



Proteomic profiling of gastric cancer with peritoneal metastasis identifies a protein signature associated with immune microenvironment and patient outcome

Yanyan Chen¹ · Guoxin Cai² · Junjie Jiang³ · Chao He¹ · Yiran Chen¹ · Yongfeng Ding⁴ · Jun Lu¹ · Wenyi Zhao² · Yan Yang¹ · Yiqin Zhang⁵ · Guanghao Wu⁶ · Haiyong Wang¹ · Zhan Zhou² · Lisong Teng¹

Received: 15 January 2023 / Accepted: 24 February 2023 / Published online: 17 March 2023

© The Author(s) 2023

Abstract

Background Peritoneal metastasis (PM) frequently occurs in patients with gastric cancer (GC) and is a major cause of mortality. Risk stratification for PM can optimize decision making in GC treatment.

Methods A total of 25 GC patients (13 with synchronous, 6 with metachronous PM and 6 PM-free) were included in this study. Quantitative proteomics by high-depth tandem mass tags labeling and whole-exome sequencing were conducted in primary GC and PM samples. Proteomic signature and prognostic model were established by machine learning algorithms in PM and PM-free GC, then validated in two external cohorts. Tumor-infiltrating immune cells in GC were analyzed by CIBERSORT.

Results Heterogeneity between paired primary and PM samples was observed at both genomic and proteomic levels. Compared to primary GC, proteome of PM samples was enriched in RNA binding and extracellular exosomes. 641 differently expressed proteins (DEPs) between primary GC of PM group and PM-free group were screened, which were enriched in extracellular exosome and cell adhesion pathways. Subsequently, a ten-protein signature was derived based on DEPs by machine learning. This signature was significantly associated with patient prognosis in internal cohort and two external proteomic datasets of diffuse and mixed type GC. Tumor-infiltrating immune cell analysis showed that the signature was associated with immune microenvironment of GC.

Conclusions We characterized proteomic features that were informative for PM progression of GC. A protein signature associated with immune microenvironment and patient outcome was derived, and it could guide risk stratification and individualized treatment.

Keywords Gastric cancer · Peritoneal metastasis · Proteomics · Machine learning

Yanyan Chen, Guoxin Cai and Junjie Jiang contributed equally to this manuscript.

✉ Lisong Teng
lsteng@zju.edu.cn

¹ Department of Surgical Oncology, The First Affiliated Hospital, Zhejiang University School of Medicine, 79# Qingchun Road, Hangzhou 310003, China

² Institute of Drug Metabolism and Pharmaceutical Analysis and Zhejiang Provincial Key Laboratory of Anti-Cancer Drug Research, College of Pharmaceutical Sciences, Zhejiang University, Hangzhou, China

³ Department of Gastroenterology, Affiliated Hangzhou First People's Hospital, Zhejiang University School of Medicine, Hangzhou, China

⁴ Department of Medical Oncology, The First Affiliated Hospital, Zhejiang University School of Medicine, Hangzhou, China

⁵ Department of Informatics, The First Affiliated Hospital of Wenzhou Medical University, Wenzhou, China

⁶ School of Clinical Medicine, Hangzhou Normal University Medical College, Hangzhou, China

Background

Gastric cancer (GC) is the fifth most commonly diagnosed malignancy and the fourth leading cause of cancer-related mortality worldwide [1]. Peritoneal metastasis (PM) of GC is an aggressive disease with poor prognosis, having limited response to palliative chemotherapy [2]. PM is occasionally found at the time of initial GC diagnosis or during intended radical surgery (synchronous PM). Even after radical surgery, PM occurs as peritoneal recurrence (metachronous PM), accounting for up to 50% of all recurrences [3, 4]. Emerging intraperitoneal treatment could provide survival benefit in locally advanced GC or GC patients with PM [5]. Nevertheless, considering the potential complications of these treatments, more precise evaluation of PM risk in GC patients is crucial to identify those who would benefit from intraperitoneal treatment, optimizing decision making to balance benefits and overtreatment.

Next-generation sequencing enables the high-throughput and systematic analysis of genetic alterations, offering novel insights into molecular basis of oncogenesis and heterogeneity. Importantly, The Cancer Genome Atlas (TCGA) classified GC as four subtypes: Epstein-Barr virus (EBV), microsatellite instability (MSI), chromosomal instability (CIN), and genomically stable (GS) [6]. Our team previously purposed a molecular classification that differentiates GC subtypes associated with prognosis and metastasis patterns towards liver metastasis or PM [7]. However, genomic heterogeneity between primary and metastatic tumor in GC could be extensive [8]. Therefore, to investigate the molecular background of PM propensity and to identify patients with high PM risk, study including primary lesions and corresponding peritoneal metastases remain required.

As proteins are regarded as “executors of life”, proteomic studies provided insight into cancer mechanisms and potential clinical implications. Proteomic profiling of diffuse GC identified three subtypes with different prognosis [9]. Li et al. uncovered proteomic signatures for progression of gastric lesions and risk of early GC [10]. Recent proteomic study of GC undergoing chemotherapy or targeted therapy identified proteomic features to predict therapeutic response [11]. Nevertheless, proteomic characteristics of GC with PM and their clinical relevance have not been extensively studied.

In recent years, cancer immunotherapy using immune checkpoint inhibitors (ICIs) has emerged as promising treatment for specific subgroups of GC [12, 13]. However, the clinical efficacy of ICIs against PM of GC is still unclear. Previous studies have suggested that the response to anti-PD-1 therapy tended to be less prominent in GC with PM [14, 15]. Since the tumor immune

microenvironment (TIM) is highly associated with immunotherapy response and prognosis [16], investigation of TIM of GC with PM could potentially reveal relevance to ICI response.

In this study, we sought to investigate the molecular features of GC with PM based on in-depth proteome profiling and whole-exome sequencing of GC. By machine learning, we identified a proteomic signature associated with PM and subsequently generated a prognostic signature, which was validated in external proteomic datasets. This proteomic signature was also associated with TIM and potentially predictive for immunotherapy response. Our findings may have translational significance for guiding decision making in GC management.

Methods

Patient samples

A total of 25 GC patients who underwent gastrectomy at the First Affiliated Hospital of Zhejiang University School of Medicine between January 2018 and January 2022 were retrospectively enrolled. 13 patients with synchronous PM underwent palliative surgery for GI bleeding or obstruction. 12 patients underwent curative surgery, among whom 6 developed peritoneal recurrence (metachronous) as the first metastatic site, 6 did not develop PM within 24 months (non-PM group). Patients received standard adjuvant chemotherapy or perioperative systemic treatment (with or without immunotherapy). Follow-up data were obtained by phone and the out-patient clinical database (last follow-up, November 2022). Patient-written consents were obtained from study participants, informing the use for proteomic and genomic profiling and publication. Fresh-frozen, FFPE samples and malignant ascites were collected in accordance with ethical guidelines in the First Affiliated Hospital, Zhejiang University School of Medicine.

DNA extraction and whole-exome sequencing

For fresh-frozen samples, QIAamp DNA Mini Kit (Qiagen) was used to isolate genomic DNA from tumor tissues and matched normal mucosa according to the manufacturer's instructions. For FFPE tumor tissues, DNA was extracted using QIAamp DNA FFPE Tissue Kit (Qiagen). DNA degradation and contamination were monitored on 1% agarose gels. Qubit® DNA Assay Kit in Qubit® 2.0 Fluorometer (Invitrogen, USA) was used to quantify DNA concentration. For malignant ascites, tumor content was centrifuged for 20 min at 2000g and sediment tumor cells were used for DNA and protein extraction.

Whole-exome library construction was generated using the Agilent SureSelect Human All Exon V6 Kit (Agilent Technologies, Santa Clara, CA, USA). The index-coded samples were clustered on a cBot Cluster Generation System using HiSeq PE Cluster Kit (Illumina). The DNA libraries were sequenced on Illumina HiSeq platform (Illumina, San Diego, California, USA) and 150 bp paired-end reads were generated. We first conducted data quality control and then performed all downstream bioinformatics analyses based on the high-quality clean data. The paired-end clean reads were aligned to the Human Genome Reference Consortium build 37 (GRCh37) using BWA v.0.7.8 [17]. Mapped reads were then de-duplicated using Sambamba tools (v0.4.7) [18].

