CANCER INNOVATION

ORIGINAL ARTICLE

Identification of collagen subtypes of gastric cancer for distinguishing patient prognosis and therapeutic response

Di Wang[1] 🟢 | **Jing Zhang**[2] | **Jianchao Wang**[2] | **Zhonglin Cai**[3] | **Shanfeng Jin**[1] | **Gang Chen**[2]

[1]Department of Molecular Pathology, Clinical Oncology School of Fujian Medical University, Fujian Cancer Hospital, Fuzhou, China

[2]Department of Pathology, Clinical Oncology School of Fujian Medical University, Fujian Cancer Hospital, Fuzhou, China

[3]Department of Urology, Gongli Hospital of Shanghai Pudong New Area, Shanghai, China

Correspondence
Gang Chen, Department of Pathology, Clinical Oncology School of Fujian Medical University, Fujian Cancer Hospital, Fuzhou 350014, China.
Email: naichengang@126.com

Abstract

Background: Gastric cancer is a highly heterogeneous disease, presenting a major obstacle to personalized treatment. Effective markers of the immune checkpoint blockade response are needed for precise patient classification. We, therefore, divided patients with gastric cancer according to collagen gene expression to indicate their prognosis and treatment response.

Methods: We collected data for 1250 patients with gastric cancer from four cohorts. For the TCGA-STAD cohort, we used consensus clustering to stratify patients based on expression levels of 44 collagen genes and compared the prognosis and clinical characteristics between collagen subtypes. We then identified distinct transcriptomic and genetic alteration signatures for the subtypes. We analyzed the associations of collagen subtypes with the responses to chemotherapy, immunotherapy, and targeted therapy. We also established a platform-independent collagen-subtype predictor. We verified the findings in three validation cohorts (GSE84433, GSE62254, and GSE15459) and compared the collagen subtyping method with other molecular subtyping methods.

Results: We identified two subtypes of gastric adenocarcinoma: a high-expression collagen subtype (CS-H) and a low-expression collagen subtype (CS-L). Collagen subtype was an independent prognostic factor, with better

---

Abbreviations: AUC, area under the curve; BP, biological process; CAF, cancer-associated fibroblast; CNV, copy number variation; CS-H, high-expression collagen subtype; CS-L, low-expression collagen subtype; DEG, differentially expressed gene; EMT, epithelial mesenchymal transition; FDR, false discovery rate; GDSC, Genomics of Drug Sensitivity in Cancer; GEO, Gene Expression Omnibus; GO, Gene Ontology; GSEA, gene set enrichment analysis; GSVA, gene set variation analysis; IC$_{50}$, half-maximal inhibitory concentration; ICB, immune checkpoint blockade; KEGG, Kyoto Encyclopedia of Genes and Genomes; LASSO, least absolute shrinkage and selection operator; MDSC, myeloid-derived suppressor cell; MSigDB, Molecular Signatures Database; MSI-H, microsatellite instability–high; OS, overall survival; PAC, proportion of ambiguous clustering; PCA, principal component analysis; PI3K, phosphoinositide 3-kinase; PLAGE, pathway level analysis of gene expression; RFE, recursive feature elimination; ROC, receiver operating characteristic; ssGSEA, single-sample gene set enrichment analysis; TAM M2, tumor-associated macrophage M2 type; TCGA, The Cancer Genome Atlas; TIDE, Tumor Immune Dysfunction and Exclusion; TMB, tumor mutational burden; TME, tumor microenvironment.

---

overall survival in the CS-L subgroup. The inflammatory response, angiogenesis, and phosphoinositide 3-kinase (PI3K)/Akt pathways were transcriptionally active in the CS-H subtype, while DNA repair activity was significantly greater in the CS-L subtype. *PIK3CA* was frequently amplified in the CS-H subtype, while *PIK3C2A*, *PIK3C2G*, and *PIK3R1* were frequently deleted in the CS-L subtype. CS-H subtype tumors were more sensitive to fluorouracil, while CS-L subtype tumors were more sensitive to immune checkpoint blockade. CS-L subtype was predicted to be more sensitive to HER2-targeted drugs, and CS-H subtype was predicted to be more sensitive to vascular endothelial growth factor and PI3K pathway-targeting drugs. Collagen subtyping also has the potential to be combined with existing molecular subtyping methods for better patient classification.

Conclusions: We classified gastric cancers into two subtypes based on collagen gene expression and validated these subtypes in three validation cohorts. The collagen subgroups differed in terms of prognosis, clinical characteristics, transcriptome, and genetic alterations. The subtypes were closely related to patient responses to chemotherapy, immunotherapy, and targeted therapy.

K E Y W O R D S
biomarker, cancer classification, collagen, gastric cancer, phosphatidylinositol 3-kinase

1 | INTRODUCTION

Gastric cancer is the fifth most common cancer and the fourth leading cause of cancer-related mortality globally [1]. East Asia, Eastern Europe, and South America are hotspots of incidence for gastric cancer [1, 2]. It is a highly heterogeneous disease in terms of genetics, histopathology, and treatment response, which presents a major obstacle to personalized treatment [3, 4]. The efficacy of immune checkpoint blockade (ICB) for the treatment of gastric cancer has recently been well validated; however, the existing biomarker [microsatellite instability–high (MSI-H)] cannot accurately predict ICB response, and 50% of MSI-H patients are resistant to programmed cell death protein-1 (PD-1) blockade [5]. Furthermore, classification systems based on genomic status have not been effectively translated into treatment stratification or therapeutic strategies [2, 6]. There is thus an urgent need for alternative molecular subtypes that are more clinically applicable and able to predict therapeutic response.

The tumor microenvironment (TME) has received extensive attention in recent years because of its ability to modulate tumor growth and progression and influence treatment response [7–9]. The TME comprises nontumor cells, including fibroblasts, immune cells, and endothelial cells, and the extracellular matrix [9]. Collagens are the main component of the extracellular matrix, and they are remodeled in tumors by cancer-associated fibroblasts (CAFs) [10, 11]. Collagens can regulate tumor behavior and promote immunotherapy resistance in cancer through CD8$^+$ T-cell exhaustion [12, 13] and may thus be biomarkers for the diagnosis, prognosis, and classification of a variety of tumors [10, 14, 15]. The collagen signature is also a marker of lymph node metastasis, peritoneal metastasis, and prognosis in patients with gastric cancer [16–18].

In this study, we identified two molecular subtypes of gastric cancer based on the expression patterns of collagen genes, with distinct clinical features and prognoses. Angiogenesis, DNA repair, and the inflammatory response differed significantly between the collagen subtypes, and the phosphoinositide 3-kinase (PI3K)/Akt pathway was significantly upregulated in patients with high collagen expression. Importantly, the collagen subtype was closely related to patient responses to chemotherapy, immunotherapy, and targeted therapy.

2 | METHODS

2.1 | Patients and data sets

Data for 1250 patients with gastric cancer were collected from four cohorts: The Cancer Genome Atlas (TCGA)-STAD [3], GSE84433 [19], GSE62254 [20], and GSE15459 [21]

(Table S1). For the TCGA-STAD cohort, as the derivation cohort, multiomics data, including messenger RNA expression, protein expression, somatic mutations, copy number data, and clinical information, were downloaded from GDC PanCanAtlas Publications. The drug responses of patients from TCGA were curated by Ding et al. [22].

The other three transcriptome data sets (GSE84433, GSE62254, and GSE15459) were used as validation cohorts and downloaded from the Gene Expression Omnibus (GEO), and clinical information was obtained from corresponding reports.

PRJEB25780, which contains tumor transcriptome data for 45 patients treated with PD-1 blockade, was downloaded from the European Nucleotide Archive (Table S1) [23]. RNA sequence reads were aligned to the human reference genome (GRCh38, GENCODE Release 38) using STAR [24], and RSEM was used to quantify gene expression [25]. We used a regularized logarithm ("rlog" function from DESeq2) to normalize gene quantification for collagen subtyping analysis [26]. Analysis of each cohort is depicted in Figure S1.

2.2 | Curation of a collagen gene set

The collagen gene set was derived from the collagen gene group (HUGO Gene Nomenclature Committee, ID: 490) and contained 44 genes after removing non-protein-coding genes (Table S2).

2.3 | Collagen subtyping

Consensus clustering of collagen genes at the transcript level was performed using ConsensusClusterPlus v1.50.0 (parameters: maxK = 6, pItem = 0.8, pFeature = 1, reps = 10,000, seed = 31415) [27]. Ward.D and the Euclidean distance were used as the clustering algorithm and distance metric, respectively. The proportion of ambiguous clustering (PAC) and consensus matrix were adopted to assess the optimal cluster number (k) [28].

2.4 | Transcriptome analysis

Differential expression analysis was performed using limma, and genes with a false discovery rate (FDR) ≤ 0.01 were defined as differentially expressed genes (DEGs) [29]. We used clusterProfiler [30] or fgsea [31] to perform gene set enrichment analysis (GSEA) of the Molecular Signatures Database (MSigDB) [32], REAC-TOME [33], and Kyoto Encyclopedia of Genes and Genomes (KEGG) [34] gene sets (parameters: nperm = 100,000, minSize = 10, maxSize = 500), and FDR ≤ 0.05 was considered to indicate statistical significance. Gene Ontology (GO) biological process (BP) enrichment analysis was performed using clusterProfiler, and FDR ≤ 0.05 was regarded as statistically significant. The single-sample GSEA (ssGSEA) algorithm from the GSVA package was applied to calculate the ssGSEA scores to evaluate the expression activity of the gene set [35]. We used the RTN package [36] to evaluate the regulon activity of gastric cancer driver genes [37] at the transcriptome level. The data in Figure S4 were visualized via Pathview [38].

2.5 | Genetic alteration analysis

The mutation file (mc3.v0.2.8.PUBLIC.maf.gz) and copy number file (broad.mit.edu_PANCAN_Genome_Wide_SNP_6_whitelisted.seg) were downloaded for genetic alteration analysis. GISTIC 2.0 software was used to analyze the copy number file and the command line parameters from the GDC documentation were used (https://docs.gdc.cancer.gov/Data/Bioinformatics_Pipelines/CNV_Pipeline/) [39]. We then utilized maftools to analyze mutation data and copy number variation (CNV) analysis results from GISTIC separately to obtain the gene-level alteration profile of each patient [40].

