ARTICLE

Cellular and Molecular Biology

A novel high-risk subpopulation identified by CTSL and ZBTB7B in gastric cancer

Kaisa Cui [ID][1,2], Surui Yao[1], Bingxin Liu[1,2], Shengbai Sun[1,2], Liang Gong[2,3], Qilin Li[4], Bojian Fei[5 ✉] and Zhaohui Huang [ID][1,2 ✉]

BACKGROUND: Gastric cancer (GC) is characterised by a heterogeneous tumour microenvironment (TME) that is closely associated with the response to treatment, especially immunotherapies. However, most previous GC molecular subtyping systems need complex gene signatures and examination methods, restricting their clinical applications. Thus, we developed a new TME-based molecular subtype using only two genes.
METHODS: Nine independent GC cohorts at the tissue- or single-cell level with more than 2000 patients were used in this study, including data we examined by single-cell sequencing, quantitative RT-PCR and immunochemistry/immunofluorescence staining. Nine different methods, five existing molecular subtypes and a series of signatures were used to evaluate the TME and molecular characteristics of GC.
RESULTS: We established a CTSL/ZBTB7B subtyping system and uncovered the novel CTSL^High^ZBTB7B^Low^ high-risk subgroup, but characterised by relative higher immune cell infiltration and lower tumour purity. This subgroup demonstrate higher levels of immune checkpoints and more enrichment of cancer-related pathways compared with other cases.
CONCLUSIONS: We identified a high-risk subpopulation with unique TME features based on expressions of CTSL and ZBTB7B, suggesting a counterbalancing phenotype between immunostimulatory and immunosuppressive mechanisms. This subtyping system could be used to select treatment and management strategies for GC.

BACKGROUND

Gastric cancer (GC) is one of the most important cancer types worldwide with high morbidity and mortality, especially in Eastern Asian countries, including China [1]. High heterogeneity in the GC tumour microenvironment (TME) obstruct the development of efficient therapeutic approaches, leading to terrible survival rates [2]. The highly complex and heterogeneous TME contains not only cancer cells, but also a variety of immune cells that regulate tumour growth and metastasis [3]. The TME is also a key factor affecting the efficiency of cancer immunotherapies, especially immune checkpoint (IC) inhibitors, because ICs and their ligands frequently exist in the TME and regulate anti-tumour immune responses [4]. Prior studies on the Immunoscore demonstrated that high immune cell infiltrations predicts favourable survival in gastrointestinal cancer [5, 6]. However, a recent study identified a unique subtype of high intratumoural CD8 T cell infiltration and a high density of tumour-associated macrophages that expressed CD274 in colorectal cancer [7]. Moreover, we also revealed a unique subpopulation of regulatory T cell with intense CTLA4 expression. This subgroup showed high-risk of recurrence and death with an immune overdrive TME phenotype, including high

levels of immune infiltration, IC expression, MSI/dMMR status and TGF-β signalling in colorectal cancer [8]. These tumour tissues pre-existing in the TME have an abundance of immune responses and IC activation, which is necessary for IC inhibitor-based immunotherapy. Hence, the identification of this novel phenotype is urgently needed in other cancer types for personalised cancer immunotherapies, including GC.

TME-based immune molecular subtyping has been developed for cancer precision therapy in several cancer types, including GC [9, 10]. However, the signatures of these molecular subtyping systems usually contain many marker genes and require complex examination methods. Furthermore, novel technologies used in these systems, such as single-cell or in situ barcode sequencing, are expensive and leave much to be desired for large-scale diagnostic use [11]. In this scenario, it is necessary to develop new molecular subtyping systems with only a few markers that can be detected quickly and economically. Intriguingly, recent studies identified a series of subpopulations in CRC, bladder cancer and glioma by only single or double marker genes, such as CD8A/CD274, Siglec15, FOXP3/CTLA4, VTCN1/CD274 and CD8A/IDO1 stratifications [7, 8, 12–14]. However, as one of most

¹Wuxi Cancer Institute, Affiliated Hospital of Jiangnan University, 214062 Wuxi, Jiangsu, China. ²Laboratory of Cancer Epigenetics, Wuxi School of Medicine, Jiangnan University, 214122 Wuxi, Jiangsu, China. ³Key Laboratory of Carbohydrate Chemistry & Biotechnology, Ministry of Education, School of Biotechnology, Jiangnan University, 214122 Wuxi, Jiangsu, China. ⁴Computer Vision Lab, Department of Electrical Engineering, California Institute of Technology, Pasadena, CA 91125, USA. ⁵Department of Surgical Oncology, Affiliated Hospital of Jiangnan University, 214122 Wuxi, Jiangsu, China. ✉email: wx4yfbj@163.com; zhaohuihuang@jiangnan.edu.cn

prevalent types of gastrointestinal cancer, GC still lacks a TME molecular subtyping system identified by few marker genes.

In this study, we identified two genes CTSL and ZBTB7B, to classify GC patients based on multiple independent cohorts, including public datasets and our validation cohort/sample). We revealed the unique clinical, TME and molecular characteristics of tumours identified by CTSL and ZBTB7B in data of clinical GC tissues and single-cell GC data. Moreover, these subtypes were validated by a series of wet-lab experiments and a Pan-cancer analysis.

METHODS
Data and resources
The Cancer Genome Atlas (TCGA) RNA-Seq data of 20 cancer types and clinical data were downloaded from the Genomic Data Commons Data Portal (https://portal.gdc.cancer.gov/). Primary tumour samples were used in this research. Overall survival (OS) data of each TCGA cancer cohort was obtained from the integrated TCGA Pan-cancer clinical data resource [15]. Abbreviations and patient numbers of each TCGA cancer type were included in Supplementary Table 1. Seven independent GC datasets GSE66229, GSE26253, GSE26942, GSE84437, GSE15459, GSE183904 and GSE163558 were available on Gene Expression Omnibus (GEO, http://www.ncbi.nlm.nih.gov/geo). Information of each GEO GC cohort was included in Supplementary Table 2. A total of 120 GC samples (for the tissue level expression examination) and one fresh GC sample (for the single-cell level expression examination) were collected from Affiliated Hospital of Jiangnan University with informed consent, and this project was approved by the Clinical Research Ethics Committees of Affiliated Hospital of Jiangnan University (Supplementary Table 3).

Evaluation of immune infiltrates
Nine different methods were used to evaluate immune infiltrates. ESTIMATE, 22 immune cell types (LM22) of CIBERSORTx and Pan-cancer immunogenomic based on single sample gene set enrichment analysis (ssGSEA) were used as we previously described [8,16–19]. Five cell type quantification methods, including TIMER, QUANTISEQ, XCELL, MCPCOUN-TER and EPIC were calculated by TIMER 2.0 (http://timer.cistrome.org/) [20–25]. ImmuCellAI was used to estimate the abundance of immune cells from gene expression dataset (http://bioinfo.life.hust.edu.cn/ImmuCellAI#!/) [26]. Lists of Pan-cancer immunogenomic signature and aggregation schemes for each method were defined as shown in Supplementary Tables 4 and 5, respectively. A single-cell RNA-seq examination was performed by experimental personnel in the laboratory of GENECHEM according to 10× Genomics Single Cell Protocols. Seurat package in R software was used to perform single-cell cluster and annotation [27]. Cell types from each patient were annotated according to expression of EPCAM (cancer cell), MS4A1 (B cell), MS4A2 (mast cell), CD3E (T cell), CD68 (macrophage), DCN (fibroblast), TNFRSF17 (plasma cell), PLVAP (endothelial cell), CSF3R (neutrophil) and RGS5 (pericyte).

Molecular subtype analysis
Six immune subtypes were identified in more than 10000 tumours across 33 TCGA cancer types by performing an extensive immunogenomic analysis, including C1 (wound healing), C2 (IFN-g dominant), C3 (inflammatory), C4 (lymphocyte depleted), C5 (immunologically quiet) and C6 (TGF-β dominant) subtypes [10]. GC was divided into four molecular classification in previous TCGA study: Epstein–Barr virus (EBV), microsatellite unstable (MSI), genomically stable (GS) and chromosomal instability (CIN) tumours [9]. Two distinct molecular subtypes: mesenchymal phenotype (MP) and epithelial phenotype (EP) were identified in GC by analysing genomic and proteomic data [28]. Four clinically relevant molecular subtypes were established in GC data of Asian Cancer Research Group (ACRG), including EMT, MSI, MSS/TP53+ and MSS/TP53- subtypes [29]. Four immune/fibrotic TME subtypes were detected and conserved in human cancers: immune-enriched/fibrotic (IE/F), immune-enriched/non-fibrotic (IE), fibrotic (F) and immune-depleted (D) [30].

Enrichment analysis
Hallmark TGF-β signalling, Go biological adhesion, Hallmark EMT, Hallmark angiogenesis, KEGG pathways in cancer and Vecchi gastric cancer advanced vs early up signatures were obtained from Molecular Signatures Database v7.0 (https://www.gsea-msigdb.org/gsea/msigdb/index.jsp). Cancer-associated extracellular matrix (ECM) genes were acquired from prior research [31]. ssGSEA of GSVA package in R software was used to calculate these signatures in each GC cohort according to previous analysis [19, 32].

