

Integrated single cell and bulk transcriptomic analysis reveals disulfidptosis related malignant epithelial states associated with prognosis in gastric cancer

Zhao He

Dalian Medical University

Shuang Li

Dalian Medical University

Jirui Gu

Dalian Medical University

Jiahui Liu

Dalian Medical University

Xu Huang

Dalian Medical University

Jiaping Jing

Dalian Medical University

Qi Sun

Dalian Medical University

Miao Wei

Dalian Medical University

Jue Wang

Dalian Medical University

Dongdong Fan

Dalian Medical University

Zhiping Chen

Dalian Medical University

Jian Li

lj2122018@dmu.edu.cn

Dalian Medical University

Liang Cao

Dalian Medical University

Research Article

Keywords: gastric cancer, disulfidptosis, regulated cell death, single-cell RNA sequencing, tumor microenvironment, prognostic signature

Posted Date: October 27th, 2026

DOI: <https://doi.org/10.21203/rs.3.rs-10992120/v1>

License: © ⓘ This work is licensed under a Creative Commons Attribution 4.0 International License.
[Read Full License](#)

Additional Declarations: No competing interests reported.

Abstract

Background: Disulfidptosis is a recently described form of regulated cell death linked to cystine uptake, redox imbalance, and actin-cytoskeleton collapse, but the malignant epithelial states associated with disulfidptosis-related transcriptional programs in gastric cancer remain unclear.

Methods: We integrated GSE183904 single-cell RNA sequencing with TCGA-STAD and two independent bulk cohorts (GSE84437 and GSE62254/ACRG). Malignant epithelial cells were identified by inferred copy-number variation, stratified using a 14-gene disulfidptosis-related score, and evaluated for pathway, stemness, and cell-cell communication features. A broader 60-gene program and a six-gene LASSO-Cox risk model were assessed in bulk cohorts together with tumor-microenvironment and predicted drug-sensitivity analyses.

Results: Among 158,641 cells, 6,011 malignant epithelial cells were identified. High-score cells showed higher SLC7A11 expression and enrichment of cell-cycle, mitotic-spindle, and cytoskeletal programs, whereas low-score cells displayed stronger stemness-related features. The six-gene risk score stratified overall survival in TCGA-STAD and both independent cohorts and remained independently associated with survival in 319 TCGA-STAD patients with complete clinical data. High-risk tumors showed greater stromal enrichment, lower inferred tumor purity, and differential predicted sensitivity to several targeted agents.

Conclusions: This analysis links disulfidptosis-related transcriptional states at single-cell resolution with reproducible patient-level prognostic features in gastric cancer. The findings are computational and hypothesis-generating and require experimental and prospective validation.

Introduction

Gastric cancer remains a major global health burden and a leading cause of cancer-related mortality [1,2]. Although advances in endoscopic screening, surgery, perioperative chemotherapy, molecularly targeted therapy, and immune checkpoint blockade have improved outcomes for selected patients, many patients present with locally advanced or metastatic disease and continue to experience poor long-term survival [2,3]. Disease progression is shaped not only by tumor-cell genomic alterations but also by dynamic interactions among malignant epithelial cells, immune cells, stromal cells, endothelial cells, and extracellular-matrix components.

Large-scale molecular profiling has revealed substantial heterogeneity in gastric cancer. The Cancer Genome Atlas classified gastric adenocarcinoma into Epstein-Barr virus-positive, microsatellite instability, genomically stable, and chromosomal instability subtypes [4], whereas the Asian Cancer Research Group identified clinically relevant molecular subtypes associated with distinct outcomes [5]. These studies have transformed the molecular classification of gastric cancer; however, bulk genomic and transcriptomic profiles represent averaged signals from mixed malignant and non-malignant cell populations and therefore cannot directly resolve the cellular origin of specific molecular programs.

Single-cell RNA sequencing (scRNA-seq) enables intratumoral heterogeneity to be examined at cellular resolution. Gastric cancer single-cell atlases have identified diverse epithelial-lineage states, cancer-associated fibroblast populations, immune-cell subsets, and microenvironmental programs [6]. Integrating single-cell data with bulk patient cohorts can connect cell-state biology with clinical outcomes and improve the biological interpretation of prognostic biomarkers.

Disulfidptosis is a recently described form of regulated cell death induced by disulfide stress. In *SLC7A11*-high cells exposed to glucose limitation, increased cystine uptake and impaired reducing capacity promote intracellular disulfide accumulation, aberrant disulfide bonding within the actin cytoskeleton, and cytoskeletal collapse [7,8]. Because malignant cells frequently encounter nutrient limitation, redox imbalance, altered amino-acid transport, and cytoskeletal remodeling, disulfidptosis-related programs may define biologically relevant tumor-cell states.