2.6 | Prediction of treatment response

We used the Tumor Immune Dysfunction and Exclusion (TIDE) response prediction module to predict patient response to ICB based on expression profiles [41, 42]. TIDE predicts the ICB response by evaluating multiple transcriptomic biomarkers, including the TIDE score, T-cell dysfunction score, T-cell exclusion score, and scores for three cell types—myeloid-derived suppressor cells (MDSCs), tumor-associated macrophages M2 type (TAM M2s), and CAFs—reported to restrict T-cell infiltration in tumors. A higher TIDE score was associated with a worse ICB response.

We also used the pRRophetic package to impute patient responses to chemotherapy and multiple targeted drugs. A drug-response predictive model based on gene expression was constructed via pRRophetic based on drug response data and baseline gene expression levels from cancer cell lines in the Genomics of Drug Sensitivity in Cancer (GDSC) database. We therefore inputted gene expression data for the patients with gastric cancer and used the predictive model to estimate patient sensitivity

to specific drugs, measured as the half-maximal inhibitory concentration (IC_{50}) [43].

2.7 | Construction and validation of collagen-subtype predictor

The TCGA-STAD cohort was used as the training cohort, and the three GEO cohorts were used as validation cohorts. We first performed feature selection for the collagen subtype in the training cohort based on the expression of 44 collagen genes, using the Boruta algorithm [44], recursive feature elimination (RFE) algorithm [45], and least absolute shrinkage and selection operator (LASSO) algorithm [46]. Three signatures were obtained. In the context of the transcriptome, we then calculated the score for each signature using four unsupervised methods provided by the GSVA package—gene set variation analysis (GSVA), ssGSEA, z score, and pathway level analysis of gene expression (PLAGE)—and determined the optimal threshold in the training cohort [35, 47]. We finally evaluated the performance of these 12 models in three validation cohorts and selected the model with the highest area under the curve (AUC) as the collagen-subtype predictor.

2.8 | Pancancer collagen subtyping

For pancancer analysis, we first excluded unqualified samples based on merged sample quality annotations and retained cancer types with a sample size ≥100. We used the same analysis pipeline as for the TCGA-STAD analysis to perform collagen subtyping and survival analysis for the retained cancer types.

2.9 | Statistical analyses

Intergroup differences in continuous variables were analyzed using Student's *t*-test or Wilcoxon's rank sum test. Associations between two categorical variables were analyzed using Fisher's exact test. Prognostic associations between collagen subtypes and clinical features were evaluated via Kaplan–Meier curves, log-rank tests, and Cox proportional hazards regression models. The pROC package was used to construct the receiver operating characteristic (ROC) curve of the model in each cohort and obtain the AUC, and the optimal threshold was determined using the "closest.topleft" method in the training cohort [47]. All statistical analyses were performed using R, version 4.1.1.

3 | RESULTS

3.1 | Consensus clustering of collagen gene expression identified two subtypes of gastric cancer

We stratified gastric cancers based on the expression levels of 44 collagen genes by unsupervised consensus clustering of RNA-seq data from TCGA-STAD. The PAC was smallest when the cluster number was 2, which was determined as the optimal cluster number (Figure 1a). A heatmap of the consensus matrix showed that the two subtypes could be clearly separated (Figure 1b). The expression of collagen genes in the blue-subtype tumors was relatively low, and this subtype was named the low-expression collagen subtype (CS-L), while the orange subtype was named the high-expression collagen subtype (CS-H) (Figure 1c). The assignments of these subtypes were validated using principal component analysis (PCA), which confirmed the robust difference in expression patterns between these two collagen subtypes (Figure 1d). CS-H subtype tumors showed increased transcriptional activity of collagen gene sets (Figure 1e) and protein expression levels of type VI collagen (Figure 1f), consistent with the transcription level of collagen genes.

To assess the clinical importance of the collagen subtype, we examined the correlations between the collagen subtype and overall survival (OS) time. The collagen subtype was an independent prognostic factor, and CS-L patients had a better prognosis than CS-H patients (Figure 1g,h). CS-H was associated with worse survival, AJCC stage, grade, and T stage than CS-L, according to Fisher's exact tests, suggesting that highly malignant gastric cancer is associated with high collagen expression (Figure 1i). We further confirmed the robustness of the collagen-subtype classification in the GSE84433, GSE62254, and GSE15459 cohorts (Figure 2; Figure S2). We thus identified two subgroups of gastric cancer based on collagen gene expression and determined the clinical significance of the classification system.

3.2 | Collagen subtypes were transcriptionally distinct in terms of cancer hallmarks

We further characterized the transcriptome differences between the two collagen subtypes by differential expression analysis of TCGA-STAD RNA-Seq data. We compared CS-H and CS-L samples and identified 8605 DEGs, including 3571 downregulated and 5034 upregulated genes (Figure 3a). Enrichment analysis of these
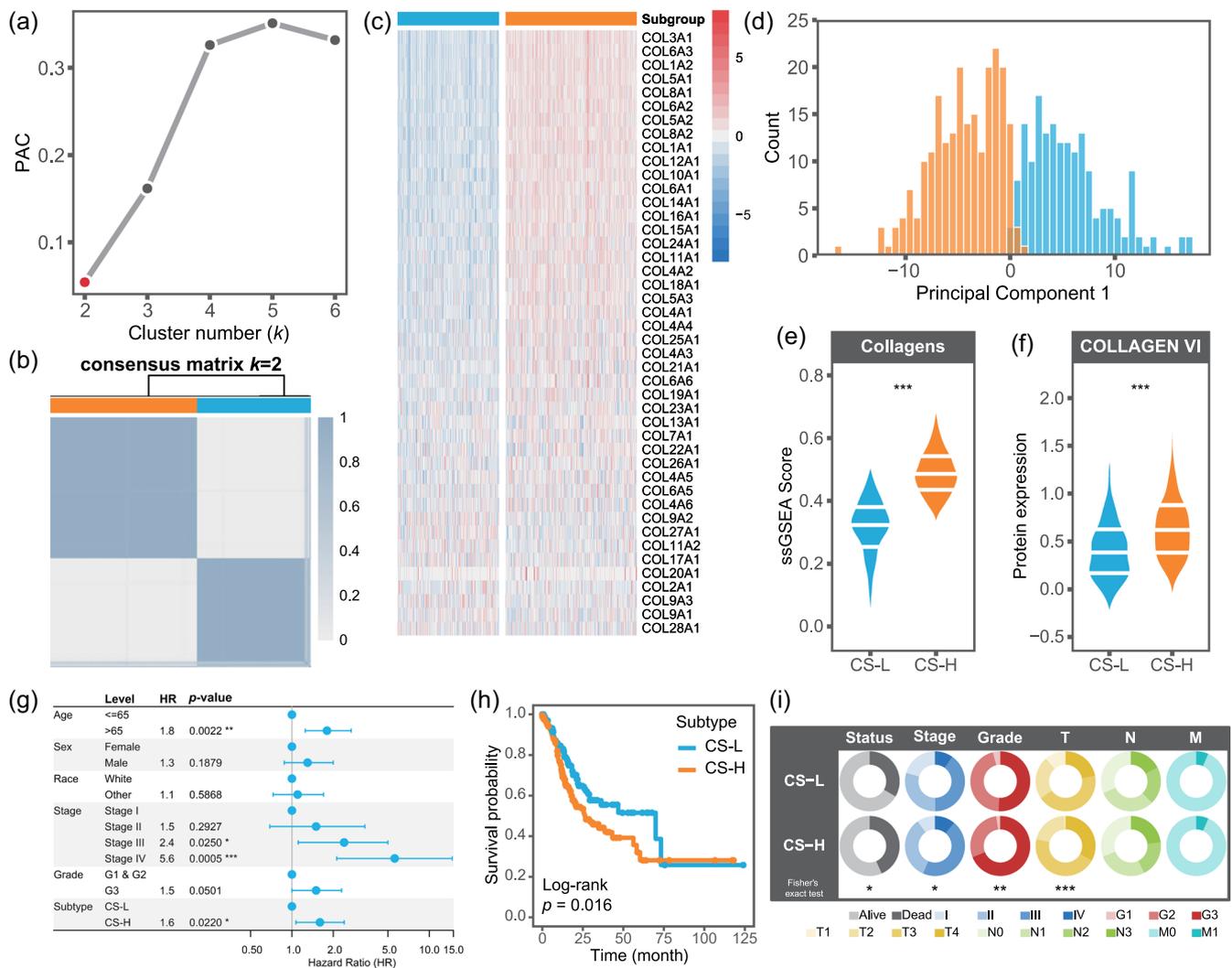
FIGURE 1 Identification of two collagen subtypes in gastric cancer from the TCGA-STAD cohort. (a) PAC curve for each k. $k = 2$ with the lowest PAC was the optimal k. (b) Consensus matrix for $k = 2$. (c) Heatmap of expression patterns of 44 collagen genes. Each column represents one patient in the cohort. (d) PCA results for gastric cancer samples. The first principal component scores showed that CS-H- and CS-L-subtype tumors clustered separately. (e) Distribution of collagen ssGSEA scores between the two collagen subtypes. ***$p < 0.001$. (f) Distribution of type VI collagen protein expression levels between the two collagen subtypes. ***$p < 0.001$. (g) Forest plot with hazard ratios of clinical variables and collagen subtypes according to multivariate Cox proportional analysis of OS. (h) Kaplan–Meier curves for OS in patients with the two collagen subtypes. (i) Comparison of clinical characteristics between the two collagen subtypes. *$p < 0.05$, **$p < 0.01$, and ***$p < 0.001$. OS, overall survival; PAC, proportion of ambiguous clustering; TCGA, The Cancer Genome Atlas.

genes revealed subtype-specific BPs. Upregulated genes in the CS-H subgroup were enriched in angiogenesis, extracellular matrix organization, calcium ion homeostasis, and leukocyte migration (Figure 3b).

