019.04.052.

[43] R. Zhao, B. He, Q. Bie, J. Cao, H. Lu, Z. Zhang, J. Liang, L. Wei, H. Xiong, B. Zhang, AQP5 complements LGR5 to determine the fates of gastric cancer stem cells through regulating ULK1 ubiquitination, J. Exp. Clin. Cancer Res. 41 (2022) 322, https://doi.org/10.1186/s13046-022-02532-w.
[44] L. Sun, C. Huang, M. Zhu, S. Guo, Q. Gao, Q. Wang, B. Chen, R. Li, Y. Zhao, M. Wang, Z. Chen, B. Shen, W. Zhu, Gastric cancer mesenchymal stem cells regulate PD-L1-CTCF enhancing cancer stem cell-like properties and tumorigenesis, Theranostics 10 (2020) 11950–11962, https://doi.org/10.7150/thno.49717.
[45] K.W. McCracken, E.M. Catá, C.M. Crawford, K.L. Sinagoga, M. Schumacher, B.E. Rockich, Y.-H. Tsai, C.N. Mayhew, J.R. Spence, Y. Zavros, J.M. Wells, Modelling human development and disease in pluripotent stem-cell-derived gastric organoids, Nature 516 (2014) 400–404, https://doi.org/10.1038/nature13863.
[46] S. Han, J. Fink, D.J. Jörg, E. Lee, M.K. Yum, L. Chatzeli, S.R. Merker, M. Josserand, T. Trendafilova, A. Andersson-Rolf, C. Dabrowska, H. Kim, R. Naumann, J.-H. Lee, N. Sasaki, R.L. Mort, O. Basak, H. Clevers, D.E. Stange, A. Philpott, J.K. Kim, B.D. Simons, B.-K. Koo, Defining the identity and dynamics of adult gastric isthmus stem cells, Cell Stem Cell 25 (2019), https://doi.org/10.1016/j.stem.2019.07.008.
[47] X. Liu, C. Cheng, Y. Cai, Y. Gu, Y. Wu, K. Chen, Z. Wu, Pan-cancer analyses reveal the regulation and clinical outcome association of PCLAF in human tumors, Int. J. Oncol. 60 (2022), https://doi.org/10.3892/ijo.2022.5356.
[48] X. Liu, Y. Cai, C. Cheng, Y. Gu, X. Hu, K. Chen, Y. Wu, Z. Wu, PCLAF promotes neuroblastoma G1/S cell cycle progression via the E2F1/PTTG1 axis, Cell Death Dis. 13 (2022) 178, https://doi.org/10.1038/s41419-022-04635-w.
[49] F. Ma, C. Zhi, M. Wang, T. Li, S.A. Khan, Z. Ma, Z. Jing, C. Bo, Q. Zhou, S. Xia, S. Huang, S. Huang, Z. Zhang, H. Jia, X. Cui, M. Yao, T. Ji, Dysregulated NF-κB signal promotes the hub gene PCLAF expression to facilitate nasopharyngeal carcinoma proliferation and metastasis, Biomed. Pharmacother. 125 (2020) 109905, https://doi.org/10.1016/j.biopha.2020.109905.
[50] L. Liu, Z. Lv, M. Wang, D. Zhang, D. Liu, F. Zhu, HBV enhances sorafenib resistance in hepatocellular carcinoma by reducing ferroptosis via SRSF2-mediated abnormal PCLAF splicing, Int. J. Mol. Sci. 24 (2023), https://doi.org/10.3390/ijms24043263.
[51] X. Wang, Y.-S. Jung, S. Jun, S. Lee, W. Wang, A. Schneider, Y. Sun Oh, S.H. Lin, B.-J. Park, J. Chen, K. Keyomarsi, J.-I. Park, PAF-Wnt signaling-induced cell plasticity is required for maintenance of breast cancer cell stemness, Nat. Commun. 7 (2016) 10633, https://doi.org/10.1038/ncomms10633.
[52] T. Wang, C.W. Ong, J. Shi, S. Srivastava, B. Yan, C.L. Cheng, W.P. Yong, S.L. Chan, K.G. Yeoh, B. Iacopetta, M. Salto-Tellez, Sequential expression of putative stem cell markers in gastric carcinogenesis, Br. J. Cancer 105 (2011) 658–665, https://doi.org/10.1038/bjc.2011.287.
[53] S. Chen, J.-H. Hou, X.-Y. Feng, X.-S. Zhang, Z.-W. Zhou, J.-P. Yun, Y.-B. Chen, M.-Y. Cai, Clinicopathologic significance of putative stem cell marker, CD44 and CD133, in human gastric carcinoma, J. Surg. Oncol. 107 (2013) 799–806, https://doi.org/10.1002/jso.23337.
[54] D.S. Chen, I. Mellman, Elements of cancer immunity and the cancer-immune set point, Nature 541 (2017) 321–330, https://doi.org/10.1038/nature21349.
