scientific reports

Check for updates

OPEN

Prognostic molecular subtype reveals the heterogeneity of tumor immune microenvironment in gastric cancer

Hui Dai[1], Jing Ren[1], Chun Wang[1], Jianfei Huang[2] & Xudong Wang[3]✉

Gastric cancer (GC) remains a leading cause of cancer-related deaths and exhibits considerable heterogeneity among patients. Thus, accurate classifications are essential for predicting prognosis and developing personalized therapeutic strategies. To address this, we retrospectively analyzed multi-omics data from 359 GC samples, incorporating transcriptomic RNA (mRNA), DNA methylation, mutation data, and clinical parameters. Using ten clustering algorithms, we integrated these datasets to classify GC into molecular subtypes. The robustness of our clustering approach was externally validated using an independent cohort generated from different sequencing technologies, and we characterized the heterogeneity of each subtype. Our analysis identified three distinct molecular subtypes of GC, designated CS1, CS2, and CS3. These subtypes exhibited significant differences in survival outcomes, activation of cancer-related pathways, immune microenvironment composition, genomic alterations, and responses to immunotherapy and chemotherapy. Notably, Cathepsin V (CTSV) was significantly downregulated in the immunologically active and highly responsive CS3 subtype, while it was upregulated in the immunologically exhausted CS2 subtype. These findings suggest that CTSV could serve as both a prognostic marker and a molecular classifier. Furthermore, this study provides the first evidence of CTSV's high expression in GC and its potential role in tumor progression. The novel clustering approach, based on ten clustering algorithms and comprehensive analysis of multi-omics data in gastric cancer, can guide prognosis, characterize different clinical and biological features, and elucidate the tumor immune microenvironment, providing insights into the intratumor heterogeneity of GC and potential novel therapeutic strategies. Additionally, CTSV emerges as a prognostic marker linked to tumor immunity and disease progression, which lays the foundation for improved stratification strategies and the development of targeted therapeutic approaches in GC.

Keywords Gastric cancer, Clustering algorithms, Molecular subtypes, Prognosis, Immune microenvironment, Immunotherapy

Gastric cancer (GC) is the fifth most common malignancy and the third leading cause of cancer-related mortality worldwide[1,2]. It is a heterogeneous disease with distinct epidemiological and histopathological characteristics, leading to variations in clinical presentation and treatment response[3]. Among all gastric tumors, approximately 90% are malignant, with gastric adenocarcinoma accounting for 95% of these cases[4]. Radical surgery remains the primary curative approach for early-stage, resectable GC[5]. However, for advanced disease, chemotherapy and targeted therapy have been shown to improve clinical outcomes[6,7]. Despite these advancements, challenges such as drug toxicity, difficulties in identifying patients most likely to benefit from treatment, and the development of drug resistance continue to hinder substantial improvements in prognosis[5,8]. In recent years, immunotherapy has emerged as a promising treatment strategy, complementing surgery, chemotherapy, radiotherapy, and targeted therapy. Compared to conventional treatments, immunotherapy has demonstrated durable efficacy with a more favorable toxicity profile, offering new therapeutic opportunities for patients with GC[8–10].

[1]Medical School, Nantong University, Nantong 226001, Jiangsu, China. [2]Department of Clinical Biobank, Affiliated Hospital of Nantong University, Nantong 226001, Jiangsu, China. [3]Department of Laboratory Medicine, Affiliated Hospital of Nantong University, No. 20, Xisi Road, Nantong 226001, Jiangsu, China. ✉email: wangxudong816@163.com

Traditionally, tumor classification was based on cellular or tissue origin, leading to a "one-size-fits-all" approach in pathology and treatment. However, the widespread adoption of sequencing technologies has revealed significant heterogeneity in the genomic, transcriptomic, and epigenetic profiles of tumors, highlighting the limitations of conventional classification systems[11]. Initially, GC was categorized into two subtypes, intestinal and diffuse, based on the Lauren classification. Then, the World Health Organization (WHO) refined this system, classifying GC into four subtypes: papillary, tubular, mucinous, and poorly cohesive. However, despite these efforts, neither classification has proven sufficient for guiding personalized treatment[12]. In 2014, the Cancer Genome Atlas (TCGA) consortium introduced a molecular and genomic-based classification system for GC, identifying four distinct subtypes: Epstein-Barr virus-positive tumors (EBV+), microsatellite instability tumors (MSI), genomically stable tumors (GS) and chromosomally unstable tumors (CIN)[13]. EBV + GC is characterized by lymphoid stroma or lymphoepithelioma-like histology and frequently harbors mutations in PIK3 CA and ARID1 A. The MSI subtype is defined by mutations or promoter methylation in mismatch repair genes, displaying a hypermethylation profile similar to that of EBV + GC. Notably, MSI status is associated with favorable postoperative survival in resectable primary GC[14]. GS GC predominantly exhibits diffuse histology with relatively low levels of genetic aberrations, whereas CIN GC is typically associated with intestinal histology and widespread DNA copy number alterations, including receptor tyrosine kinase gene amplifications[13]. In 2015, the Asian Cancer Research Group (ACRG) categorized GC into four subtypes: MSI, MSS/EMT, MSS/TP53−, and MSS/TP53+[15].

A comprehensive understanding of human health and disease requires an in-depth analysis of molecular complexity and variability across multiple biological dimensions[16]. The integrated analysis of multi-omics datasets offers a powerful approach for improving prognostic accuracy and refining predictive models of disease phenotypes, ultimately enhancing therapeutic interventions and preventive strategies[17,18]. In this study, we applied machine learning and ten clustering algorithms to identify GC subtypes based on transcriptomic and DNA methylation profiles. Three distinct subtypes, designated as CS1, CS2 and CS3 were identified, with each exhibiting unique survival outcomes, clinicopathological characteristics, tumor immune microenvironment composition, genomic mutation profiles, and differential responses to immunotherapy and chemotherapy, thereby presenting a framework for risk stratification and personalized treatment strategies. Additionally, Cathepsin V (CTSV) was found to be significantly downregulated in the immunologically active CS3 subtype, which is associated with a favorable prognosis, while its expression was elevated in the immunologically silent CS2 subtype, which is linked to poor prognosis. These findings suggest that CTSV may serve as both a prognostic biomarker and a molecular classifier for GC subtypes. Furthermore, its high expression pattern and role in GC progression have been confirmed, highlighting its potential as a key regulator in tumor biology.

Materials and methods

Extraction and preprocessing of multi-omics data

The dataset for GC was obtained from The Cancer Genome Atlas (TCGA)[19], which offers comprehensive multi-omics data for analysis. We collected gene expression profiles encompassing mRNA and DNA methylation data, along with clinicopathologic information using the "TCGAbiolinks" R package. Transcriptomic data were transformed using log2(TPM + 1) to normalize the data. Samples were filtered to exclude formalin-fixed, recurrent, and metastatic tissues, retaining only those with sample identifiers ending in "01 A" or "11 A". Additionally, we acquired an external independent validation cohort, specifically GSE84437, GSE26253, GSE62254 and GSE15459[20–22], from the GEO database and quantified them using the Affymetrix Human Genome U133 Plus 2.0 Array. Affymetrix probe IDs were annotated with gene symbols based on the GPL6947, GPL8432 and GPL570 platforms. For DNA methylation data, we performed logit transforms on β-values before conducting ComBat adjustment, followed by reverse logit transformation post-adjustment[23]. Afterwards, we utilized the R package ChAMP to comprehensively filter the methylation matrix. Specifically, we removed probes with detection pvalues > 0.01, probes with < 3 beads in at least 5% of samples per probe, all non-CpG probes, all SNP-related probes, all multi-hit probes, and probes located on sex chromosomes[23,24].

Identification of multi omics-based cancer subtypes by integrative analysis

The MOVICS package, a unified analytic pipeline, integrates 10 state-of-the-art multi-omics clustering algorithms to facilitate integrative clustering and visualization of researches on cancer typing[25]. These algorithms include CIMLR, iClusterBayes, MoCluster, COCA, ConsensusClustering, IntNMF, LRAcluster, NEMO, PINSPlus, and SNF[25]. These algorithms have been rigorously appraised for properties in prior research[26,27]. Based on optimal average statistic value of Gaps-statistics and CPI analysis provided by package MOVICS, we found an optimal clustering number. Subsequently, integrative clustering of the TCGA cohort was performed using the Bayesian latent variable model implemented in the R package MOVICS[25,28].

Assessment of signaling pathways activation

The subtype-specific upregulated and downregulated genes among subgroups were identified with the 'limma' R package[29]and functionally annotated based on the Gene Ontology (GO) with the 'clusterProfiler' R package (adjusted p value < 0.05)[30]. Single-sample gene-set enrichment analysis (ssGSEA) package "GSVA" was made use of to analyze 50 HALLMARK gene sets and immune and stromal signatures gene sets to reveal the activation of biological pathways. Gene sets representing GO biological processes and HALLMARK were sourced from the Molecular Signatures Database (MSigDB) (https://www.gsea-msigd4b.org/gsea/msigdb/index.jsp)[31]. The activated status of the immune signature was appraised by gene sets extracted from previous studies[32]. The normalized enrichment score (NES) of dissimilar subtypes were calculated from the gene sets associated with immune and stromal features to demonstrate distinctness in the immune activation status.

Assessment of immune cell infiltration and immune microenvironment

The Tracking tumor Immuno-Phenotype algorithm (TIP) (http://biocc.hrbmu.edu.cn/TIP/index.jsp) was utilized to evaluate the anti-tumor immune status. Infiltration rates of 22 immunocytes were analyzed using CIBERSORT (http://cibersort.stanford.edu/). CIBERSORT employs support vector regression (SVR), enhancing deconvolution performance through feature selection and robust mathematical optimization techniques[33].

The ESTIMATE algorithm leverages the unique transcriptional profiles of cancer samples to infer tumor cellularity and the composition of infiltrating normal cells[34]. It has been applied to various cancer types, including gastric cancer[35], colon cancer[36], clear cell renal cell carcinoma[37], glioblastoma[38], and Hepatocellular carcinoma[39]. Moreover, tumor microenvironment cell populations and ratios were analyzed using different algorithms such as EPIC, MCPcounter and QuanTIseq. The T-cell exhaustion (TEX) state was calculated and quantified using ssGSEA based on previously studied Exhausted signature genes[40].

Role of the subtypes in conjecturing the effect of immunotherapy

Tumor Immune Dysfunction and Exclusion (TIDE) is a computational model that incorporates two principal mechanisms of tumor immune evasion[41]. The TIDE score, calculated using TIDE algorithms available at http://tide.dfci.harvard.edu/, was exploited to forecast the response to immune checkpoint inhibitors targeting PD1, PD-L1, and CTLA4[41]. The Cancer Immunome Atlas database (TCIA,https://tcia.at/) was applied in downloading the immunophenoscores (IPS) of TCGA-GC. For the GEO cohort, IPS scores are calculated based on Z-scores from four categories of genes related to immunogenicity, providing a quantitative measure of immune therapy response[42].

Evaluation of genomic mutation and copy number variation

Tumor mutation burden (TMB) refers to the count of mutations per million bases, while fraction genome alteration (FGA) represents the proportion of gene fragments with incremental or lost copy numbers in the whole genome. The mutation spectrum and copy number variation of STAD were acquired from the TCGA database utilizing the 'TCGAbiolinks' R package, and then mutational differences among GC isoforms were compared and visualized employing the 'Maftools' R package[43].

Drug sensitivity prediction

For drug susceptibility analysis, we attained the analyzed dataset of chemotherapeutic drugs from the Genomics of Drug Sensitivity in Cancer database (GDSC, https://www.cancerrxgene.org/)[44]. The GDSC database serves as the largest public repository for information on cancer cell drug susceptibility and molecular markers of drug response. It hosts and annotates extensive datasets on drug susceptibility in cancer cells and correlates this information with detailed genomic data to facilitate the discovery of molecular biomarkers of drug response[44]. The 'pRRophetic' R package was occupied in conjecturing the half-maximal inhibitory concentration (IC50) values of chemotherapeutic drugs, indicative of their effectiveness in inhibiting specific biological or biochemical processes[45].

Tissue microarray and immunohistochemistry

Protein levels of CTSV were assessed by immunohistochemistry (IHC). The tissue microarray chips totally contained 181 samples of gastric adenocarcinoma and 60 normal tissues samples. The key patient characteristics included overall survival time, survival status, M stage, N stage, T stage, and AJCC stages. The tissue microarray chips were subjected to baking, dehydration, xylene dewaxing, ethanol rehydration, antigen repair, and afterwards incubated with the primary antibody overnight at 4 ° C. On the second day, secondary antibodies (HRP labeled anti mouse antibodies, DAKO) were incubated and re stained with diaminobenzidine (DAB) and hematoxylin after washing in PBS. For each sample on the microarray, the entire tissue is evaluated and classified based on the degree of staining. The antibodies used in immunohistochemistry included antiCTSV (1:150, T56745, Abmart).

Cell culture

The GC cell lines HGC- 27 and AGS used were purchased from the Chinese Academy of Sciences (Shanghai, China). The cells were cultured in RPMI 1640 medium (Gibco, USA) containing 10% fetal bovine serum, 1% penicillin, and streptomycin in a 37 ° C incubator with 5% CO2.

RNA extraction and quantitative real-time polymerase chain reaction (qRT–PCR).