GSEA of hallmark gene sets in MSigDB showed that epithelial mesenchymal transition (EMT), inflammatory response, hypoxia, and KRAS signaling were enriched in the CS-H subgroup, and DNA repair and MYC targets were enriched in the CS-L subgroup (Figure 3c). We then validated cancer-related hallmarks through ssGSEA, which confirmed that EMT, angiogenesis, inflammatory response, hypoxia, and apoptosis were significantly

enriched in the CS-H subgroup, while DNA repair was significantly enriched in the CS-L subgroup (Figure 3d). These results were confirmed in the GSE84433, GSE62254, and GSE15459 cohorts (Figure S3). We further analyzed the activity of gastric cancer driver genes as regulators in the transcriptional regulatory network and found that the regulon activity of driver genes exhibited collagen-subtype specificity (Figure 3e). Regulon activity of *PIK3CA* was significantly increased in the CS-H subgroup (Figure 3e). Overall, these results indicate that tumors with different collagen subtypes exhibit biological differences at the transcriptome level.
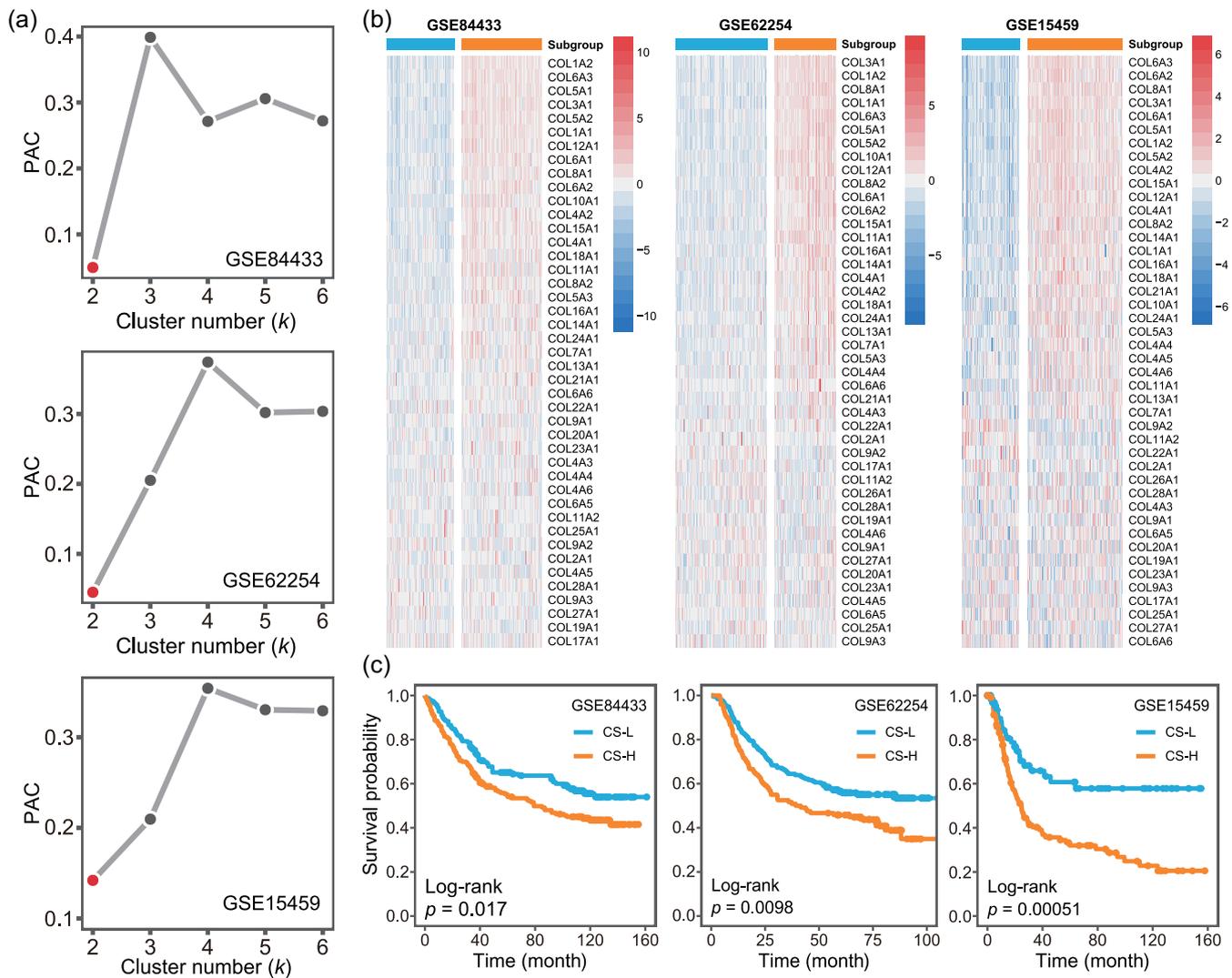
FIGURE 2 Confirmation of collagen subtypes in the validation cohorts (GSE84433, GSE62254, and GSE15459). (a) PAC curves for each k of the validation cohorts. The optimal k for all cohorts was 2. (b) Heatmap of 44 collagen gene expression patterns in the validation cohorts. Each column represents one patient in the corresponding cohort. (c) Kaplan–Meier curves for OS in patients with the two collagen subtypes. OS, overall survival.

3.3 | Genetic alterations led to differences in PI3K/Akt pathway activity between subtypes

We explored the difference in genetic alterations between the collagen subtypes by analyzing somatic mutation and CNV data in the TCGA-STAD data set via maftools. In terms of somatic mutations, the tumor mutational burden (TMB) was significantly greater in the CS-L subgroup than in the CS-H subgroup (Figure 4a). The mutation frequencies of *TP53*, *PCLO*, *LAMA1*, and *PEG3* were significantly greater in the CS-L subgroup compared with the CS-H subgroup, while the mutation frequency of *CDH1* was significantly greater in the CS-H subgroup. Notably,

there was no significant difference in *PIK3CA* mutation frequency between the collagen subtypes (Figure 4b). We then obtained gene-level CNVs through maftools and found that *PIK3CA* was frequently amplified in the CS-H subgroup, while *PIK3C2A*, *PIK3C2G*, and *PIK3R1* were frequently deleted in the CS-L subgroup (Figure 4c). These CNVs reshaped the expression of the corresponding genes (Figure 4d): amplification of *PIK3CA* increased its expression and deletion of *PIK3C2A* and *PIK3R1* decreased their expression (Figure 4d). As core genes of the PI3K pathway, pathway activity was significantly reduced in the overall-deleted patients compared with the overall-amplified patients (Figure 4e). These subtype-specific changes resulted in high
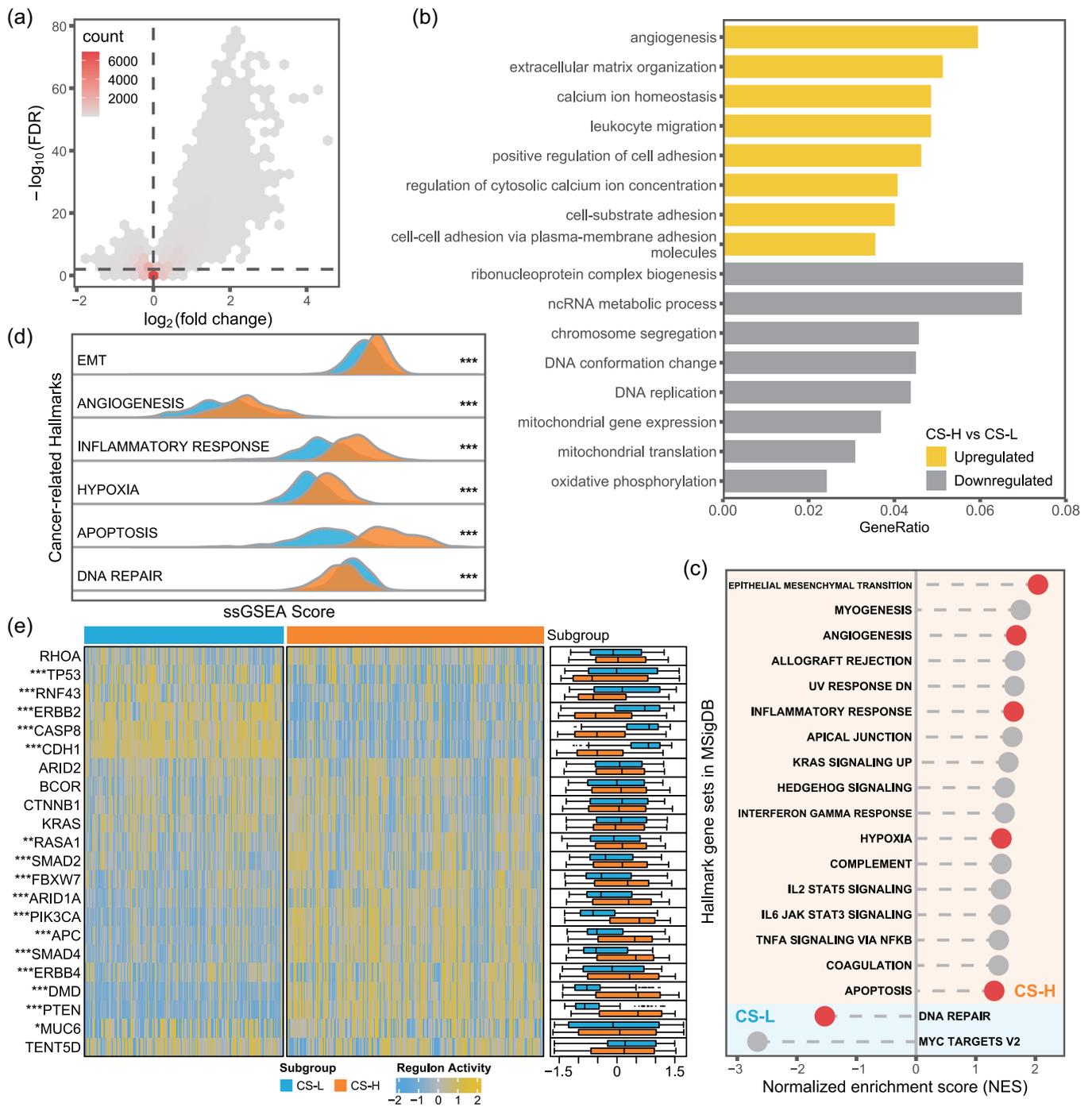
FIGURE 3 Transcriptomic features of collagen subtypes in gastric cancer from the TCGA-STAD cohort. (a) Volcano plot of DEGs between the CS-H and CS-L subtypes. Genes with an FDR ≤ 0.01 were defined as DEGs. (b) Significantly enriched GO BP terms of upregulated and downregulated DEGs (FDR ≤ 0.05). (c) Significant hallmark gene sets in the MSigDB associated with collagen subtypes revealed by GSEA (FDR ≤ 0.05). (d) Comparison of ssGSEA scores for cancer-related hallmarks between collagen subtypes. Statistical analysis was performed using Wilcoxon's rank sum test. ***$p < 0.001$. (e) Heatmap showing regulon activity profiles for gastric cancer driver genes; Wilcoxon's rank sum test used for comparisons between subtypes. *$p < 0.05$, **$p < 0.01$, and ***$p < 0.001$. BP, biological process; DEG, differentially expressed gene; FDR, false discovery rate; GO, Gene Ontology; TCGA, The Cancer Genome Atlas.