Identification of somatic single-nucleotide variants (SNVs) was conducted by muTect (v1.1.4), and the somatic InDels were detected by Strelka (v1.0.13). ANNOVAR (ANNOVAR_2015Mar22) was used to annotate variant call format files [19].

Protein extraction and quantitative proteome profiling

The protein extraction was performed according to Zhang et al. [20]. Samples were minced and lysed in lysis buffer (8 M Urea, 100 mM Tris Hydrochloride, pH 8.0) containing protease and phosphatase Inhibitors (Thermo Scientific) followed by sonication. The lysate was centrifuged and the supernatant was collected. Extracts from sample was reduced with 10 mM dithiothreitol at 56 °C for 60 min and alkylated with 10 mM iodoacetamide at room temperature in the dark for additional 60 min. The samples were digested with trypsin. The digested peptides were desalinated using C18 column (50% acetonitrile and 0.1% formic acid) and resolved in 100 mM triethylamine buffer. Finally, the samples were labeled using the TMT[®] Mass Tagging Kits and Reagents (Thermo scientific, USA) following the manufacturer's instructions.

Mass Spectrometry (MS) detection was carried out as previously described [20]. Digested samples were analyzed on Q Exactive HF-X Hybrid Quadrupole-Orbitrap Mass Spectrometer (Thermo Scientific, USA) coupled with a high-performance liquid chromatography system (EASY nLC 1200, Thermo Fisher). MS raw files generated by LC-MS/MS were searched against the NCBI human Refseq protein database using Proteome Discoverer 2.4 (PD2.4, Thermo) software. Then peptide spectrum matches with reliability > 99%, or proteins containing at least one unique peptide, with FDR < 1% were identified.

Copy number analysis

Somatic copy number variations (SCNVs) were identified using CNVkit [21]. Then GISTIC 2.0 (v 2.0.22) [22] was

used to identify the genome regions with significant alterations and screen out the recurrent CNV region.

Analysis of clonal architecture

The R package SciClone [23] was used to infer the clonal and subclonal architecture of somatic mutations by analyzing the variant allele frequencies in an individual sample using the Bayesian binomial mixture model.

Identification of differentially expressed proteins

Analysis of differently expressed proteins (DEPs) between tumor and normal was conducted by (a) R package samr (v3.0) with an FDR threshold at 0.05 and a fold change (FC) threshold at 1.5. The results were run under 100 times permutations resampling and visualized by VolcanoR [24]; (b) Referring to GFold [25], the DEPs of each sample pair were derived by the following steps. First, for each sample pair, log foldchanges of each gene were fitted to t-distribution. Two-tail test was performed to get DEPs with a *P* value threshold of 0.05. The overlaps of (a) and (b) were regarded as the final DEPs.

Identification of PM-associated proteins

Our previously developed tool MATTE (v1.2.3) [26] was used to identify the differentiated module between PM samples and without-PM samples based on DEPs. After clustering, all differentially expressed modules were compared by their signal-to-noises (SNRs). The module with the highest SNR was thought as the PM-associated module. External transcriptome data [27] was used for filtering proteins. The classifier for PM samples identification was built by Xgboost (v1.7) [28] and simplified by scikit-learn (v1.2) [29]. Specifically, we first fitted GB-tree classifier by data with previously filtered proteins. By taking the area under the curve (AUC) in fivefold cross validation as the metric, 32 proteins got best score. To further simplify the model, we listed all combinations of proteins and scored them as previously. Each protein's frequency in the top 500 protein combinations of all was ranked. Ten proteins with the highest frequency in the combinations were selected for the final model.

Establishment of prognostic model

For previously selected ten proteins, a GB-linear classifier was fitted in our cohort. Then weight of each protein was extracted. The sum of weighted protein expression was calculated as PM risk score. Subjects were stratified into high, moderate and low-risk levels by PM risk score.

Immune cells infiltration enumerations

Estimation of immune cell types infiltration level was based on sample-level normalization proteome data. Then, the estimation was performed by TIMER web server [30] and the result of CIBERSORT algorithm [31] was used for analysis.

Statistical analysis

The statistical analysis was performed by SPSS 21.0 software. Two continuous variables were compared using Student's *t* test. Survival data was analyzed by Kaplan–Meier curves with log-rank test. Univariate and multivariate Cox regression analysis were performed to calculate the hazard ratio and 95% confidence interval. A *P* value of <0.05 was considered significant.

Results

Patient characteristics, proteome profiling and whole exome sequencing

We enrolled 19 GC patients who developed synchronous ($n = 13$) or metachronous ($n = 6$) PM without other distant metastasis, and 6 GC patients with stage IIIB or IIIC disease who did not develop peritoneal recurrence for at least 24 months after curative surgery (PM-free group). Primary tumor tissue (T), peritoneal metastases (M), and their matching adjacent normal tissues (N) were collected for proteome profiling and/or whole-exome sequencing (Fig. 1A). The overall survival showed significant difference between PM and PM-free groups (Fig. 1B). The detailed information is shown in Table S1. By high-depth tandem mass tags (TMT) labeling for quantitative proteomics, a total of 20 pairs of tumor and normal tissues were measured and the results showed good consistency in proteome identification and quantification (Fig. S1), resulting in the identification of 9852 proteins. To increase reliability, we selected 6799 proteins that were detected with at least 2 unique peptides and normalization to fraction of total (FOT) $> 10^{-5}$. Z-score normalization (\log_2 of relative abundance scaled by subjects' SD) was performed for comparative analysis.

To obtain genetic background of our cohort, we conducted whole-exome sequencing. The mean coverage of sequencing was 263-fold in tumor and 122-fold in normal samples. A total of 3439 somatic non-synonymous mutations were detected in all sequenced cases. *TP53* (63%), *SPTA1* (21%), *PREX2* (16%), *ZFH4* (21%), *LRP1B* (21%) were identified as significantly mutated genes (Fig. 1C). All GC cases were identified as microsatellite stable.

Genomic and proteomic features in PM in comparison with primary GC

To investigate the heterogeneity between primary GC and PM, we compared multiple genomic features in 9 pairs of samples. All paired samples shared some alterations, confirming that they had the same tumor origin. The proportions of somatic mutations shared with paired PM ranged from 9 to 62% in primary GC (Fig. S2). Mutations of *TP53* and *ZFH4* amplifications were generally shared between primary tumor and PM, while *CSMD3* mutations, CNVs were frequently discrepant, such as potentially targetable amplifications of *CCNE1*, *MYC* and *KRAS* (Fig. 1D). Notably, compared with primary tumor, PM samples harbored more recurrent CNV (Fig. 1E). Mutational signature and neoantigen were generally concordant in paired primary and PM samples (Figs. S3–4). The number of subclones was frequently discrepant in PM and primary GC, reflecting different intratumor heterogeneity. Notably, number of subclones in PM increased in 4 out of 5 patients with metachronous PM (Fig. 1F). The above results confirmed high genomic heterogeneity between primary GC and PM samples.

Four PM samples derived from malignant ascites were analyzed at proteome level (two samples were collected from Meta-3 at different timepoints). The protein abundance between paired primary GC/PM tissue showed relatively low correlation (Fig. 2A). Compared with primary GC samples, we identified 118 proteins enriched in PM samples (Table S2). A large number of these proteins distributed in extracellular region, including exosome and secretory granule (Fig. 2B). Pathway analysis revealed that these proteins were enriched in RNA binding, polysomal ribosome and extracellular exosome (Fig. 2C). We further focused on proteins that showed the same expression trend in primary and PM samples. PLG, SP1 and FABP3 were consistently up- or downregulated in two pairs of primary GC/PM, and there were multiple altered pathways overlapped across cases, including focal adhesion and complement pathway (Fig. 2D). Specifically, Meta-3M2 (24 months after radical surgery, during second-line treatment) had more altered proteins associated with signaling pathways than Meta-3M1 (10 months after radical surgery, first-line chemotherapy completed), suggesting that more mechanical changes might be induced through tumor progression and additional treatment.