Quantitative RT-PCR (qRT-PCR), haematoxylin–eosin (HE), immunohistochemistry (IHC) and multicolour immunofluorescence
Total RNA was purified from GC tissues and was reverse transcribed into cDNA as we previously described [8]. QRT-PCR was used to quantitate the relative mRNA expression of CTSL and ZBTB7B using UltraSYBR Mixture (CWBIO, China). β-Actin was used as an internal control. Primers are listed in Supplementary Table 6. HE and IHC staining assays were performed based on our previously described method [8, 33]. IHC staining was applied to determine the relative protein expression of CTSL and ZBTB7B using anti-cathepsin L (CTSL, Santa Cruz, sc-32320, 1:100) and anti-TH-POK (ZBTB7B, Santa Cruz, sc-376250, 1:100) anti-bodies, respectively. Multiple fluorescent immunohistochemical staining kit (absin, abs50012) was used for immunofluorescence labelling as previously described [34]. Primary antibodies including anti-cathepsin L (Abcam, ab133641, 1:100) and anti-CD68 (Santa Cruz, sc-20060, 1:50).

Statistical analysis, code availability and visualisation
OS of ICs, IC score, CTSL or ZBTB7B was evaluated by Kaplan–Meier survival analysis and log-rank test as previously describe (Supplementary Fig. 1) [8, 32]: samples were divided into high and low expression groups, P and hazard ratio (HR) values of OS of based two groups were examined according to Kaplan–Meier survival analysis. The lowest log-rank P value was select from the 10th to 90th percentiles of the samples. Pscore and HRscore was defined as –log (P value) and $\log_2$ (HR value), respectively. Patients' survival was analysed by R software 4.1.0 and Graphpad prism 9. Cox regression model analyses were conducted using R software 4.1.0 and Graphpad prism 9. T test or Mann–Whitney test was used in the two-group comparison and the Kruskal–Wallis test was used in the three-group comparison using Graphpad prism 9. All reported P values were two-sided. Group comparison and Chi-squared test analyses were performed using GraphPad Prism 9. Analyses were performed using R software 4.1.0 with GSVA (1.28.0) package (ssGSEA) for Pan-cancer immunogenomic and enrichment score calculation; survival (2.44-1.1) package for two-group OS analysis and COX regression model analysis; ESTIMATE (1.0.13) package for ESTIMATE ImmuneScore and TumorPurity estimation; Seurat (4.0.3) packages for cluster and annotation in single-cell GC samples. A P value <0.05 was regarded as statistically significant. Figures were designed, analysed and visualised by GraphPad Prism 9 and R software 4.1.0.

RESULTS
CTSL and ZBTB7B expression were intensely associated with IC expression and survival in GC
Activation of ICs induces an immunosuppressive TME and promotes tumour progression, suggesting that high expression of ICs is associated with poor disease outcomes [35]. Thus, we performed survival analyses for nine ICs in the TCGA GC cohort according to a previously TME study [14]. First, we explored the expression levels of these ICs and identified a co-expression pattern in the TCGA GC cohort (Supplementary Fig. 2a). Then, we grouped patients based on each IC expression level as we previously described [8]. The optimal cutoff of each IC expression was applied to best stratify patients into two groups, respectively (Supplementary Fig. 1). Interestingly, most ICs were associated with a favourable OS or did not show an obvious relationship with survival (Pscore < 1.3, equivalent to log-rank test P > 0.05), except that PDCD1LG2 showed a relative weaker relationship to poor survival (P = 0.036, Supplementary Fig. 2b–j). Then, we estimated whether combined factors integrated by ICs could serve as a good indicator for a poor prognosis in GC. IC score was generated by ssGSEA based on expressions of these nine ICs. High IC score group was significantly associated with favourable survival and showed higher ICs expression compared with low IC score group (Supplementary Fig. 3). These analyses revealed that it may inappropriate to use ICs directly as a basis for stratification and
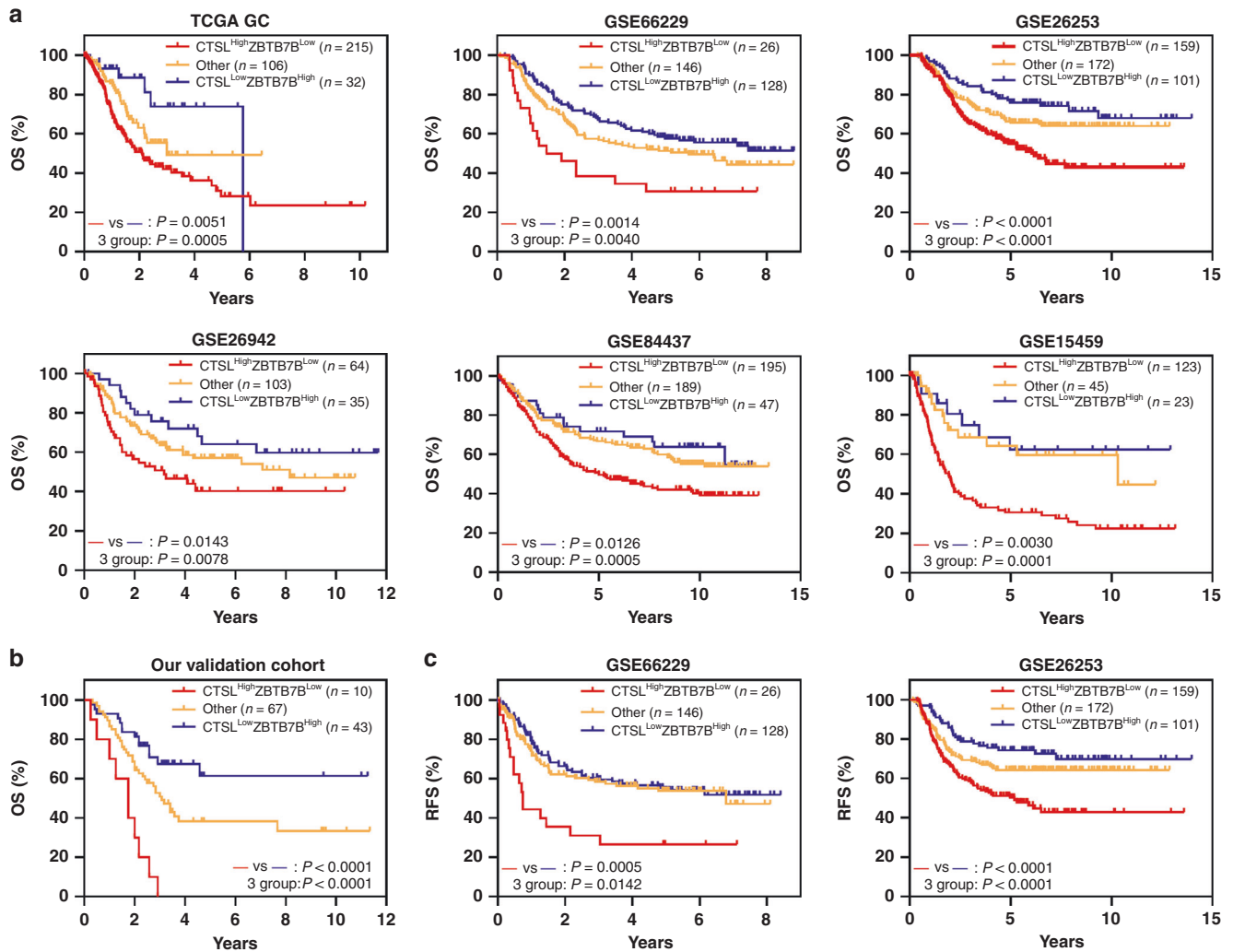
Fig. 1 Stratification based on the expression levels of CTSL and ZBTB7B was associated with GC prognosis. a, b Kaplan–Meier plots showing the OS for each subgroup of CTSL/ZBTB7B stratification system in TCGA GC, GSE66229, GSE26253, GSE26942, GSE84437, GSE15459 **a** and our validation **b** cohorts. **c** Kaplan–Meier plots showing the relapse-free survival (RFS) for each CTSL/ZBTB7B stratification system in GSE66229 and GSE26253 cohorts.

prognosis evaluation in GC. It is necessary to develop indirect methods to associate the ICs expression with poor prognosis in GC.

Thus, we used a systemic biological approach to identify two genes (CTSL and ZBTB7B) that representing positive or negative correlations, respectively (Supplementary Fig. 4a). First, 60483 genes were found in the TCGA GC RNA-Seq data, and 4755 genes were identified after reducing the background noise. Then, we analysed the correlations of these genes to IC scores one by one. A total of 456 and 34 genes were identified as IC positively and negatively correlated genes, respectively. There are two paths in the next set of analyses. For the IC negatively correlated genes, only ZBTB7B was identified as favourable survival-related. Of the 456 IC positively correlated genes, 18 candidate genes were significantly associated with patient poor survival, and we further identified seven upregulated genes. CTSL showed the highest HR and was highly correlated with the IC score. As expected, the expression level of CTSL increased as the IC score and expression levels of each IC increased, whereas the expression of ZBTB7B decreased as the IC score and expression levels of each IC increased (Supplementary Fig. 4b). Furthermore, high expression of CTSL or ZBTB7B was associated with poor or favourable OS, respectively, in multiple independent GC cohorts (Supplementary

Figs. 4c, d and 5). Moreover, both the expression of both CTSL and ZBTB7B was dysregulated in GC (Supplementary Fig. 4e, f). To further validate these findings, an independent cohort of 120 GC patients was collected and subjected to a qRT-PCR assay. Consistent with above observations in public GC cohorts, CTSL was upregulated in GC and associated with poor survival, whereas ZBTB7B was downregulated in GC and associated with favourable survival (Supplementary Fig. 6). These results demonstrate the potential of CTSL and ZBTB7B expression to associate ICs with survival outcomes of GC.