For the purpose of extracting RNA, stomach cancer cells were lysed using TRIzol RNA separation reagent (Invitrogen, Carlsbad, CA, USA), and then RNA was precipitated with isopropanol. The colloidal precipitate was dried until transparent, and then DEPC water was added to obtain the original RNA solution. The system was configured based on the obtained RNA concentration, and cDNA was obtained through reverse transcription. Subsequently, qRT PCR reaction system was configured, and gene expression was quantified by the ABI 7500 quantitative PCR instrument using SYBR Green Master Mix. The expression levels were calculated with the $2-\Delta\Delta Ct$ method, and GAPDH acted as the internal reference for normalization. Primer sequences are listed below: CTSV-F: TTTGCTGATGTGGGAGTGTAGTT; CTSV-R: CTGAGTGGATGTTTCCTTATTGC;

Protein extraction and western blotting analysis

For the purpose of extracting proteins, stomach cancer cells were dissolved in a RIPA buffer containing PMSF and protease inhibitors. We used identical amounts of protein in SDS-PAGE, which was then transferred to PVDF membranes, blocked with 5% skim milk for two hours at room temperature, and incubated at 4℃ overnight. The following primary antibodies are used including rabbit antiCTSV(1:1000, T56745, Abmart), mouse anti-GAPDH (1:10000, 60004 - 1-lg, Proteintech). Secondary antibodies (HRP-conjugated goat anti-mouse or goat anti-rabbit) were obtained from Abcam. The GAPDH expression level was used to normalize the data.

Transfection of cells with small interfering RNA (siRNA)
Cells were seeded in culture well plates one day in advance. The transfection reagent Lipofectamine 3000 (Invitrogen, Carlsbad, CA, USA) and siRNA (RIBOBIO, China) were respectively diluted with basal medium according to the manufacturer's instructions, and then combined into a transfection mixture. After incubation for 20 min, the mixture was added to the cell culture well plates. The complete medium was replaced after incubation for 6–8 h. RNA was extracted at 48 h and proteins were extracted at 72 h. The knockdown efficiency was examined by qRT-PCR and Western blotting respectively. The siRNA sequences are as follows: siCTSV- 1: GGACUGUUCGCGUCCUCAATT, UUGAGGACGCGAACAGUCCTT; siCTSV- 2: CGUCCUUCCAGUUC UACAATT, UUGUAGAACUGGAAGGACGTT; siCTSV- 3: GGUUGCUUUCGAAACCAGATT, UCUGGUU UCGAAAGCAACCTT.

Colony formation assay The transfected cells were evenly seeded into 6-well plates at a density of 800 cells per well, and then placed in a CO_2 incubator for culturing. After about 2 weeks of culturing, the cells were fixed with 4% paraformaldehyde for 30 min and then stained with crystal violet for 10 min. After being washed with phosphate-buffered saline (PBS), the cell colonies were counted.

CCK8 cell proliferation assay 48 h after siRNA transfection, cells were seeded into 96-well plates at a density of 3×10^3 cells per well. Then 10 μL of the enhanced CCK8 reagent (Beyotime, C0043) was added to each well, after the cells adhered to the well surface. After incubation in the cell culture incubator for 2–4 h, the absorbance at 450 nm (OD value) was detected using a multi-functional microplate reader.

Transwell assay Transwell plates (8.0-mm in diameter) (Corning, NY, USA) were placed in 24-well plates for cell migration assays. For cell invasion assays, the chambers needed to be paved with a mixture of Matrigel and serum-free medium at the bottom and placed in the incubator to solidify.

5×10^4 cells (for migration) and 8×10^4 cells (for invasion) were seeded in the chambers and cultured with medium without FBS. 500 μL of medium containing 20% FBS was added to the 24-well plates. After incubation, the cells were fixed with 4% paraformaldehyde, stained with crystal violet, and after washing, the cells that did not penetrate the upper chamber were wiped off for quantification.

Nude mouse tumor xenograft model Athymic nude mice (BALB/c Nude, 4 weeks old) were obtained from the Experimental Animal Center of Nantong University. The animal experiment design in this project strictly follows the principles of animal experiment welfare ethics review. Nude mice were divided into two groups: control and knockdown group. Each group had a quantity of 4. The environmental conditions for raising nude mice are SPF level.

After one week of acclimatization, 1×10^7 cells were subcutaneously injected into the axilla of each mouse to induce tumor formation. When the tumor volume reached approximately 200 mm^3, 20μg 2′-O-Me and 5′-cholesterol-modified CTSV siRNA (siCTSV- 2: sense(5'– 3'): CGUCCUUCCAGUUCUACAATT; antisense(5'– 3'): UUGUAGAACUGGAAGGACGTT) (siCTSV group) or negative siRNA (si-nc group) diluted in PBS was mixed with Lipofectamine 3000 and injected into multiple sites of the tumor every three days. After injection, the mice were continued to be fed and observed for 3–4 weeks. Tumor volume was measured every three days. Before the tumor volume reached approximately 2000 mm^3, mice anesthetized with isoflurane were euthanized by cervical dislocation method. Animals were euthanized and tumor tissue was removed for volume and weight testing. Mouse carcasses were sealed in white transparent self-sealing bags and sent to the Laboratory Animal Center for centralized harmless disposal.

Ethics declarations
All methods were carried out in accordance with relevant guidelines and regulations. All methods are reported in accordance with ARRIVE guidelines (https://arriveguidelines.org). The study was conducted in keeping with the Declaration of Helsinki (as revised in 2013). The studies involving humans were approved by Medical Ethics Committee of Affiliated Hospital of Nantong University (No. 2019-L053). Written informed consent was obtained from all patients for the use of their tissues for research purposes. The animal experiments were approved by the Institutional Animal Care and Use Committee of the Affiliated Hospital of Nantong University (No. S20241118 - 011).

Statistical analysis
Data are presented as mean ± SD. Statistical analyses were performed using R software and GraphPad Prism. Pearson correlation coefficients were used to assess the relationship between two continuous variables. Categorical variables were compared using Pearson χ^2 tests. Differences between two or multiple groups were analyzed using student t-tests or one-way ANOVA. The log-rank test was employed for Kaplan-Meier survival analysis, and the "survminer" package was used to determine the optimal cut-off value. All statistical tests were two-sided, with $p < 0.05$ considered statistically significant ($^*p < 0.05$; $^{**}p < 0.01$; $^{***}p < 0.001$; $^{****}p < 0.0001$).

Result

Identification of three GC subtypes using multi-clustering algorithms

Based on the optimal average statistic value of Gaps-statistics and CPI analysis, three optimal subtypes for gastric cancer patients were identified in this study (Fig. 1A). A total of 10 multi-omics ensemble clustering algorithms were applied to these three preset subtypes and the results were combined. It was found favorable consistency among the three subtypes using the 10 algorithms (Fig. 1B). Meanwhile, the quality of the subtypes using silhouette analysis was evaluated, and the advanced silhouette width indicated the robustness of three subtypes (0.70 vs. 0.38 vs. 0.31) (Fig. 1C). To further characterize the molecular landscape of these subtypes, transcriptomic mRNA and DNA methylation profiles were visualized. The heatmap revealed distinct molecular

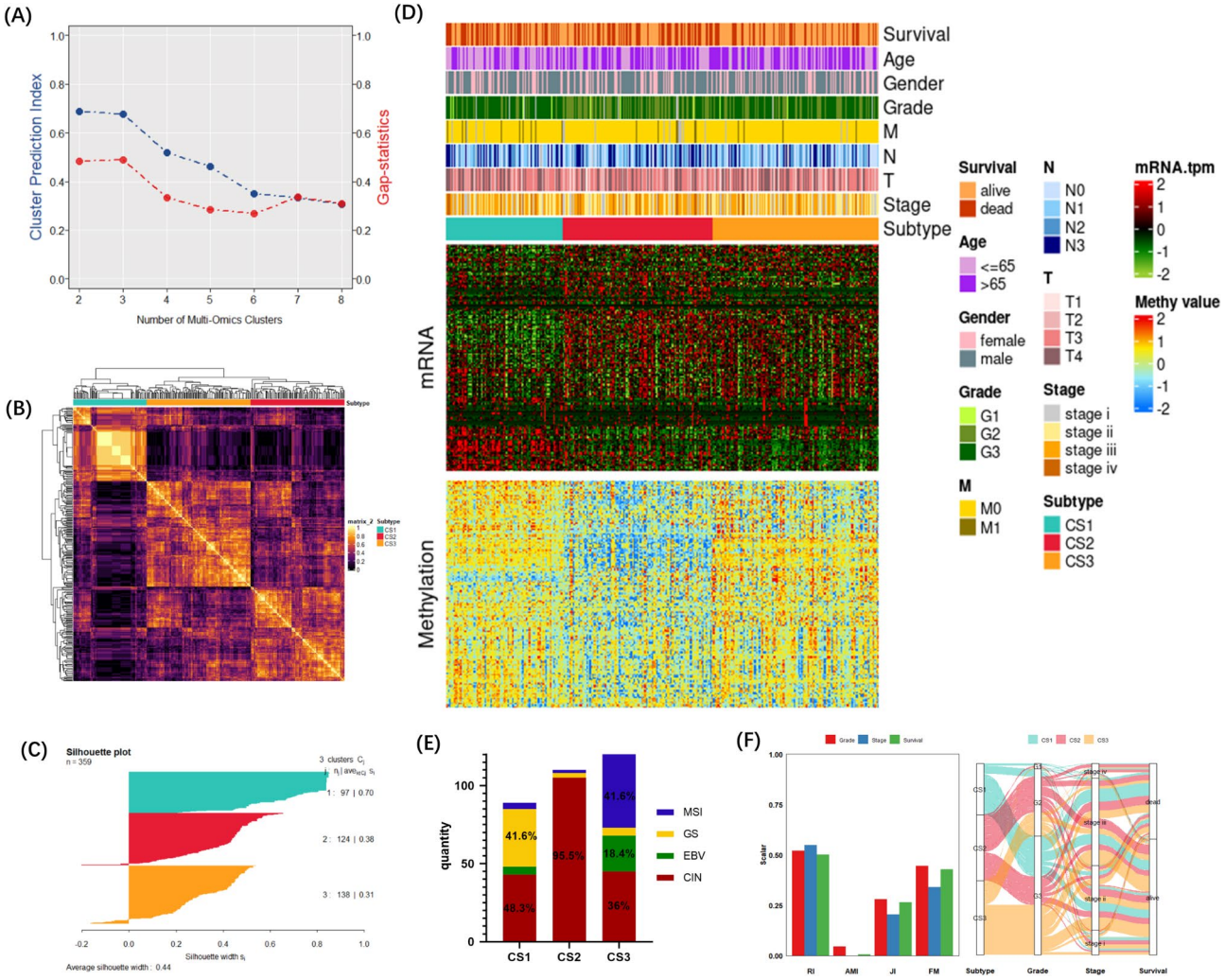

Fig. 1. Stratification of gastric cancer according to integrative multi-omics profiling. (**A**) Determination of the optimal count of multi-omics clusters. (**B**) Heatmap depicting molecular clustering derived from multi-omics data. (**C**) Silhouette analysis assessing the cohesion and separation of the identified clusters. (**D**) Comprehensive heatmap delineating the intricate molecular characteristics across mRNA and DNA methylation profiles within three integrative consensus subtypes. (**E**) Comparisons between TCGA molecular classification and three clusters. (**F**) Quantitative silhouette scoring of sample similarity based on consensus ensemble outcomes, accompanied by an alluvial diagram illustrating the flow distribution among distinct subtypes and the association of these subtypes with varied clinical attributes. The bar charts of different colors represent "Grade" (red), "Stage" (blue), and "Survival" (green). The height of the bar chart reflects the scalar scores of the corresponding indicators under different evaluation indicators (RI, AMI, JI, FM).

patterns across the three subtypes, highlighting their differences at the molecular level (Fig. 1D). A comparison between TCGA molecular classification and the present subtypes revealed that most GSs were observed in the CS1 subtype, and 95.5% of patients in the CS2 subtype belonged to the CIN subtype, while most EBV + and MSI subtypes were observed in the CS3 subtype (Fig. 1E).