Characteristic alterations in proteome of GC associated with PM propensity

As illustrated in Fig. 3A, to screen the candidate proteins associated with PM propensity in primary GC, we first selected proteins differentially expressed between tumor and normal tissues ($n = 20$), and then compared their

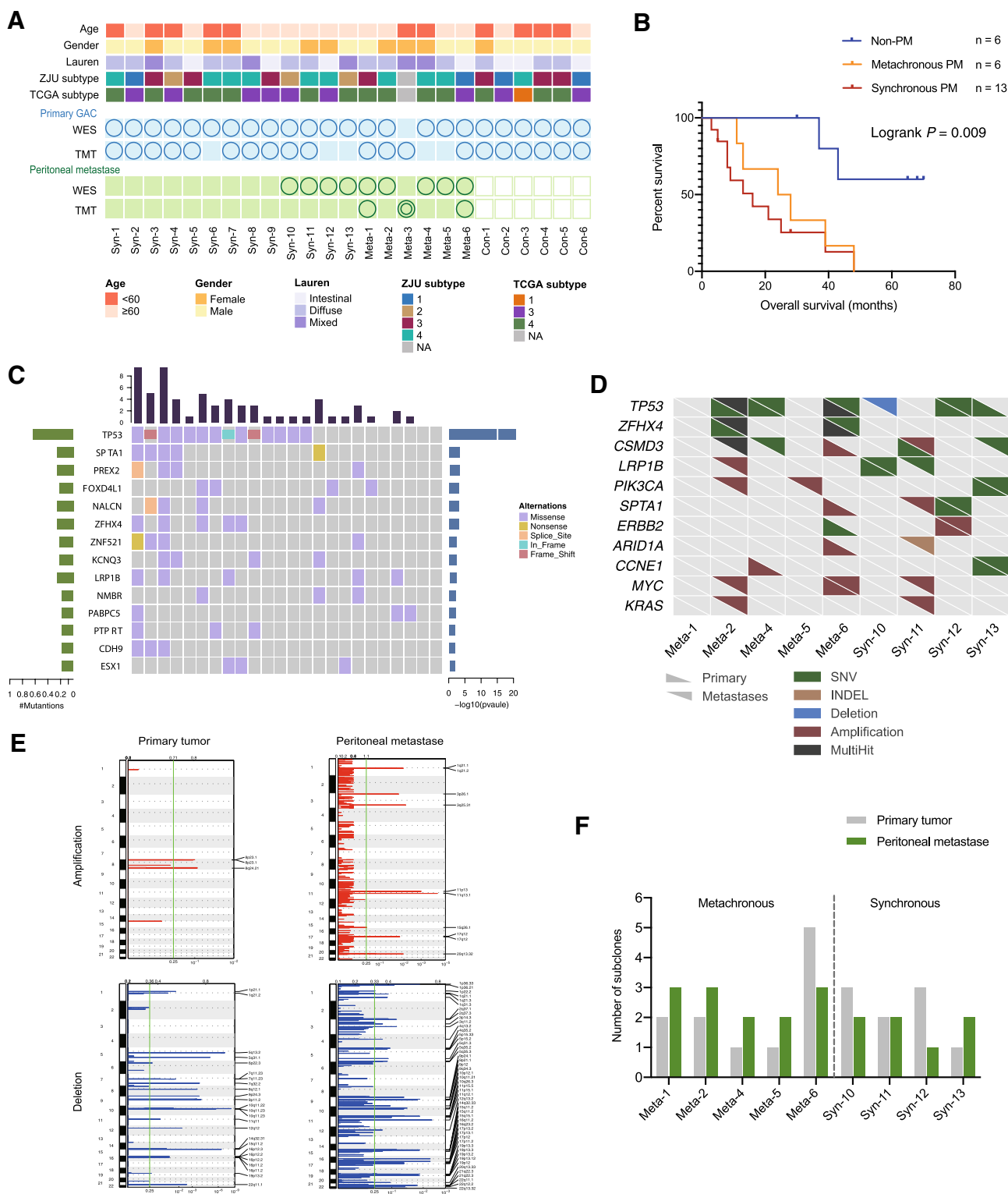


Fig. 1 Genomic background of primary GC and paired PM samples. **A** Clinicopathological information and molecular classifications of GC cohort; Circles represent available data. (Dual circles represent two samples collected at different timepoints), **B** Survival analysis of three patient groups, **C** Landscape of somatic mutations of the cohort. Genes with variants in at least 3 patients were depicted. Bars on top and to the right of the graph show the number of non-synonymous

mutations in each subject and gene, respectively, **D** Comparison of key somatic mutations and CNV between paired primary GC and PM, **E** Somatic CNV of primary GC and PM samples, **F** Number of subclones of each paired primary GC/PM. *Syn* synchronous metastasis, *Meta* metachronous metastasis, *Con* control group, *WES* whole-exome sequencing, *TMT* tandem mass tags

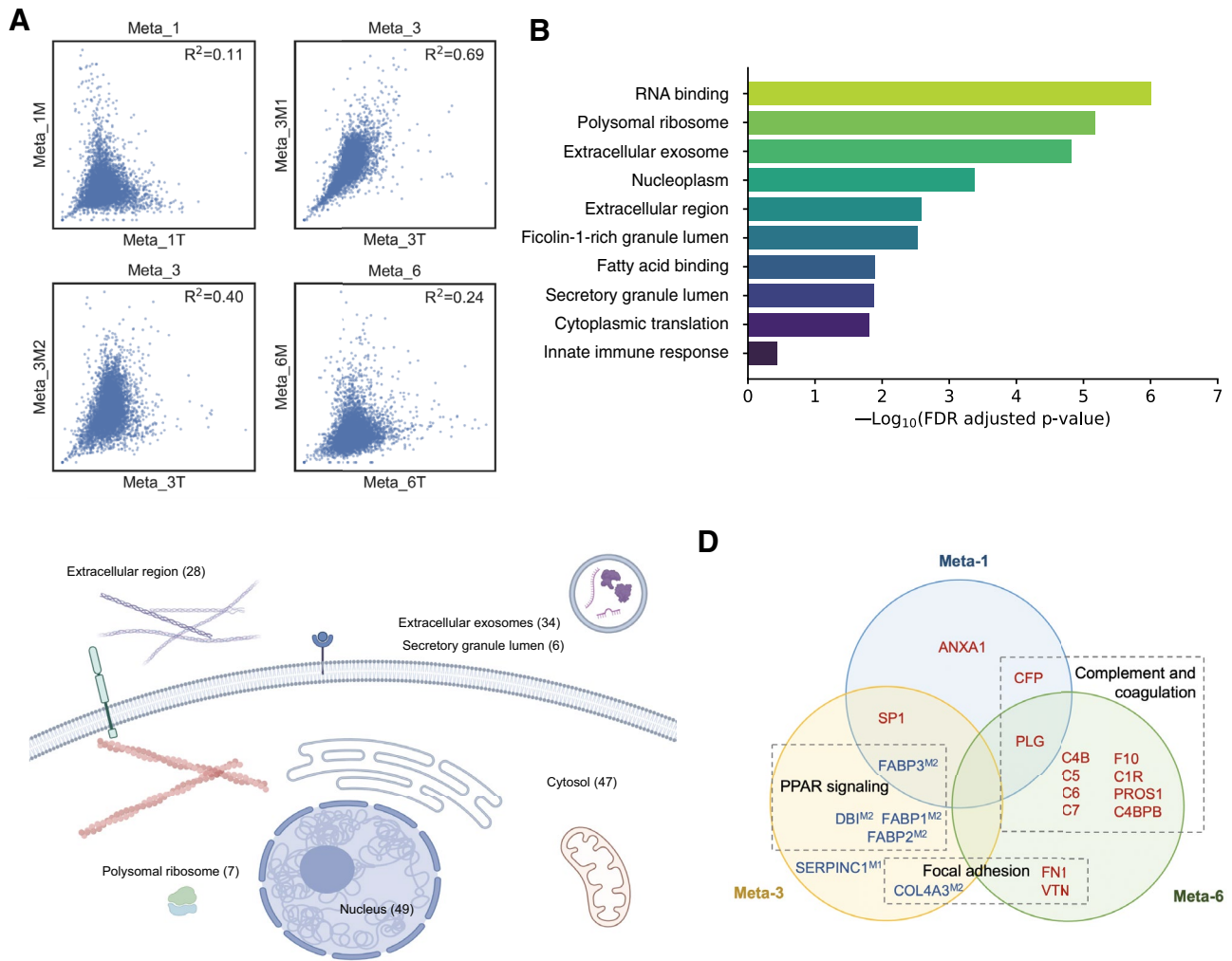


Fig. 2 Proteomic features of paired primary GC and PM. **A** Scatterplot showing correlation of protein abundance in paired primary GC/PM samples, **B** Top ranked pathways that are significantly enriched in PM as compared with primary GC, **C** Subcellular distribution of proteins