Although disulfidptosis-related genes have been associated with prognosis in several cancers, the cellular context of these signals in gastric cancer remains insufficiently resolved. A bulk-tumor score cannot determine which malignant epithelial states carry the underlying program or whether cell-level features correspond to reproducible patient-level associations. We therefore integrated public single-cell and bulk transcriptomic cohorts to characterize a disulfidptosis-related malignant epithelial state axis, examine its relationship with stemness and predicted microenvironmental communication, and test whether related transcriptional features are associated with survival across independent patient cohorts. Because the study is computational, the resulting scores, communication estimates, and pharmacogenomic predictions are interpreted as hypothesis-generating rather than as direct evidence of disulfidptosis or treatment response.

Results

A single-cell atlas reveals remodeling of the gastric cancer tumor microenvironment

We analyzed the GSE183904 scRNA-seq dataset to construct a cellular atlas of the gastric cancer tumor microenvironment. After quality control, 158,641 cells were retained and visualized by uniform manifold approximation and projection (UMAP). Marker-based annotation identified 17 major cell populations, including epithelial cells, chief/mucous cells, pit cells, gland mucous cells, parietal cells, CD8⁺ T cells, CD4⁺ T cells, regulatory T cells, NK/gamma delta T cells, B cells, plasma cells, mast cells, monocytes, macrophages, fibroblasts/cancer-associated fibroblasts, endothelial cells, and smooth muscle/pericyte-like cells (Fig. 1).

Canonical marker expression supported the major lineage annotations. Descriptive comparison of cell-type proportions between normal and tumor samples indicated marked compositional remodeling in tumors, including expansion of macrophage/monocyte-lineage and fibroblast/cancer-associated fibroblast populations and redistribution of epithelial-lineage compartments. Because cell recovery

differs across samples and tissues, these proportional differences were interpreted as descriptive features of the analyzed dataset rather than direct estimates of absolute tissue abundance.

Copy-number variation inference identifies malignant epithelial cells and defines disulfidptosis-related score groups

To focus on malignant epithelial cells, we inferred large-scale copy-number variation (CNV) patterns using non-epithelial populations as references. Tumor epithelial cells displayed broad chromosomal gains and losses, whereas reference populations showed comparatively stable profiles. Based on CNV burden together with epithelial marker expression, 6,011 cells were classified as malignant epithelial cells for downstream analyses (Fig. 2A).

Reclustering revealed multiple malignant epithelial transcriptional programs, including tumor-epithelial, glandular/mucous, cell-cycle-active, mesenchymal-like, and enteroendocrine-like states (Fig. 2B). Broad EPCAM and KRT18 expression supported their epithelial origin (Fig. 2C). A prespecified 14-gene core subset defined for this study was used to calculate the cell-level score from the extended 60-gene disulfidptosis-related panel (Supplementary Table S1 (ESM 1)). Malignant cells were divided by score tertiles into low-score ($n = 2,004$), medium-score ($n = 2,003$), and high-score ($n = 2,004$) groups (Fig. 2D-F). SLC7A11 expression was increased in high-score cells, supporting the biological relevance of the stratification while not demonstrating that the cells were actively undergoing disulfidptosis (Fig. 2G).

High disulfidptosis-related scores define a proliferative and cytoskeleton-dependent malignant epithelial state

We compared high-score and low-score malignant epithelial cells to characterize the transcriptional programs associated with the disulfidptosis-related score. Differential expression analysis identified 1,890 upregulated genes and 847 downregulated genes in the high-score group (Fig. 3A). Representative genes related to cytoskeletal architecture, glucose uptake, and cystine transport, including ACTB, ACTG1, MYH9, SLC7A11, and SLC2A3, were significantly increased in high-score malignant cells (Fig. 3B). These findings suggest that high-score cells represent a state with strong dependence on glucose metabolism, amino-acid uptake, and actin-cytoskeleton homeostasis rather than simply a more aggressive phenotype.

Hallmark gene-set enrichment analysis showed positive enrichment of E2F targets, the G2/M checkpoint, and mitotic-spindle programs in high-score cells, whereas TNF-alpha/NF-kappaB and KRAS-related programs showed relative negative enrichment (Fig. 3C). Gene Ontology analysis highlighted mitotic-spindle organization, spindle-assembly checkpoints, microtubule cytoskeleton organization, actin

binding, and microtubule-associated processes, and KEGG analysis similarly identified cell-cycle and cytoskeleton-related pathways (Fig. 3D,E).

Taken together, the high-score state is best interpreted as a proliferative, metabolically engaged, and cytoskeleton-dependent malignant epithelial state with putative susceptibility to disulfide stress. This interpretation avoids equating a transcriptomic score with experimentally observed disulfidptotic cell death.

Low-score malignant epithelial cells exhibit stronger stemness-related features

Low-score malignant epithelial cells showed higher composite stemness, pluripotency, and cancer stem-cell marker scores than high-score cells (Fig. 4A-C). Epithelial-mesenchymal transition (EMT) scores also differed among groups, although the magnitude of variation was smaller than that observed for the stemness-related scores (Fig. 4D).