CS3 subtype defined as a subset of GC with favorable prognosis

The identified subtypes demonstrated significant associations with key clinical features, including age ($P = 0.001$) and tumor grade ($P < 0.001$). Patients classified under the CS1 subtype had a lower average age and a higher proportion of grade G3 tumors, which are commonly associated with poor prognosis (Table 1; Fig. 1F). Survival analysis further revealed statistically significant differences in overall survival (OS) and progression-free survival (PFS) across the three subtypes. Notably, patients in the CS3 subtype exhibited the most favorable prognosis, with significantly improved OS ($P = 0.003$) and PFS ($P = 0.004$) compared to the other subtypes (Fig. 2A). To further evaluate the prognostic potential of these subtypes, univariate and multivariate Cox regression analyses were conducted. Given that well-established clinicopathological factors such as age, gender, tumor grade, nodal involvement (N), tumor invasion depth (T), and AJCC stage are widely recognized as prognostic indicators in clinical practice and GC research, these variables were incorporated into the Cox regression model. Additionally,

	level	CS1	CS2	CS3	*p*
n		97	124	138	
Age (%)	<= 65 > 65	61 (62.9) 36 (37.1)	57 (46.3) 66 (53.7)	53 (39.0) 83 (61.0)	0.001
Gender (%)	Female Male	33 (34.0) 64 (66.0)	39 (31.5) 85 (68.5)	50 (36.2) 88 (63.8)	0.717
Grade (%)	G1 G2 G3	4 (4.3) 14 (15.1) 75 (80.6)	3 (2.5) 66 (54.5) 52 (43.0)	2 (1.5) 44 (32.4) 90 (66.2)	< 0.001
M (%)	M0 M1	84 (90.3) 9 (9.7)	110 (95.7) 5 (4.3)	131 (97.0) 4 (3.0)	0.101
N (%)	N0 N1 N2 N3	29 (30.5) 25 (26.3) 16 (16.8) 25 (26.3)	31 (25.6) 33 (27.3) 28 (23.1) 29 (24.0)	51 (37.8) 35 (25.9) 27 (20.0) 22 (16.3)	0.306
T (%)	T1 T2 T3 T4	2 (2.1) 18 (18.6) 44 (45.4) 33 (34.0)	5 (4.0) 31 (25.0) 59 (47.6) 29 (23.4)	10 (7.2) 24 (17.4) 62 (44.9) 42 (30.4)	0.258
Stage (%)	Stage I Stage II Stage III Stage IV	10 (10.6) 29 (30.9) 46 (48.9) 9 (9.6)	15 (12.4) 39 (32.2) 54 (44.6) 13 (10.7)	21 (15.6) 50 (37.0) 58 (43.0) 6 (4.4)	0.439
Race (%)	Asian Black/African American White	28 (29.5) 0 (0.0) 67 (70.5)	26 (23.2) 7 (6.2) 79 (70.5)	31 (25.8) 4 (3.3) 85 (70.8)	0.130
OS.status(%)	Alive Dead	51 (52.6) 46 (47.4)	68 (54.8) 56 (45.2)	97 (70.3) 41 (29.7)	0.008
PFS.status(%)	Alive Dead	54 (55.7) 43 (44.3)	77 (62.1) 47 (37.9)	103 (74.6) 35 (25.4)	0.007

Table 1. Correlation analysis between three subtypes and clinicopathological characteristics.

the EBV + and MSI subtypes, previously identified in the TCGA classification as favorable prognostic groups, were also included[13]. The univariate Cox regression analysis demonstrated a strong association between CS subtypes and OS ($P < 0.001$). Further multivariate Cox analysis revealed that after adjusting for common clinicopathological variables, only the CS subtype and tumor grade remained significant prognostic factors for GC patients. Among them, the CS3 subtype was identified as the only significant protective factor (hazard ratio [HR]: 0.557, $P < 0.05$) (Fig. 2B). Given that transcriptomic profiling is one of the most widely utilized molecular approaches in cancer research, 30 highly upregulated mRNAs were identified as classifiers for each subtype in the TCGA-STAD cohort. These gene markers were subsequently applied to external datasets to independently predict GC subtypes, thereby validating the reproducibility of the classification (Fig. 2C). Notably, strong concordance was observed between the identified subtypes in the TCGA cohort and those predicted in the external validation cohort using the nearest template prediction (NTP) approach (Fig. 2D and E). After that, four external GC cohorts from the GEO database (GSE84437, GSE26253, GSE62254 and GSE15459) were recruited for the following investigation. NTP classifies every external sample as a determined subtype (Fig. 2F). Consistently, patients classified under the CS3 subtype exhibited the most favorable prognosis across all external datasets (Fig. 2G).

Distinct signaling pathway activation in CS subtypes

To investigate the functional and molecular differences among the three GC subtypes, GSEA was performed in the TCGA-STAD cohort and subsequently validated in the GEO cohort. The analysis identified subtype-specific GO biological function terms and Kyoto Encyclopedia of Genes and Genomes (KEGG) pathways (Fig. 3A). The validation results obtained from the GEO cohort demonstrated strong consistency with these findings, further supporting the robustness of the subtype classification (Fig. 3B). To assess immune-related functional differences among the subtypes, we analyzed the GSEA enrichment scores of immune function-related pathways. The results revealed that CS2 tumors exhibited a pronounced negative correlation with immune-related pathways compared to CS1 and CS3 (Fig. 3C). Furthermore, significant variations were observed in the enrichment levels of hallmark pathways across the three CS subtypes. For instance, the CS1 subtype was enriched in immune-associated pathways, including IL2_STAT5 signaling, IL6_JAK_STAT3 signaling, and epithelial-mesenchymal transition. In contrast, CS2 tumors displayed enhanced activation of cell cycle-related pathways, such as DNA repair, E2 F targets, and G2M checkpoints, suggesting a more proliferative and less immune-responsive phenotype. Notably, CS3 tumors exhibited strong immune relevance, with significant enrichment in pathways related to INF-α response, INF-γ response, and inflammatory response, indicating a highly immunologically active tumor microenvironment (TME) (Fig. 3D and E).

Differences in the TME landscape and immunotherapy responses among CS subtypes

The distinct enrichment of immune-related pathways across the CS subtypes suggests substantial heterogeneity in the TME. To characterize these differences, we first used the TIP algorithm to evaluate the tumor

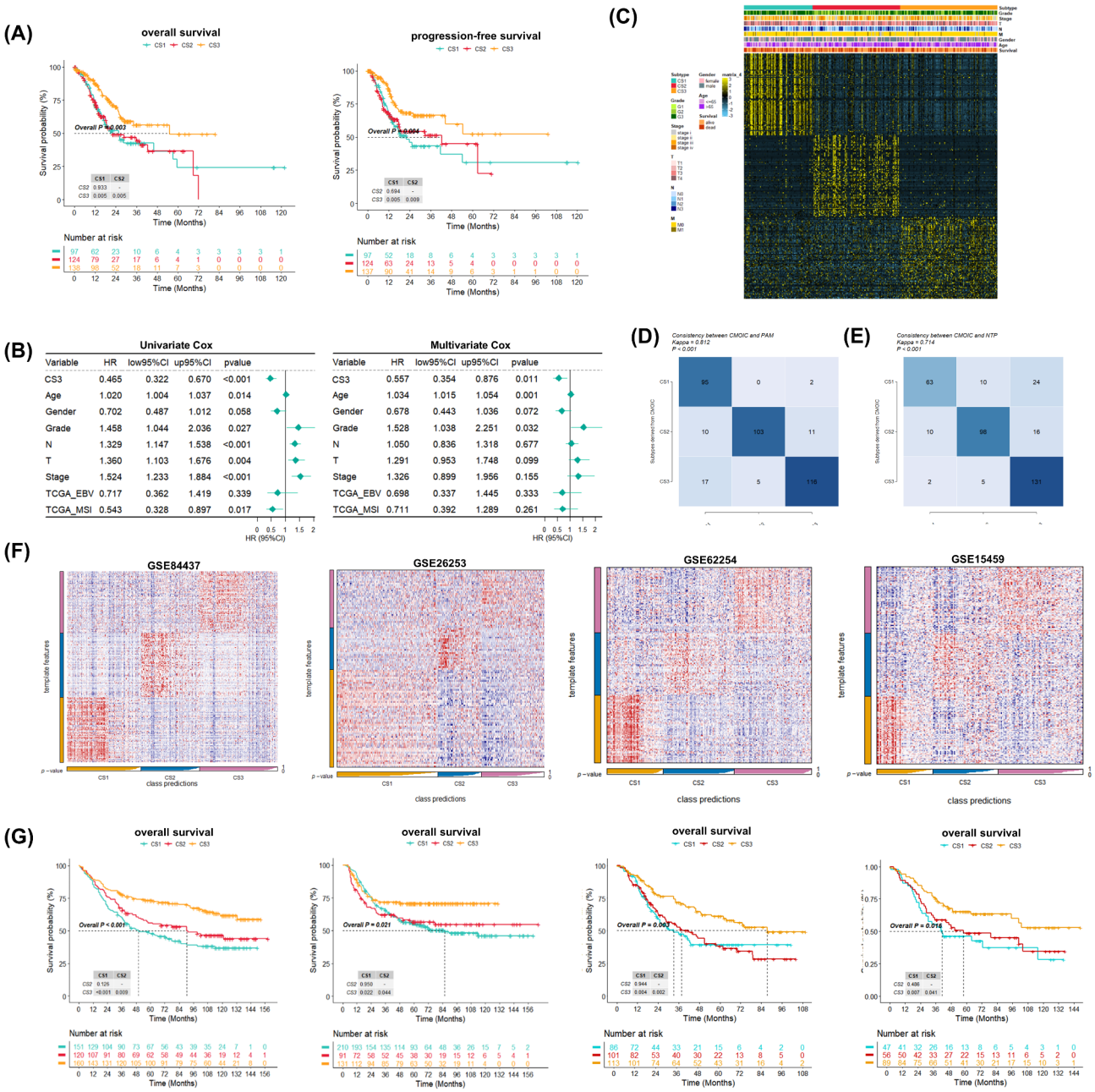

Fig. 2. CS3 subtype defines a subset of GC with favorable prognosis. (**A**) Kaplan-Meier survival plots delineating overall survival (left) and progression-free survival (right) for STAD patients stratified by integrative consensus subtypes. (**B**) Univariate (left) and multivariate (right) cox regression analysis to estimate the prognostic significance and differential prognostic influence. Age is a continuous variable. CS subtype, gender, grade, T stage, N stage, AJCC stage, and TCGA subtypes are categorical variables. CS3 is the hazard ratio (HR) and p-value obtained compared to CS1 and CS2 subtypes. (**C**) Upregulated biomarker heatmap for three subtypes. (**D**) Consistency between the preset subtype and the TCGA cohort subtype. (**E**) Consistency between the preset subtype and the marker genes utilized for external validation cohorts (NTP). (**F**) Heatmap visualizing the transcript expression patterns of upregulated biomarkers within the nearest template predicted cancer subtypes of GSE84437, GSE26253, GSE62254, GSE15459 cohorts. (**G**) Kaplan-Meier survival curves depicting overall survival using NTP in GSE84437, GSE26253, GSE62254, GSE15459 cohorts.

immunophenotypes of the three subtypes across the seven steps of the tumor immune cycle in both the TCGA and GEO cohorts. Notably, CS3 tumors exhibited the most active antitumor immune status, which could promote antitumor immune cell recruitment, whereas CS2 tumors displayed an immune-quiescent phenotype, suggesting minimal immune activation (Figs. 4A and S1 A). To further investigate immune cell infiltration, multiple immuno-evaluation methods, including ESTIMATE, CIBERSORT, EPIC, MCPcounter, and QuanTIseq, were applied. The ESTIMATE algorithm revealed significant differences in overall immune infiltration levels

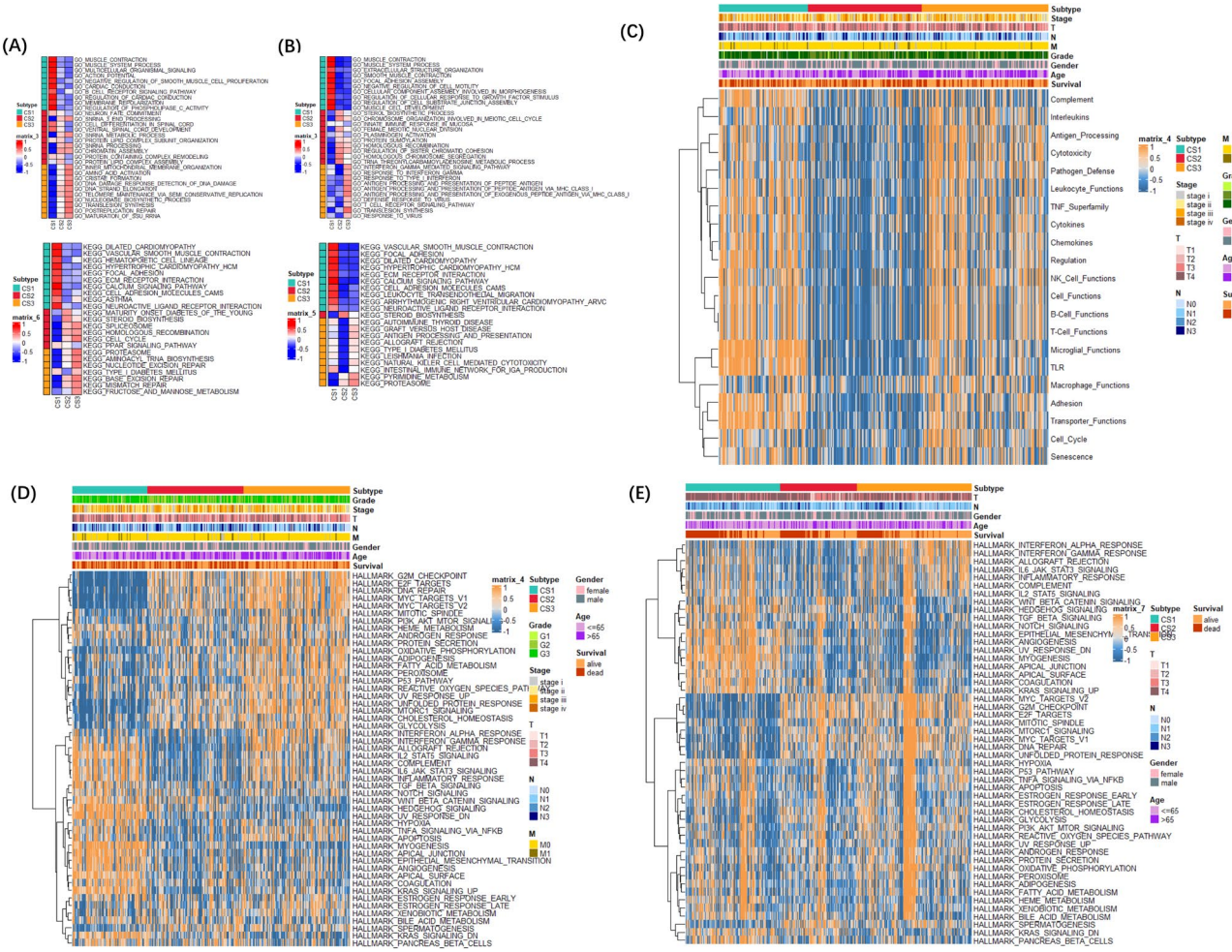

Fig. 3. Distinct activity of tumor-associated pathways in GC clusters. (A) GSEA enrichment analysis showing the subtype-specific GO biological function items and KEGG pathways 46–48 in TCGA-STAD cohort. (B) GSEA enrichment analysis showing the subtype-specific GO biological function items and KEGG athways 46–48 in GEO cohort. (C) Heatmap of the enrichment score related to immune pathways. (D) Heatmap of differentially activated hallmark pathway in TCGA-STAD cohort based on GSVA analysis. (E) Heatmap of differentially activated hallmark pathway in GEO cohort based on GSVA analysis.