Stratification based on the expression of CTSL and ZBTB7B was associated with GC prognosis
We next applied the optimal cut-offs for CTSL and ZBTB7B to stratify patients into different risk subgroups. When dichotomising patients using the optimal cut-offs of CTSL and ZBTB7B in the TCGA GC cohort, the CTSL^{High}ZBTB7B^{Low} subgroup was associated with a worst OS, while the CTSL^{Low}ZBTB7B^{High} subgroup showed a best OS (Fig. 1a). To validate the existence of these distinctive subgroups in GC, additional GEO datasets and our validation cohort were analysed, and confirmed these results in the TCGA GC cohort (Fig. 1a, b and Supplementary Fig. 7). Notably, survival prediction among combination of CTSL/ZBTB7B was better than
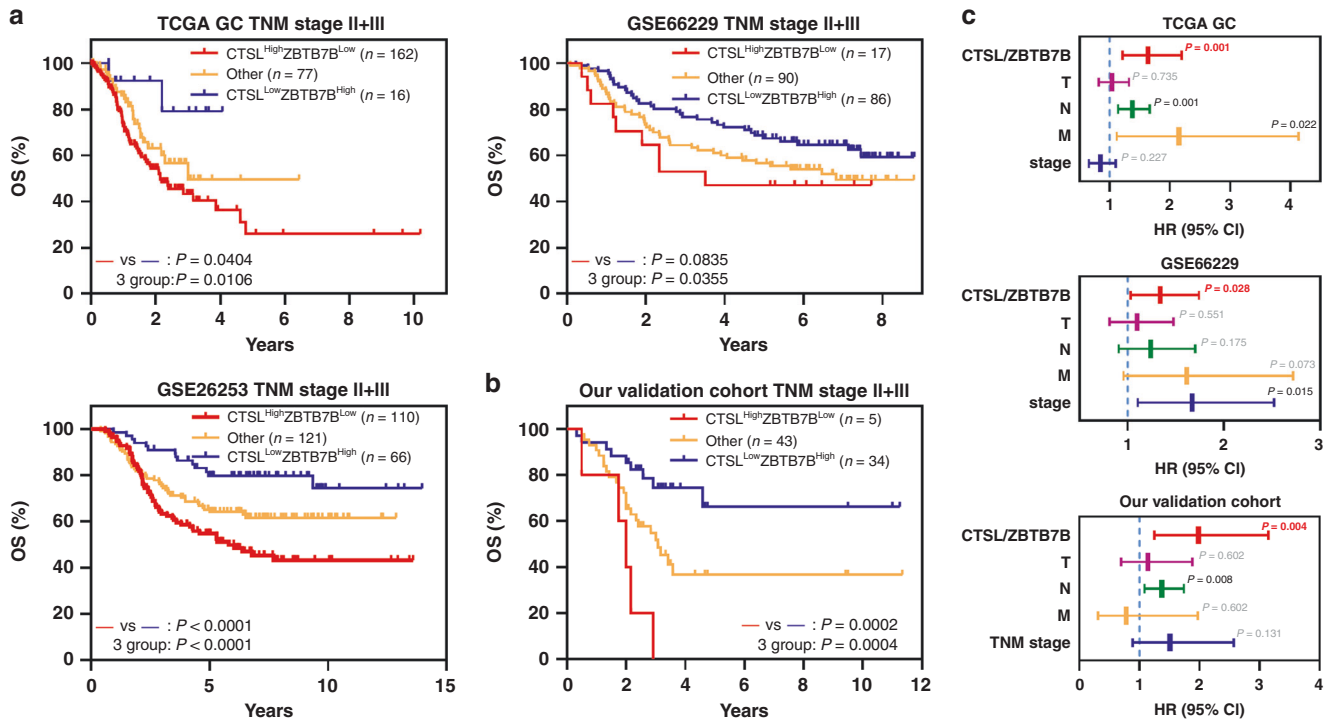
Fig. 2 Prognostic value of CTSL/ZBTB7B stratification according to TNM stage. a, b Kaplan–Meier plots showing the OS for each subgroup of CTSL/ZBTB7B stratification system in stages II–III samples of TCGA GC, GSE66229, GSE26253 **a** and our validation **b** cohorts. **c** Multivariate COX regression models showing the effect to OS of CTSL/ZBTB7B stratification system and TNM stage indexes in TCGA GC, GSE66229 and our validation cohorts.

single CTSL or ZBTB7B according to *P* values. To further explore the clinical prognostic value for each subpopulation of the CTSL/ ZBTB7B stratification, we estimated the recurrence risk in the three subgroups. Our results revealed that the CTSL^{High}ZBTB7B^{Low} subgroup had a higher risk of relapse, while the CTSL^{Low} ZBTB7B^{High} subgroup showed a lower risk of relapse in two GEO datasets (Fig. 1c). Collectively, these data reveal that the CTSL^{High}ZBTB7B^{Low} and CTSL^{Low} ZBTB7B^{High} subgroups were intensely correlated with survival and recurrence in GC patients.

The TNM staging system has been widely accepted as a powerful predictor of survival and treatment response in human cancers. To evaluate the prognostic value of the CTSL/ZBTB7B system, Kaplan–Meier and stratification analyses were performed using in GC patients with TNM II + III stage, and the CTSL^{High}ZBTB7B^{Low} subgroup still showed poor outcomes in comparison with the CTSL^{Low}ZBTB7B^{High} subgroup in public datasets and validation GC cohort (Fig. 2a, b). Next, Cox regression models were developed to integrate the CTSL/ZBTB7B stratification and TNM staging indexes, and no collinearity was observed among these factors (tolerance >0.1 and VIF < 10). The CTSL/ ZBTB7B stratification remained an independent prognostic variable for GC patients in both public and validation cohorts (Fig. 2c). In summary, these data suggest that the stratification based on the expression levels of CTSL and ZBTB7B was significantly associated with the GC prognosis.

The CTSL^{High}ZBTB7B^{Low} high-risk subpopulation featured by high immune infiltrates and low tumour purity
We investigated the TME characteristic in each subgroup using ESTIMATE method (Fig. 3a, b). The CTSL^{High}ZBTB7B^{Low} subgroup demonstrated the highest immune infiltrates and lowest tumour purity, while the CTSL^{Low}ZBTB7B^{High} subgroup showed the lowest immune infiltrate in GC cohorts. Moreover, we observed that total and major immune cell types were also highly infiltrated in the CTSL^{High}ZBTB7B^{Low} subgroup, and were poorly infiltrated in the

CTSL^{Low}ZBTB7B^{High} subgroup in most public GC cohorts by using other evaluation approaches (Supplementary Figs. 8–15). For instance, CD4 and CD8 T cell levels were highest in the CTSL^{High}ZBTB7B^{Low} subgroup from most public GC cohorts according to TIMER, Pan-cancer immunogenomic ssGSEA and EPIC, whereas the CTSL^{Low}ZBTB7B^{High} subgroup demonstrated contrary results. CEACAM5 and KRT20 are considered gastro-intestinal cancer markers, and we found the CTSL^{High}ZBTB7B^{Low} subgroup demonstrated the lowest expression levels of CEACAM5 and KRT20 in most public GC cohorts, while the CTSL^{Low} ZBTB7B^{High} subgroup showed the highest levels of these two markers (Supplementary Fig. 16) [36–39].