expression of PI3K in the PI3K/Akt pathway in the CS-H subtype (Figure S4). PI3K/Akt pathway activity was therefore upregulated in the CS-H subtype compared with the CS-L subtype (Figure 4f). These results indicate distinct genetic alteration patterns between collagen subtypes, with CNV leading to activation of the PI3K/Akt pathway in the CS-H subtype.
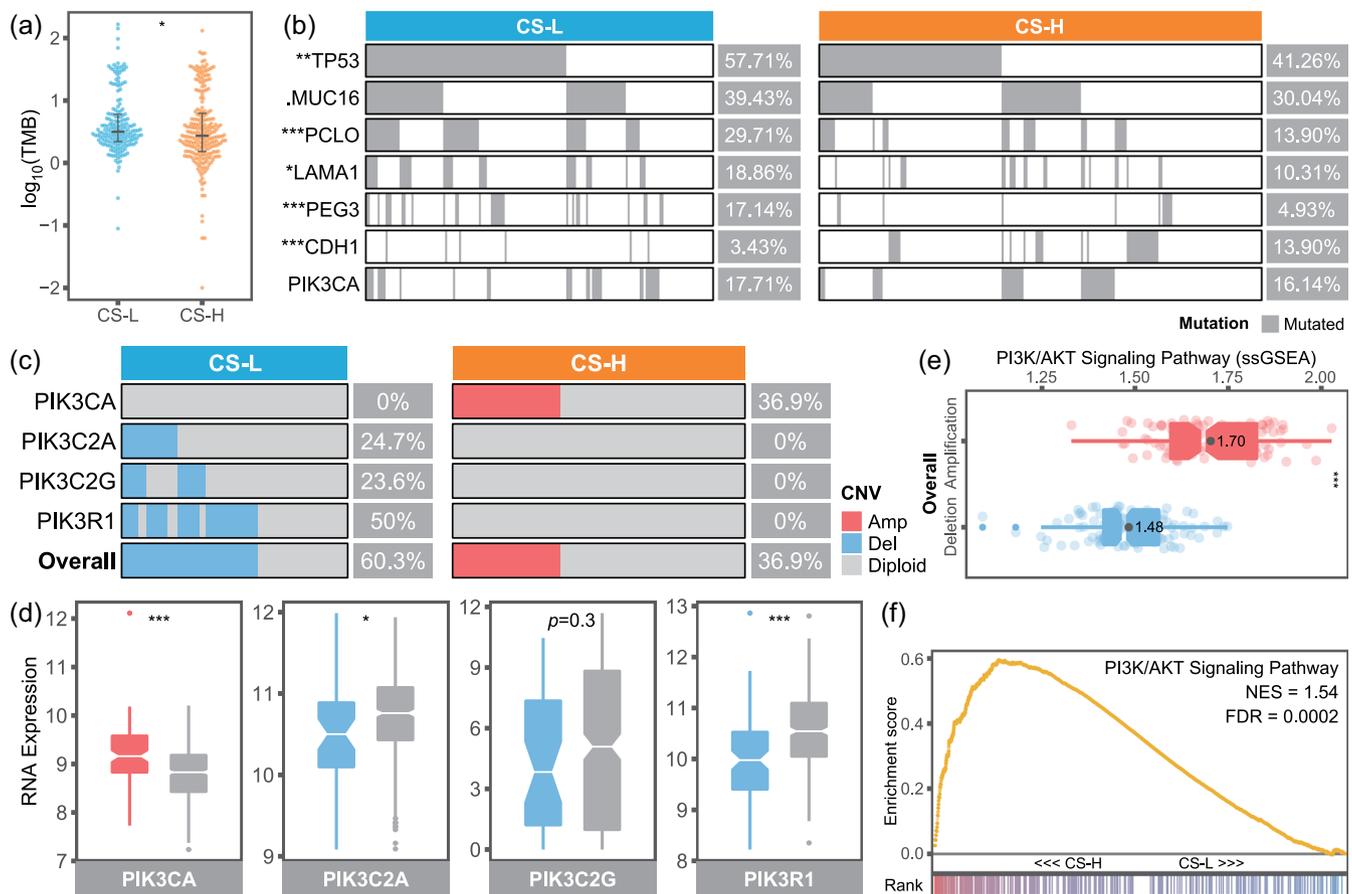
FIGURE 4 Collagen subtype-specific genetic alterations and their impact on the PI3K pathway in the TCGA-STAD cohort. (a) Distribution of TMB between the two collagen subtypes. Statistical analysis performed using Wilcoxon's rank sum test. *$p < 0.05$. (b) Comparison of mutated genes between collagen subtypes. Statistical analysis performed using Fisher's exact test. $p < 0.1$, *$p < 0.05$, **$p < 0.01$, and ***$p < 0.001$. (c) CNV profiles of *PIK3* genes between collagen subtypes. Blue represents deletion and red represents amplification. (d) Boxplot showing RNA expression of *PIK3* genes grouped by CNV status of the corresponding gene. Student's *t*-test used for comparison. *$p < 0.05$ and ***$p < 0.001$. (e) Comparison of ssGSEA scores for PI3K/Akt pathway between overall-deleted and overall-amplified patients. Statistical analysis performed using Wilcoxon's rank sum test. ***$p < 0.001$. (f) GSEA plot for PI3K/Akt pathway. CNV, copy number variation; TCGA, The Cancer Genome Atlas; TMB, tumor mutational burden.

3.4 | Collagen subtypes correlated with therapeutic response

We explored the relationship between collagen subtype and therapeutic response by evaluating multiple markers and patient treatment responses in the TCGA-STAD cohort. Chemotherapy is the first-line therapy for gastric cancer, and we therefore compared the response to chemotherapy between patients with different collagen subtypes. The response rate to fluorouracil was significantly greater in the CS-H compared with the CS-L group (Figure 5a). pRRophetic analysis revealed that CS-H tumors had significantly lower estimated IC_{50} values for the chemotherapeutic drugs cisplatin, docetaxel, doxorubicin, etoposide, and methotrexate (Figure 5b), indicating that patients with CS-H tumors were more sensitive to these chemotherapeutic drugs.

We subsequently used TIDE to estimate multiple transcriptomic biomarkers and predict patient response to ICB therapy. The T-cell dysfunction and exclusion scores were significantly greater in the CS-H subgroup than in the CS-L subgroup, consistent with the TIDE score results (Figure 5c). This suggested more serious dysfunction and exclusion of cytotoxic T lymphocytes in CS-H tumors and indicated that these tumors were less likely to respond to ICB. Notably, among the three types of tumor-infiltrating restricted T cells, CAFs exhibited greater differences between collagen subtypes than MDSCs and M2 TAMs (Figure 5c). When multiple biomarkers were combined, TIDE predicted that CS-L-subtype patients had a significantly greater ICB response rate than CS-H-subtype patients (Figure 5d). We verified this finding in the PRJEB25780 cohort of patients who received PD-1 blockade (pembrolizumab) and showed
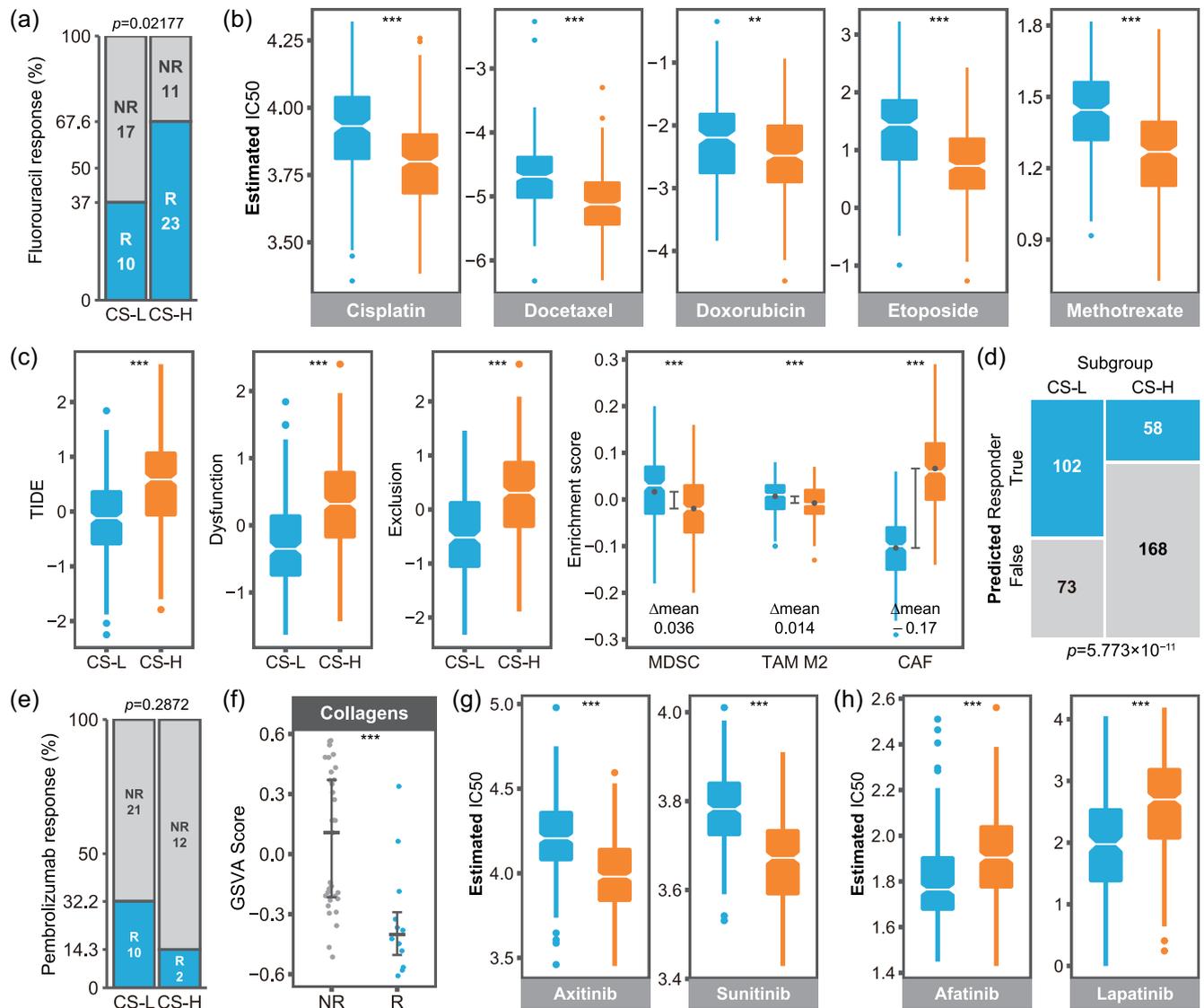
FIGURE 5 Differences in therapeutic responses between patients with two collagen subtypes. (a) Response rates to fluorouracil between collagen subtypes in the TCGA-STAD cohort. Statistical analysis performed using Fisher's exact test. R, responder; NR, nonresponder. (b) Boxplot showing estimated IC_{50} values of chemotherapeutic drugs for two collagen subtypes in the TCGA-STAD cohort. Student's *t*-test used for comparison. ***p < 0.01 and ***p < 0.001. (c) Comparison of TIDE, dysfunction, exclusion, MDSC, TAM M2, and CAF enrichment scores between collagen subtypes in the TCGA-STAD cohort. Statistical analysis performed using Student's *t*-test. ***p < 0.001. (d) Response rates to ICB predicted by TIDE between collagen subtypes in the TCGA-STAD cohort. Statistical analysis performed using Fisher's exact test. (e) Response rates to pembrolizumab between collagen subtypes in the PRJEB25780 cohort. Statistical analysis performed using Fisher's exact test. (f) Collagen GSVA scores in responders and nonresponders to pembrolizumab in the PRJEB25780 cohort. Statistical analysis performed using Wilcoxon's rank sum test. ***p < 0.001. (g) Boxplot showing estimated IC_{50} values of drugs targeting VEGFR in two collagen subtypes in the TCGA-STAD cohort. Student's *t*-test used for comparison. ***p < 0.001. (h) Boxplot showing estimated IC_{50} values for drugs targeting HER2 in two collagen subtypes in the TCGA-STAD cohort. Student's *t*-test used for comparison. ***p < 0.001. CAF, cancer-associated fibroblast; TCGA, The Cancer Genome Atlas.