among the three subtypes. CS1 demonstrated higher immune and stromal cell infiltration, along with increased ESTIMATE scores, while the immune-silent CS2 subtype exhibited the lowest immune infiltration (Figs. 4B and S1B). Previous studies have shown that high stromal scores were associated with poor prognosis, while there was no significant correlation between immune scores and prognosis in gastric cancer[49,50], which was consistent with the adverse prognosis of CS1 subtype. To elucidate why the CS1 subtype, despite its elevated antitumor activity and immune infiltration, exhibits poorer prognosis, we further examined the proportions of specific immune cell populations. Analyses across multiple algorithms consistently demonstrated that CS3 tumors had the highest abundance of antitumor immune cells, including CD8 + T cells, activated CD4 + T cells, and follicular helper T cells. In contrast, immunosuppressive cells such as regulatory T cells (Tregs), M2 macrophages, and cancer-associated fibroblasts (CAFs) were most abundant in the CS1 subtype (Figs. 4C–E and S1 C–E). Moreover, the CS1 subtype exhibited the highest M2/M1 macrophage ratio (Figs. 4F and S1 F) and T cell exhaustion scores (Figs. 4G and S1G). Collectively, these findings suggest that the CS1 subtype, despite its significant immune and stromal infiltration, are characterized by an immunosuppressive TME, as evidenced by high CAF infiltration, an elevated M2/M1 ratio, and increased T cell exhaustion. By contrast, CS3 tumors exhibit a highly active antitumor immune state, with abundant infiltration of antitumor immune cells, minimal CAF presence, the lowest M2/M1 ratio, and the least T cell exhaustion, further supporting their favorable prognosis.

Immunotherapy has emerged as a breakthrough in cancer treatment, including GC, and is now considered an effective therapeutic option following surgery, chemotherapy, radiotherapy, and targeted therapy[5,9]. Given the distinct immunological landscapes of the three subtypes and the limited subset of patients who benefit from immunotherapy, it is crucial to determine the differential susceptibility of these subtypes to immune-based treatments. The expression levels of immune checkpoint genes are important biomarkers for predicting immunotherapy responsiveness. For instance, high PD-L1 expression on tumor cells can facilitate immune

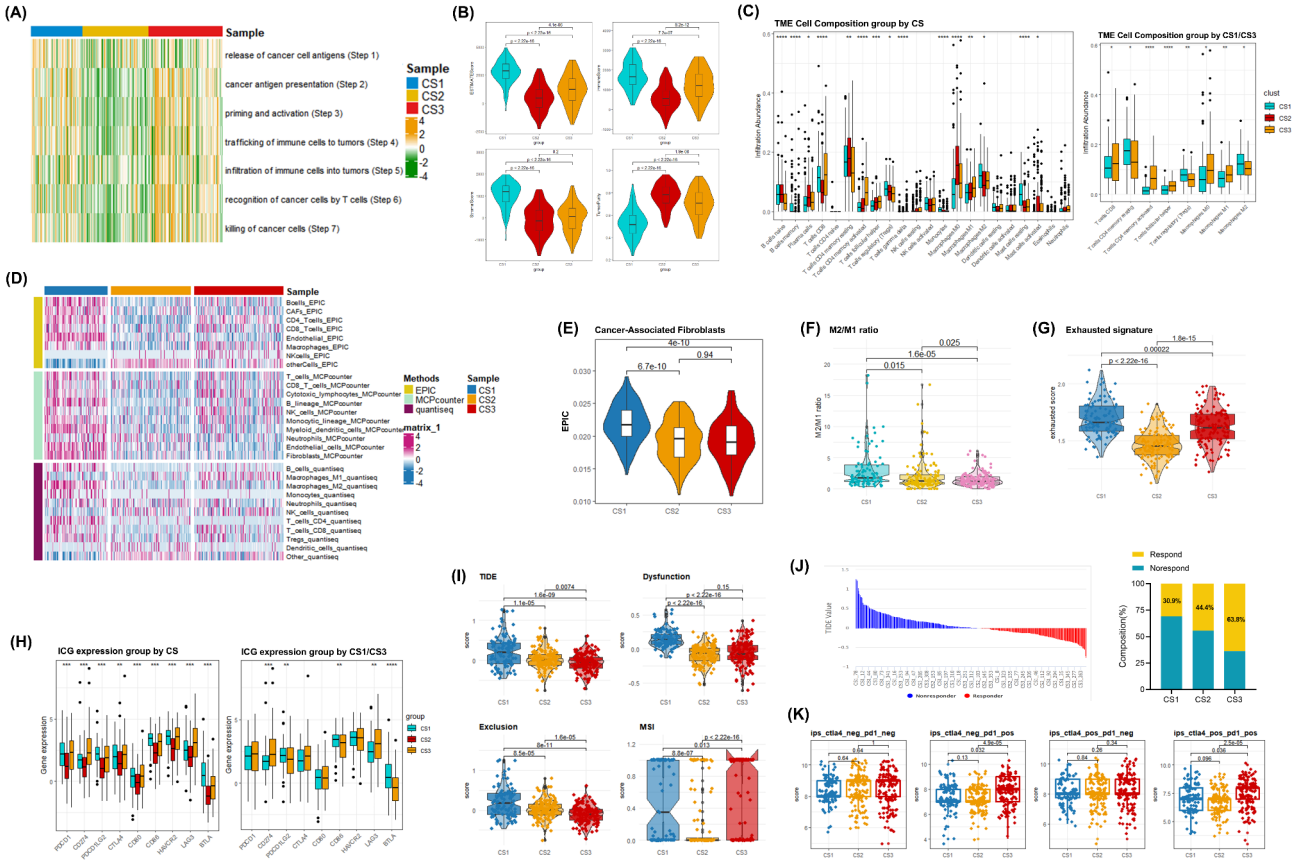

Fig. 4. Differences in tumor microenvironment landscape and immunotherapy response among CS subtypes. (**A**) Heatmap visualizing differences in anti-tumor immune status assessed by TIP algorithm across three subtypes. (**B**) Violin plots showing ESTIMATE Score, Immune Score, and Stromal Score based on ESTIMATE algorithm across subtypes. (**C**) Box plots illustrating the proportions of various immune cell infiltrations evaluated by Cibersort among three subtypes (left) and between CS1 and CS3 subtype (right). (**D**) Heatmap depicting differences in immune cell abundance assessed by EPIC, MCPcounter, and quanTIseq algorithms across three subtypes. (**E**) Differences in tumor-associated fibroblast expression levels among three subtypes. (**F**) Variations in M2/M1 ratios among three subtypes. (**G**) Differences in exhausted T cell signature scores based on GSEA across three subtypes. (**H**) Box plots of immune checkpoint gene expression among three subtypes (left) and between CS1 and CS3 subtype (right). (**I**) TIDE scores, exhaustion scores, exclusion scores, and MSI Expression Signature scores assessed by TIDE algorithm across subtypes. (**J**) Predictions of potential responders to immunotherapy by TIDE algorithm (left) and quantification of the proportion of potential responders (right). (**K**) Distribution of immunotherapy response potential among different subtypes based on immune phenotype score (IPS). * $p < 0.05$; ** $p < 0.01$; *** $p < 0.001$; **** $p < 0.0001$.

evasion by binding to PD- 1 on T cells, thereby inhibiting T cell activity and promoting tumor immune escape. Immune checkpoint inhibitors (ICIs) restore T cell function by blocking this inhibitory interaction, thereby enhancing the efficacy of immunotherapy. To assess potential immunotherapy responsiveness, we first examined the expression levels of key immune checkpoint genes. Compared to the other two subtypes, CS3 tumors exhibited significantly higher expression of most immune checkpoint genes in the TCGA-STAD cohort, a trend that was even more pronounced in the GEO cohort (Figs. 4H and S1H). Subsequently, the TIDE algorithm was used to evaluate the potential responses of patients with the three subtypes of gastric cancer to immunotherapy. CS3 tumors exhibited the least immune dysfunction and exclusion, along with a higher MSI phenotype (Figs. 4I and S1I), and the majority of potential responders to immunotherapy belonged to this subtype (TCGA: CS1: 30.9% vs. CS2: 44.4% vs. CS3: 63.8%; GEO: CS1: 21.9% vs. CS2: 54.2% vs. CS3: 70.0%) (Figs. 4J and S1 J). Microsatellite instability (MSI) testing contributes to the understanding of tumor genetic characteristics, prognosis assessment, and is crucial in cancer research and clinical practice. What's more, it as a vital genomic biomarker analysis, is prominently known for its role in identifying patients with a high likelihood of benefiting from immune checkpoint inhibitors[5,51,52]. Furthermore, the Immunophenotype Score (IPS) results indicated that, compared to CS1 and CS2 subtypes, patients in the CS3 subtype had significantly higher IPS scores for anti-PD- 1 and anti-PD- 1 combined with anti-CTLA4 immunotherapy, reflecting stronger immunotherapy responsiveness (Figs. 4K and S1 K).

Assessment of mutational landscapes and prediction of chemotherapy sensitivity

To further elucidate the genomic differences among the three CS subtypes, the mutational landscapes of each subtype were analyzed. The overall distribution of gene mutations with significant differences among the subtypes is presented in Fig. 5A. Notably, tumor suppressor genes, such as TP53, were extensively mutated in CS2 tumors (CS1: 25% vs. CS2: 70% vs. CS3: 38%, $P < 0.0001$), while oncogenes, such as MUC16, exhibited significantly higher mutation rates in CS3 tumors (CS1: 18% vs. CS2: 23% vs. CS3: 44%, $P < 0.0001$) (Fig. 5A). Given that MUC16 mutations in GC are associated with improved prognosis and increased tumor mutational burden (TMB), this observation provides a potential explanation for the favorable prognosis observed in the CS3 subtype[53–55] Additionally, mutations in LRP1B, TTN, and ARID1 A were more frequently detected in the CS3 subtype (LRP1B: CS1: 18% vs. CS2: 22% vs. CS3: 41%, $P < 0.0001$; TTN: CS1: 31% vs. CS2: 53% vs. CS3: 64%, $P < 0.0001$; ARID1 A: CS1: 16% vs. CS2: 7% vs. CS3: 51%, $P < 0.0001$) (Fig. 5A). Previous studies have reported that LRP1B, TTN and ARID1 A mutations are associated with enhanced responses to immune checkpoint inhibitors (ICIs)[55–60]. Given that tumors with higher immunogenicity, such as those with elevated TMB levels, are known to derive long-term clinical benefits from ICIs[61–64], we assessed TMB across the subtypes and found significantly higher TMB levels in CS3 tumors ($P < 0.001$, Fig. 5B). These findings collectively support the classification of CS3 as a subtype more likely to benefit from immunotherapy. Chromosomal instability was also assessed by calculating the fraction of genome altered (FGA) score, which indicated that CS2 tumors exhibited significantly greater chromosomal instability, characterized by more frequent copy number gains or losses, compared to the other two subtypes ($P < 0.001$, Fig. 5C). To further explore potential therapeutic strategies for each CS subtype, chemotherapy drug sensitivities were predicted using data from the Genomics of Drug Sensitivity in Cancer (GDSC) database by evaluating half-maximal inhibitory concentrations (IC50). The analysis suggested that CS3 tumors were more responsive to conventional chemotherapy agents, including 5-fluorouracil, Paclitaxel, Cisplatin, and Oxaliplatin (Fig. 5D). In contrast, CS1 tumors exhibited greater sensitivity to AZD8055, BMS-754,807, Foretinib, and SB216763 (Fig. 5E), while CS2 tumors demonstrated higher sensitivity to Linsitinib, SB505124, and TAF1 (Fig. 5F).