To further validate these findings, we performed a single-cell analysis using primary tumour cases from the GSE183904, GSE163558 and our validation sample (Supplementary Figs. 17–19), and found expression of CTSL or ZBTB7B were almost existed in multiple cell types (Fig. 3c–e left/middle). Interestingly, GC cell population expressed relatively higher ZBTB7B, whereas CTSL expression levels were higher in non-malignant cell types, especially in macrophages. Moreover, CTSL showed the highest expression level in macrophages, while ZBTB7B showed the highest expression level in cancer cells (Fig. 3c–e right). Besides, we defined those cells with detectable expression of CTSL or ZBTB7B as CTSL^{+} or ZBTB7B^{+} cells, respectively. The ratio of the percentage of CTSL^{+} to that of ZBTB7B^{+} cells in each case was negatively correlated with cancer cell proportion in the GSE183904 GC dataset, but positively correlated with immune cells proportion (Fig. 3f). Finally, results of HE and IHC staining in validation cohort also revealed higher immune cell infiltration and lower tumour purity in the CTSL^{High} ZBTB7B^{Low} subgroup than in the CTSL^{Low} ZBTB7B^{High} subgroup (Fig. 3g). We noted that a recent study about TME suggest that macrophages derived from bone marrow or purified from E0771 tumours overexpressed CTSL [40]. Thus, multicolour immuno-fluorescence staining was performed in GC tissues and confirmed
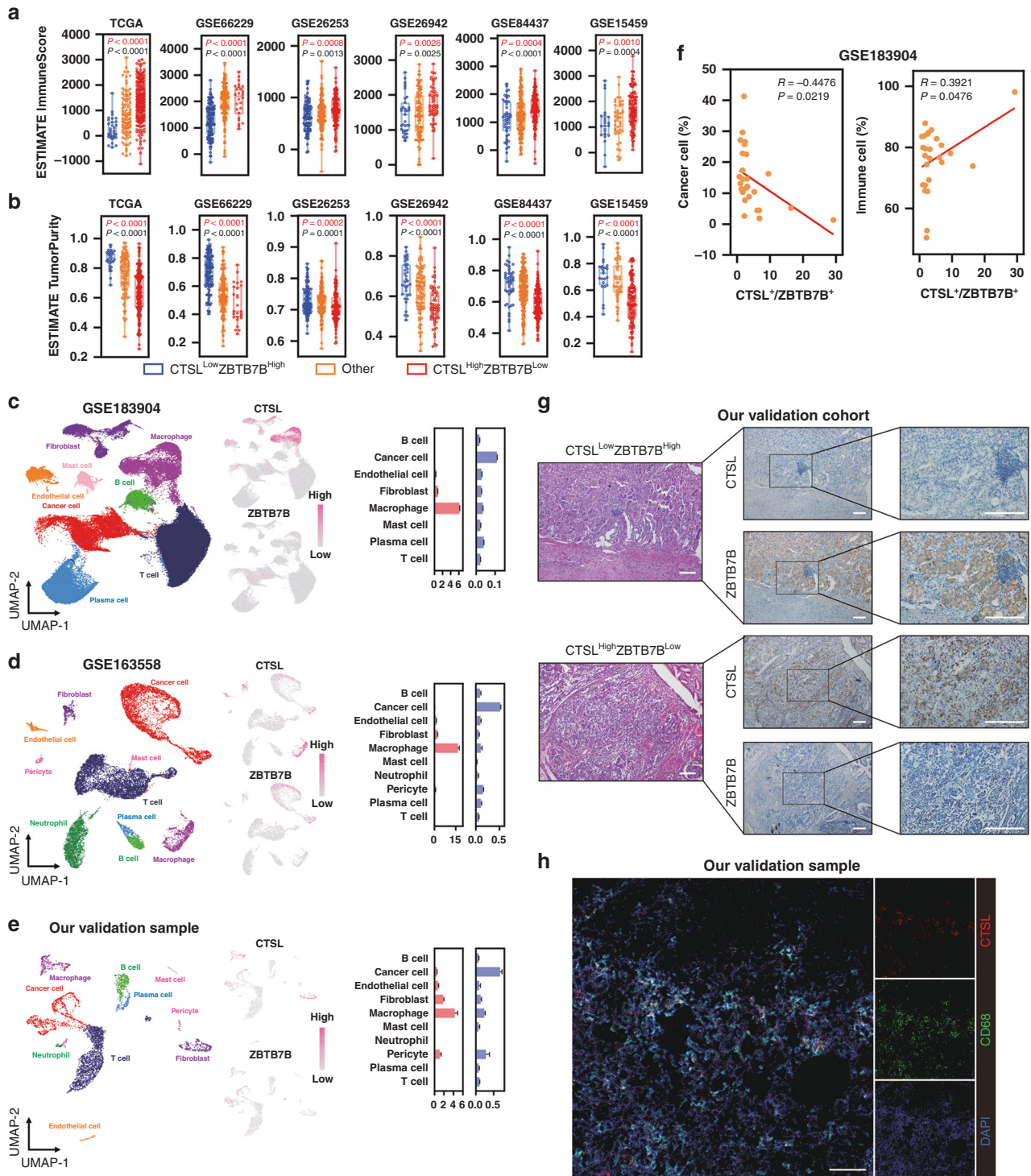
Fig. 3 The CTSL^High^ZBTB7B^Low^ risk subpopulation feature by high immune infiltrates and low tumour purity. a, b Box plots showing the total immune infiltrates **a** and tumour purity **b** of ESTIMATE for each subgroup of CTSL/ZBTB7B stratification system in TCGA GC, GSE66229, GSE26253, GSE26942, GSE84437 and GSE15459 cohorts. **c** Left: UMAP plots showing the cancer cell, immune cell and other cell type cluster in the GSE183904 GC dataset. Middle: UMAP plots showing expression of CTSL and ZBTB7B in cells of the GSE183904 GC dataset. Right: Bar plots showing expression levels of CTSL and ZBTB7B in each cell type of the GSE183904 GC dataset. Data are shown as means ± SEM. **d** Similar to **c**, but in the GSE163558 GC dataset. **e** Similar to **c**, but in our validation sample. **f** Scatter plots showing the Pearson correlation between cancer/immune cell proportion and ratio of CTSL^+^ to ZBTB7B^+^ cells in the GSE183904 GC dataset. **g** Representative pictures for HE staining and CSTL/ZBTB7B IHC staining in our GC validation samples. Scale bar for images = 100 μm. **h** Representative pictures for multicolour immunofluorescence staining in GC validation sample. Red: CTSL; Green: CD68; Blue: DAPI. Scale bar for image = 50 μm.

that CTSL was co-expressed to macrophage marker CD68 (Fig. 3h). In summary, our data at both GC bulk tissue and single-cell levels indicate that the CTSL^{High}ZBTB7B^{Low} high-risk subgroup is characterised by a TME with high immune cell infiltration and low tumour purity.

Exploration of potential reasons accounting for the poor survival in CTSL^{High}ZBTB7B^{Low} patients

To answer why the CTSL^{High}ZBTB7B^{Low} cases with high immune infiltration and low tumour purity but show the worst survival, we analysed the expression levels of ICs, because the IC signals could induce an immunosuppressive TME [35]. Notably, the CTSL^{High}ZBTB7B^{Low} subgroup expressed the highest levels of ICs in GC cohorts (Fig. 4a). To further evaluate the TME characteristics in different CTSL/ZBTB7B subgroups, we investigated the proportions of different immune subtypes, GC molecular classification, EP/MP subtypes, ACRG subtypes and TME subtypes in each subgroup of this stratification system using the TCGA GC cohort [9, 10, 28–30]. The CTSL^{High}ZBTB7B^{Low} subgroup showed higher proportions of C2, C3, C6, EBV, GS, MP, EMT, IE/F and IE subtypes than other subgroups (Fig. 4b). Note that the CTSL^{Low}ZBTB7B^{High} subgroup does not overlap with the C6 or EBV subtype. Intriguingly, C2 and C3 cases had strong T cell signals, and C6 cases displayed the highest TGF-β signature. In addition, EBV subtypes displayed PD-L1/2 (CD274/PDCD1LG2) overexpression and high levels of immune cell signalling. GS cases are characterised by higher cell adhesion. MP and EMT subtypes showed active EMT biological process. IE/F and IE subtypes were distinguished by high levels of immune infiltrate. In particular, IE/F features with cancer-associated fibroblast (CAF) activation, high IC expression and a high response to immunotherapy. High levels of the TGF-β signature, PD-L1/2, cell adhesion and EMT are considered tumour-promoting phenotypes and significantly associated with a poor prognosis. CAF have been reported to promote tumorigenesis and metastasis by secreting TGF-β and other mechanisms. Cancer-associated ECM genes dysregulation is correlated with the activation of the TGF-β signalling in CAFs and is associated with immunosuppression in otherwise immunologically active tumours [31]. Thus, we further analysed these signatures. Both TGF-β signalling and cancer-associated ECM genes signature levels were the highest in the CTSL^{High}ZBTB7B^{Low} subgroup and the lowest in the CTSL^{Low}ZBTB7B^{High} subgroup (Supplementary Fig. 20a, b). Besides, we estimated gene signatures of adhesion and observed the highest adhesion levels in the CTSL^{High}ZBTB7B^{Low} subgroup compared with other GC subgroups (Supplementary Fig. 20c). Notably, the CTSL^{High}ZBTB7B^{Low} subgroup showed the highest EMT and angiogenesis activity, with the CTSL^{Low}ZBTB7B^{High} subgroup showing the lowest (Supplementary Fig. 20d, e). Then, the CTSL^{High}ZBTB7B^{Low} subgroup had the highest levels of CAFs in the multiple public GC cohorts (Supplementary Fig. 21). Moreover, the CTSL^{High}ZBTB7B^{Low} subgroup exhibited the highest levels of pathway in cancer and advanced tumour progression signatures (Supplementary Fig. 20f, g). Taken together, the CTSL^{High}ZBTB7B^{Low} patients featured by a comparatively higher immunosuppressive TME and enriched cancer/tumour metastasis-related pathways, which further explains why the CTSL^{High}ZBTB7B^{Low} subgroup demonstrates poor clinical outcomes despite high immune cell infiltration and low tumour purity.