UMAP visualization showed heterogeneous distributions of composite stemness scores and disulfidptosis-related score groups across the malignant epithelial compartment rather than confinement to a single cluster (Fig. 4E,F). Thus, the score appears to capture a functional axis: high-score cells favor proliferative and cytoskeletal programs, whereas low-score cells exhibit stronger stemness-related features.

Cell-cell communication inference suggested that high-score malignant cells had stronger outgoing interactions with several immune and stromal populations, including fibroblasts and macrophages, whereas low-score cells showed relatively enhanced malignant-cell-to-malignant-cell communication (Fig. 4G,H). These results are computational predictions of ligand-receptor activity and do not establish direct physical or causal interactions.

A six-gene disulfidptosis-related signature stratifies overall survival across cohorts

In TCGA-STAD, patients with higher bulk disulfidptosis-related scores calculated from the extended 60-gene panel had longer overall survival than patients with lower scores (log-rank $P = 0.0042$; Fig. 5A). This association indicates that the overall gene-set program has prognostic relevance at the bulk-tumor level.

A prespecified 41-gene candidate subset of the extended panel was evaluated in the TCGA-STAD training cohort using cross-validated LASSO-Cox regression (Supplementary Table S1). The selected model at $\lambda_{\min} = 0.0374$ comprised ACTG1, CFL2, PRDX4, GPX4, PRDX2, and KEAP1 (Fig. 5B-D). ACTG1, CFL2, PRDX4, and GPX4 had positive coefficients, whereas PRDX2 and KEAP1 had negative coefficients.

Patients were stratified into high- and low-risk groups using the cohort-specific median risk score. The risk-ranked expression heatmap of the six model genes is shown in Fig. 5E. High-risk patients had significantly shorter overall survival in TCGA-STAD (log-rank $P = 2.17 \times 10^{-6}$; Fig. 5F). Using cohort-specific median stratification, the prognostic association was reproduced in GSE84437 ($P = 0.00014$; Fig. 5G) and GSE62254/ACRG (hazard ratio = 5.87, 95% confidence interval 3.08-11.19; $P < 0.001$; Fig. 5H).

The 14-gene cell-level score, the 60-gene bulk disulfidptosis-related score, and the six-gene prognostic risk score address distinct analytical questions and are not interchangeable. The first represents a study-defined core transcriptional program related to disulfidptosis at single-cell resolution, the second summarizes the broader pathway program in bulk tumors, and the third is a survival-weighted combination selected from 41 candidate genes. Accordingly, a high disulfidptosis-related score should not be interpreted as a high prognostic risk score.

High-risk tumors show stromal enrichment, lower purity, and differential predicted drug sensitivity

Functional enrichment analysis of genes associated with the high-risk group highlighted immune- and inflammation-related biological processes, including immune response, inflammatory response, leukocyte migration, cytokine production, and chemotaxis. Molecular-function and pathway analyses further implicated cytokine/chemokine activity and signaling pathways such as cytokine-cytokine receptor interaction, chemokine signaling, Toll-like receptor signaling, NF-kappaB signaling, PI3K-Akt signaling, and MAPK signaling (Fig. 6A). These findings suggest that the risk signature is also linked to immune-inflammatory and cytokine-signaling programs at the bulk-tumor level.

ESTIMATE analysis showed higher StromalScore and ESTIMATEScore values in high-risk tumors, together with lower tumor purity; the difference in ImmuneScore was less pronounced (Fig. 6B). These results suggest that the risk signature captures, at least in part, a stroma-enriched tumor ecosystem.

Somatic mutation profiling summarized the mutation landscape of the high-risk TCGA-STAD group (Fig. 6C). Predicted drug-sensitivity analysis showed that high-risk patients had lower predicted log-transformed IC50 values for lapatinib, gefitinib, erlotinib, dasatinib, and sapitinib, suggesting greater predicted sensitivity to these agents. No significant differences were observed for sorafenib, bortezomib, or vorinostat (Fig. 6D). These findings are pharmacogenomic predictions and require experimental and clinical validation before therapeutic interpretation.

The risk score is independently associated with overall survival and supports nomogram construction

Univariable and multivariable Cox regression analyses were conducted in 319 TCGA-STAD patients with complete survival and clinical information. After adjustment for age, sex, and AJCC stage, the six-gene risk score remained significantly associated with overall survival (Fig. 6E).

A nomogram integrating the risk score, age, sex, and AJCC stage was constructed to estimate 1-, 3-, and 5-year overall survival probabilities (Fig. 6F). The model achieved a concordance index of 0.734. Calibration curves showed overall agreement between predicted and observed survival probabilities at the evaluated time points (Fig. 6G). These findings support the incremental prognostic value of the risk score but do not establish clinical utility without independent prospective validation.

Discussion

This analysis addressed a specific cross-scale question: what cellular phenotype is represented by a disulfidptosis-related transcriptional program in gastric cancer, and whether related features retain patient-level prognostic associations across independent cohorts. The principal contribution is therefore not the six-gene model alone, but the integration of malignant epithelial cell-state characterization with reproducible bulk-cohort associations. This design provides a cellular context for disulfidptosis-related signals that would be obscured in bulk-tumor expression profiles.