that the response rate of CS-L patients was greater than that of CS-H patients (Figure 5e), and the collagen scores of nonresponders were greater than those of responders (Figure 5f). SubMap analysis [48] showed that CS-H subtype was associated with a lack of response (Figure S5a).

We also estimated the IC_{50} of the targeted drugs in each patient. Angiogenesis (Figure 3c,d) and the PI3K/Akt pathway (Figure 4f) were significantly upregulated in the CS-H group, and multiple proangiogenic factors, including vascular endothelial

growth factor (VEGFR), were highly expressed (Figure S5b, c). These findings suggest that angiogenesis and the PI3K pathway could be therapeutic targets for CS-H subtype tumors. HER2 expression was significantly greater in CS-L-subtype tumors (Figure S5d). Correspondingly, CS-H-subtype tumors were more sensitive to drugs targeting VEGFR, including axitinib and sunitinib (Figure 5g), and drugs targeting the PI3K pathway (Figure S5e), while CS-L-subtype tumors were more sensitive to drugs targeting HER2, including afatinib and lapatinib (Figure 5h). Notably, the treatment responses of the two subtypes in the GSE84433, GSE62254, and GSE15459 cohorts were consistent with those in the TCGA cohort (Figure S6–S8). Overall, these results showed that collagen subtype was associated with responses to chemotherapy, ICB therapy, and targeted therapy.

3.5 | Collagen-subtype predictor excelled on multiple platforms

Considering the clinical significance of the collagen subtype and the need for convenient clinical application, we established a collagen-subtype predictor (Figure 6a). With respect to the TCGA-STAD cohort, we used three machine learning algorithms—Boruta, RFE, and LASSO—to perform feature selection for the collagen genes and

obtained 30, 24, and 17 genes, respectively (Table S3). Twelve scores were calculated for these three signatures using GSVA, ssGSEA, z score, and PLAGE, and the subtype-prediction performances of these 12 scores (models) were evaluated in the validation cohorts (Figure 6b). Among them, the score computed by PLAGE based on 17-gene signatures from LASSO performed best in the three validation cohorts (average AUC = 0.996). The optimal threshold for discrimination was 0.003: samples with a score ≥ 0.003 were assigned to the CS-L subtype, and those with a score < 0.003 were assigned to the CS-H subtype. This collagen-subtype predictor was validated in three cohorts tested according to two types of gene arrays, and all the studies showed excellent discriminative power (Kappa coefficients of 0.9323, 0.8711, and 0.871, respectively) (Figure 6c; Tables S4–S6). These results showed that the collagen-subtype predictor could accurately classify patients with gastric cancer according to collagen subtypes, independent of the platform.

3.6 | Comparison of collagen and other molecular subtypes

We explored the relationship between collagen subtypes and previously reported molecular subtypes of gastric cancer [20, 49]. The TCGA study revealed five molecular subtypes of gastrointestinal adenocarcinoma: EBV, MSI, HM-SNV, CIN, and GS [49]. Based on gene expression
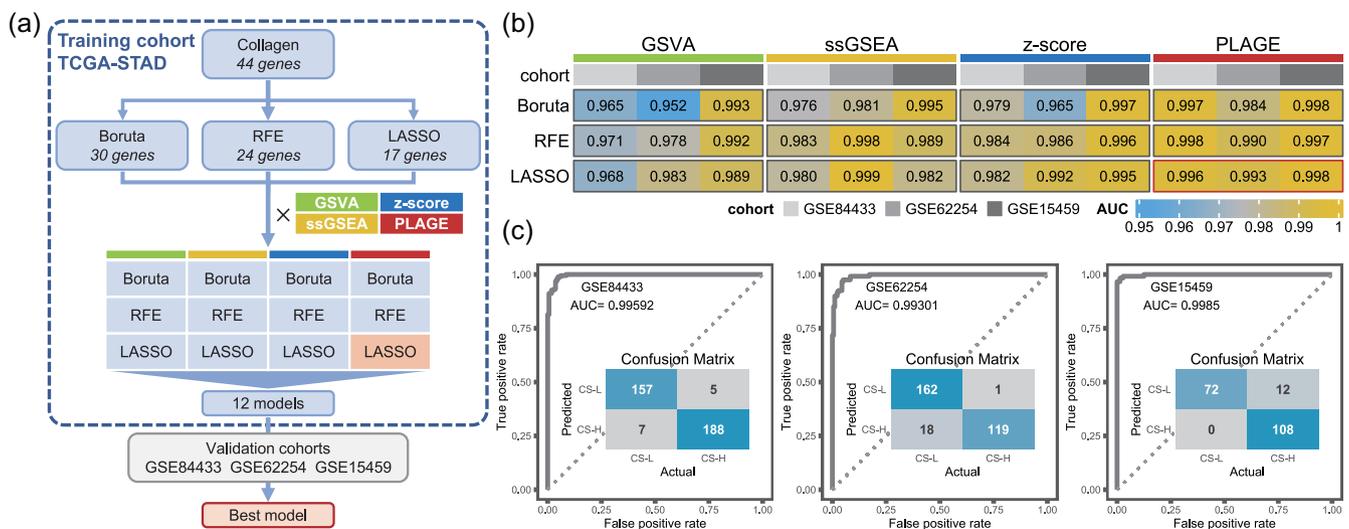

FIGURE 6 Training and validation of collagen-subtype predictor. (a) Schematic diagram of collagen-subtype predictor construction. (b) Performance of the 12 scores (models) in the validation cohorts. Performance was evaluated according to the area under the curve (AUC). The score computed by PLAGE on 17-gene signatures from LASSO was the best model and was therefore called the collagen-subtype predictor. (c) ROC curves and confusion matrices for collagen-subtype prediction in the GSE84433, GSE62254, and GSE15459 cohorts. ROC, receiver operating characteristic.

signatures, the Asian Cancer Research Group (ACRG) defined four molecular subtypes of gastric cancer: MSI, MSS/EMT, MSS/TP53$^+$, and MSS/TP53$^-$ [20].

In the TCGA-STAD cohort, we compared the distribution of TCGA subtypes between collagen subtypes and found that the TCGA GS subtype was enriched in the CS-H subgroup (Figure 7a). For the ACRG classification approach, we explored differences between collagen subtypes in the GSE62254 cohort and showed that ACRG MSS/TP53$^+$ and MSS/TP53$^-$ subtypes were enriched in the CS-L subgroup, while most ACRG MSS/EMT patients were enriched in the CS-H subgroup (Figure 7b). Furthermore, the prognosis of CS-L patients was significantly better than that of CS-H patients for MSI gastric cancers according to the TCGA and ACRG classifications (Figure 7c,d). These findings suggest that collagen subtyping may potentially be combined with existing molecular subtyping methods for better patient classification.