C2, CIN, MP and D cases occupied the largest proportion in the TCGA GC cohort based on their matched molecular subtype. Interestingly, the CTSL/ZBTB7B stratification system still stratified C2, CIN, MP or D cases into different subgroups, respectively. The CTSL^{High}ZBTB7B^{Low} subgroup was associated with poor 5-year OS compared with the CTSL^{Low}ZBTB7B^{High} subgroup in C2, CIN or MP GC patients, respectively (Fig. 4c). These data demonstrates that CSTL/ZBTB7B stratification has additional and independent prognostic implications beyond these classification.

Pan-cancer prognostic significance of the CTSL/ZBTB7B stratification system

To estimate the prognostic effect of the CTSL/ZBTB7B stratification system in human cancer types, we analysed more than 7000 primary tumour cases of other 19 cancer types from the TCGA RNA-Seq dataset as described in our previous study [8]. The CTSL^{High}ZBTB7B^{Low} subpopulation showed worse survival than the CTSL^{Low}ZBTB7B^{High} subpopulation in THCA, BLCA, HNSC and OV, whereas kidney cancers, PAAD, melanoma and PRAD showed the opposite results (Fig. 5a). Because the CTSL^{High}ZBTB7B^{Low} subgroup only comprised six cases (1%) in THCA, we selected BLCA, HNSC and OV for subsequent analyses. The CTSL^{High}ZBTB7B^{Low} subgroup had the worst survival, with the CTSL^{Low}ZBTB7B^{High} subgroup showing the best prognosis, though three-group $P = 0.0608$ in OV (Fig. 5b). Further analyses demonstrated that the CTSL^{High}ZBTB7B^{Low} subpopulation showed the highest immune infiltration, lowest tumour purity and IC activation in these three cancer types (Fig. 5c, d). These data support the proposal that the CTSL/ZBTB7B stratification system could identify different prognostic subgroups with high immune infiltrates and low tumour purity in certain cancer types.

DISCUSSION

GC is a heterogeneous disease featured by a complicated TME and molecular characteristics. Thus, several TME-based immune molecular subtypes have been constructed for patient selection to apply precision therapy [9, 10]. Currently, the main anti-cancer drugs for GC treatment comprise chemotherapy and targeted therapy. Recently, IC inhibitors related immunotherapy has emerged as a revolutionary and promising treatment approach in some cancer types [35]. In this study, two genes CTSL and ZBTB7B were identified based on the expression of ICs in the TCGA GC cohort. Based on the expression of CTSL and ZBTB7B, we further constructed a two-gene stratification system that could efficiently identify a high-risk subgroup with high immune infiltrates and low tumour purity in multiple GC cohorts and other TCGA cancer types (Fig. 6).

CTSL, a human cysteine proteases gene, is dysregulated and associated with malignant phenotypes in human cancers. A recent study showed that CTSL promotes angiogenesis and tumour growth in GC [41]. ZBTB7B, also named as ThPOK, is a CD4-lineage transcription factor gene that has been reported to suppress GC cell viability and promote the proliferation of T cells [42]. By using multiple public GC cohorts and clinical samples we collected, we found that CTSL was upregulated in GC, while ZBTB7B was downregulated in GC. According to an approach applied in our previous studies, GC patients were divided into two groups based on the expression of CTSL or ZBTB7B, and CTSL^{High} or ZBTB7B^{Low} cases exhibit poor OS [8, 32]. According to this method, a CTSL/ZBTB7B subtyping system was developed and applied to other seven GC cohorts at tissue level, including six public cohorts and our validation cohort. The CTSL^{High}ZBTB7B^{Low} subgroup was associated with the worst clinical outcomes, while the CTSL^{Low}ZBTB7B^{High} subgroup showed best clinical outcomes. These results are also consistent with those previous studies that showed CTSL as a tumour-promoting factor, while ZBTB7B as a tumour-suppressor [41, 42]. It is noted that there are different optimal cut-off peaks for CTLS or ZBTB7B in different GC cohorts, and the percentage of each CTSL/ZBTB7B subgroup was also dynamic in different GC cohorts. These dynamic percentages may reflect heterogeneity in each GC cohort, such as TME characteristics, methodological difference and region specificity. Thus, this finding needs further validation in a large prospective patient population to define a standard cut-off for the CTSL^{High}ZBTB7B^{Low} subgroup.

The TME is the crucial for studying the immunotherapy response, and an active response to IC inhibitors requires immune
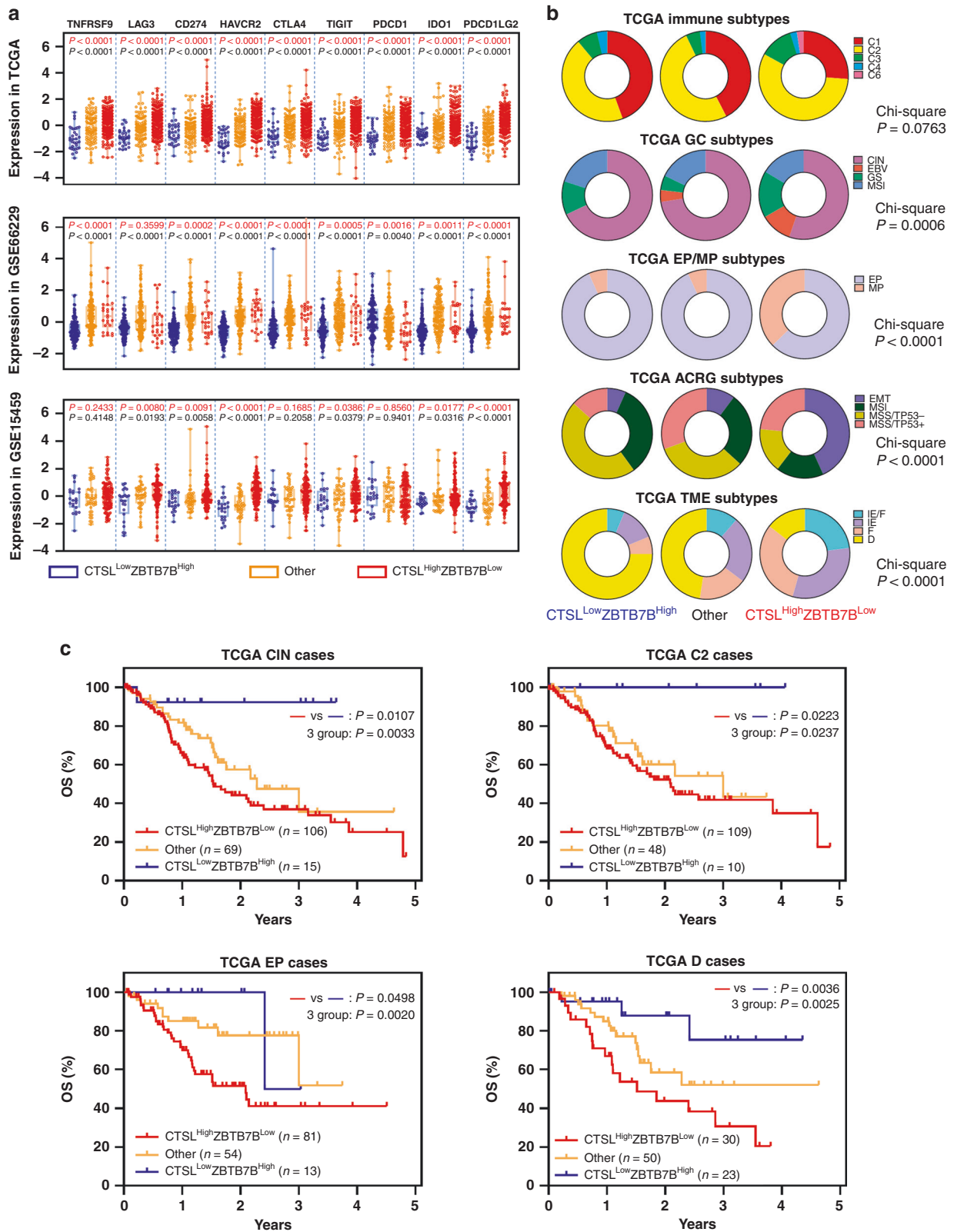
Fig. 4 Exploration of potential reasons accounting for the poor survival in CTSL^{High}ZBTB7B^{Low} patients. a Box plots showing the expression levels of nine ICs for each CTSL/ZBTB7B stratification system in TCGA GC, GSE66229 and GSE15459 cohorts. **b** Pie charts showing the distribution of immune subtypes, GC subtypes, EP/MP subtypes, ACRG subtypes and TME subtypes in each subgroup of the CTSL/ZBTB7B stratification system from TCGA GC cohort. **c** Kaplan–Meier plots showing the 5-year OS for each CTSL/ZBTB7B stratification system in cases of TCGA GC C2, CIN, EP and D subtypes.
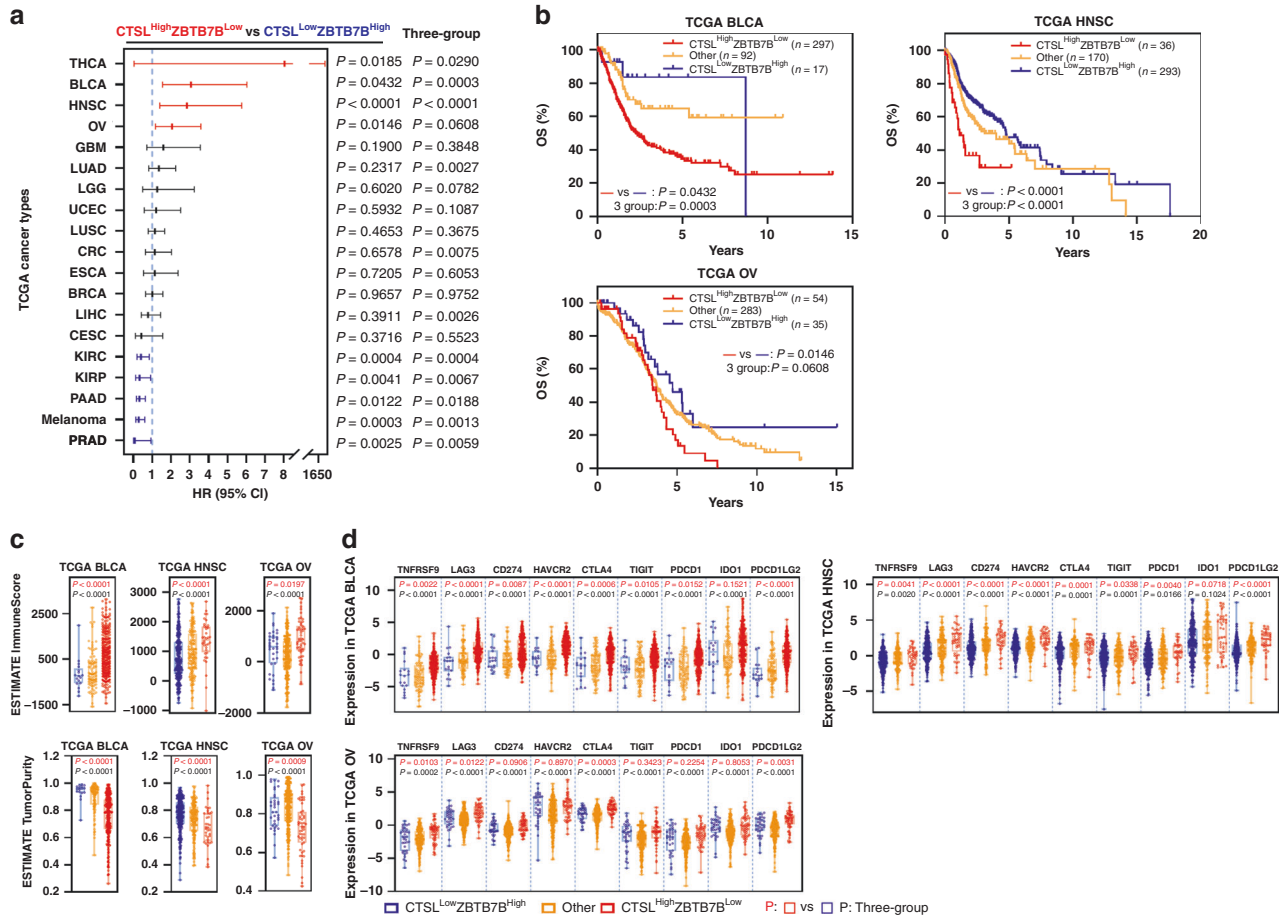
Fig. 5 The CTSL/ZBTB7B stratification system shows prognostic significance in Pan-cancer. a Forest plot showing the OS for each CTSL/ZBTB7B stratification system in TCGA Pan-cancer. Abbreviations of each TCGA cancer type are included in Supplementary Table 1. **b** Kaplan–Meier plots showing the OS for each CTSL/ZBTB7B stratification system in BLCA, HNSC and OV cohorts. **c** Box plots showing the total immune infiltrates and tumour purity of ESTIMATE in each subgroup of the CTSL/ZBTB7B stratification system from BLCA, HNSC and OV cohorts. **d** Box plots showing the expression levels of ICs in each subgroup of the CTSL/ZBTB7B stratification system from BLCA, HNSC and OV cohorts.