High-score malignant epithelial cells were enriched for cell-cycle, mitotic-checkpoint, and cytoskeletal-remodeling programs and expressed higher levels of *SLC7A11*. This pattern is compatible with the established biology of disulfidptosis, in which *SLC7A11*-dependent cystine uptake, insufficient reducing capacity, and disulfide stress destabilize the actin cytoskeleton under glucose limitation [7,8]. Nevertheless, transcriptomic scoring cannot determine whether individual cells are undergoing disulfidptosis. The score should therefore be regarded as a disulfidptosis-related state characterized by cystine transport, redox adaptation, and cytoskeletal dependence.

Low-score cells exhibited stronger stemness, pluripotency, and cancer stem-cell marker programs. This inverse relationship suggests that stem-like and cytoskeleton-dependent proliferative states may occupy different positions along a malignant epithelial state axis. Such a relationship is biologically plausible because stem-like tumor cells can adopt metabolic and cell-cycle programs distinct from those of rapidly proliferating cells [15,16].

Communication inference suggested that high-score malignant cells were more strongly connected to immune and stromal compartments, whereas low-score cells showed relatively stronger malignant-cell-to-malignant-cell communication. These findings provide hypotheses regarding ecological roles within the tumor microenvironment but remain dependent on ligand-receptor expression and require spatial or functional validation [17].

The six-gene signature comprised genes involved in actin dynamics and redox homeostasis. *ACTG1* and *CFL2* regulate cytoskeletal organization; *PRDX2*, *PRDX4*, and *GPX4* participate in antioxidant defense; and *KEAP1* regulates NRF2-associated oxidative-stress responses. The model stratified overall survival

in the training cohort and two independent cohorts using cohort-specific median cutoffs, and the risk score remained associated with prognosis after adjustment for conventional clinical variables. These findings support reproducibility of the association across datasets, but they do not establish a fixed clinical cutoff or prospective utility.

The apparently divergent associations of the bulk disulfidptosis-related score and the six-gene risk score do not necessarily represent a contradiction. The single-cell analysis used a study-defined 14-gene core score, whereas the bulk gene-set score summarized an extended 60-gene program. The prognostic model was then selected from a 41-gene candidate subset and assigned cohort-trained survival weights to six genes. These three quantities should therefore be interpreted as related but analytically distinct measures, and the six-gene risk score is not a direct measurement of disulfidptosis.

High-risk tumors were associated with higher stromal content and lower tumor purity, suggesting that the model also reflects a stroma-enriched tumor ecosystem. The risk groups also differed in predicted sensitivity to several targeted agents, with lower predicted log-transformed IC50 values for lapatinib, gefitinib, erlotinib, dasatinib, and sapitinib in the high-risk group. These pharmacogenomic predictions are hypothesis-generating and should not be interpreted as evidence of clinical treatment response without experimental or prospective validation.

Limitations

- All analyses were retrospective and based on publicly available datasets; no prospective cohort, clinical specimen, or experimental model was used for validation.
- The 14-gene cell-level score was defined for this study and should be interpreted as a disulfidptosis-related transcriptional state, not as direct evidence that individual cells were undergoing disulfidptosis.
- Cell-level comparisons may be affected by within-patient dependence. Patient-level pseudobulk or mixed-effects analyses, together with spatial validation, would strengthen inference about malignant epithelial states and cell-cell communication.
- CNV inference, ligand-receptor communication, ESTIMATE scores, and oncoPredict drug-sensitivity estimates are computational inferences and do not demonstrate causality, physical interaction, or clinical treatment response.
- The six-gene model was evaluated using cohort-specific median cutoffs. External analyses therefore test reproducibility of prognostic association rather than performance of a fixed clinical threshold; prospective validation and assay standardization are required before clinical application.

Conclusion

This integrated analysis delineates a disulfidptosis-related malignant epithelial state axis in gastric cancer. High-score cells were characterized by SLC7A11 upregulation, proliferative and cytoskeletal

programs, and stronger predicted microenvironmental communication, whereas low-score cells exhibited stronger stemness-related features. A six-gene risk score showed reproducible survival associations across three cohorts and complemented conventional clinical variables. These results provide a hypothesis-generating framework for prioritizing disulfidptosis-related cell states and candidate biomarkers for experimental and prospective validation; they do not establish active disulfidptosis or a clinically ready prognostic assay.

Methods

Data acquisition and study design

Publicly available scRNA-seq data from GSE183904 were used to characterize the gastric cancer tumor microenvironment [6,29]. TCGA-STAD served as the training cohort for bulk transcriptomic and clinical analyses [28], whereas GSE84437 and GSE62254/ACRG served as independent cohorts for prognostic assessment [30,31]. Samples lacking the survival or clinical information required for a given analysis were excluded. Analysis-specific sample sizes therefore varied according to the availability of gene-expression, survival, somatic mutation, and complete clinicopathological data. The multivariable Cox analysis and nomogram included 319 TCGA-STAD patients with complete data.