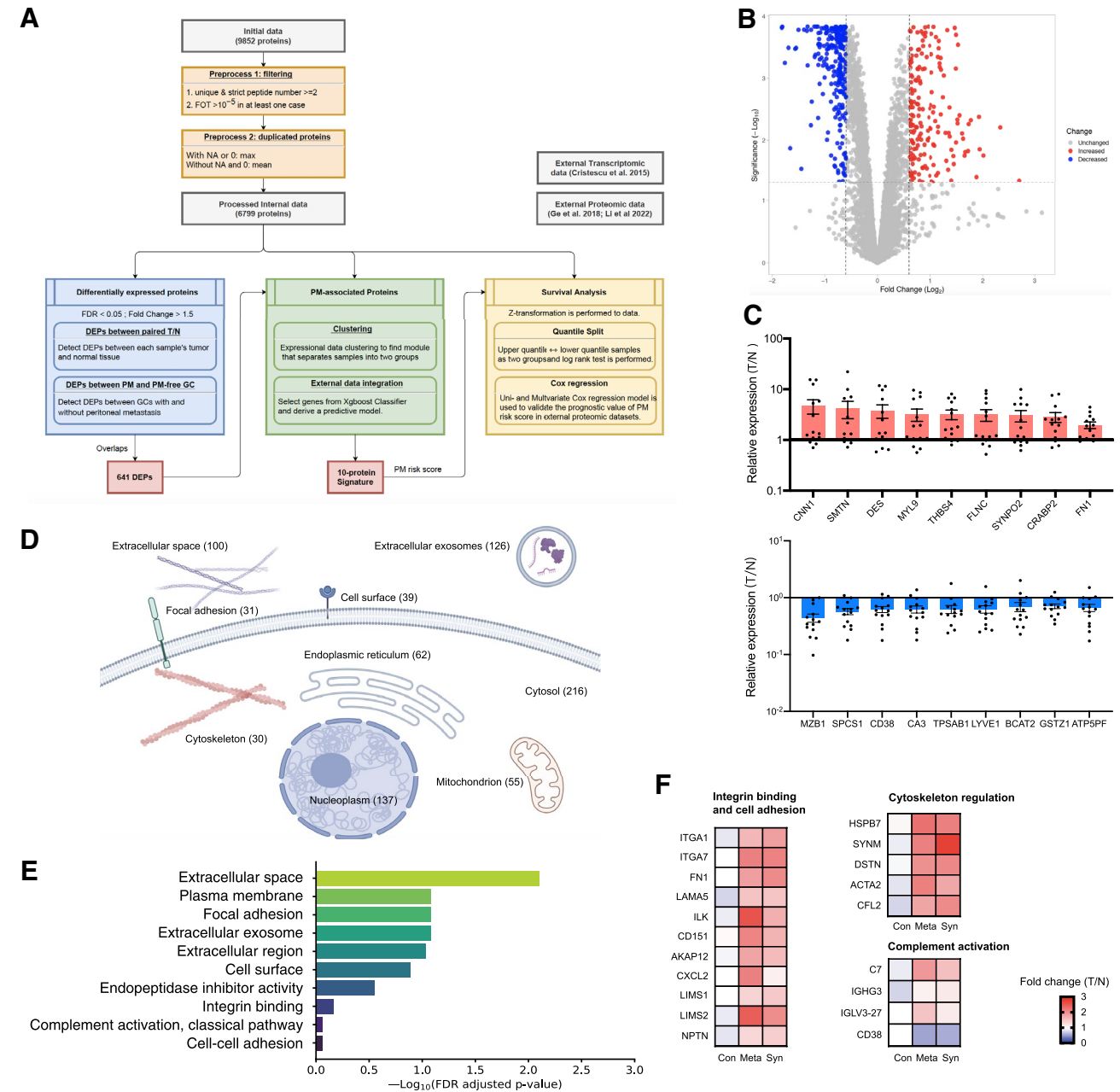
enriched in PM annotated with Gene Ontology, **D** Venn diagram presenting altered proteins (up- or downregulated in both primary and PM samples) in each subject and relevant pathways (for Meta-3M, proteins altered in superscript samples)

expression between GC with PM ($n=14$) and without PM ($n=6$), resulting in 641 differentially expressed proteins (DEPs, Fig. 3B). In addition, we analyzed significantly altered proteins in GC with PM ($n=14$). Upregulated proteins included CNN1, SMTN, DES and FN1, and down-regulated proteins included MZB1, SPCS1, CD38 and CA3 (Fig. 3C). GO analysis of the 641 DEPs revealed that extracellular exosome and extracellular space were two most significant annotations ranked by FDR (Fig. 3D, Table S3), and other significant annotations included focal adhesion and adherens junction, suggesting that GC with PM propensity were enriched in expressional changes of tumor microenvironment components. KEGG analysis revealed that these DEPs were significantly enriched in integrin binding/focal adhesion, cytoskeleton regulation

and complement activation pathways (Fig. 3E and F, Table S3).

Identification and validation of a proteomic signature associated with PM propensity and prognosis

We then attempted to obtain a protein signature associated with PM propensity in GC. First, based on our previously developed clustering method MATTE, we conducted cluster analysis that derive a PM-related module, and selected the proteins in the module with the highest signal-to-noise (Fig. 4A, $n=20$). Subsequently, an external transcriptome-based expressional data [27] was used for further screening and verification. By performing dimension reduction of expression data, PM and PM-free samples were roughly



discriminated (Fig. 4B). 10 proteins (DUOXA2, ITGA7, LIMS1, MSR3, PLCB1, RAB6B, SEMA3C, SMTN, TADA1, TBC1D14) from the PM-related module were selected to predict whether PM occurred (Fig. 4C, the selection method described in Methods). The results showed that under the fivefold cross validation, the AUC of the average

ROC curve reached 0.83 in the cohort (Fig. 4D), suggesting high accuracy of PM prediction.

The prognostic value of the 10-protein signature was further evaluated. By machine learning algorithms, a PM risk score was generated in internal cohort (Fig. S5, the weights of each protein in Tables S4). Next, we validated the model in diffuse and mixed type GC cases from two