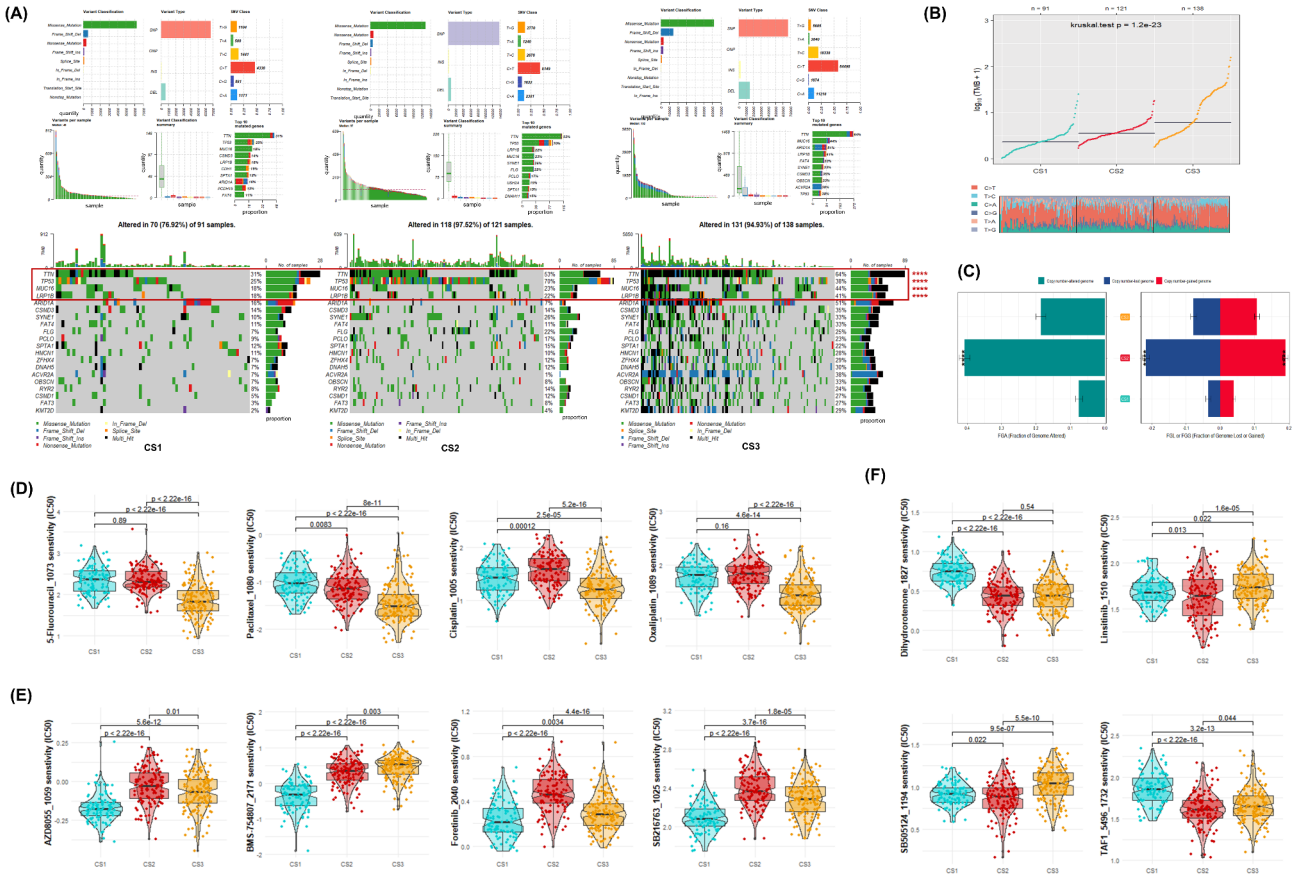

Fig. 5. Mutant landscape discrepancies and chemotherapeutics sensitivity forecast. (**A**) General information on representative gene mutations of three subtypes. (**B**) Evaluation of tumor mutation burden. (**C**) Distribution of fractional genome alterations (FGA) and the ratio of FGG to FGL. Bar charts are presented with mean ± standard error of the mean. (**D** - **F**) The average difference in IC50 values among three subtypes of representative drugs. Predictive candidate drugs for treating CS3 (**D**), CS1 (**E**), CS2 (**F**)subtype patients.

CTSV as a prognostic risk indicator with differential expression among CS subtypes

To enhance the clinical relevance of the molecular classification, we applied a log2|FC|> 1 and adjusted FDR < 0.05 as cut-off values to identify differentially expressed genes (DEGs) between the CS3 subtype and the other two subtypes (Fig. 6A). Surprisingly, we found that CTSV expression was lowest in the CS3 subtype characterized by favorable prognosis and an active tumor immune microenvironment, and highest in the CS2 subtype associated with poor prognosis and immune exhaustion (Fig. 6B). Analysis of TCGA gastric adenocarcinoma data samples showed that CTSV expression was meaningfully higher in tumor tissues compared to normal tissues (Fig. 6C), with a consistent trend observed in paired cancer and adjacent tissues (Fig. 6D). Pan-cancer expression level analysis of CTSV in normal and tumor tissues revealed significant upregulation not only in gastric cancer but also in other digestive system cancers, including colorectal cancer, esophageal cancer, and pancreatic cancer (Fig. 6E). Kaplan-Meier survival curves revealed a poor prognostic significance of high CTSV expression in the dataset (Fig. 6F). Furthermore, multivariate Cox analyses were conducted to investigate whether CTSV is an independent prognostic marker and whether there is an association between its prognostic value and the CS3 subtype. First, in a multivariate Cox analysis of the entire cohort, CTSV remained an independent prognostic

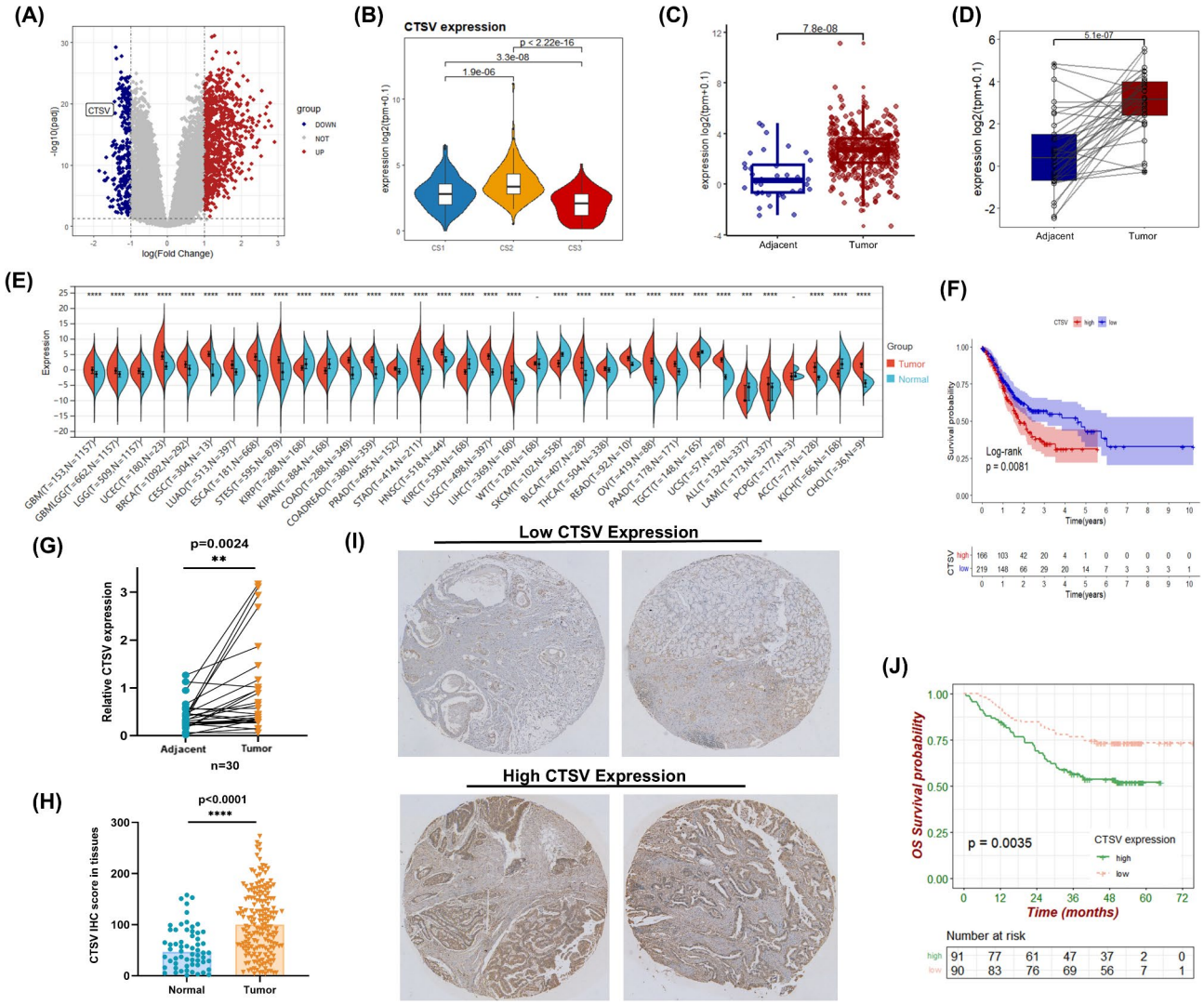

Fig. 6. High expression of CTSV in gastric cancer predicts poor prognosis. (**A**) Volcano plot showing differentially expressed genes between CS3 subtype and the other two subtypes. (**B**) Violin plots of CTSV expression levels across three CS subtypes. (**C**) mRNA expression levels of CTSV in gastric cancer and normal tissues from the TCGA database. (**D**) mRNA expression levels of CTSV in gastric cancer and paired adjacent tissues from the TCGA database. (**E**) Expression levels of CTSV in tumor and normal tissues across pan-cancer. (**F**) Kaplan-Meier curves depicting overall survival for CTSV in the TCGA cohort. (**G**) qRT-PCR detection of CTSV expression in paired gastric cancer and adjacent tissues (n = 30). (**H**) IHC scores of CTSV in normal tissues(n = 60) and CRC tissues(n = 181). (**I**) Representative images of CTSV immunohistochemical staining in high and low expression groups. (**J**) Kaplan-Meier curves depicting overall survival for CTSV based on immunohistochemical scoring (n = 181). * P < 0.05, ** P < 0.01, *** P < 0.001, and **** P < 0.0001

factor after adjusting for common clinicopathological factors, CS classification, and TCGA classification(HR = 1.89, $P < 0.001$) (Figure S2 A). Second, subgroup analyses within independent CS classifications revealed that CTSV had a significant impact on survival in the CS3 subtype, further supporting its prognostic value (HR = 2.438, $P < 0.05$) (Figures S2B–S2D). To validate the differential expression of CTSV in GC, qRT-PCR analysis was performed on RNA extracted from paired GC and adjacent tissues, which confirmed significantly higher CTSV expression in tumor tissues (Fig. 6G). Additionally, to assess the relationship between CTSV expression and prognosis, IHC was conducted on a GC tissue microarray containing 181 GC samples and 60 normal tissue samples using a multispectral pathology imaging system to score and stratify the results by median values (Fig. 6I). We found that CTSV was significantly overexpressed in GC tissues compared to normal tissues (Fig. 6H), with statistical evaluation confirming that patients with high CTSV expression had worse survival outcomes compared to those with low expression ($P = 0.0035$) (Fig. 6J). Further clinical correlation analysis revealed that high CTSV expression was significantly associated with advanced tumor staging, including higher N stage ($P < 0.001$), T stage ($P = 0.005$), and AJCC stage ($P = 0.006$) (Table 2).

.

CTSV promotes tumorigenesis and is closely associated with tumor immunity in GC

To investigate the biological role of CTSV, small interfering RNA (siRNA) targeting CTSV was transfected into HGC- 27 and AGS gastric adenocarcinoma cells, and the efficiency of knockdown was validated using qRT-PCR and western blotting (Fig. 7A and B). siCTSV- 1 and siCTSV- 2, with satisfactory knockdown efficiency, were used for subsequent biological function exploration. CCK8 cell proliferation assays and colony formation assays demonstrated significant inhibition of GC cell proliferation following CTSV knockdown (Fig. 7C and D), while transwell assays revealed reduced migration and invasion of gastric cancer cells (Fig. 7E and F). Collectively, knockdown of CTSV expression inhibits the proliferation, migration, and invasion of gastric cancer cells in vitro. To further examine the role of CTSV in tumor progression in vivo, a xenograft tumor model was established by subcutaneously injecting HGC- 27 GC cells into nude mice. After tumor formation, the mice were randomly assigned to either the si-nc group (negative control) or the siCTSV group, with injections administered every three days. After approximately four weeks, mice were euthanized by cervical dislocation, and tumors were collected for analysis (Fig. 7G). The results showed that compared to the si-nc group, the siCTSV group exhibited significantly slower tumor growth, as reflected by notable reductions in tumor volume and weight (Fig. 7H and I). These findings collectively demonstrate that CTSV actively contributes to GC progression both in vitro and in vivo.