Single-cell quality control, clustering, and visualization

Raw scRNA-seq reads were processed using Cell Ranger v7.1.0 to generate gene-barcode count matrices. Cells were retained when they contained 200-5,000 detected genes (nFeature_RNA), 500-30,000 total RNA counts (nCount_RNA), and <10% mitochondrial transcripts. Doublets were predicted and removed using DoubletFinder v2.0.3 with an expected doublet rate of 7.5% [22]. Variance stabilization and normalization were performed using SCTransform [9]. Highly variable genes were used for principal-component analysis, and JackStraw testing identified principal components 1-25 as significant. Batch effects were corrected with Harmony v0.1.1 using batch identity as a covariate [23]. A nearest-neighbor graph was constructed from the Harmony-corrected dimensions, and Leiden clustering was performed at resolutions ranging from 0.8 to 1.2 [24]. UMAP was used for two-dimensional visualization [10].

Cell-type annotation and tumor microenvironment characterization

Cell clusters were annotated using canonical marker genes and cluster-level expression patterns. Epithelial lineages were evaluated using *EPCAM*, *KRT18*, *MUC5AC*, and *ATP4A*; T cells using *CD3D*, *CD3E*, and *CD8A*; NK/gamma delta T cells using *NKG7*; macrophages and monocytes using *CD68* and *LYZ*; endothelial cells using *PECAM1* and *VWF*; fibroblasts and cancer-associated fibroblasts using *COL1A1*,

COL1A2, and *ACTA2*; and B/plasma-cell lineages using *MS4A1*, *CD79A*, *MZB1*, and *JCHAIN*. Cell-type proportions were summarized descriptively across normal and tumor samples.

Identification of malignant epithelial cells and disulfidptosis-related scoring

Large-scale CNV signals were inferred from ordered gene-expression profiles using non-epithelial immune and stromal populations as references, following the general strategy established for scRNA-seq tumor analysis [11]. Epithelial cells with broad chromosomal alterations and elevated CNV burden were classified as malignant in conjunction with epithelial marker expression. Malignant epithelial cells were then reclustered for downstream analysis.

A study-defined extended 60-gene disulfidptosis-related panel was assembled by integrating the original mechanistic framework with genes representing cystine transport, NADPH and redox homeostasis, mitochondrial metabolism, and actin-cytoskeleton regulation [7,8]. For cell-level scoring, a prespecified 14-gene core subset defined for this study was used: *SLC7A11*, *SLC3A2*, *RPN1*, *NCKAP1*, *WASF2*, *CYFIP1*, *ABI2*, *BRK1*, *RAC1*, *NDUFA11*, *NUBPL*, *OXSM*, *LRPPRC*, and *GYS1*. The complete 60-gene panel and subset membership are provided in Supplementary Table S1. Cells were ranked by score and divided into low, medium, and high groups using tertile cutoffs. The distribution of *SLC7A11* expression was examined as a biological consistency check. The score was interpreted as a disulfidptosis-related transcriptional state rather than direct evidence of cell death.

Differential expression and pathway enrichment analyses

Differential expression analysis was performed between high-score and low-score malignant epithelial cells. Multiple-testing correction was applied using the Benjamini-Hochberg procedure. Differentially expressed genes were subjected to Gene Ontology and KEGG enrichment analyses using a clusterProfiler-based workflow [12]. Hallmark gene-set enrichment analysis was performed against the Molecular Signatures Database collection [13,14]. Enrichment results were interpreted using normalized enrichment scores and adjusted P values.

Stemness scoring and cell-cell communication analysis

Composite stemness, pluripotency, cancer stem-cell marker, and EMT scores were calculated for malignant epithelial cells using predefined gene sets, including established stemness-related signatures [15,16]. Scores were compared across the three disulfidptosis-related groups. Cell-cell communication was inferred using a CellChat-based ligand-receptor framework [17], and outgoing and incoming interactions involving malignant epithelial groups and tumor-microenvironment populations were compared. Communication results were treated as expression-based predictions.

Bulk disulfidptosis-related scoring and prognostic model construction

A gene-set variation framework was used to quantify bulk disulfidptosis-related scores from the extended 60-gene panel in TCGA-STAD [18]. Patients were stratified by the median score for Kaplan-Meier analysis. A prespecified 41-gene candidate subset of the 60-gene panel was entered into LASSO-Cox regression in the TCGA-STAD training cohort [19,20]; the candidate genes and their membership in the 14-gene core set are listed in Supplementary Table S1. Ten-fold cross-validation was used to select the penalty parameter. The six-gene risk score was calculated using the following formula: Risk score = $0.015 \times \text{ACTG1} + 0.143 \times \text{CFL2} + 0.147 \times \text{PRDX4} + 0.204 \times \text{GPX4} - 0.011 \times \text{PRDX2} - 0.331 \times \text{KEAP1}$, where each term represents the corresponding gene-expression value multiplied by its fitted Cox coefficient. The same coefficient formula was applied to GSE84437 and GSE62254/ACRG, and patients were stratified using the cohort-specific median risk score. Because each cohort was stratified using its own median, these external-cohort analyses assessed the reproducibility of the prognostic association rather than the performance of a fixed clinical cutoff.