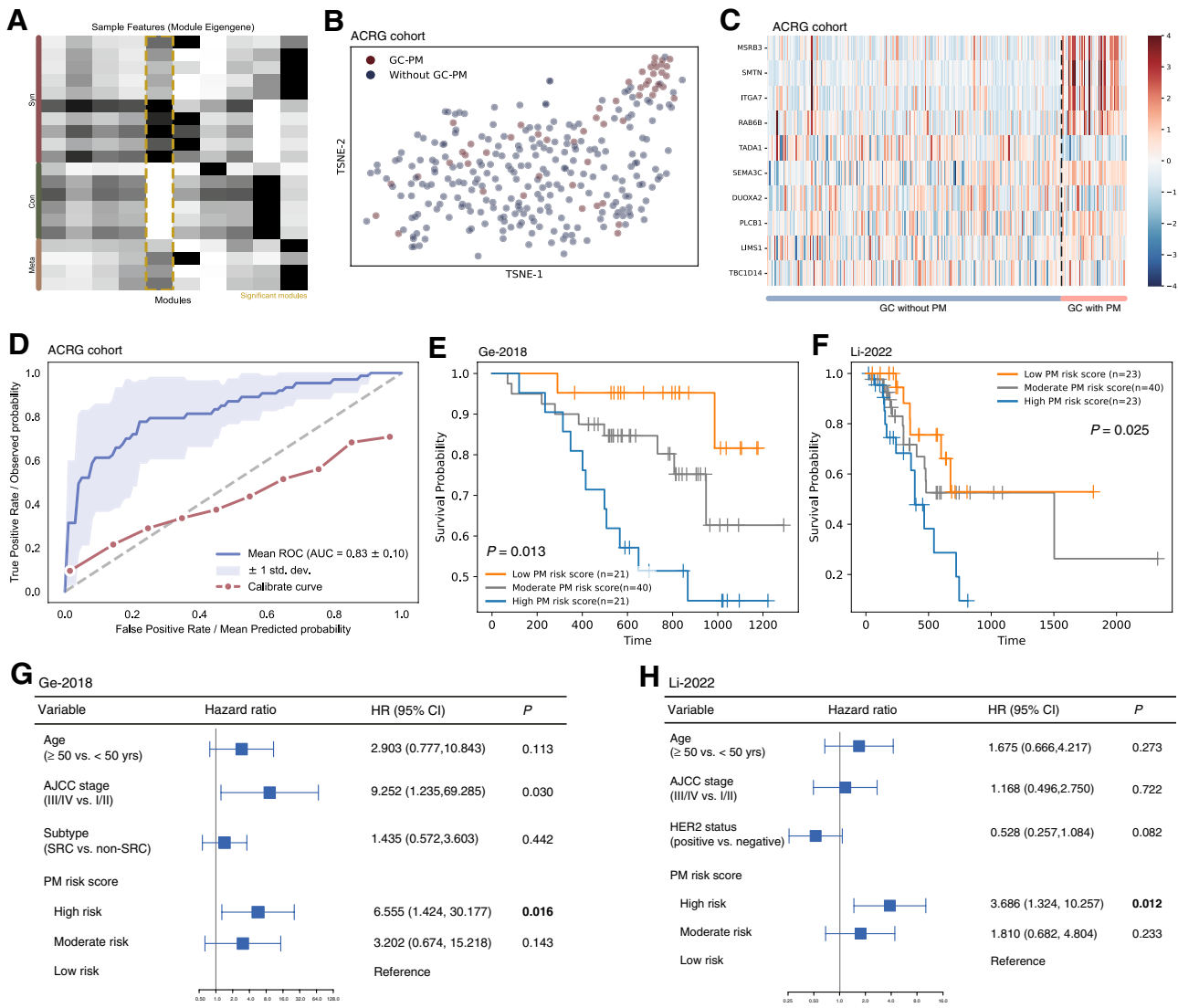


Fig. 4 Establishment of PM-associated proteomic signature and external validation. **A** MATTE clustering of PM-associated module, **B** Heatmap depicting expression of ten proteins in GC with or without PM in ACRG cohort, **C** Dimension reduction of expression data

in ACRG cohort; each dot represents one sample, **D** ROC curve of prediction model using Xgboost classifier, **E–F** Kaplan–Meier plot of Ge-2018 (**E**) and Li-2022 (**F**) cohort, **G–H** Forrest plot of multivariate Cox regression of Ge-2018 (**G**) and Li-2022 (**H**) cohort

external cohorts, Ge-2018 [9] ($n = 82$) and Li-2022 [11] cohort ($n = 86$). Patients were divided as high risk (highest quartile), moderate-risk and low risk (lowest quartile) groups according to the PM risk score, and those in high-risk group had significantly shorter overall survival than low-risk group in both cohorts ($P = 0.013$ in Ge-2018 and $P = 0.025$ in Li-2022, Fig. 4E and F). Multivariate Cox regression showed that after adjusting for clinical features such as age and AJCC staging, high PM risk score still significantly correlated to poor survival (HR = 6.555, 95% CI 1.424–30.177, $P = 0.016$ in Ge-2018 and HR = 3.686, 95% CI 1.324–10.257, $P = 0.012$ in Li-2022) (Fig. 4G and H), suggesting this signature was an independent prognosticator for overall survival. Interestingly, this prognostic

significance was absent in the whole Li-2022 cohort including Lauren intestinal subtype (Fig. S6, $n = 206$). These findings emphasized that the 10-protein signature was predictive of PM-associated prognosis.

10-protein signature is associated with TIM and treatment response

To explore the relevance between our prognostic model and TIM, we compared the proteome-based immune cell infiltration of GC samples of high-risk group and low-risk group by CIBERSORT analysis (Fig. 5A). In 22 immune cell fractions, plasma B cell, CD4+ naïve T cell and eosinophil were significantly more abundant in low-risk group, whereas M2

macrophage, regulatory T cell and resting CD4+ memory T cell were more abundant in high-risk group (Fig. 5B). Other cell fractions such as CD8+ T cell was not significantly different between two groups. These results indicated that PM risk score was associated with TIM and might be predictive for immunotherapy response.

The clinical relevance of PM risk score was shown in representative cases (Fig. 5C). Case Meta-1, 69 years old, male, evaluated as high risk, who received radical surgery and the pathological staging was pT3N3aM0, then received 8 cycles of SOX (oxaliplatin plus S-1) adjuvant chemotherapy. 7 months after surgery, tumor relapsed as large amount of malignant ascites, and responded poorly to the second-line treatment including chemotherapy and PD-1 inhibitor. The overall survival was 11 months. Case Syn-5, 70 years old, male, evaluated as low risk, whose clinical stage was cT4N2M1 (with PM) at baseline, received first-line SOX chemotherapy plus PD-1 inhibitor. 4 cycles later, the patient developed GI bleeding, and the radiological evaluation of tumor response was partial response. Then the patient received palliative surgery, in which peritoneal metastases were observed, and the pathological staging was ypT4bN2M1. After surgery the patient received another 4 cycles of SOX plus ICI and maintained stable disease, until the tumor progressed 14 months after surgery, presenting as minor pelvic ascites. The overall survival was 25 months. These cases suggested that the PM risk score might not only be associated with aggressiveness of GC, but also with treatment response.

Discussion

In the present study, we comprehensively confirmed genomic and proteomic heterogeneity between primary GC and PM, analyzed the distinct proteomic characteristics associated with PM progression, and then established a proteomic signature associated with the risk of PM and prognosis. This signature was also potentially associated with TIM and immunotherapy response in GC patients.

Currently, PM risk was estimated mainly according to TNM staging, especially T stage (T3-T4 were usually considered as with high risk), and GC of Lauren diffuse type was reported to have a significant tendency of developing PM [32]. However, due to the high heterogeneity within GC, clinicopathological features might not fully depict its aggressiveness and metastasis pattern. Our proteome-based prediction model would optimize the risk stratification of PM, as a tool for improving clinical management for GC patients.

In recent years, there are evolving multi-omics studies for PM of GC. Wang et al. described the genomic and immune landscape of peritoneal carcinomatosis of GC based on 43 PM cases, revealing two subtypes with discriminating

response rates to chemotherapy [33]. A subsequent study used single-cell transcriptome profiling of peritoneal carcinomatosis to reveal two subtypes with different intra-tumoral heterogeneity [34]. A multi-omics analysis of malignant ascitic fluid samples and tumor cell lines from 98 patients, stratified ascites-disseminated GC into two distinct molecular subtypes: “non-EMT” subtype displaying active super enhancers (SEs), and “EMT” subtype bearing TGF- β pathway activation and high expression of TEAD-1 [35]. Unlike the above studies, our study mainly focused on proteomics of primary GC with PM, attempting to reveal key proteomic features associated with PM propensity and PM-related aggressiveness.

Consistent with previous study [8], extensive genomic heterogeneity in somatic mutations, CNVs and subclone structure was observed between primary GC and paired PM, which could be a barrier to precision medicine. The proteomic difference between paired primary GC/PM was partly due to different extracellular components such as stromal content and exosomes. Proteins enriched in PM tissues were involved in RNA binding and polysomal ribosome, suggesting more active protein synthesis in metastases.

By comparing proteomic profiles between different sample groups, we observed distinct features associated with PM progression. Enriched pathways in PM samples (compared with primary GC) included RNA binding, extracellular exosome and focal adhesion. Meanwhile, DEPs between primary GC of PM group and non-PM group were enriched in extracellular exosome, focal adhesion, cytoskeleton regulation and complement activation pathways. The overlapped pathways of the two analyses suggested key biological processes of PM. Exosomes are biologically functional extracellular vesicles comprising active factors. They can mediate metastasis of GC by promoting EMT, cancer-associated fibroblast formation, pre-metastatic niche formation and immunosuppression [36]. Focal adhesion are anchoring units and interacts between the cell-extracellular matrix and cell cytoskeleton, associated with multiple oncogenic processes such as cell proliferation, cell cycle regulation migration and chemoresistance in cancer cells [37].