3.7 | Collagen subtyping was suitable for multiple tumors

We determined if the collagen subtype was clinically important in other cancer types by consensus clustering of other cancer types in TCGA. In addition to STAD, there were 17 other cancer types with two collagen subtypes indicated by different expression patterns (Figures S9a and S10). OS was better in patients with the CS-L subtype compared with the CS-H subtype among patients with bladder urothelial carcinoma, cervical squamous cell carcinoma and endocervical adenocarcinoma, kidney renal clear cell carcinoma, kidney renal papillary cell carcinoma, brain lower grade glioma, and lung adenocarcinoma, as well as all 18 tumors with two subtypes (Figure S9b). Among a group of seven tumors, differences in the expression of collagen genes were similar between the collagen subtypes (Figure S9c). These findings indicated
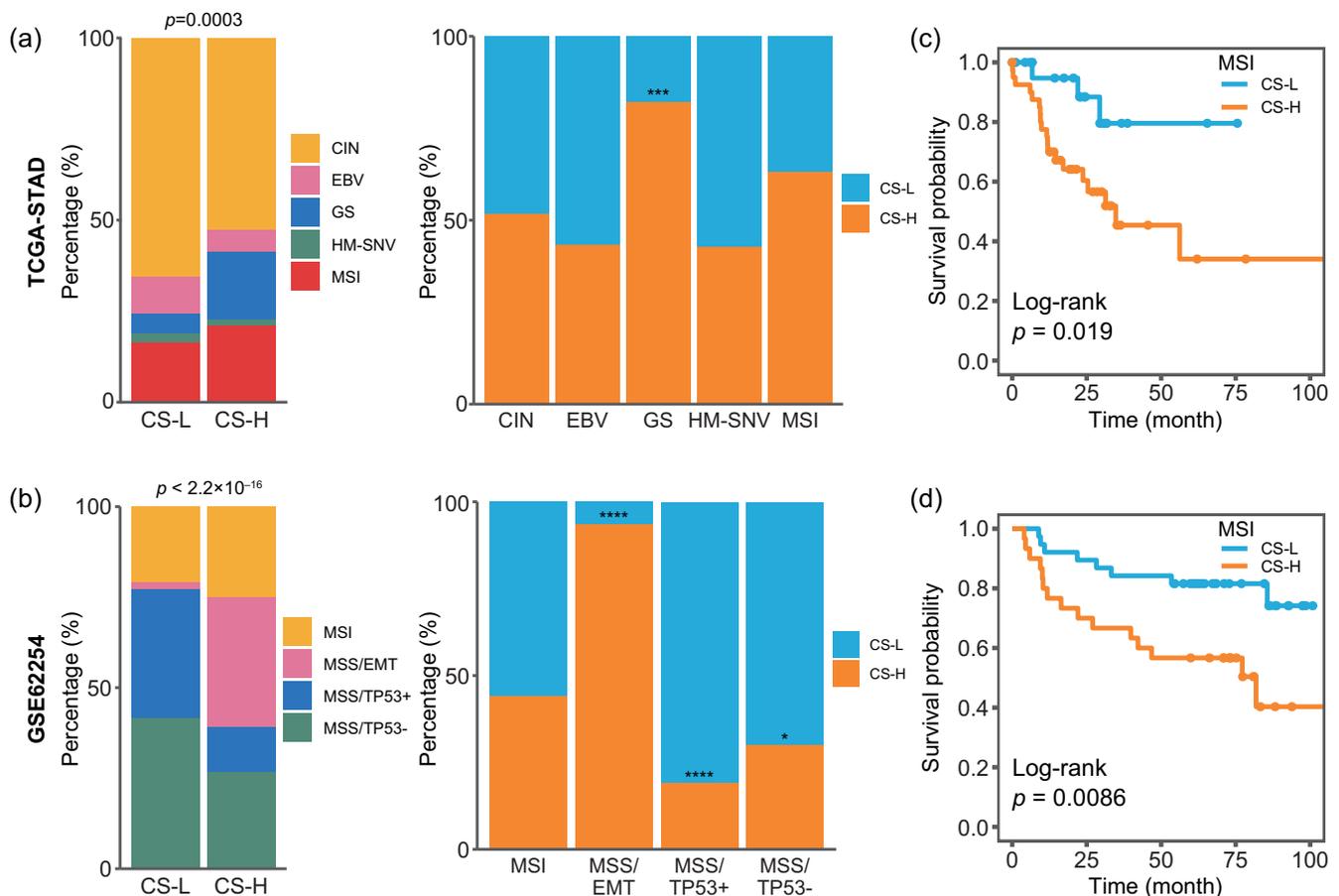

FIGURE 7 Comparisons between collagen subtyping and other subtyping methods for gastric cancer. (a) Relationships between collagen subtypes and TCGA molecular subtypes. Statistical analysis performed using Fisher's exact test. ***adjusted $p < 0.001$. (b) Relationships between collagen subtypes and ACRG molecular subtypes. Statistical analysis performed using Fisher's exact test. *adjusted $p < 0.05$, ***adjusted $p < 0.001$, and ****adjusted $p < 0.0001$. (c) Kaplan–Meier curves for OS in patients in the TCGA MSI subgroup. (d) Kaplan–Meier curves for OS in patients in the ACRG-MSI subgroup. MSI, microsatellite instability; OS, overall survival; TCGA, The Cancer Genome Atlas.

that collagen-subtype classification could have a common biological basis among cancer types. Most pathways enriched for collagen subtype-related DEGs were shared among cancer types. In particular, the upregulated genes in the seven cancer types in the CS-H subgroup were enriched in the PI3K/Akt signaling pathway (Figure S9d). These results demonstrated that collagen subtyping could be extended to multiple tumors to indicate differences in survival.

4 | DISCUSSION

For the first time, we identified two collagen subtypes of gastric cancer using consensus clustering. The two subtypes differed significantly in terms of cancer-related hallmarks and genetic alterations. Patients in the CS-L subgroup were more sensitive to immunotherapy and HER2-targeted drugs, while patients in the CS-H subgroup were more sensitive to chemotherapy, epidermal growth factor receptor-2-targeted drugs, and drugs targeting the PI3K pathway. We constructed a collagen-subtype predictor for clinical practice and showed that collagen subtyping could be combined with existing molecular subtyping methods.

Multiple classification schemes have been established for gastric cancer and have reportedly been associated with inflammatory infiltration, histopathology, treatment response, and clinical outcome [20, 49–54]. Some investigators clustered gastric cancers according to gene expression profiles [50–53]. The TCGA group yielded molecular subtypes of gastrointestinal adenocarcinomas based on mutations, copy-number alterations, and DNA methylation patterns [49], while the ACRG classified gastric cancers into four subtypes based on gene expression signatures, including EMT, MSI, cytokine signaling, cell proliferation, DNA methylation, TP53 activity, and gastric tissue [20]. Li et al. divided gastric cancer into immunity-deprived, stroma-enriched, and immunity-enriched subtypes according to the enrichment levels of 15 pathways [54]. In summary, existing molecular subtypes are based on gene expression profiles or predefined sets of gene expression signatures.

Compared with these previous classification schemes, the current study focused on the collagen component of the TME and used collagen gene expression alone to achieve a good classification. Furthermore, this classification scheme has the potential to use collagen staining instead of gene expression to quantify collagen in clinical applications. In this study, patients with different subtypes exhibited significantly different biological characteristics and genetic alteration profiles and also different treatment responses and prognoses. The collagen subtype was closely related to the responses to chemotherapy, immunotherapy, and targeted

drugs and may thus provide guidance for drug combinations. Notably, this classification scheme does not consider other omics characteristics and only divides patients into two subtypes, and its discrimination accuracy will thus be relatively low; however, this scheme could be combined with existing molecular subtypes to better classify patients. For example, the collagen subtype was still associated with prognostic differences within the MSI subgroup. The collagen subtype is thus a simple, effective, and flexible marker, emphasizing its clinical application prospects.

Collagen plays a critical role in the progression of gastric cancer and constitutes the biological basis for the collagen-subtype classification. In gastric cancer, stromal collagen deposition is increased and morphologically altered [55]. Collagen-related remodeling of the TME is associated with multiple biological pathways. In the stomach, chronic inflammation causes metaplasia, resulting in the requisite environment for the development of gastric cancer [56]. During inflammation, infiltration of macrophages induces collagen crosslinking, stromal stiffening, and fibrosis by stimulating stromal cell expression of LOX and LH2 [57]. Compared with normal tissue, type V collagen expression was increased 2- to 9-fold in chronically inflamed tissue [58]. This extracellular matrix remodeling, characterized by collagen deposition and crosslinking, is associated with tumor progression [59, 60]. Collagen subtype may thus reflect the biological characteristics of gastric carcinogenesis. *PIK3CA*, an oncogene in gastric cancer, was specifically amplified in CS-H subtype tumors and the PI3K/Akt signaling pathway was accordingly more active in CS-H-subtype tumors. This genetic change could also explain the high expression level of collagen genes. Using a genetically engineered mouse model, Wegner et al. reported that activated PI3K drove profound stromal remodeling and collagen accumulation in the prostatic epithelium [61]. Collagen remodeling within the TME is closely related to various hallmarks of cancer [62]. The inflammatory response, angiogenesis, EMT, hypoxia, and apoptosis were significantly increased while DNA repair was significantly decreased in CS-H tumors. These biological differences reflect differences in patient prognosis and treatment response.

Patients with different collagen subtypes had significantly different responses to treatment. The DNA repair pathway plays a major role in cancer cell resistance to chemotherapy drugs, and the DNA repair capacity can thus predict the tumor's response to chemotherapy drugs [63]. We showed that CS-H-subtype patients had lower transcriptional activity in the DNA repair pathway and were more sensitive to chemotherapy. In addition to chemotherapy, targeted therapies for gastric cancer are increasingly important. Based on the different expression levels of the drug targets HER2 and VEGFR, we found

that CS-L-subtype tumors were more sensitive to drugs targeting HER2, while CS-H-subtype tumors were more sensitive to drugs targeting VEGFR. Moreover, chemotherapeutic agents and sunitinib share common targets, indicating that VEGFR-targeted therapy can be used in combination with chemotherapy [64] (Figure S5f).

In CheckMate 649, nivolumab (a PD-1 inhibitor) in combination with chemotherapy showed superior survival benefit compared with chemotherapy alone, suggesting that ICB is a promising treatment for gastric cancer [65]. Moreover, screening patients who benefit from ICB has become an urgent problem. High TMB [23, 66] and *MUC16* mutation status [67] have been associated with improved clinical efficacy of ICB therapy, consistent with the significantly greater response to ICB in patients with the CS-L subtype in the present study. Banchereau et al. showed that the PI3K/Akt pathway was enriched in nonresponders to anti-PD-L1 therapy, consistent with the poor response of patients with the CS-H subtype to ICB [68]. Compared with the reported effects of *PIK3CA* mutations on immunotherapy [69], this study revealed that CNVs in PI3K genes affected PI3K pathway activity, with potential impacts on immunotherapy efficacy. Furthermore, we showed that *PIK3CA* expression was significantly positively correlated with *TGFB2* and *TGFB3* expression (Figure S5g). Based on signaling interplay between the transforming growth factor (TGF)-β receptor and the PI3K/Akt pathway in cancer, we speculate that PI3K activates TGF-β in CS-H-subtype tumors. TGF-β is a CAF activator [70], which could explain the high infiltration of CAFs in CS-H tumors. Using the single-cell atlas of gastric cancer, Kumar et al. [71] identified inhibin subunit beta A as the main component of the TGF-β pathway and as a regulator of CAFs. The poor response of patients with the CS-H subtype could be due to the high infiltration of CAFs into tumors, suggesting that these patients might need to eliminate CAFs to respond to ICB [41, 72]. Targeted therapy and immunotherapy could have combined effects and clinical benefits. In the KEYNOTE-811 trial, dual PD-1 and HER2 blockade markedly reduced tumor size and significantly improved the objective response rate in HER2-positive patients with gastric cancer [73]. This was consistent with greater sensitivity of CS-L subtype patients to PD-1 and HER2 blockade compared with the CS-H subtype, suggesting the potential value of the collagen subtype in combination-treatment decision-making.

We used multiple data sets from different regions for validation in the present study; however, there might still be potential confounding factors that can cause bias. We therefore plan to carry out a multicenter prospective clinical trial to confirm the role of the collagen subtype before further clinical application.

5 | CONCLUSIONS

Gastric cancers can be classified into two subtypes based on collagen gene expression patterns, with different transcriptome, genetic alterations, prognosis, clinical characteristics, and therapeutic responses. The collagen subtype could be used to screen patients with gastric cancer who might benefit from chemotherapy, ICB, and targeted therapy and to select suitable treatment strategies.