[55] S.J. Turley, V. Cremasco, J.L. Astarita, Immunological hallmarks of stromal cells in the tumour microenvironment, Nat. Rev. Immunol. 15 (2015) 669–682, https://doi.org/10.1038/nri3902.
[56] T.F. Gajewski, S.-R. Woo, Y. Zha, R. Spaapen, Y. Zheng, L. Corrales, S. Spranger, Cancer immunotherapy strategies based on overcoming barriers within the tumor microenvironment, Curr. Opin. Immunol. 25 (2013) 268–276, https://doi.org/10.1016/j.coi.2013.02.009.
[57] T. Di Tomaso, S. Mazzoleni, E. Wang, G. Sovena, D. Clavenna, A. Franzin, P. Mortini, S. Ferrone, C. Doglioni, F.M. Marincola, R. Galli, G. Parmiani, C. Maccalli, Immunobiological characterization of cancer stem cells isolated from glioblastoma patients, Clin. Cancer Res. 16 (2010) 800–813, https://doi.org/10.1158/1078-0432.CCR-09-2730.
[58] A. Wu, J. Wei, L.-Y. Kong, Y. Wang, W. Priebe, W. Qiao, R. Sawaya, A.B. Heimberger, Glioma cancer stem cells induce immunosuppressive macrophages/microglia, Neuro Oncol. 12 (2010) 1113–1125, https://doi.org/10.1093/neuonc/noq082.
[59] S. Pellegatta, P.L. Poliani, D. Corno, F. Menghi, F. Ghielmetti, B. Suarez-Merino, V. Caldera, S. Nava, M. Ravanini, F. Facchetti, M.G. Bruzzone, G. Finocchiaro, Neurospheres enriched in cancer stem-like cells are highly effective in eliciting a dendritic cell-mediated immune response against malignant gliomas, Cancer Res. 66 (2006) 10247–10252.
[60] P. Ge, W. Wang, L. Li, G. Zhang, Z. Gao, Z. Tang, X. Dang, Y. Wu, Profiles of immune cell infiltration and immune-related genes in the tumor microenvironment of colorectal cancer, Biomed. Pharmacother. 118 (2019) 109228, https://doi.org/10.1016/j.biopha.2019.109228.
[61] X. Liu, S. Wu, Y. Yang, M. Zhao, G. Zhu, Z. Hou, The prognostic landscape of tumor-infiltrating immune cell and immunomodulators in lung cancer, Biomed. Pharmacother. 95 (2017) 55–61, https://doi.org/10.1016/j.biopha.2017.08.003.
[62] N.Y. Chia, P. Tan, Molecular classification of gastric cancer, Ann. Oncol. 27 (2016) 763–769, https://doi.org/10.1093/annonc/mdw040.
[63] R. Cristescu, J. Lee, M. Nebozhyn, K.-M. Kim, J.C. Ting, S.S. Wong, J. Liu, Y.G. Yue, J. Wang, K. Yu, X.S. Ye, I.-G. Do, S. Liu, L. Gong, J. Fu, J.G. Jin, M.G. Choi, T. S. Sohn, J.H. Lee, J.M. Bae, S.T. Kim, S.H. Park, I. Sohn, S.-H. Jung, P. Tan, R. Chen, J. Hardwick, W.K. Kang, M. Ayers, D. Hongyue, C. Reinhard, A. Loboda, S. Kim, A. Aggarwal, Molecular analysis of gastric cancer identifies subtypes associated with distinct clinical outcomes, Nat Med 21 (2015) 449–456, https://doi.org/10.1038/nm.3850.
[64] X. Han, J. Chen, J. Wang, J. Xu, Y. Liu, TTN mutations predict a poor prognosis in patients with thyroid cancer, Biosci. Rep. 42 (2022), https://doi.org/10.1042/BSR20221168.
[65] Y. Wang, R. Jiang, H. Zhao, F. Li, Y. Li, M. Zhu, TTN-AS1 delivered by gastric cancer cell-derived exosome induces gastric cancer progression through in vivo and in vitro studies, Cell Biol. Toxicol. 39 (2023) 557–571, https://doi.org/10.1007/s10565-022-09762-w.
[66] Q.-X. Zheng, J. Wang, X.-Y. Gu, C.-H. Huang, C. Chen, M. Hong, Z. Chen, TTN-AS1 as a potential diagnostic and prognostic biomarker for multiple cancers, Biomed. Pharmacother. 135 (2021) 111169, https://doi.org/10.1016/j.biopha.2020.111169.
[67] S.T. Kim, R. Cristescu, A.J. Bass, K.-M. Kim, J.I. Odegaard, K. Kim, X.Q. Liu, X. Sher, H. Jung, M. Lee, S. Lee, S.H. Park, J.O. Park, Y.S. Park, H.Y. Lim, H. Lee, M. Choi, A. Talasaz, P.S. Kang, J. Cheng, A. Loboda, J. Lee, W.K. Kang, Comprehensive molecular characterization of clinical responses to PD-1 inhibition in metastatic gastric cancer, Nat Med 24 (2018) 1449–1458, https://doi.org/10.1038/s41591-018-0101-z.