The Asian Cancer Research Group (ACRG) categorizes GC into four subtypes using expression data. The MSS/EMT subtype, having the worst prognosis, was characterized by a gene expression signature correlated with EMT and a distinct tendency to develop PM [27]. It was the first large-scale transcriptional study that contains a considerably large number of subjects with PM. A large multi-omic study of early onset GC revealed that genes with significant survival differences showed stronger mRNA-protein correlations than genes with non-significant survival differences [38]. Therefore, although the data of ACRG was transcriptome-based, it is considerable to combine this dataset with our proteome-based data to screen key proteins with significant

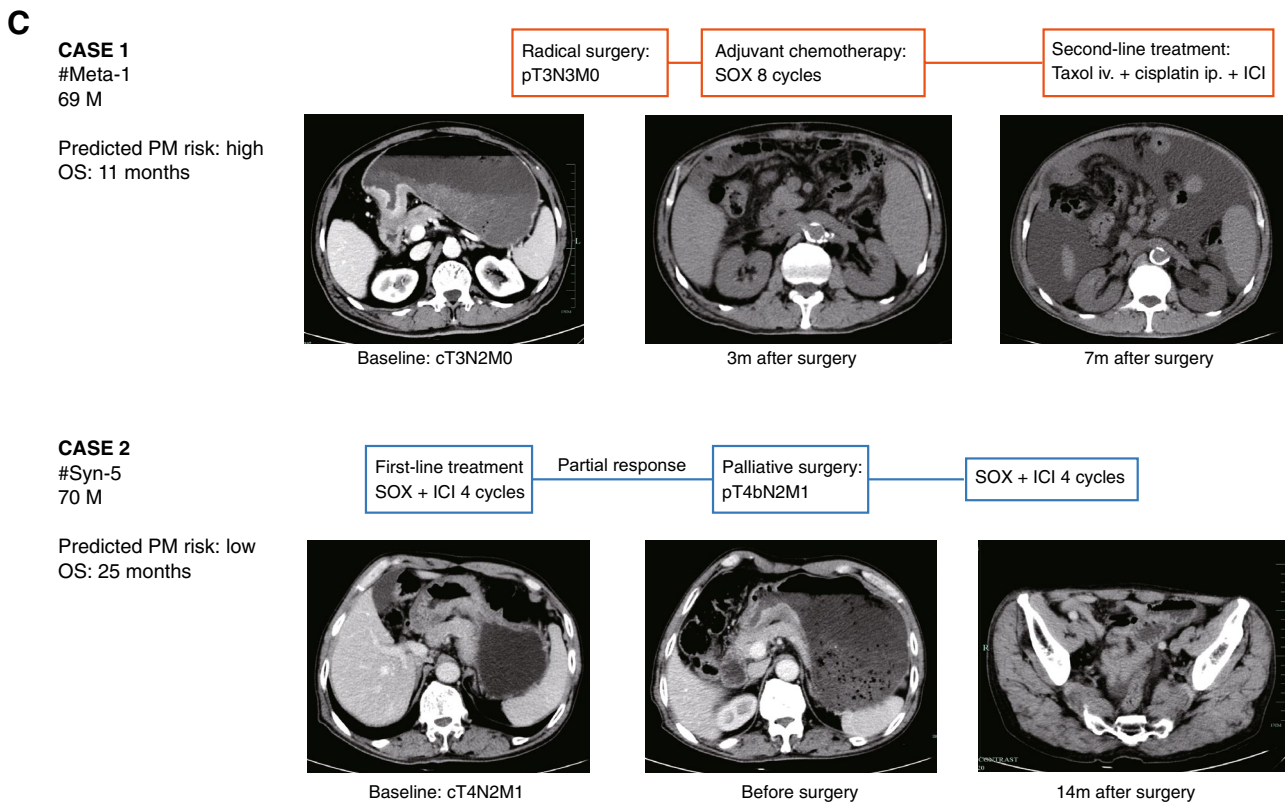
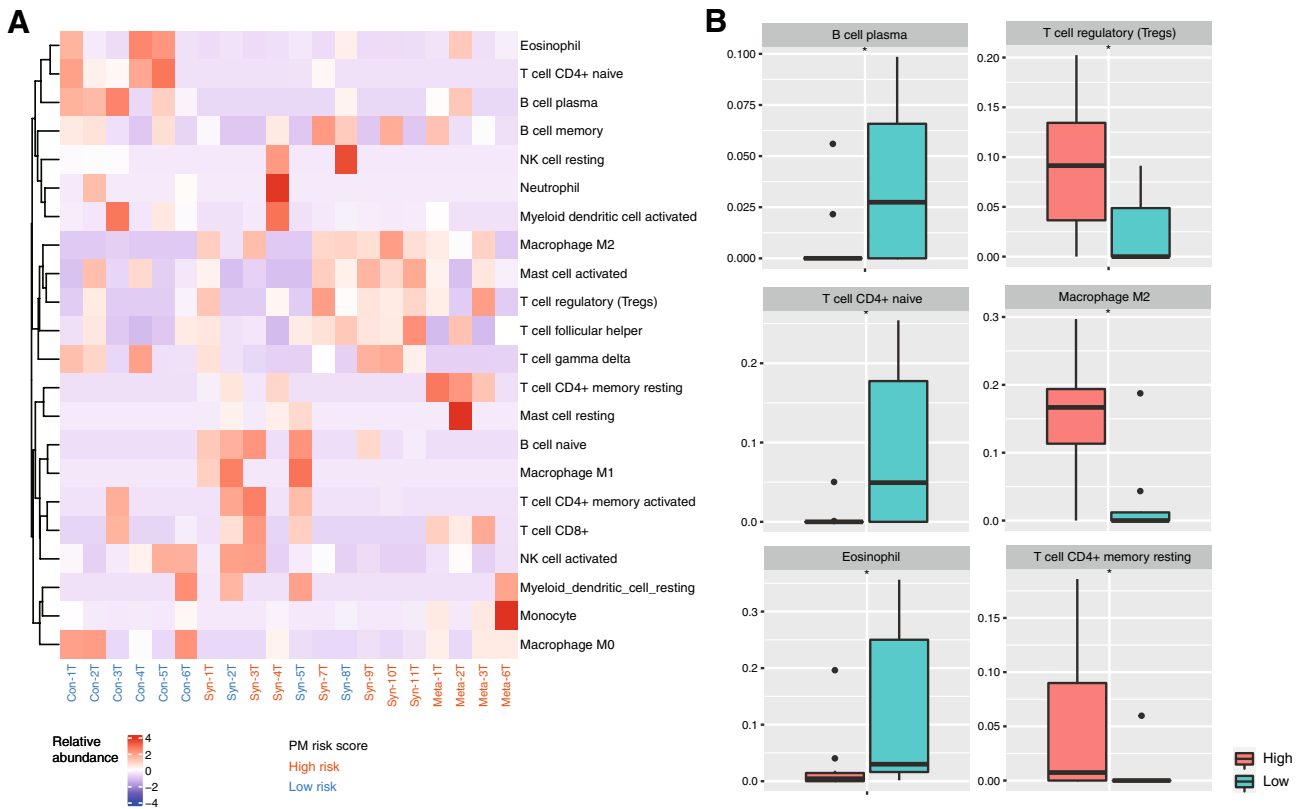


Fig. 5 PM risk score is associated with tumor immune microenvironment and treatment response. **A** Heatmap of immune cell infiltration by CIBERSORT analysis. **B** Immune cell fractions in GC with high and low PM risk score. **C** Treatment and outcome of representative GC cases with high or low PM risk. SOX, oxaliplatin plus S-1 chemotherapy; ICI, immune checkpoint inhibitor; OS, overall survival; * $P < 0.05$; ns, not significant

clinical implications, as these proteins were likely to have high mRNA-protein correlation. By this combination, we screened proteins associated with PM propensity, and the internal validation resulted in high accuracy.

Most proteins in our signature were involved in cancer-related mechanisms. For instance, ITGA7 was associated with cancer stemness in esophageal squamous cell carcinoma and correlates with poor prognosis in hepatocellular carcinoma [39, 40]. DUOX2 was maturation factor of an oxidative protein DUOX2, which promotes invasion and metastasis of colorectal cancer [41, 42]. MSRB3 was reported to promote cancer by regulating genome stability and endoplasmic reticulum stress [43]. Upregulation of TADA1 was reported to promote lung squamous cell carcinoma progression [44]. RAB6B regulates intracellular membrane trafficking pathways, and its silence inhibited AKT/JNK signaling in GC [45, 46]. SEMA3C was reported to be an oncogene in multiple tumors [47]. TBC1D14 was found to inhibit lymph node metastasis by regulating autophagy in head and neck squamous cell carcinoma [48]. The biological functions of these proteins in PM progression of GC are promising aspects in future research.