Considering the differential expression of CTSV in immunologically active CS3 tumors and immunologically exhausted CS2 tumors, we first assessed the pathway enrichment of CTSV in gastric cancer. Consistent with CS2 tumors, tumors with high CTSV expression were enriched in proliferation-related signaling pathways, such as DNA repair, E2 F targets, and G2M checkpoints. Therefore, CTSV may drive the continuous proliferation of gastric cancer cells and promote gastric cancer progression by promoting the cell cycle, ensuring normal cell division, and regulating genes related to cell growth, proliferation, and metabolism (Fig. 8A). Notably, samples with low CTSV expression were enriched in immune-related pathways such as α/γ interferon response and inflammatory response (Fig. 8B). Pan-cancer enrichment analysis based on the TCGA database presented that CTSV was negatively correlated with α/γ interferon response and inflammatory response pathways in many cancers (Fig. 8C). TIP analysis results showed that tumors with high CTSV expression displayed meaningfully weaker anti-tumor immune status, which was unfavorable for the recruitment of anti-tumor immune cells

	level	High CTSV Expression	Low CTSV Expression	*p*
n		91	90	
Age (%)	<= 65 > 65	54 (59.3) 37 (40.7)	45 (50) 45 (50)	0.266
Gender (%)	Female Male	27 (29.7) 64 (70.3)	24 (26.7) 66 (73.3)	0.776
M (%)	M0 M1 Unknown	70 (76.9) 13 (14.3) 8 (8.8)	74 (82.2) 5 (5.6) 11 (12.2)	0.127
N (%)	N0 N1 N2 N3 Unknown	17 (18.7) 16 (17.6) 19 (20.9) 35 (38.5) 4 (4.4)	41 (45.6) 18 (20.0) 7 (7.8) 17 (18.9) 7 (7.8)	< 0.001
T (%)	T1 T2 T3 T4 Unknown	13 (14.3) 11 (12.1) 23 (25.3) 40 (44.0) 4 (4.4)	30 (33.3) 14 (15.6) 13 (14.4) 25 (27.8) 8 (8.9)	0.005
Stage (%)	Stage I Stage II Stage III Stage IV Unknown	12 (13.2) 26 (28.6) 40 (44.0) 8 (8.8) 5 (5.5)	27 (30.0) 31 (34.4) 19 (21.1) 6 (6.7) 7 (7.8)	0.006

Table 2. Correlation analysis between CTSV expression and clinicopathological characteristics.

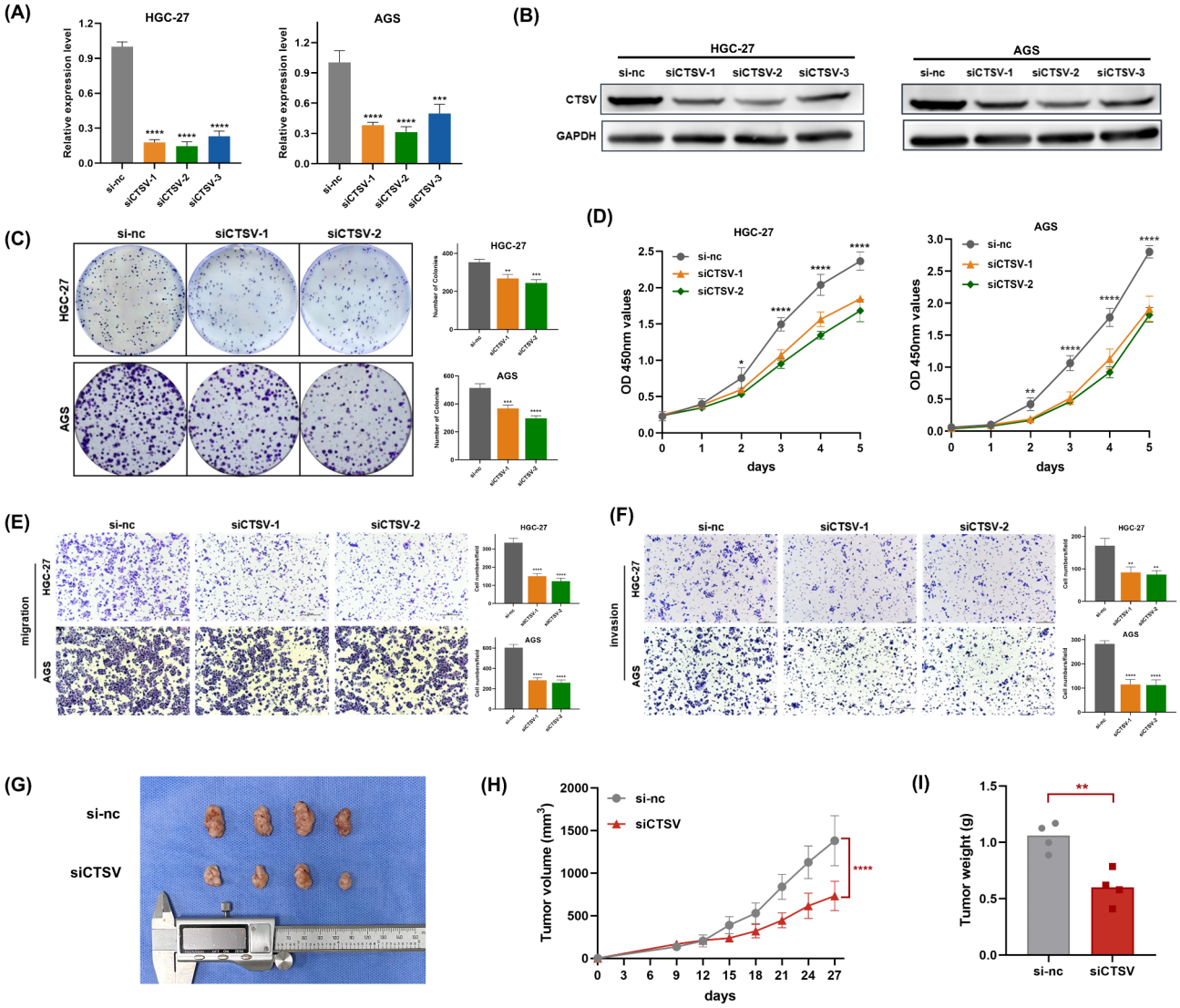

Fig. 7. CTSV promotes GC progression in vivo and in vitro. (**A**) qRT-PCR detection of knockdown efficiency of CTSV at mRNA level. (**B**) Western blotting detection of knockdown efficiency of CTSV-si at protein level. (**C**) Colony formation assay results for CTSV-si cell lines and control cell lines. (**D**) CCK8 proliferation assay measuring absorbance at 450 nm for CTSV-si cell lines and control cell lines. (**E**) Migration images of CTSV-si cell lines and control cell lines. (**F**) Invasion images of CTSV-si cell lines and control cell lines. (**G**-**I**) BALB/c Nude mice were subcutaneously injected with SW1116 cells, and tumor formation was monitored over several weeks. $^{*}P < 0.05$, $^{**}P < 0.01$, $^{***}P < 0.001$, and $^{****}P < 0.0001$.

(Fig. 8D). We then evaluated the immune infiltration levels in the CTSV expression differential groups. Clearly, the predictions based on various algorithms consistently displayed that the immune and stromal infiltration in the high CTSV expression group was significantly lower, demonstrating immune exhaustion (Fig. 8E and F). Additionally, tumors with low CTSV expression had higher TIDE scores and MSI scores, indicating a greater likelihood of benefiting from immunotherapy (Fig. 8G). Taken together, high CTSV expression in gastric cancer patients can serve as an indicator of poor prognosis in GC. Meanwhile, tumors with low CTSV expression exhibit an active anti-tumor immune status and higher immune infiltration, potentially more sensitive to immunotherapy, while tumors with high CTSV expression exhibit immune exhaustion and have a lower likelihood of benefiting from immunotherapy.

Discussion

TCGA is one of the most extensive multi-omics datasets, covering over 33 distinct cancer types[65]. This initiative aims to generate, integrate, analyze, and interpret comprehensive molecular maps of DNA, RNA, proteins, and epigenetic modifications in tumor samples, along with clinical and histopathological data[16]. Recent studies have demonstrated that integrating multi-omics datasets provides deeper and clearer insights into tumor biology. GC, known for its epidemiological and histopathological heterogeneity, presents significant challenges in clinical

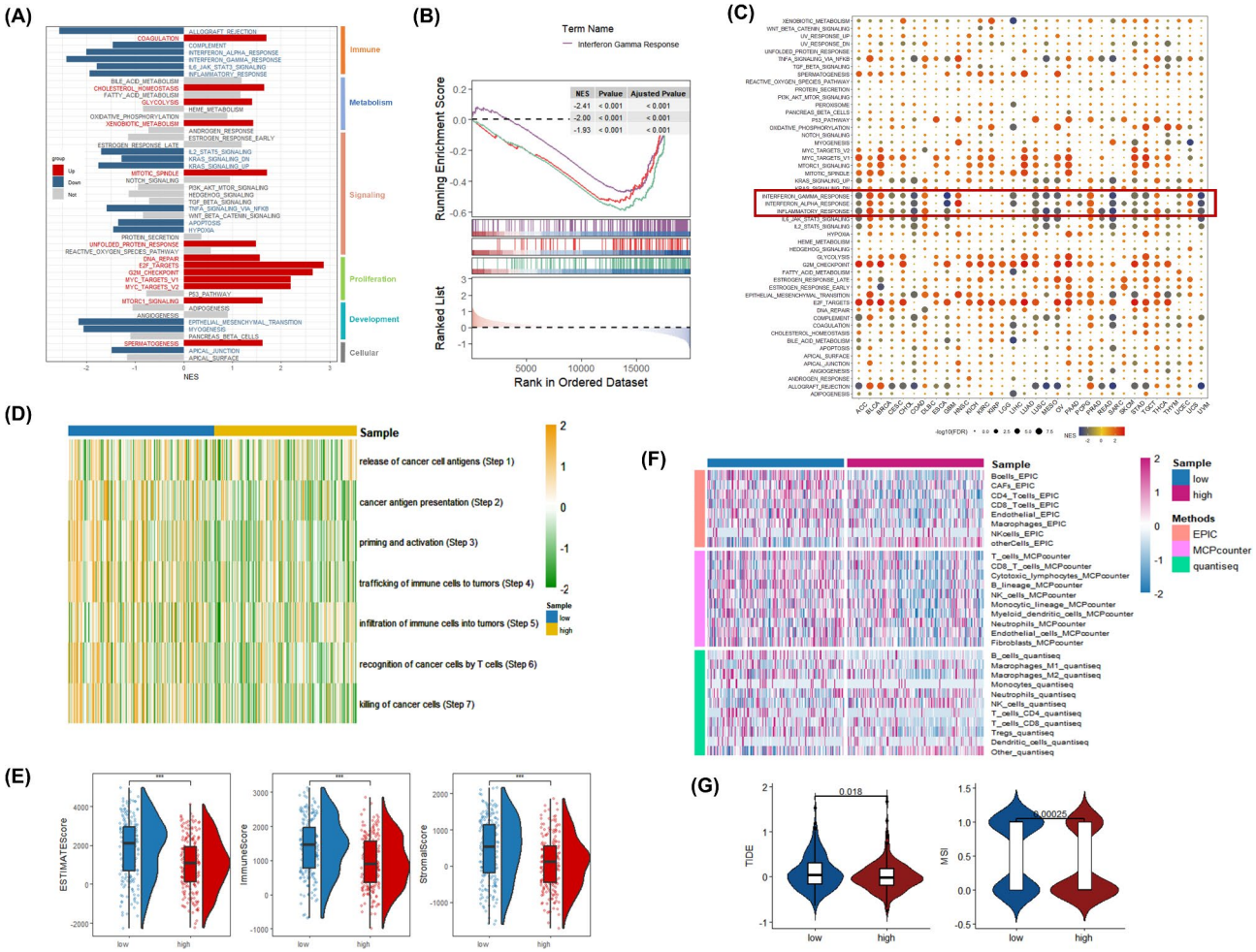

Fig. 8. Differences in pathway enrichment and immune infiltration landscape of CTSV differentially expressed tumors. (**A**) Bar graph illustrating the enrichment levels of hallmark gene sets between CTSV-high group and CTSV-low group. Red indicates significantly enriched hallmark gene sets in CTSV-high group, and blue indicates significant enrichment in CTSV-low group. (**B**) GSEA enrichment analysis results for α/γ interferon response and inflammatory response pathways in CTSV differentially expressed groups. (**C**) GSEA enrichment analysis results for pan-cancer hallmark pathways in CTSV differentially expressed groups. (**D**) Differences in anti-tumor immune status evaluated by TIP algorithm in CTSV differentially expressed groups. (**E**) ESTIMATE Score, Immune Score, and Stromal Score evaluated by ESTIMATE algorithm in CTSV differentially expressed groups. (**F**) Differences in immune cell abundance evaluated by EPIC, MCPcounter, and quanTIseq algorithms in CTSV differentially expressed groups. (**G**) TIDE scores and MSI scores evaluated by TIDE algorithm in CTSV differentially expressed groups. * P < 0.05, ** P < 0.01, *** P < 0.001, and **** P < 0.0001.

drug resistance and the development of effective targeted therapies. Since the molecular subtyping of cancers plays a crucial role in deciphering tumor heterogeneity and guiding clinical treatment strategies, establishing comprehensive and robust prognostic models and drug selection frameworks is essential for improving prognosis prediction and facilitating personalized therapeutic approaches for GC patients. In this study, we applied ten clustering algorithms to integrate multi-omics data, constructing a prognostic subtyping system for GC based on mRNA and DNA methylation expression profiles. This classification system effectively captured distinct molecular characteristics, with CS3 tumors exhibiting significantly better OS and PFS than CS1 and CS2 tumors. The robustness and prognostic value of this classification system were further validated in an independent external cohort, confirming its potential as a promising tool for risk stratification.