Tumor microenvironment, mutation, predicted drug sensitivity, and nomogram analyses

The ESTIMATE algorithm was used to derive ImmuneScore, StromalScore, ESTIMATEScore, and inferred tumor purity from bulk expression data [21]. Somatic mutation data were summarized using maftools to characterize the mutation landscape [25]. Predicted log-transformed half-maximal inhibitory concentration (IC50) values for lapatinib, gefitinib, erlotinib, sorafenib, dasatinib, sapitinib, bortezomib, and vorinostat were estimated from bulk transcriptomic profiles using the oncoPredict R package [26], with the preprocessed GDSC2 cell-line expression matrix (GDSC2_Expr) and drug-response matrix (GDSC2_Res) used as the training reference. Predicted values were compared between the low- and high-risk groups using the Wilcoxon rank-sum test. Multiple-testing correction was performed using the Benjamini-Hochberg procedure, and adjusted P values were reported.

Univariable and multivariable Cox proportional-hazards models were used to assess the association of the risk score with overall survival after adjustment for age, sex, and AJCC stage. A nomogram integrating these variables was constructed to estimate 1-, 3-, and 5-year overall survival. Model discrimination was summarized using the concordance index, and calibration curves compared predicted with observed survival probabilities [27].

Statistical analysis

Continuous variables were compared between two groups using the Wilcoxon rank-sum test and among three or more groups using the Kruskal-Wallis test unless otherwise specified. Pairwise comparisons

were performed following an overall multi-group test where applicable. Survival distributions were estimated using the Kaplan-Meier method and compared using the log-rank test. Cox models were used to estimate hazard ratios and 95% confidence intervals. All tests were two-sided, and $P < 0.05$ was considered statistically significant. Adjusted P values are reported for analyses involving multiple testing.

Use of generative AI

During manuscript preparation, ChatGPT (OpenAI) was used for English-language editing and wording refinement. The tool was not used to generate or modify research data or statistical results. All scientific content, interpretations, and conclusions were reviewed and verified by the authors, who take full responsibility for the manuscript.

Declarations

Acknowledgements

The authors thank The Cancer Genome Atlas (TCGA), the Gene Expression Omnibus (GEO), the Genomics of Drug Sensitivity in Cancer (GDSC), and the investigators of the original studies for making their data publicly available.

Author contributions

Z.H. and S.L. conceived and designed the study and contributed equally to this work. Z.H., S.L., J.R.G., and J.H.L. collected, curated, and processed the data and performed the bioinformatics analyses. Z.H. and X.H. prepared and organized the figures. J.P.J. and Q.S. contributed to the interpretation of the results. M.W., J.W., D.D.F., and Z.P.C. contributed to discussion and interpretation of the findings and reviewed the manuscript. Z.H. drafted the original manuscript. S.L., J.L., and L.C. critically revised the manuscript. J.L. and L.C. supervised and guided the study. All authors reviewed and approved the final manuscript.

Funding

This work was supported by the Liaoning Provincial Department of Science and Technology Project (No. 2026JH2/101300040).

Data availability

The datasets analyzed in this study are publicly available. The single cell RNA sequencing dataset GSE183904 used for tumor microenvironment and malignant epithelial cell state analyses is available from the NCBI Gene Expression Omnibus at <https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE183904>. TCGA STAD served as the bulk transcriptomic and clinical training cohort and is available through the National Cancer Institute Genomic Data Commons project record at <https://portal.gdc.cancer.gov/projects/TCGA-STAD>. The independent prognostic validation cohorts GSE84437 and GSE62254, also referred to as the ACRG cohort, are available at <https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE84437> and <https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE62254>, respectively. The preprocessed GDSC2 cell-line expression and drug-response matrices used as the oncoPredict training reference were obtained from the oncoPredict training-data repository at <https://osf.io/temyk/>. Hallmark gene sets used for gene set enrichment analysis were obtained from the Molecular Signatures Database at <https://www.gsea-msigdb.org/gsea/msigdb/genesets.jsp?collection=H>. No new primary datasets were generated in this study. Processed data supporting the reported analyses are available from the corresponding authors upon reasonable request.

Code availability

The analysis scripts used to generate the reported results are available from the corresponding author upon reasonable request. No custom software was developed for this study.

Ethical Approval and Accordance

Not applicable. This study analyzed only publicly available, de-identified datasets and did not involve the recruitment of human participants, collection of human tissues, or use of animals. Therefore, formal ethics committee approval was not required.

Consent to participate

Not applicable. This study used publicly available, de-identified datasets and did not involve direct participation of identifiable human participants. Therefore, informed consent to participate was not required.

Consent to publish

Not applicable.

Competing interests

The authors declare no competing interests.