Our prognostic model was derived from a PM-associated protein signature by machine learning, intending to stratify GC patients by PM-related survival. It is well established that Lauren subtype was significantly correlated to metastasis or recurrence pattern [32]. Therefore, we validated our prognostic model using two independent external proteomic datasets of diffuse/mixed GC, in which the OS could reflect PM risk to some extent. Ge-2018 was a proteomic study of 84 diffuse GC, in which 8 patients had available report of first recurrence site of peritoneal seeding. Li-2022 was a proteomic study including 206 GC patients (86 with diffuse/mixed type GC) undergoing chemotherapy or HER2 targeted therapy. Although the two cohorts had discrepant inclusion criteria, validation of our prognostic model yielded consistent results, suggesting the robustness of this model. The prognostic value of this model was absent in intestinal GC or the whole cohort, because intestinal GC is more likely to develop hematogenous metastasis, which highly affects OS. The predictive value of this model for PM-related survival in intestinal type GC still requires further cohort validation.

Tumor-infiltrating immune cells are correlated with the progression of cancer and patients' outcome, and also affect tumor responses to immunotherapy [49]. The infiltration of macrophages M2 and Tregs was significantly negatively

correlated with prognosis of colorectal cancer patients [50]. Infiltration of plasma cells as effectors of humoral immune response, were correlated with longer survival in GC [51] and breast cancer [52]. Activated CD4 + memory cells, rather than resting cells, play important role in anti-tumor immunity. Taken together, our findings revealed that GC with high PM risk tended to harbor a more immunosuppressive microenvironment, and might respond poorly to immunotherapy. Although preoperative chemotherapy could remodel TIM [53], no significant difference was observed between GC with and without preoperative chemotherapy in our cohort (Fig. S7). The two representative cases both received combined chemotherapy and immunotherapy. Meta-1 was predicted as high PM risk, and was diagnosed as locally advanced stage, but peritoneal recurrence developed shortly after radical surgery, leading to a poor survival. In contrast, Syn-5, predicted as low PM risk, although first diagnosed as advanced stage, had a much longer survival. The distinct prognosis might be partly due to their different TIM and responses to immunotherapy.

Strengths of our study included proteomic profiling of both primary GC and PM samples, machine learning and modeling integrating large-scale expressional data, and a panel of proteins associated with both PM propensity and patient outcome, which was validated at protein level in two independent cohorts. We acknowledge several limitations. First, attempting to reveal key features associated with PM, our proteomic data had a small sample size, especially that of PM samples, because the clinical scenario to acquire both primary and PM samples was rare. Second, lacking required clinical information, we have not been able to obtain external proteomic data to directly validate the predictive value of our model to PM risk or PM-free survival, but only validated its prognostic value in diffuse/mixed type GC. Although PM is the most frequent metastatic pattern in diffuse/mixed type GC, future studies of large cohorts including all Lauren subtypes of GC are warranted to validate the model.

In summary, we depicted molecular features of primary GC and PM tissues by multidimensional proteomic and genomic analysis, and characterized proteomic features that were informative for the mechanisms of PM progression. A proteomic signature was derived by machine learning algorithms to predict PM propensity and prognosis in GC. The signature was associated with TIM and potentially with immunotherapy response. Our findings could guide prophylactic intraperitoneal treatment in locally advanced GC patients with high PM risk and improve individualized disease management.

Supplementary Information The online version contains supplementary material available at <https://doi.org/10.1007/s10120-023-01379-0>.

Acknowledgements The authors thank staffs from Novogene Bioinformatics Technology Co. Ltd., Beijing, China for assistance with

bioinformatics analyses. This study was supported by National Natural Science Foundation of China (No. 31971371), the Project of the Regional Diagnosis and Treatment Centre of the Health Planning Committee (No. JBZX-201903) and Science and Technology Plan Project of Jinhua (2021-4-165).

Data availability The data in this study are available from the author for correspondence upon reasonable request.

Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.

References

1. Ferlay J, Colombet M, Soerjomataram I, Mathers C, Parkin DM, Pineros M, Znaor A, Bray F. Estimating the global cancer incidence and mortality in 2018: GLOBOCAN sources and methods. *Int J Cancer*. 2019;144(8):1941–53.
2. Wei J, Yang Y, Du J, Zou Z, Zhang Y, Wu N, Wang Q, Yuan L, Liu B. Peritoneal metastasis in relation to outcome and therapeutic strategy in gastric cancer. *Transl Cancer Res*. 2017;6(1):149–56.
3. Chen Y, Zhou Q, Wang H, Zhuo W, Ding Y, Lu J, Wu G, Xu N, Teng L. Predicting peritoneal dissemination of gastric cancer in the era of precision medicine: molecular characterization and biomarkers. *Cancers (Basel)*. 2020;12(8):2236–55.
4. Ikoma N, Chen HC, Wang X, Blum M, Estrella JS, Fournier K, Mansfield P, Ajani J, Badgwell BD. Patterns of initial recurrence in gastric adenocarcinoma in the era of preoperative therapy. *Ann Surg Oncol*. 2017;24(9):2679–87.
5. Zhang J-F, Lv L, Zhao S, Zhou Q, Jiang C-G. Hyperthermic intraperitoneal chemotherapy (HIPEC) combined with surgery: a 12-year meta-analysis of this promising treatment strategy for advanced gastric cancer at different stages. *Ann Surg Oncol*. 2022;29(5):3170–86.
6. Cancer Genome Atlas Research N: Comprehensive molecular characterization of gastric adenocarcinoma. *Nature* 2014;513(7517):202–9.
7. Wang H, Ding Y, Chen Y, Jiang J, Chen Y, Lu J, et al. A novel genomic classification system of gastric cancer via integrating multidimensional genomic characteristics. *Gastric Cancer*. 2021;24(6):1227–41.
8. Pectasides E, Stachler MD, Derks S, Liu Y, Maron S, Islam M, et al. Genomic heterogeneity as a barrier to precision medicine in gastroesophageal adenocarcinoma. *Cancer Discov*. 2018;8(1):37–48.
9. Ge S, Xia X, Ding C, Zhen B, Zhou Q, Feng J, et al. A proteomic landscape of diffuse-type gastric cancer. *Nat Commun*. 2018;9(1):1012.
10. Li X, Zheng NR, Wang LH, Li ZW, Liu ZC, Fan H, et al. Proteomic profiling identifies signatures associated with progression of precancerous gastric lesions and risk of early gastric cancer. *EBioMedicine*. 2021;74: 103714.
11. Li Y, Xu C, Wang B, Xu F, Ma F, Qu Y, et al. Proteomic characterization of gastric cancer response to chemotherapy and targeted therapy reveals new therapeutic strategies. *Nat Commun*. 2022;13(1):5723.
12. Smyth EC, Nilsson M, Grabsch HI, van Grieken NC, Lordick F. Gastric cancer. *Lancet*. 2020;396(10251):635–48.
13. Kim ST, Cristescu R, Bass AJ, Kim KM, Odegaard JI, Kim K, et al. Comprehensive molecular characterization of clinical responses to PD-1 inhibition in metastatic gastric cancer. *Nat Med*. 2018;24(9):1449–58.
14. Hagi T, Kurokawa Y, Kawabata R, Omori T, Matsuyama J, Fujitani K, et al. Multicentre biomarker cohort study on the efficacy of nivolumab treatment for gastric cancer. *Br J Cancer*. 2020;123(6):965–72.
15. Kang YK, Boku N, Satoh T, Ryu MH, Chao Y, Kato K, et al. Nivolumab in patients with advanced gastric or gastro-oesophageal junction cancer refractory to, or intolerant of, at least two previous chemotherapy regimens (ONO-4538-12, ATTRACTION-2): a randomised, double-blind, placebo-controlled, phase 3 trial. *Lancet*. 2017;390(10111):2461–71.
16. Lei X, Lei Y, Li JK, Du WX, Li RG, Yang J, Li J, Li F, Tan HB. Immune cells within the tumor microenvironment: Biological functions and roles in cancer immunotherapy. *Cancer Lett*. 2020;470:126–33.
17. Li H, Durbin R. Fast and accurate short read alignment with burrows-wheeler transform. *Bioinformatics*. 2009;25(14):1754–60.
18. Tarasov A, Vilella AJ, Cuppen E, Nijman IJ, Prins P. Sambamba: fast processing of NGS alignment formats. *Bioinformatics*. 2015;31(12):2032–4.
19. Wang K, Li M, Hakonarson H. ANNOVAR: functional annotation of genetic variants from high-throughput sequencing data. *Nucleic Acids Res*. 2010;38(16):e164.
20. Zhang H, Liu T, Zhang Z, Payne SH, Zhang B, McDermott JE, et al. Integrated proteogenomic characterization of human high-grade serous ovarian cancer. *Cell*. 2016;166(3):755–65.
21. Talevich E, Shain AH, Botton T, Bastian BC. CNVkit: genome-wide copy number detection and visualization from targeted DNA sequencing. *PLoS Comput Biol*. 2016;12(4): e1004873.
22. Mermel CH, Schumacher SE, Hill B, Meyerson ML, Beroukhi R, Getz G. GISTIC2.0 facilitates sensitive and confident localization of the targets of focal somatic copy-number alteration in human cancers. *Genome Biol*. 2011;12(4):R41.
23. Miller CA, White BS, Dees ND, Griffith M, Welch JS, Griffith OL, et al. SciClone: inferring clonal architecture and tracking the spatial and temporal patterns of tumor evolution. *PLoS Comput Biol*. 2014;10(8): e1003665.
24. Goedhart J, Luijsterburg MS. VolcanoR is a web app for creating, exploring, labeling and sharing volcano plots. *Sci Rep*. 2020;10(1):20560.
25. Feng J, Meyer CA, Wang Q, Liu JS, Shirley Liu X, Zhang Y. GFOLD: a generalized fold change for ranking differentially expressed genes from RNA-seq data. *Bioinformatics*. 2012;28(21):2782–8.
26. Cai G, Zhou Z, Gu X: MATTE: anti-noise module alignment for phenotype-gene-related. *bioRxiv* 2022(2022.05.29.493935).
27. 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(5):449–56.
28. Chen T, Guestrin C XGBoost: a scalable tree boosting system. *Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining* 2016:785–94.
29. Pedregosa F, Varoquaux G, Gramfort A, Michel V, Thirion B, Grisel O, Blondel M. Scikit-learn: machine learning in python. *J Mach Learn Res*. 2011;12:2825–30.