AUTHOR CONTRIBUTIONS
Di Wang: Conceptualization (equal); data curation (equal); formal analysis (lead); investigation (lead); methodology (lead); resources (equal); software (lead); validation (lead); visualization (lead); writing—original draft (lead); writing—review and editing (equal). **Jing Zhang**: Data curation (equal); resources (equal). **Jianchao Wang**: Writing—review and editing (equal). **Zhonglin Cai**: Data curation (equal); resources (equal). **Shanfeng Jin**: Data curation (equal); resources (equal). **Gang Chen**: Conceptualization (equal); funding acquisition (lead); project administration (lead); supervision (lead).

CONFLICT OF INTEREST STATEMENT
The authors declare no conflict of interest.

DATA AVAILABILITY STATEMENT
The TCGA-STAD cohort is available at the GDC PanCanAtlas Publications (https://gdc.cancer.gov/about-data/publications/pancanatlas). The microarray data (GSE84433, GSE62254, and GSE15459) are available at the NCBI Gene Expression Omnibus (https://www.ncbi.nlm.nih.gov/geo/). The PRJEB25780 cohort is available at the European Nucleotide Archive (https://www.ebi.ac.uk/ena/).

ETHICS STATEMENT
The need for ethical approval was waived because the study used only publicly available data and materials.

INFORMED CONSENT

Not applicable.

ORCID

Di Wang http://orcid.org/0000-0003-2059-7054

REFERENCES

1. Sung H, Ferlay J, Siegel RL, Laversanne M, Soerjomataram I, Jemal A, et al. Global cancer statistics 2020: GLOBOCAN estimates of incidence and mortality worldwide for 36 cancers in 185 countries. CA Cancer J Clin. 2021;71(3):209–49. https://doi.org/10.3322/caac.21660
2. Smyth EC, Nilsson M, Grabsch HI, van Grieken NC, Lordick F. Gastric cancer. The Lancet. 2020;396(10251):635–48. https://doi.org/10.1016/S0140-6736(20)31288-5
3. The Cancer Genome Atlas Research Network. Comprehensive molecular characterization of gastric adenocarcinoma. Nature. 2014;513(7517):202–9. https://doi.org/10.1038/nature13480
4. Röcken C, Amallraja A, Halske C, Opasic L, Traulsen A, Behrens HM, et al. Multiscale heterogeneity in gastric adenocarcinoma evolution is an obstacle to precision Medicine. Genome Med. 2021;13(1):177. https://doi.org/10.1186/s13073-021-00975-y
5. Kwon M, An M, Klempner SJ, Lee H, Kim KM, Sa JK, et al. Determinants of response and intrinsic resistance to PD-1 blockade in microsatellite instability-high gastric cancer. Cancer Discovery. 2021;11(9):2168–85. https://doi.org/10.1158/2159-8290.CD-21-0219
6. Grunberg N, Pevsner-Fischer M, Goshen-Lago T, Diment J, Stein Y, Lavon H, et al. Cancer-associated fibroblasts promote aggressive gastric cancer phenotypes via heat shock factor 1-mediated secretion of extracellular vesicles. Cancer Res. 2021;81(7):1639–53. https://doi.org/10.1158/0008-5472.CAN-20-2756
7. Hinshaw DC, Shevde LA. The tumor microenvironment innately modulates cancer progression. Cancer Res. 2019;79(18):4557–66. https://doi.org/10.1158/0008-5472.can-18-3962
8. Wu T, Dai Y. Tumor microenvironment and therapeutic response. Cancer Lett. 2017;387:61–8. https://doi.org/10.1016/j.canlet.2016.01.043
9. Quante M, Varga J, Wang TC, Greten FR. The gastrointestinal tumor microenvironment. Gastroenterology. 2013;145(1):63–78. https://doi.org/10.1053/j.gastro.2013.03.052
10. Nissen NI, Karsdal M, Willumsen N. Collagens and cancer associated fibroblasts in the reactive stroma and its relation to cancer biology. J Exp Clin Cancer Res. 2019;38(1):115. https://doi.org/10.1186/s13046-019-1110-6
11. Sleeboom JJF, van Tienderen GS, Schenke-Layland K, van der Laan LJW, Khalil AA, Verstegen MMA. The extracellular matrix as hallmark of cancer and metastasis: from biomechanics to therapeutic targets. Sci Transl Med. 2024;16(728):eadg3840. https://doi.org/10.1126/scitranslmed.adg3840
12. Kaur A, Ecker BL, Douglass SM, Kugel III CH, Webster MR, Almeida FV, et al. Remodeling of the collagen matrix in aging skin promotes melanoma metastasis and affects immune cell motility. Cancer Discovery. 2019;9(1):64–81. https://doi.org/10.1158/2159-8290.CD-18-0193
13. Peng DH, Rodriguez BL, Diao L, Chen L, Wang J, Byers LA, et al. Collagen promotes anti-PD-1/PD-L1 resistance in cancer through LAIR1-dependent CD8$^+$ T cell exhaustion. Nat Commun. 2020;11(1):4520. https://doi.org/10.1038/s41467-020-18298-8
14. Ogawa Y, Masugi Y, Abe T, Yamazaki K, Ueno A, Fujii-Nishimura Y, et al. Three distinct stroma types in human pancreatic cancer identified by image analysis of fibroblast subpopulations and collagen. Clin Cancer Res. 2021;27(1):107–19. https://doi.org/10.1158/1078-0432.CCR-20-2298
15. Martins Cavaco AC, Dâmaso S, Casimiro S, Costa L. Collagen biology making inroads into prognosis and treatment of cancer progression and metastasis. Cancer Metastasis Rev. 2020;39(3):603–23. https://doi.org/10.1007/s10555-020-09888-5
16. Chen D, Chen G, Jiang W, Fu M, Liu W, Sui J, et al. Association of the collagen signature in the tumor micro-environment with lymph node metastasis in early gastric cancer. JAMA Surgery. 2019;154(3):e185249. https://doi.org/10.1001/jamasurg.2018.5249
17. Chen D, Liu Z, Liu W, Fu M, Jiang W, Xu S, et al. Predicting postoperative peritoneal metastasis in gastric cancer with serosal invasion using a collagen nomogram. Nat Commun. 2021;12(1):179. https://doi.org/10.1038/s41467-020-20429-0
18. Wang J, Liu Z, Lin L, Wu Z, Gao X, Cai X, et al. Collagen-related gene expression level predicts the prognosis and immune therapy response. Gastric Cancer. 2023;26(6):891–903. https://doi.org/10.1007/s10120-023-01416-y
19. Yoon SJ, Park J, Shin Y, Choi Y, Park SW, Kang SG, et al. Deconvolution of diffuse gastric cancer and the suppression of CD34 on the BALB/c nude mice model. BMC Cancer. 2020;20(1):314. https://doi.org/10.1186/s12885-020-06814-4
20. 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. Nature Med. 2015;21(5):449–56. https://doi.org/10.1038/nm.3850
21. Ooi CH, Ivanova T, Wu J, Lee M, Tan IB, Tao J, et al. Oncogenic pathway combinations predict clinical prognosis in gastric cancer. PLoS Genet. 2009;5(10):e1000676. https://doi.org/10.1371/journal.pgen.1000676
22. Ding Z, Zu S, Gu J. Evaluating the molecule-based prediction of clinical drug responses in cancer. Bioinformatics. 2016;32(19):2891–5. https://doi.org/10.1093/bioinformatics/btw344
23. 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. Nature Med. 2018;24(9):1449–58. https://doi.org/10.1038/s41591-018-0101-z
24. Dobin A, Davis CA, Schlesinger F, Drenkow J, Zaleski C, Jha S, et al. STAR: ultrafast universal RNA-seq aligner. Bioinformatics. 2013;29(1):15–21. https://doi.org/10.1093/bioinformatics/bts635
25. Li B, Dewey CN. RSEM: accurate transcript quantification from RNA-Seq data with or without a reference genome. BMC Bioinformatics. 2011;12:323. https://doi.org/10.1186/1471-2105-12-323
26. Love MI, Huber W, Anders S. Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. Genome Biol. 2014;15(12):550. https://doi.org/10.1186/s13059-014-0550-8
27. Wilkerson MD, Hayes DN. ConsensusClusterPlus: a class discovery tool with confidence assessments and item tracking.

Bioinformatics. 2010;26(12):1572–3. https://doi.org/10.1093/bioinformatics/btq170

28. Şenbabaoğlu Y, Michailidis G, Li JZ. Critical limitations of consensus clustering in class discovery. Sci Rep. 2014;4:6207. https://doi.org/10.1038/srep06207

29. Ritchie ME, Phipson B, Wu D, Hu Y, Law CW, Shi W, et al. Limma powers differential expression analyses for RNA-sequencing and microarray studies. Nucleic Acids Res. 2015;43(7):e47. https://doi.org/10.1093/nar/gkv007

30. Wu T, Hu E, Xu S, Chen M, Guo P, Dai Z, et al. clusterProfiler 4.0: a universal enrichment tool for interpreting omics data. The Innovation. 2021;2(3):100141. https://doi.org/10.1016/j.xinn.2021.100141

31. Korotkevich G, Sukhov V, Budin N, Shpak B, Artyomov MN, Sergushichev A. Fast gene set enrichment analysis. bioRxiv. 2021;060012. https://doi.org/10.1101/060012

32. Liberzon A, Birger C, Thorvaldsdóttir H, Ghandi M, Mesirov JP, Tamayo P. The molecular signatures database hallmark gene set collection. Cell Systems. 2015;1(6):417–25. https://doi.org/10.1016/j.cels.2015.12.004

33. Gillespie M, Jassal B, Stephan R, Milacic M, Rothfels K, Senff-Ribeiro A, et al. The reactome pathway knowledgebase 2022. Nucleic Acids Res. 2022;50(D1):D687–92. https://doi.org/10.1093/nar/gkab1028

34. Kanehisa M, Furumichi M, Sato Y, Ishiguro-Watanabe M, Tanabe M. KEGG: integrating viruses and cellular organisms. Nucleic Acids Res. 2021;49(D1):D545–51. https://doi.org/10.1093/nar/gkaa970

35. Hänzelmann S, Castelo R, Guinney J. GSVA: gene set variation analysis for microarray and RNA-seq data. BMC Bioinformatics. 2013;14:7. https://doi.org/10.1186/1471-2105-14-7

36. Castro MAA, de Santiago I, Campbell TM, Vaughn C, Hickey TE, Ross E, et al. Regulators of genetic risk of breast cancer identified by integrative network analysis. Nature Genet. 2016;48(1):12–21. https://doi.org/10.1038/ng.3458