The TME undergoes continuous interactions and co-evolves with tumor cells, ultimately facilitating tumor progression and growth. A comprehensive analysis of clinical and immune features at the subtype level revealed distinct omics differences among CS subtypes, each exhibiting unique prognostic characteristics. Among the three subtypes, CS3 was found to be significantly enriched in immune-related pathways, including α/γ interferon response and inflammatory response, and also exhibited higher infiltration of anti-tumor immune cells, elevated anti-tumor activity, and higher MSI scores, all of which indicate a more immunologically active tumor phenotype. At the genetic level, CS3 tumors harbor higher mutation rates and TMB in key genes such as MUC16, LRP1B,

and TTN. Notably, in GC, mutations in MUC16[53–55], TTN, and LRP1B have been associated with improved prognosis and a favorable response to ICI therapy[55–57]. These mutations contribute to an increased TMB, and both high TMB and MSI have been validated as biomarkers for predicting immunotherapy responsiveness in clinical practice[61–64]. Collectively, these findings support the hypothesis that patients with the CS3 subtype are more likely to benefit from immunotherapy. In contrast, CS2 tumors exhibit heightened activation of cell cycle-related pathways but display a profound depletion of immune function pathways. This subtype is characterized by a suppressed anti-tumor immune status with minimal immune cell infiltration, making it immunologically quiescent. Although CS1 tumors demonstrate higher immune infiltration, they also contain the highest proportion of immunosuppressive cells, including Tregs, M2 macrophages, and CAFs. CAFs secrete cytokines, chemokines, growth factors, and exosomes, which engage in dynamic crosstalk with tumor-infiltrating immune cells, contributing to an immunosuppressive TME that allows tumor cells to evade immune surveillance mechanisms[66,67]. Additionally, CS1 tumors exhibit the highest M2/M1 macrophage ratio and T-cell exhaustion scores, suggesting that although this subtype presents an inflamed immune environment, the presence of high immunosuppressive cell populations and exhausted T cells counteracts its immune activation. This results in an inflamed yet immunosuppressive TME, which may limit the efficacy of immune-based therapies.

According to the TCGA classification, among the four molecular subtypes, EBV + and MSI tumors exhibit a greater likelihood of responding to immunotherapy[13,68]. EBV + tumors display high immunogenicity due to viral infection, which induces strong immune responses and is frequently accompanied by mutations in PIK3 CA, ARID1 A, and BCOR. Additionally, PD-L1 (CD274) expression is relatively high in EBV + GC, a feature that correlates with increased CD8 + T-cell infiltration in the TME[69,70]. Similarly, MSI tumors, which are characterized by defective DNA mismatch repair mechanisms, exhibit a high mutational burden, primarily affecting genes such as PIK3 CA, ERBB2, and ERBB3[13,71]. In the subtype comparison, most EBV + and MSI subtypes were classified into CS3 subtype, which is consistent with the better immunotherapy responsiveness of CS3 and the higher expression of PD-L1 (CD274) in it. In terms of genomic characteristics, CS3 tumors also exhibited the highest immunogenicity and the most frequent mutations in PIK3 CA, ARID1 A, and BCOR among three CS subtypes. These similarities provide validation for the reliability of our analysis. Furthermore, the high infiltration of CD8 + and activated CD4 + T cells in CS3 tumors, to some extent, explains the enrichment of the IFN-γ immune response signature in them. On one hand, IFN-γ induces anti-tumor immunity, and on the other hand, it induces the expression of PD-L1 in both tumor cells and immune cells[72,73], providing evidence for the favorable survival outcome and high responsiveness to immune checkpoint inhibitors (ICIs) of CS3. Our CS subtypes not only capture the genomic features that define TCGA subtypes, but also considers the activation status of cancer-related pathways, the landscape of the tumor immune microenvironment, and the expression patterns of key genes like CTSV. The absence of CTSV expression in CS3 tumors may contribute to distinct biological behaviors, including higher anti-tumor immune cell infiltration, enhanced anti-tumor activity, and a greater likelihood of responding to immunotherapy.

Through the screening of subtype-differentiating genes, we identified CTSV as a key target molecule that exhibited significant phenotypic differences across the three CS subtypes. Notably, CTSV expression was highest in the immunologically silent CS2 subtype, which is associated with poor prognosis, and lowest in the immunologically active CS3 subtype, which is linked to a better prognosis. CTSV has been reported to play an essential role in multiple cancers, including colorectal cancer[74,75], pancreatic cancer[76,77], breast ductal cancer[78,79], bladder cancer[80], liver cancer[81], cutaneous squamous cell carcinoma[82,83], lung cancer[84,85]and other processes. In lung cancer, CTSV drives cancer progression by shaping the immunosuppressive environment and adhesion molecules cleavage, and suppresses T cell activity and encourages immunological escape from the tumor[84]. In liver cancer tissue, immunohistochemical staining results displayed that high expression of CTSV had a poor prognosis and is relevant to macrophage infiltration, as well as the effectiveness of liver cancer treatment[86]. Although ongoing studies have consistently shown that CTSV promotes multiple tumor development, its expression patterns and underlying mechanisms in gastric cancer remain unclear. Our analysis and experiments confirmed the high expression of CTSV in GC. Moreover, high CTSV expression levels were meaningfully correlated with poor prognosis and advanced pathological N and T stages. Functional analyses demonstrated that silencing CTSV expression significantly inhibited the proliferation, migration, and invasion of GC cells, indicating its potential role in tumor progression. Furthermore, the immune characteristics of tumors with high CTSV expression closely resembled those of CS2 tumors, displaying an immunologically silent phenotype with low immune infiltration. In contrast, tumors with low CTSV expression exhibited high enrichment of anti-tumor immune pathways and elevated immune cell infiltration, consistent with the immunologically active characteristics of CS3 tumors.

Our study has several advantages and novel findings. Firstly, we employed ten clustering algorithms to achieve a more comprehensive and accurate identification of molecular subtypes. This comprehensive approach enabled us to cross-validate and enhance the robustness of the three identified molecular subtypes (CS1, CS2, and CS3), minimizing the potential biases that might be introduced by a single algorithm, thereby providing more reliable subtyping results. Secondly, we conducted large-scale multi-cohort validation. The validation of molecular subtyping results is of utmost importance for the credibility of the study. We validated the clustering results across four independent GEO cohorts (GSE84437, GSE26253, GSE62254, and GSE15459), which altogether included a total of 1355 GC samples. Furthermore, the new molecular subtypes have distinct clinicopathological and immunological characteristics, and providing insights into the intratumor heterogeneity and potential therapeutic strategies for GC. The CS1 subtype is characterized by an inflamed yet immunosuppressive microenvironment. The CS2 subtype presents a unique pattern of immune exhaustion along with the upregulation of CTSV. The CS3 subtype shows an immunologically active state and a significant downregulation of CTSV, all of which have not been reported in previous subtyping studies. In addition, our study is the first to demonstrate the role of CTSV

as a potential prognostic marker and subtype classifier in GC, focusing not only on the promoting effect of CTSV on GC progression in vitro and in vivo, but also on its unique differences among the molecular subtypes of GC.

In summary, this study integrated mRNA and DNA methylation data using ten clustering algorithms to establish a molecular subtyping system for GC with validated prognostic and treatment selection potential through multi-omics cross-validation in external cohorts. A comprehensive analysis of clinical and immune features revealed distinct omics differences among CS subtypes, supporting their risk stratification capability. CTSV was identified as a key molecule that can effectively indicate prognosis and identify phenotypic features. For the first time, we determined its high expression pattern in GC and its role in GC development. The differences in the immunological microenvironment between CTSV differential expression groups also suggest a close association with GC tumor immunity.

Data availability

The datasets extracted and analysed during the current study are available from TCGA (https://portal.gdc.cancer.gov/) and GEO (https://www.ncbi.nlm.nih.gov/geo/) datasets. The original available statements demonstrated in the study are summarized in the article materials. Additional inquiries should be submitted to the corresponding authors.

Received: 18 December 2024; Accepted: 31 March 2025
Published online: 25 April 2025

References

1. Smyth, E. C., Nilsson, M., Grabsch, H. I., van Grieken, N. C. & Lordick, F. Gastric cancer. *Lancet* **396** (10251), 635–648 (2020).
2. Green, P. H., O'Toole, K. M., Slonim, D., Wang, T. & Weg, A. Increasing incidence and excellent survival of patients with early gastric cancer: experience in a united States medical center. *Am. J. Med.* **85** (5), 658–661 (1988).
3. Lei, Z. et al. Identification of molecular subtypes of gastric cancer with different responses to PI3-kinase inhibitors and 5-fluorouracil. *Gastroenterology* **145** (3), 554–565 (2013).
4. Schwartz, G. K. Invasion and metastases in gastric cancer: in vitro and in vivo models with clinical correlations. *Semin Oncol.* **23** (3), 316–324 (1996).
5. Guan, W. L., He, Y. & Xu, R. H. Gastric cancer treatment: recent progress and future perspectives. *J. Hematol. Oncol.* **16** (1), 57 (2023).
6. Wagner, A. D. et al. Chemotherapy for advanced gastric cancer. *Cochrane Database Syst. Rev.* **8**, (8), (2017). Cd004064.
7. Song, Z., Wu, Y., Yang, J., Yang, D. & Fang, X. Progress in the treatment of advanced gastric cancer. *Tumour Biol.* **39** (7), 1010428317714626 (2017).
8. Li, X., Xu, J., Xie, J. & Yang, W. Research progress in targeted therapy and immunotherapy for gastric cancer. *Chin. Med. J. (Engl).* **135** (11), 1299–1313 (2022).
9. Li, K., Zhang, A., Li, X., Zhang, H. & Zhao, L. Advances in clinical immunotherapy for gastric cancer. *Biochim. Biophys. Acta Rev. Cancer.* **1876** (2), 188615 (2021).
10. Coutzac, C., Pernot, S., Chaput, N. & Zaanan, A. Immunotherapy in advanced gastric cancer, is it the future? *Crit. Rev. Oncol. Hematol.* **133**, 25–32 (2019).
11. Palucka, A. K. & Coussens, L. M. The basis of oncoimmunology. *Cell* **164** (6), 1233–1247 (2016).
12. Chia, N. Y. & Tan, P. Molecular classification of gastric cancer. *Ann. Oncol.* **27** (5), 763–769 (2016).
13. Comprehensive molecular characterization of gastric adenocarcinoma. *Nature* **513**, (7517), 202–209. (2014).
14. Pietrantonio, F. et al. Individual patient data Meta-Analysis of the value of microsatellite instability as a biomarker in gastric cancer. *J. Clin. Oncol.* **37** (35), 3392–3400 (2019).
15. Cristescu, R. et al. Molecular analysis of gastric cancer identifies subtypes associated with distinct clinical outcomes. *Nat. Med.* **21** (5), 449–456 (2015).
16. Subramanian, I., Verma, S., Kumar, S., Jere, A. & Anamika, K. Multi-omics data integration, interpretation, and its application. *Bioinform Biol. Insights.* **14**, 1177932219899051 (2020).
17. Hasin, Y., Seldin, M. & Lusis, A. Multi-omics approaches to disease. *Genome Biol.* **18** (1), 83 (2017).
18. Yan, J., Risacher, S. L., Shen, L. & Saykin, A. J. Network approaches to systems biology analysis of complex disease: integrative methods for multi-omics data. *Brief. Bioinform.* **19** (6), 1370–1381 (2018).
19. Wang, Z., Jensen, M. A. & Zenklusen, J. C. A practical guide to the cancer genome atlas (TCGA). *Methods Mol. Biol.* **1418**, 111–141 (2016).
20. Mak, T. K. et al. The cancer-associated fibroblast-related signature predicts prognosis and indicates immune microenvironment infiltration in gastric cancer. *Front. Immunol.* **13**, 951214 (2022).
21. Huo, J., Xie, W., Fan, X. & Sun, P. Pyroptosis, apoptosis, and necroptosis molecular subtype derived prognostic signature universal applicable for gastric cancer-A large sample and multicenter retrospective analysis. *Comput. Biol. Med.* **149**, 106037 (2022).
22. Peng, K. et al. A 16-mRNA signature optimizes recurrence-free survival prediction of stages II and III gastric cancer. *J. Cell. Physiol.* **235** (7–8), 5777–5786 (2020).
23. Tian, Y. et al. ChAMP: updated methylation analysis pipeline for illumina beadchips. *Bioinformatics* **33** (24), 3982–3984 (2017).
24. Zhou, W., Laird, P. W. & Shen, H. Comprehensive characterization, annotation and innovative use of infinium DNA methylation BeadChip probes. *Nucleic Acids Res.* **45**, (4), e22. (2017).
25. Lu, X., Meng, J., Zhou, Y., Jiang, L. & Yan, F. MOVICS: an R package for multi-omics integration and visualization in cancer subtyping. *Bioinformatics* **36** (22–23), 5539–5541 (2021).
26. Pierre-Jean, M., Deleuze, J. F., Le Floch, E. & Mauger, F. Clustering and variable selection evaluation of 13 unsupervised methods for multi-omics data integration. *Brief. Bioinform.* **21** (6), 2011–2030 (2020).
27. Rappoport, N. & Shamir, R. Multi-omic and multi-view clustering algorithms: review and cancer benchmark. *Nucleic Acids Res.* **46** (20), 10546–10562 (2018).
28. Mo, Q. et al. A fully bayesian latent variable model for integrative clustering analysis of multi-type omics data. *Biostatistics* **19** (1), 71–86 (2018).
29. Ritchie, M. E. et al. K., Limma powers differential expression analyses for RNA-sequencing and microarray studies. *Nucleic Acids Res.* **43**, (7), e47. (2015).
30. Wu, T. et al. clusterProfiler 4.0: A universal enrichment tool for interpreting omics data. *Innovation (Camb)* 2, (3), 100141. (2021).
31. Liberzon, A. et al. The molecular signatures database (MSigDB) hallmark gene set collection. *Cell. Syst.* **1** (6), 417–425 (2015).
32. Meng, J. et al. Tumor immune microenvironment-based classifications of bladder cancer for enhancing the response rate of immunotherapy. *Mol. Ther. Oncolytics.* **20**, 410–421 (2021).