30. Li T, Fu J, Zeng Z, Cohen D, Li J, Chen Q, Li B, Liu XS. TIMER2.0 for analysis of tumor-infiltrating immune cells. *Nucleic Acids Res.* 2020;48(W1):W509–14.
31. Newman AM, Liu CL, Green MR, Gentles AJ, Feng W, Xu Y, Hoang CD, Diehn M, Alizadeh AA. Robust enumeration of cell subsets from tissue expression profiles. *Nat Methods.* 2015;12(5):453–7.
32. Lee JH, Chang KK, Yoon C, Tang LH, Strong VE, Yoon SS. Lauren histologic type is the most important factor associated with pattern of recurrence following resection of gastric adenocarcinoma. *Ann Surg.* 2018;267(1):105–13.
33. Wang R, Song S, Harada K, Ghazanfari Amlashi F, Badgwell B, Pizzi MP, et al. Multiplex profiling of peritoneal metastases from gastric adenocarcinoma identified novel targets and molecular subtypes that predict treatment response. *Gut.* 2020;69(1):18–31.
34. Wang R, Dang M, Harada K, Han G, Wang F, Pool Pizzi M, et al. Single-cell dissection of intratumoral heterogeneity and lineage diversity in metastatic gastric adenocarcinoma. *Nat Med.* 2021;27(1):141–51.
35. Tanaka Y, Chiwaki F, Kojima S, Kawazu M, Komatsu M, Ueno T, et al. Multi-omic profiling of peritoneal metastases in gastric cancer identifies molecular subtypes and therapeutic vulnerabilities. *Nat Cancer.* 2021;2(9):962–77.
36. Chen X, Wang H, Huang Y, Chen Y, Chen C, Zhuo W, Teng L. Comprehensive roles and future perspectives of exosomes in peritoneal metastasis of gastric cancer. *Front Oncol.* 2021;11: 684871.
37. Mishra YG, Manavathi B. Focal adhesion dynamics in cellular function and disease. *Cell Signal.* 2021;85: 110046.
38. Mun D-G, Bhin J, Kim S, Kim H, Jung JH, Jung Y, et al. Proteogenomic characterization of human early-onset gastric cancer. *Cancer Cell.* 2019;35(1):111–24.e10.
39. Ming XY, Fu L, Zhang LY, Qin YR, Cao TT, Chan KW, Ma S, Xie D, Guan XY. Integrin alpha7 is a functional cancer stem cell surface marker in oesophageal squamous cell carcinoma. *Nat Commun.* 2016;7:13568.
40. Ge JC, Wang YX, Chen ZB, Chen DF. Integrin alpha 7 correlates with poor clinical outcomes, and it regulates cell proliferation, apoptosis and stemness via PTK2-PI3K-Akt signaling pathway in hepatocellular carcinoma. *Cell Signal.* 2020;66: 109465.
41. Wu Y, Konate MM, Lu J, Makhlof H, Chuaqui R, Antony S, et al. IL-4 and IL-17A cooperatively promote hydrogen peroxide production, oxidative DNA damage, and upregulation of dual oxidase 2 in human colon and pancreatic cancer cells. *J Immunol.* 2019;203(9):2532–44.
42. Zhang X, Han J, Feng L, Zhi L, Jiang D, Yu B, et al. DUOX2 promotes the progression of colorectal cancer cells by regulating the AKT pathway and interacting with RPL3. *Carcinogenesis.* 2021;42(1):105–17.
43. Morel AP, Ginestier C, Pommier RM, Cabaud O, Ruiz E, Wicinski J, et al. A stemness-related ZEB1-MSRB3 axis governs cellular pliancy and breast cancer genome stability. *Nat Med.* 2017;23(5):568–78.
44. Wu Y, Li L, Wang Q, Zhang L, He C, Wang X, Liu H. LINC00511 promotes lung squamous cell carcinoma proliferation and migration via inhibiting miR-150-5p and activating TADA1. *Transl Lung Cancer Res.* 2020;9(4):1138–48.
45. Jiang X, Yang L, Gao Q, Liu Y, Feng X, Ye S, Yang Z. The role of RAB GTPases and its potential in predicting immunotherapy response and prognosis in colorectal cancer. *Front Genet.* 2022;13: 828373.
46. Zhao L, Xue M, Zhang L, Guo B, Qin Y, Jiang Q, et al. MicroRNA-4268 inhibits cell proliferation via AKT/JNK signalling pathways by targeting Rab6B in human gastric cancer. *Cancer Gene Ther.* 2020;27(6):461–72.
47. Zhang D, Lindstrom A, Kim EJ, Hwang CI, Hall ML, Lin TY, Li Y. SEMA3C supports pancreatic cancer progression by regulating the autophagy process and tumor immune microenvironment. *Front Oncol.* 2022;12: 890154.
48. Lu T, Li Y, Pan M, Yu D, Wang Z, Liu C, Hu G. TBC1D14 inhibits autophagy to suppress lymph node metastasis in head and neck squamous cell carcinoma by downregulating macrophage erythroblast attacher. *Int J Biol Sci.* 2022;18(5):1795–812.
49. Petitprez F, Meylan M, de Reynies A, Sautes-Fridman C, Fridman WH. The tumor microenvironment in the response to immune checkpoint blockade therapies. *Front Immunol.* 2020;11:784.
50. Ye L, Zhang T, Kang Z, Guo G, Sun Y, Lin K, et al. Tumor-infiltrating immune cells act as a marker for prognosis in colorectal cancer. *Front Immunol.* 2019;10:2368.
51. Xie S, Mo P, Li N, Cai W, Ruan J, Ye J, Mao J. Tumor-infiltrating lymphocyte-based risk score for predicting prognosis in gastric cancer. *Front Oncol.* 2020;10: 522015.
52. Xu Q, Chen S, Hu Y, Huang W. Landscape of immune microenvironment under immune cell infiltration pattern in breast cancer. *Front Immunol.* 2021;12: 711433.
53. Kim R, An M, Lee H, Mehta A, Heo YJ, Kim KM, et al. Early tumor-immune microenvironmental remodeling and response to first-line fluoropyrimidine and platinum chemotherapy in advanced gastric cancer. *Cancer Discov.* 2022;12(4):984–1001.

Publisher's Note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.