37. Bailey MH, Tokheim C, Porta-Pardo E, Sengupta S, Bertrand D, Weerasinghe A, et al. Comprehensive character-ization of cancer driver genes and mutations. Cell. 2018; 173(2):371–385.e18. https://doi.org/10.1016/j.cell.2018.02.060

38. Luo W, Brouwer C. Pathview: an R/Bioconductor package for pathway-based data integration and visualization. Bioinformatics. 2013;29(14):1830–1. https://doi.org/10.1093/bioinformatics/btt285

39. Mermel CH, Schumacher SE, Hill B, Meyerson ML, Beroukhim 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. https://doi.org/10.1186/gb-2011-12-4-r41

40. Mayakonda A, Lin DC, Assenov Y, Plass C, Koeffler HP. Maftools: efficient and comprehensive analysis of somatic variants in cancer. Genome Res. 2018;28(11):1747–56. https://doi.org/10.1101/gr.239244.118

41. Fu J, Li K, Zhang W, Wan C, Zhang J, Jiang P, et al. Large-scale public data reuse to model immunotherapy response and resistance. Genome Med. 2020;12(1):21. https://doi.org/10.1186/s13073-020-0721-z

42. Jiang P, Gu S, Pan D, Fu J, Sahu A, Hu X, et al. Signatures of T cell dysfunction and exclusion predict cancer immunotherapy response. Nature Med. 2018;24(10):1550–8. https://doi.org/10.1038/s41591-018-0136-1

43. Geeleher P, Zhang Z, Wang F, Gruener RF, Nath A, Morrison G, et al. Discovering novel pharmacogenomic biomarkers by imputing drug response in cancer patients from large genomics studies. Genome Res. 2017;27(10):1743–51. https://doi.org/10.1101/gr.221077.117

44. Kursa MB, Rudnicki WR. Feature selection with the boruta package. J Stat Softw. 2010;36(11):1–13. https://doi.org/10.18637/jss.v036.i11

45. Kuhn M. Building predictive models in R Using the caret package. J Stat Softw. 2008;28(5):1–26. https://doi.org/10.18637/jss.v028.i05

46. Friedman J, Hastie T, Tibshirani R. Regularization paths for generalized linear models via coordinate descent. J Stat Softw. 2010;33(1):1–22. https://doi.org/10.18637/jss.v033.i01

47. Robin X, Turck N, Hainard A, Tiberti N, Lisacek F, Sanchez JC, et al. pROC: an open-source package for R and S+ to analyze and compare ROC curves. BMC Bioinformatics. 2011;12:77. https://doi.org/10.1186/1471-2105-12-77

48. Hoshida Y, Brunet JP, Tamayo P, Golub TR, Mesirov JP. Subclass mapping: identifying common subtypes in independent disease data sets. PLoS One. 2007;2(11):e1195. https://doi.org/10.1371/journal.pone.0001195

49. Liu Y, Sethi NS, Hinoue T, Schneider BG, Cherniack AD, Sanchez-Vega F, et al. Comparative molecular analysis of gastrointestinal adenocarcinomas. Cancer Cell. 2018;33(4): 721–735.e8. https://doi.org/10.1016/j.ccell.2018.03.010

50. Kim B, Bang S, Lee S, Kim S, Jung Y, Lee C, et al. Expression profiling and subtype-specific expression of stomach cancer. Cancer Res. 2003;63(23):8248–55.

51. Shah MA, Khanin R, Tang L, Janjigian YY, Klimstra DS, Gerdes H, et al. Molecular classification of gastric cancer: a new paradigm. Clin Cancer Res. 2011;17(9):2693–701. https://doi.org/10.1158/1078-0432.CCR-10-2203

52. Tan IB, Ivanova T, Lim KH, Ong CW, Deng N, Lee J, et al. Intrinsic subtypes of gastric cancer, based on gene expression pattern, predict survival and respond differently to chemotherapy. Gastroenterology. 2011;141(2):476–485.e11. https://doi.org/10.1053/j.gastro.2011.04.042

53. Lei Z, Tan IB, Das K, Deng N, Zouridis H, Pattison S, et al. Identification of molecular subtypes of gastric cancer with different responses to PI_3-kinase inhibitors and 5-fluorouracil. Gastroenterology. 2013;145(3):554–65. https://doi.org/10.1053/j.gastro.2013.05.010

54. Li L, Wang X. Identification of gastric cancer subtypes based on pathway clustering. npj Precision Oncology. 2021;5(1):46. https://doi.org/10.1038/s41698-021-00186-z

55. Zhou ZH, Ji CD, Xiao HL, Zhao HB, Cui YH, Bian XW. Reorganized collagen in the tumor microenvironment of gastric cancer and its association with prognosis. J Cancer. 2017;8(8):1466–76. https://doi.org/10.7150/jca.18466

56. Busada JT, Ramamoorthy S, Cain DW, Xu X, Cook DN, Cidlowski JA. Endogenous glucocorticoids prevent gastric metaplasia by suppressing spontaneous inflammation. J Clin Invest. 2019;129(3):1345–58. https://doi.org/10.1172/JCI123233

57. Maller O, Drain AP, Barrett AS, Borgquist S, Ruffell B, Zakharevich I, et al. Tumour-associated macrophages drive

stromal cell-dependent collagen crosslinking and stiffening to promote breast cancer aggression. Nat Mater. 2021;20(4):548–59. https://doi.org/10.1038/s41563-020-00849-5

58. Narayanan AS, Engel LD, Page RC. The effect of chronic inflammation on the composition of collagen types in human connective tissue. Coll Relat Res. 1983;3(4):323–34. https://doi.org/10.1016/s0174-173x(83)80014-4

59. Levental KR, Yu H, Kass L, Lakins JN, Egeblad M, Erler JT, et al. Matrix crosslinking forces tumor progression by enhancing integrin signaling. Cell. 2009;139(5):891–906. https://doi.org/10.1016/j.cell.2009.10.027

60. Acerbi I, Cassereau L, Dean I, Shi Q, Au A, Park C, et al. Human breast cancer invasion and aggression correlates with ECM stiffening and immune cell infiltration. Integr Biol. 2015;7(10):1120–34. https://doi.org/10.1039/c5ib00040h

61. Wegner KA, Mueller BR, Unterberger CJ, Avila EJ, Ruetten H, Turco AE, et al. Prostate epithelial-specific expression of activated PI3K drives stromal collagen production and accumulation. J Pathol. 2020;250(2):231–42. https://doi.org/10.1002/path.5363

62. Xu S, Xu H, Wang W, Li S, Li H, Li T, et al. The role of collagen in cancer: from bench to bedside. J Transl Med. 2019;17(1):309. https://doi.org/10.1186/s12967-019-2058-1

63. Nagel ZD, Kitange GJ, Gupta SK, Joughin BA, Chaim IA, Mazzucato P, et al. DNA repair capacity in multiple pathways predicts chemoresistance in glioblastoma multiforme. Cancer Res. 2017;77(1):198–206. https://doi.org/10.1158/0008-5472.CAN-16-1151

64. Liu H, Zhang W, Zou B, Wang J, Deng Y, Deng L. DrugCombDB: a comprehensive database of drug combinations toward the discovery of combinatorial therapy. Nucleic Acids Res. 2019;48(D1):D871–81. https://doi.org/10.1093/nar/gkz1007

65. Janjigian YY, Shitara K, Moehler M, Garrido M, Salman P, Shen L, et al. First-line nivolumab plus chemotherapy versus chemotherapy alone for advanced gastric, gastro-oesophageal junction, and oesophageal adenocarcinoma (CheckMate 649): a randomised, open-label, phase 3 trial. The Lancet. 2021; 398(10294):27–40. https://doi.org/10.1016/S0140-6736(21)00797-2

66. Wang F, Wei XL, Wang FH, Xu N, Shen L, Dai GH, et al. Safety, efficacy and tumor mutational burden as a biomarker of overall survival benefit in chemo-refractory gastric cancer treated with toripalimab, a PD-1 antibody in phase Ib/II clinical trial NCT02915432. Ann Oncol. 2019;30(9):1479–86. https://doi.org/10.1093/annonc/mdz197

67. Zhang L, Han X, Shi Y. Association of MUC16 mutation with response to immune checkpoint inhibitors in solid tumors. JAMA Network Open. 2020;3(8):e2013201. https://doi.org/10.1001/jamanetworkopen.2020.13201

68. Banchereau R, Leng N, Zill O, Sokol E, Liu G, Pavlick D, et al. Molecular determinants of response to PD-L1 blockade across tumor types. Nat Commun. 2021;12(1):3969. https://doi.org/10.1038/s41467-021-24112-w

69. Zeng D, Wu J, Luo H, Li Y, Xiao J, Peng J, et al. Tumor microenvironment evaluation promotes precise checkpoint immunotherapy of advanced gastric cancer. J Immunother Cancer. 2021;9(8):e002467. https://doi.org/10.1136/jitc-2021-002467

70. Sahai E, Astsaturov I, Cukierman E, DeNardo DG, Egeblad M, Evans RM, et al. A framework for advancing our understanding of cancer-associated fibroblasts. Nat Rev Cancer. 2020;20(3):174–86. https://doi.org/10.1038/s41568-019-0238-1

71. Kumar V, Ramnarayanan K, Sundar R, Padmanabhan N, Srivastava S, Koiwa M, et al. Single-cell Atlas of lineage states, tumor microenvironment, and subtype-specific expression programs in gastric cancer. Cancer Discovery. 2022;12(3): 670–91. https://doi.org/10.1158/2159-8290.cd-21-0683

72. Bejarano L, Jordão MJC, Joyce JA. Therapeutic targeting of the tumor microenvironment. Cancer Discovery. 2021;11(4): 933–59. https://doi.org/10.1158/2159-8290.cd-20-1808

73. Janjigian YY, Kawazoe A, Yañez P, Li N, Lonardi S, Kolesnik O, et al. The KEYNOTE-811 trial of dual PD-1 and HER2 blockade in HER2-positive gastric cancer. Nature. 2021;600(7890):727–30. https://doi.org/10.1038/s41586-021-04161-3

SUPPORTING INFORMATION

Additional supporting information can be found online in the Supporting Information section at the end of this article.

How to cite this article: Wang D, Zhang J, Wang J, Cai Z, Jin S, Chen G. Identification of collagen subtypes of gastric cancer for distinguishing patient prognosis and therapeutic response. Cancer Innov. 2024;3:e125. https://doi.org/10.1002/cai2.125