33. Chen, B., Khodadoust, M. S., Liu, C. L., Newman, A. M. & Alizadeh, A. A. Profiling tumor infiltrating immune cells with CIBERSORT. *Methods Mol. Biol.* **1711**, 243–259 (2018).
34. Yoshihara, K. et al. Inferring tumour purity and stromal and immune cell admixture from expression data. *Nat. Commun.* **4**, 2612 (2013).
35. Zhou, L., Huang, W., Yu, H. F., Feng, Y. J. & Teng, X. Exploring TCGA database for identification of potential prognostic genes in stomach adenocarcinoma. *Cancer Cell. Int.* **20**, 264 (2020).
36. Alonso, M. H. et al. Comprehensive analysis of copy number aberrations in microsatellite stable colon cancer in view of stromal component. *Br. J. Cancer*. **117** (3), 421–431 (2017).
37. Xu, W. H. et al. Prognostic value and immune infiltration of novel signatures in clear cell renal cell carcinoma microenvironment. *Aging (Albany NY)*. **11** (17), 6999–7020 (2019).
38. Jia, D. et al. Mining TCGA database for genes of prognostic value in glioblastoma microenvironment. *Aging (Albany NY)*. **10** (4), 592–605 (2018).
39. Xiang, S. et al. Identification of prognostic genes in the tumor microenvironment of hepatocellular carcinoma. *Front. Immunol.* **12**, 653836 (2021).
40. Zheng, Y. et al. Immune suppressive landscape in the human esophageal squamous cell carcinoma microenvironment. *Nat. Commun.* **11** (1), 6268 (2020).
41. Jiang, P. et al. Signatures of T cell dysfunction and exclusion predict cancer immunotherapy response. *Nat. Med.* **24** (10), 1550–1558 (2018).
42. Charoentong, P. et al. Pan-cancer Immunogenomic analyses reveal Genotype-Immunophenotype relationships and predictors of response to checkpoint Blockade. *Cell. Rep.* **18** (1), 248–262 (2017).
43. Mayakonda, A., Lin, D. C., Assenov, Y., Plass, C. & Koeffler, H. P. Maftools: efficient and comprehensive analysis of somatic variants in cancer. *Genome Res.* **28** (11), 1747–1756 (2018).
44. Yang, W. et al. Genomics of drug sensitivity in cancer (GDSC): a resource for therapeutic biomarker discovery in cancer cells. *Nucleic Acids Res.* **41**, (2013). (Database issue), D955-61.
45. Geeleher, P., Cox, N. & Huang, R. S. pRRophetic: an R package for prediction of clinical chemotherapeutic response from tumor gene expression levels. *PLoS One* **9**, (9), e107468. (2014).
46. Kanehisa, M., Furumichi, M., Sato, Y., Matsuura, Y. & Ishiguro-Watanabe, M. KEGG: biological systems database as a model of the real world. *Nucleic Acids Res.* **53** (D1), D672–d677 (2025).
47. Kanehisa, M. Toward Understanding the origin and evolution of cellular organisms. *Protein Sci.* **28** (11), 1947–1951 (2019).
48. Kanehisa, M. & Goto, S. KEGG: Kyoto encyclopedia of genes and genomes. *Nucleic Acids Res.* **28** (1), 27–30 (2000).
49. Ren, N., Liang, B. & Li, Y. Identification of prognosis-related genes in the tumor microenvironment of stomach adenocarcinoma by TCGA and GEO datasets. *Biosci Rep* 40, (10). (2020).
50. Zhu, X. et al. Potential Prognostic Value and Mechanism of Stromal-Immune Signature in Tumor Microenvironment for Stomach Adenocarcinoma. *Biomed Res Int* 2020, 4673153. (2020).
51. Park, R., Da Silva, L. L. & Saeed, A. Immunotherapy Predictive Molecular Markers in Advanced Gastroesophageal Cancer: MSI and Beyond. *Cancers (Basel)* 13, (7). (2021).
52. Ozer, M. et al. Neoadjuvant Immunotherapy for Patients with dMMR/MSI-High Gastrointestinal Cancers: A Changing Paradigm. *Cancers (Basel)* 15, (15). (2023).
53. Li, X., Pasche, B., Zhang, W. & Chen, K. Association of MUC16 mutation with tumor mutation load and outcomes in patients with gastric cancer. *JAMA Oncol.* **4** (12), 1691–1698 (2018).
54. Smyth, E. C. & Fitzgerald, R. C. MUC16 mutations and prognosis in gastric cancer: A little goes a long way. *JAMA Oncol.* **4** (12), 1698–1699 (2018).
55. Yang, Y. et al. MUC4, MUC16, and TTN genes mutation correlated with prognosis, and predicted tumor mutation burden and immunotherapy efficacy in gastric cancer and pan-cancer. *Clin. Transl Med.* **10**, (4), e155. (2020).
56. Brown, L. C. et al. LRP1B mutations are associated with favorable outcomes to immune checkpoint inhibitors across multiple cancer types. *J Immunother Cancer* 9, (3). (2021).
57. Jia, Q., Wang, J., He, N., He, J. & Zhu, B. Titin mutation associated with responsiveness to checkpoint blockades in solid tumors. *JCI Insight* 4, (10). (2019).
58. Jiang, T., Chen, X., Su, C., Ren, S. & Zhou, C. Pan-cancer analysis of ARID1A alterations as biomarkers for immunotherapy outcomes. *J. Cancer*. **11** (4), 776–780 (2020).
59. Li, J. et al. Epigenetic driver mutations in ARID1A shape cancer immune phenotype and immunotherapy. *J. Clin. Invest.* **130** (5), 2712–2726 (2020).
60. Vokshi, B. H. & Toska, E. Mutant ARID1A: igniting cancer immunotherapy. *Trends Immunol.* **45** (8), 565–567 (2024).
61. Goodman, A. M. et al. Tumor mutational burden as an independent predictor of response to immunotherapy in diverse cancers. *Mol. Cancer Ther.* **16** (11), 2598–2608 (2017).
62. Palmeri, M. et al. Real-world application of tumor mutational burden-high (TMB-high) and microsatellite instability (MSI) confirms their utility as immunotherapy biomarkers. *ESMO Open.* **7** (1), 100336 (2022).
63. Samstein, R. M. et al. Tumor mutational load predicts survival after immunotherapy across multiple cancer types. *Nat. Genet.* **51** (2), 202–206 (2019). L. G. T.
64. Jardim, D. L., Goodman, A., de Melo Gagliato, D. & Kurzrock, R. The challenges of tumor mutational burden as an immunotherapy biomarker. *Cancer Cell.* **39** (2), 154–173 (2021).
65. Weinstein, J. N. et al. The cancer genome atlas Pan-Cancer analysis project. *Nat. Genet.* **45** (10), 1113–1120 (2013).
66. Mao, X. et al. Crosstalk between cancer-associated fibroblasts and immune cells in the tumor microenvironment: new findings and future perspectives. *Mol. Cancer*. **20** (1), 131 (2021).
67. Pei, L. et al. Roles of cancer-associated fibroblasts (CAFs) in anti- PD-1/PD-L1 immunotherapy for solid cancers. *Mol. Cancer*. **22** (1), 29 (2023).
68. Derks, S. et al. Abundant PD-L1 expression in Epstein-Barr Virus-infected gastric cancers. *Oncotarget* **7** (22), 32925–32932 (2016).
69. Xing, X. et al. Analysis of PD1, PDL1, PDL2 expression and T cells infiltration in 1014 gastric cancer patients. *Oncoimmunology* **7**, (3), e1356144. (2018).
70. Thompson, E. D. et al. Patterns of PD-L1 expression and CD8 T cell infiltration in gastric adenocarcinomas and associated immune stroma. *Gut* **66** (5), 794–801 (2017).
71. Le, D. T. et al. PD-1 Blockade in tumors with Mismatch-Repair deficiency. *N Engl. J. Med.* **372** (26), 2509–2520 (2015).
72. Mandai, M. et al. Dual faces of IFNγ in cancer progression: A role of PD-L1 induction in the determination of Pro- and antitumor immunity. *Clin. Cancer Res.* **22** (10), 2329–2334 (2016).
73. Castro, F., Cardoso, A. P., Gonçalves, R. M., Serre, K. & Oliveira, M. J. Interferon-Gamma at the crossroads of tumor immune surveillance or evasion. *Front. Immunol.* **9**, 847 (2018).
74. Santamaría, I. et al. Cathepsin L2, a novel human cysteine proteinase produced by breast and colorectal carcinomas. *Cancer Res.* **58** (8), 1624–1630 (1998).
75. Wang, C. H. et al. Cathepsin V mediates the Tazarotene-induced gene 1-induced reduction in invasion in colorectal cancer cells. *Cell. Biochem. Biophys.* **78** (4), 483–494 (2020).
76. Al-Hashimi, A. et al. Significance of nuclear cathepsin V in normal thyroid epithelial and carcinoma cells. *Biochim. Biophys. Acta Mol. Cell. Res.* **1867** (12), 118846 (2020).

77. Yasuda, Y. et al. Cathepsin V, a novel and potent elastolytic activity expressed in activated macrophages. *J. Biol. Chem.* **279** (35), 36761–36770 (2004).
78. Sereesongsaeng, N., McDowell, S. H., Burrows, J. F., Scott, C. J. & Burden, R. E. Cathepsin V suppresses GATA3 protein expression in luminal A breast cancer. *Breast Cancer Res.* **22** (1), 139 (2020).
79. Toss, M. et al. Prognostic significance of cathepsin V (CTSV/CTSL2) in breast ductal carcinoma in situ. *J. Clin. Pathol.* **73** (2), 76–82 (2020).
80. Xia, Y., Ge, M., Xia, L., Shan, G. & Qian, H. CTSV (cathepsin V) promotes bladder cancer progression by increasing NF-κB activity. *Bioengineered* **13** (4), 10180–10190 (2022).
81. Jing, J., Wang, S., Ma, J., Yu, L. & Zhou, H. Elevated CTSL2 expression is associated with an adverse prognosis in hepatocellular carcinoma. *Int. J. Clin. Exp. Pathol.* **11** (8), 4035–4043 (2018).
82. Chen, N., Seiberg, M. & Lin, C. B. Cathepsin L2 levels inversely correlate with skin color. *J. Invest. Dermatol.* **126** (10), 2345–2347 (2006).
83. Hagemann, S. et al. The human cysteine protease cathepsin V can compensate for murine cathepsin L in mouse epidermis and hair follicles. *Eur. J. Cell. Biol.* **83** (11-12), 775–780 (2004).
84. Zhu, L., Zeng, Q., Wang, J., Deng, F. & Jin, S. Cathepsin V drives lung cancer progression by shaping the immunosuppressive environment and adhesion molecules cleavage. *Aging (Albany NY).* **15** (23), 13961–13979 (2023).
85. Yang, L. et al. Glycosylated cathepsin V serves as a prognostic marker in lung cancer. *Front. Oncol.* **12**, 876245 (2022).
86. Liu, J. et al. Cathepsin V is correlated with the prognosis and tumor microenvironment in liver cancer. *Mol. Carcinog.* **63** (3), 400–416 (2024).

Acknowledgements
We would like to extend our sincere gratitude to Research Center of Clinical Medicine for providing the experimental equipment. We appreciate all those who provided beneficial suggestions and comments to help improve the quality of our paper.

Author contributions
Conceptualization, H.D. and X.W.; methodology, H.D. and J.R.; validation, H.D.; writing—original draft preparation, H.D., J.R. and C.W.; writing—review and editing, J.H.; visualization, H.D.; Resources, J.H. All authors have read and agreed to the published version of the manuscript.

Funding
This research was supported by the National Natural Science Foundation of China (81972015).

Declarations

Competing interests
The authors declare no competing interests.

Ethics approval and consent to participate
The study was conducted in accordance with the Declaration of Helsinki (as revised in 2013). The studies involving humans were approved by Medical Ethics Committee of Affiliated Hospital of Nantong University.

Additional information
Supplementary Information The online version contains supplementary material available at https://doi.org/10.1038/s41598-025-96686-0.

Correspondence and requests for materials should be addressed to X.W.