References

1. Sung, H. et al. Global cancer statistics 2020: GLOBOCAN estimates of incidence and mortality worldwide for 36 cancers in 185 countries. *CA Cancer J. Clin.* **71**, 209-249 (2021).
2. Smyth, E. C., Nilsson, M., Grabsch, H. I., van Grieken, N. C. T. & Lordick, F. Gastric cancer. *Lancet* **396**, 635-648 (2020).
3. Janjigian, Y. Y. et al. First-line nivolumab plus chemotherapy versus chemotherapy alone for advanced gastric, gastro-oesophageal junction, and oesophageal adenocarcinoma: the CheckMate 649 study. *Lancet* **398**, 27-40 (2021).
4. The Cancer Genome Atlas Research Network. Comprehensive molecular characterization of gastric adenocarcinoma. *Nature* **513**, 202-209 (2014).
5. Cristescu, R. et al. Molecular analysis of gastric cancer identifies subtypes associated with distinct clinical outcomes. *Nat. Med.* **21**, 449-456 (2015).
6. Kumar, V. et al. Single-cell atlas of lineage states, tumor microenvironment, and subtype-specific expression programs in gastric cancer. *Cancer Discov.* **12**, 670-691 (2022).
7. Liu, X. et al. Actin cytoskeleton vulnerability to disulfide stress mediates disulfidptosis. *Nat. Cell Biol.* **25**, 404-414 (2023).
8. Zheng, T., Liu, Q., Xing, F., Zeng, C. & Wang, W. Disulfidptosis: a new form of programmed cell death. *J. Exp. Clin. Cancer Res.* **42**, 137 (2023).
9. Hafemeister, C. & Satija, R. Normalization and variance stabilization of single-cell RNA-seq data using regularized negative binomial regression. *Genome Biol.* **20**, 296 (2019).
10. McInnes, L., Healy, J. & Melville, J. UMAP: uniform manifold approximation and projection for dimension reduction. Preprint at <https://arxiv.org/abs/1802.03426> (2018).
11. Tirosh, I. et al. Dissecting the multicellular ecosystem of metastatic melanoma by single-cell RNA-seq. *Science* **352**, 189-196 (2016).
12. Yu, G., Wang, L. G., Han, Y. & He, Q. Y. clusterProfiler: an R package for comparing biological themes among gene clusters. *OMICS* **16**, 284-287 (2012).
13. Subramanian, A. et al. Gene set enrichment analysis: a knowledge-based approach for interpreting genome-wide expression profiles. *Proc. Natl Acad. Sci. USA* **102**, 15545-15550 (2005).
14. Liberzon, A. et al. The Molecular Signatures Database Hallmark gene set collection. *Cell Syst.* **1**, 417-425 (2015).
15. Malta, T. M. et al. Machine learning identifies stemness features associated with oncogenic dedifferentiation. *Cell* **173**, 338-354.e15 (2018).
16. Batlle, E. & Clevers, H. Cancer stem cells revisited. *Nat. Med.* **23**, 1124-1134 (2017).
17. Jin, S. et al. Inference and analysis of cell-cell communication using CellChat. *Nat. Commun.* **12**, 1088 (2021).

18. Hanzelmann, S., Castelo, R. & Guinney, J. GSVA: gene set variation analysis for microarray and RNA-seq data. *BMC Bioinformatics* **14**, 7 (2013).
19. Tibshirani, R. The lasso method for variable selection in the Cox model. *Stat. Med.* **16**, 385-395 (1997).
20. Friedman, J., Hastie, T. & Tibshirani, R. Regularization paths for generalized linear models via coordinate descent. *J. Stat. Softw.* **33**, 1-22 (2010).
21. Yoshihara, K. et al. Inferring tumour purity and stromal and immune cell admixture from expression data. *Nat. Commun.* **4**, 2612 (2013).
22. McGinnis, C. S., Murrow, L. M. & Gartner, Z. J. DoubletFinder: doublet detection in single-cell RNA sequencing data using artificial nearest neighbors. *Cell Syst.* **8**, 329-337.e4 (2019).
23. Korsunsky, I. et al. Fast, sensitive and accurate integration of single-cell data with Harmony. *Nat. Methods* **16**, 1289-1296 (2019).
24. Traag, V. A., Waltman, L. & van Eck, N. J. From Louvain to Leiden: guaranteeing well-connected communities. *Sci. Rep.* **9**, 5233 (2019).
25. 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**, 1747-1756 (2018).
26. Maeser, D., Gruener, R. F. & Huang, R. S. oncoPredict: an R package for predicting in vivo or cancer patient drug response and biomarkers from cell-line screening data. *Brief. Bioinform.* **22**, bbab260 (2021).
27. Harrell, F. E. Jr. *Regression Modeling Strategies*. 2nd ed. (Springer, Cham, 2015).
28. National Cancer Institute. Genomic Data Commons Data Portal: TCGA-STAD. <https://portal.gdc.cancer.gov/projects/TCGA-STAD> (accessed 28 July 2026).
29. Gene Expression Omnibus. GSE183904. <https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE183904> (accessed 28 July 2026).
30. Gene Expression Omnibus. GSE84437. <https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE84437> (accessed 28 July 2026).
31. Gene Expression Omnibus. GSE62254. <https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE62254> (accessed 28 July 2026).