infiltration [43, 44]. In this study, nine independent methods (ESTIMATE, TIMER, CIBERSORTx, ssGSEA based on Pan-cancer immunogenomic, QUANTISEQ, XCELL, MCPCOUNTER, EPIC and ImmuCellAI) were used to evaluate the immune-related activity of each CTSL/ZBTB7B subgroup. Interestingly, the CTSL^{High}ZBTB7B^{Low} subgroup was positively correlated with immune cell infiltrates in most GC cohorts, while the CTSL^{Low}ZBTB7B^{High} subgroup showed negative correlations. CTSL^{High}ZBTB7B^{Low} cases demonstrated the lowest, while CTSL^{Low}ZBTB7B^{High} cases showed the highest tumour purity. These TME phenotypes seem contradict the functions of the two genes according to recent studies that CTSL promotes tumour progression, whereas ZBTB7B inhibits cell proliferation in GC [41, 42]. However, these studies were mainly focused on malignant phenotypes in GC cell lines and did not focus on the TME of GC tissues. Thus, we need to further explore why does the CTSL^{High}ZBTB7B^{Low} subpopulation have high immune infiltration and low tumour purity phenotypes, but a poor prognosis. Our analyses of single-cell data from the public and we collected sample revealed that CTSL expression levels were higher in nonmalignant cell types, especially in macrophages, while ZBTB7B was mainly expressed in cancer cells. This result further explained our findings at the bulk tissue level from GC cohorts. HE, IHC and multicolour immunofluorescence results in our validation samples also confirm these results to a certain extent. Interestingly, a recent study about TME reported that tumour-associated macrophages derived from bone marrow or

purified from E0771 tumours overexpressed CTSL and harboured hyperactive cysteine protease activity in their lysosomes, resulting in decreased CD8 T cell activation and enhanced tumour growth [40]. In our study, CTSL/ZBTB7B is the top poor/favourable clinical outcomes gene in our identification analysis, and represents immune infiltrates/tumour purity level, respectively. These two genes were used to develop an effective and readily implemented subtyping system to identify high-risk patients with high immune infiltrates and low tumour purity in GC. Nevertheless, the interaction between CTSL and ZBTB7B for poor prognosis in GC is yet to be determined and may further require dynamic single-cell analysis and some transgenetic mouse model of immune contexture in the future. Nonetheless, our findings do not contradict, but rather extend the function of CTSL and ZBTB7B, also suggesting that some tumour-promoting genes affect tumorigenesis and progression by regulating the functions of nonmalignant cell populations. The second possible explanation is that expression levels of IC genes were highest in the CTSL^{High}ZBTB7B^{Low} subgroup compared with the CTSL^{Low}ZBTB7B^{High} subgroup according to analyses in GC cohorts. ICs provide negative signals that limit T cell immune responses and are used by tumour cells to mediate tumour escape [35]. Third, parts of cell types are associated with immunosuppression phenotypes. For instance, Tregs and CAFs were enriched in the CTSL^{High}ZBTB7B^{Low} subgroup of GC cohorts. Tregs exhibit immunosuppressive activities. CAFs have pro-tumorigenic
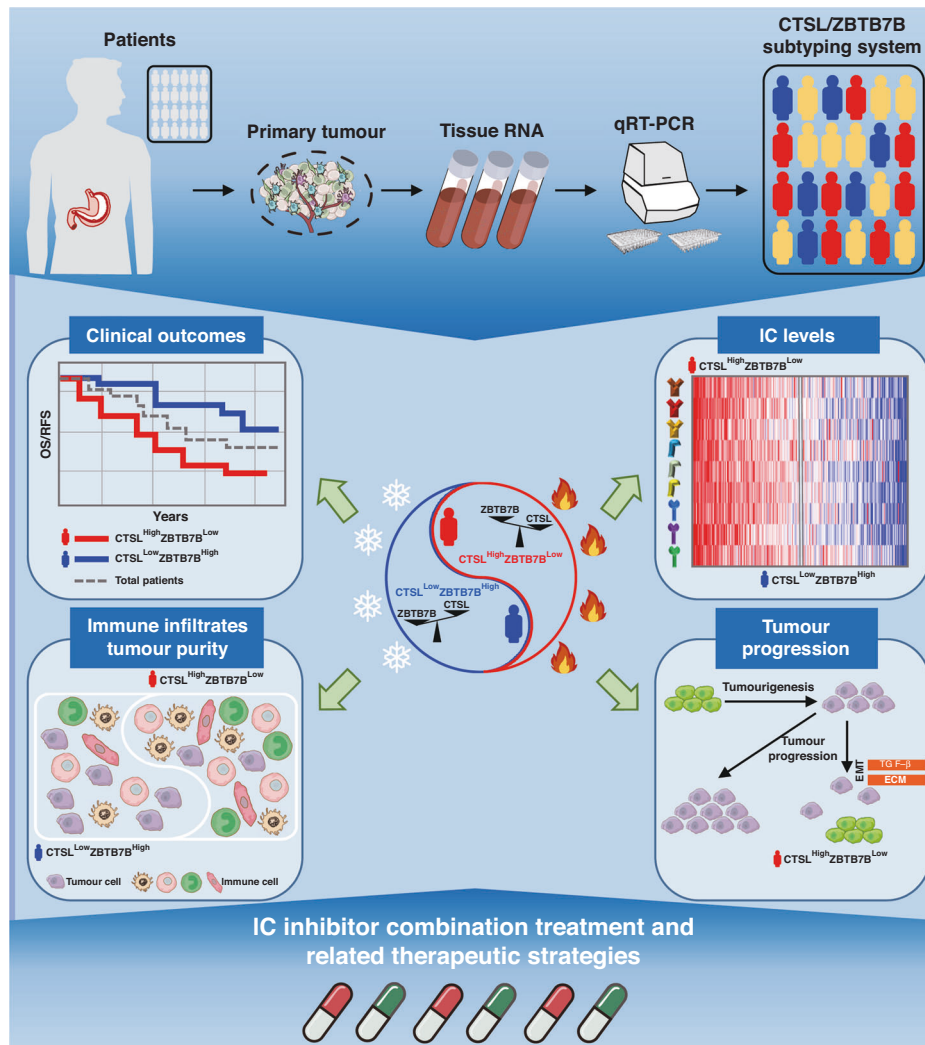
Fig. 6 Schema showing the flow chart, characteristics and application values of CTSL/ZBTB7B subtyping system for human cancer.

and immunosuppressive properties in most cases [35]. Fourth, immunosuppressive and cancer-related pathways were active in the CTSL^{High}ZBTB7B^{Low} subgroup. Prior studies explored and identified several molecular subtyping systems in TCGA, and we found that the CTSL^{High}ZBTB7B^{Low} high-risk subgroup encompassed higher proportions of C2, C3, C6, EBV, GS, MP, EMT, IE/F and IE subtypes than the CTSL^{Low}ZBTB7B^{High} subgroup. These molecular subtypes are associated with a series of cancer-related pathways or phenotypes, such as a high TGF-β signature, ECM, high IC expression, cell adhesion and active EMT process. In particular, increased TGF-β signalling is one of the primary mechanisms mediating immune escape. The TGF-β signalling also plays an important role in inducing EMT, which is critical in tumour invasion and metastasis [8, 45]. EMT also contributes to immune evasion via multiple routes, including the establishment of the TME, TGF-β and decreased sensitivity to immune effector cells [11]. Besides, cancer-associated ECM genes are associated with the activation of the TGF-β signalling in CAFs [31]. These features of the CTSL^{High}ZBTB7B^{Low} high-risk subgroup seems similar to an immune overdrive TME phenotype in previous studies, suggesting a counterbalancing phenotype between immunostimulatory and immunosuppressive mechanisms [7, 8]. Nonetheless, previous immune overdrive TME phenotypes were usually identified based on unique T cell populations-related TME, such high intratumoural CD8 T cell infiltration and a high density of tumour-associated macrophages that expressed CD274, as well as Tregs with

overexpressed CTLA4 [7, 8]. In this study, CTSL and ZBTB7B were identified at tissue levels and mainly expressed in different cell types, which were dissimilar to previous immune overdrive TME phenotypes.

The CTSL/ZBTB7B subtyping system displayed certain interesting aspects, advantages, and significance. First, nine GC datasets were used in this study, including Western and Eastern GC populations, and they were analysed by multiple approaches of gene expression examination, such as RNA-Seq (bulk and single-cell), microarray, qRT-PCR, IHC and multicolour immunofluorescence. Moreover, nine independent immune algorithms were used in analyses. These factors make the conclusions of this research more robust and reliable. Second, an accurate prognostic estimate is a critical prerequisite for selecting appropriate treatment strategies, such as TNM stage. The CTSL/ZBTB7B stratification remain an independent prognostic variable for survival after adjusting for TNM stage. Besides, this stratification has additional and independent prognostic implications beyond a series previous molecular subtypes. Third, the CTSL/ZBTB7B subtyping system can identify high immune infiltration and low tumour purity GC patients with relative higher ICs level, which may provide a guideline for the GC patients' selection in immunotherapies. The effectiveness of immunomodulatory strategies lies in the presence of anti-tumour immune responses. Patients of the CTSL^{High}ZBTB7B^{Low} subpopulation demonstrated poor clinical outcomes, high immune cell infiltration and levels of ICs, suggesting that anti-tumour immune responses

pre-exist in the TME, which could provide fertile ground for effective IC inhibitor-based monotherapy or combination therapy [7, 8, 11, 46]. Moreover, the CTSL^{High}ZBTB7B^{Low} high-risk subpopulation showed high TGF-β signalling and an active EMT process, hinting that targeting TGF-β and EMT pathway could augment the effectiveness of IC inhibitor-based therapies in these patients, such as simultaneous targeting of TGF-β/PD-L1 [47]. Besides, the CTSL^{High}ZBTB7B^{Low} high-risk cases probably benefit from cysteine proteases inhibit related therapy [40]. Besides, we performed a Pan-cancer analysis to explore the prognostic effect of the CTSL/ZBTB7B stratification system in human cancer types. BLCA, HNSC and OV showed similar results to GC, while kidney cancer and melanoma, etc. showed contrasting results. The reason behind this could be strongly tumour heterogeneous TME among cancer types. These findings suggest that the CTSL/ZBTB7B stratification system could be used for certain cancer types and need further validation in extra cohorts. Fourth, most previous molecular subtyping systems contain many marker genes and require complex examination methods. Additionally, novel techniques are costly and far from suitable for large-scale diagnostic use [11]. Thus, new molecular subtypes containing only a few marker genes have been identified in some cancer types, such as CD8A/CD274, Siglec15, FOXP3/CTLA4, VTCN1/CD274 and CD8A/IDO1 stratifications [7, 8, 12–14]. In this study, our subtyping system only needs to examine the expression levels of CTSL and ZBTB7B by performing qRT-PCR, which is commonly used in cancer diagnosis and treatment.

Collectively, our comprehensive analyses uncover a new CTSL/ZBTB7B subtyping system that could classify patients into high immune infiltrates and low tumour purity high-risk subgroup in GC. Notably, combined evaluation of CTSL and ZBTB7B expression is a strong predictor of survival and recurrence in GC patients. This study further provides a personalised prognostic method and may contribute to precision cancer therapies. Patients from the CTSL^{High}ZBTB7B^{Low} subpopulation may benefit from more active IC inhibitor-based combination treatment, which will also help in the development of related therapeutic strategies.

DATA AVAILABILITY
Available of public GC datasets are described in the 'Methods' section. The data that support this study are available from the corresponding author upon reasonable request.

REFERENCES
1. Sung H, Ferlay J, Siegel RL, Laversanne M, Soerjomataram I, Jemal A, et al. Global Cancer Statistics 2020: GLOBOCAN Estimates of Incidence and Mortality Worldwide for 36 Cancers in 185 Countries. CA Cancer J Clin. 2021;71:209–49.
2. Ramezankhani R, Solhi R, Es HA, Vosough M, Hassan M. Novel molecular targets in gastric adenocarcinoma. Pharm Ther. 2021;220:107714.
3. Giraldo NA, Sanchez-Salas R, Peske JD, Vano Y, Becht E, Petitprez F, et al. The clinical role of the TME in solid cancer. Br J Cancer. 2019;120:45–53.
4. Toor SM, Sasidharan Nair V, Decock J, Elkord E. Immune checkpoints in the tumor microenvironment. Semin Cancer Biol. 2020;65:1–12.
5. Jiang Y, Zhang Q, Hu Y, Li T, Yu J, Zhao L, et al. ImmunoScore signature: a prognostic and predictive tool in gastric cancer. Ann Surg. 2018;267:504–13.
6. Pages F, Mlecnik B, Marliot F, Bindea G, Ou FS, Bifulco C, et al. International validation of the consensus Immunoscore for the classification of colon cancer: a prognostic and accuracy study. Lancet. 2018;391:2128–39.
7. Fakih M, Ouyang C, Wang C, Tu TY, Gozo MC, Cho M, et al. Immune overdrive signature in colorectal tumor subset predicts poor clinical outcome. J Clin Investig. 2019;129:4464–76.
8. Cui K, Yao S, Zhang H, Zhou M, Liu B, Cao Y, et al. Identification of an immune overdrive high-risk subpopulation with aberrant expression of FOXP3 and CTLA4 in colorectal cancer. Oncogene. 2021;40:2130–45.
9. Cancer Genome Atlas Research Network. Comprehensive molecular characterization of gastric adenocarcinoma. Nature 2014;513:202–9.
10. Thorsson V, Gibbs DL, Brown SD, Wolf D, Bortone DS, Ou Yang TH, et al. The immune landscape of cancer. Immunity 2018;48:812.e14–30.e14.
11. Galon J, Bruni D. Approaches to treat immune hot, altered and cold tumours with combination immunotherapies. Nat Rev Drug Discov. 2019;18:197–218.
12. Zhang R, Li T, Wang W, Gan W, Lv S, Zeng Z, et al. Indoleamine 2, 3-dioxygenase 1 and CD8 expression profiling revealed an immunological subtype of colon cancer with a poor prognosis. Front Oncol. 2020;10:594098.
13. Hu J, Yu A, Othmane B, Qiu D, Li H, Li C, et al. Siglec15 shapes a non-inflamed tumor microenvironment and predicts the molecular subtype in bladder cancer. Theranostics. 2021;11:3089–108.
14. Chen D, Li G, Ji C, Lu Q, Qi Y, Tang C, et al. Enhanced B7-H4 expression in gliomas with low PD-L1 expression identifies super-cold tumors. J Immunother Cancer. 2020;8:e000154.
15. Liu J, Lichtenberg T, Hoadley KA, Poisson LM, Lazar AJ, Cherniack AD, et al. An integrated TCGA pan-cancer clinical data resource to drive high-quality survival outcome analytics. Cell. 2018;173:400–16.e11.
16. Yoshihara K, Shahmoradgoli M, Martinez E, Vegesna R, Kim H, Torres-Garcia W, et al. Inferring tumour purity and stromal and immune cell admixture from expression data. Nat Commun. 2013;4:2612.
17. Newman AM, Steen CB, Liu CL, Gentles AJ, Chaudhuri AA, Scherer F, et al. Determining cell type abundance and expression from bulk tissues with digital cytometry. Nat Biotechnol. 2019;37:773–82.
18. Charoentong P, Finotello F, Angelova M, Mayer C, Efremova M, Rieder D, et al. Pan-cancer immunogenomic analyses reveal genotype-immunophenotype relationships and predictors of response to checkpoint blockade. Cell Rep. 2017;18:248–62.
19. Hanzelmann S, Castelo R, Guinney J. GSVA: gene set variation analysis for microarray and RNA-seq data. BMC Bioinformatics. 2013;14:7.
20. Sturm G, Finotello F, Petitprez F, Zhang JD, Baumbach J, Fridman WH, et al. Comprehensive evaluation of transcriptome-based cell-type quantification methods for immuno-oncology. Bioinformatics. 2019;35:i436–45.
21. Li B, Severson E, Pignon JC, Zhao H, Li T, Novak J, et al. Comprehensive analyses of tumor immunity: implications for cancer immunotherapy. Genome Biol. 2016;17:174.
22. Finotello F, Mayer C, Plattner C, Laschober G, Rieder D, Hackl H, et al. Molecular and pharmacological modulators of the tumor immune contexture revealed by deconvolution of RNA-seq data. Genome Med. 2019;11:34.
23. Aran D, Hu Z, Butte AJ. xCell: digitally portraying the tissue cellular heterogeneity landscape. Genome Biol. 2017;18:220.
24. Becht E, Giraldo NA, Lacroix L, Buttard B, Elarouci N, Petitprez F, et al. Estimating the population abundance of tissue-infiltrating immune and stromal cell populations using gene expression. Genome Biol. 2016;17:218.
25. Racle J, de Jonge K, Baumgaertner P, Speiser DE, Gfeller D. Simultaneous enumeration of cancer and immune cell types from bulk tumor gene expression data. Elife. 2017;6:e26476.
26. Miao YR, Zhang Q, Lei Q, Luo M, Xie GY, Wang H, et al. ImmuCellAI: a unique method for comprehensive T-cell subsets abundance prediction and its application in cancer immunotherapy. Adv Sci. 2020;7:1902880.
27. Hao Y, Hao S, Andersen-Nissen E, Mauck WM 3rd, Zheng S, Butler A, et al. Integrated analysis of multimodal single-cell data. Cell. 2021;184:3573–87.e29.
28. Oh SC, Sohn BH, Cheong JH, Kim SB, Lee JE, Park KC, et al. Clinical and genomic landscape of gastric cancer with a mesenchymal phenotype. Nat Commun. 2018;9:1777.
29. Cristescu R, Lee J, Nebozhyn M, Kim KM, Ting JC, Wong SS, et al. Molecular analysis of gastric cancer identifies subtypes associated with distinct clinical outcomes. Nat Med. 2015;21:449–56.
30. Bagaev A, Kotlov N, Nomie K, Svekolkin V, Gafurov A, Isaeva O, et al. Conserved pan-cancer microenvironment subtypes predict response to immunotherapy. Cancer Cell. 2021;39:845–65.e7.
31. Chakravarthy A, Khan L, Bensler NP, Bose P, De Carvalho DD. TGF-beta-associated extracellular matrix genes link cancer-associated fibroblasts to immune evasion and immunotherapy failure. Nat Commun. 2018;9:4692.
32. Cui K, Liu C, Li X, Zhang Q, Li Y. Comprehensive characterization of the rRNA metabolism-related genes in human cancer. Oncogene. 2020;39:786–800.
33. Gong L, Li Y, Cui K, Chen Y, Hong H, Li J, et al. Nanobody-engineered natural killer cell conjugates for solid tumor adoptive immunotherapy. Small. 2021;17:e2103463.
34. Gu Q, Li J, Chen Z, Zhang J, Shen H, Miao X, et al. Expression and prognostic significance of PD-L2 in diffuse large B-cell lymphoma. Front Oncol. 2021;11:664032.
35. Riera-Domingo C, Audige A, Granja S, Cheng WC, Ho PC, Baltazar F, et al. Immunity, hypoxia, and metabolism-the menage a trois of cancer: implications for immunotherapy. Physiol Rev. 2020;100:1–102.
36. Miettinen M. Keratin 20: immunohistochemical marker for gastrointestinal, urothelial, and Merkel cell carcinomas. Mod Pathol. 1995;8:384–8.
37. Tot T. Cytokeratins 20 and 7 as biomarkers: usefulness in discriminating primary from metastatic adenocarcinoma. Eur J Cancer. 2002;38:758–63.
38. Zhou J, Fan X, Chen N, Zhou F, Dong J, Nie Y, et al. Identification of CEACAM5 as a biomarker for prewarning and prognosis in gastric cancer. J Histochem Cytochem. 2015;63:922–30.

39. Wang W, Seeruttun SR, Fang C, Chen J, Li Y, Liu Z, et al. Prognostic significance of carcinoembryonic antigen staining in cancer tissues of gastric cancer patients. Ann Surg Oncol. 2016;23:1244–51.
40. Cui C, Chakraborty K, Tang XA, Schoenfelt KQ, Hoffman A, Blank A, et al. A lysosome-targeted DNA nanodevice selectively targets macrophages to attenuate tumours. Nat Nanotechnol. 2021;16:1394–402.
41. Pan T, Jin Z, Yu Z, Wu X, Chang X, Fan Z, et al. Cathepsin L promotes angiogenesis by regulating the CDP/Cux/VEGF-D pathway in human gastric cancer. Gastric Cancer. 2020;23:974–87.
42. Xia L, Jiang L, Chen Y, Zhang G, Chen L. ThPOK transcriptionally inactivates TNFRSF12A to increase the proliferation of T cells with the involvement of the NF-kB pathway. Cytokine. 2021;148:155658.
43. Li L, Wang X. Identification of gastric cancer subtypes based on pathway clustering. NPJ Precis Oncol. 2021;5:46.
44. Ren F, Zhao Q, Zhao M, Zhu S, Liu B, Bukhari I, et al. Immune infiltration profiling in gastric cancer and their clinical implications. Cancer Sci. 2021;112:3569–84.
45. Batlle E, Massague J. Transforming growth factor-beta signaling in immunity and cancer. Immunity. 2019;50:924–40.
46. Melaiu O, Lucarini V, Giovannoni R, Fruci D, Gemignani F. News on immune checkpoint inhibitors as immunotherapy strategies in adult and pediatric solid tumors. Semin Cancer Biol. 2022;79:18–43.
47. Lan Y, Moustafa M, Knoll M, Xu C, Furkel J, Lazorchak A, et al. Simultaneous targeting of TGF-beta/PD-L1 synergizes with radiotherapy by reprogramming the tumor microenvironment to overcome immune evasion. Cancer Cell. 2021;39:1388–403.e10.

AUTHOR CONTRIBUTIONS
KC, ZH and BF designed the study. KC, BL and QL performed bioinformatics analyses, proofread and visualisation. SY, BL, SS and BF performed GC samples collective and information maintenance. KC and ZH designed the wet-lab experiments. SY, BL and LG performed the wet-lab experiments. KC, SY and BL analysed and visualised the wet-lab experiment results. KC and BL performed graphic abstract. All authors discussed the results. KC, ZH and BF wrote the manuscript.

COMPETING INTERESTS
The authors declare no competing interests.

ETHICS APPROVAL AND CONSENT TO PARTICIPATE
This study was approved by the Clinical Research Ethics Committees of Affiliated Hospital of Jiangnan University and written informed consent was obtained from all the participants.

CONSENT FOR PUBLICATION
Not applicable.

ADDITIONAL INFORMATION
Supplementary information The online version contains supplementary material available at https://doi.org/10.1038/s41416-022-01936-x.

Correspondence and requests for materials should be addressed to Bojian Fei or Zhaohui Huang.

Reprints and permission information is available at http://www.nature.com/reprints