Figures

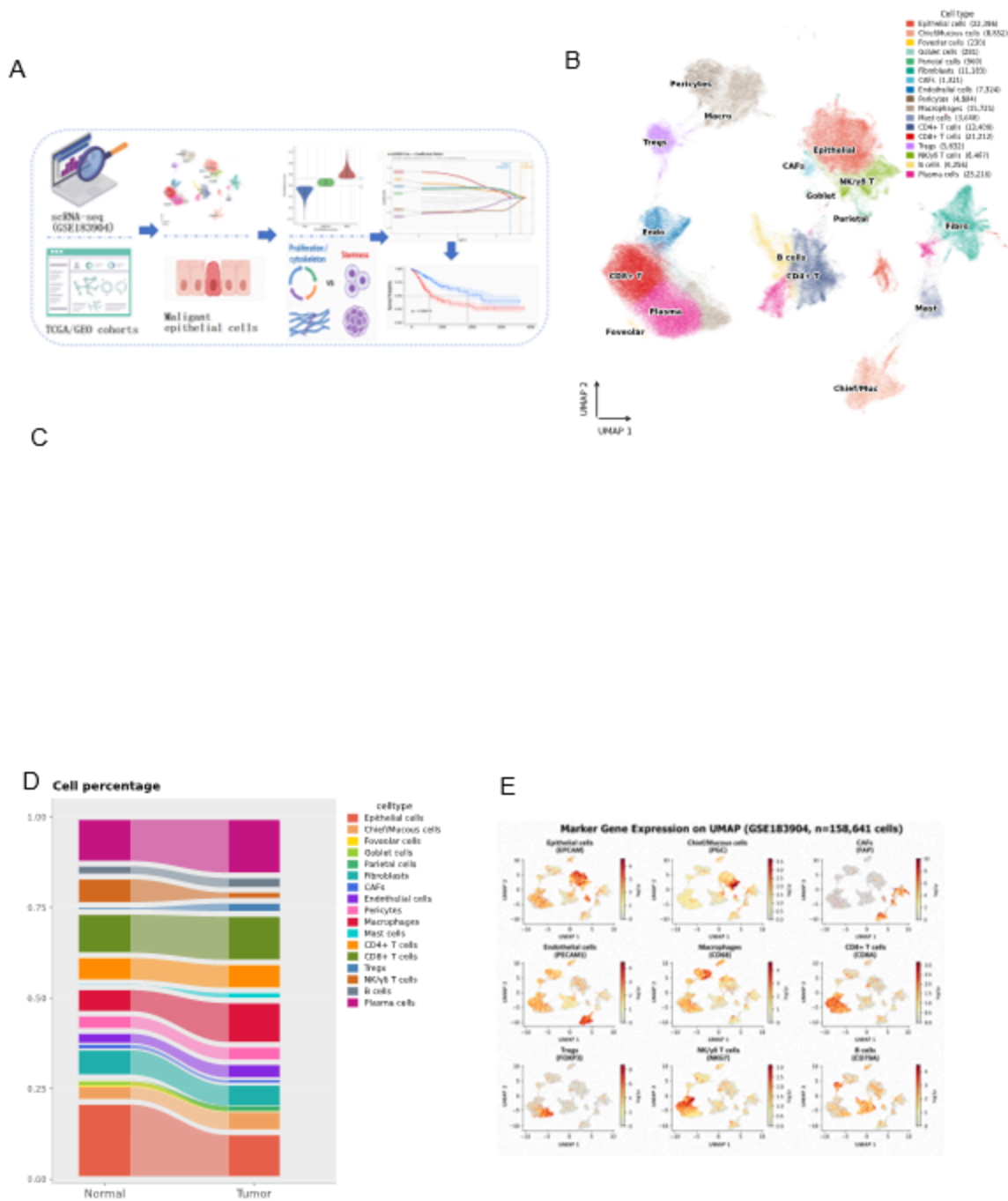


Figure 1

Single-cell atlas of gastric cancer and tumor microenvironment remodeling. (A) Overview of the scRNA-seq analysis workflow. (B) UMAP visualization of 158,641 high-quality cells annotated into 17 major cell populations. (C) Dot plot of representative marker genes across annotated cell types; dot size denotes the proportion of expressing cells and color denotes scaled average expression. (D) Descriptive comparison of cell-type proportions between normal and tumor samples. (E) UMAP feature plots of

(C) Expression of EPCAM and KRT18 in malignant epithelial cells. (D) Distribution of the study-defined 14-gene core disulfidptosis-related score across the low-, medium-, and high-score groups. (E) UMAP distribution of the three score groups. (F) Continuous disulfidptosis-related scores projected onto UMAP space. (G) SLC7A11 expression in malignant epithelial cells. CNV, copy-number variation

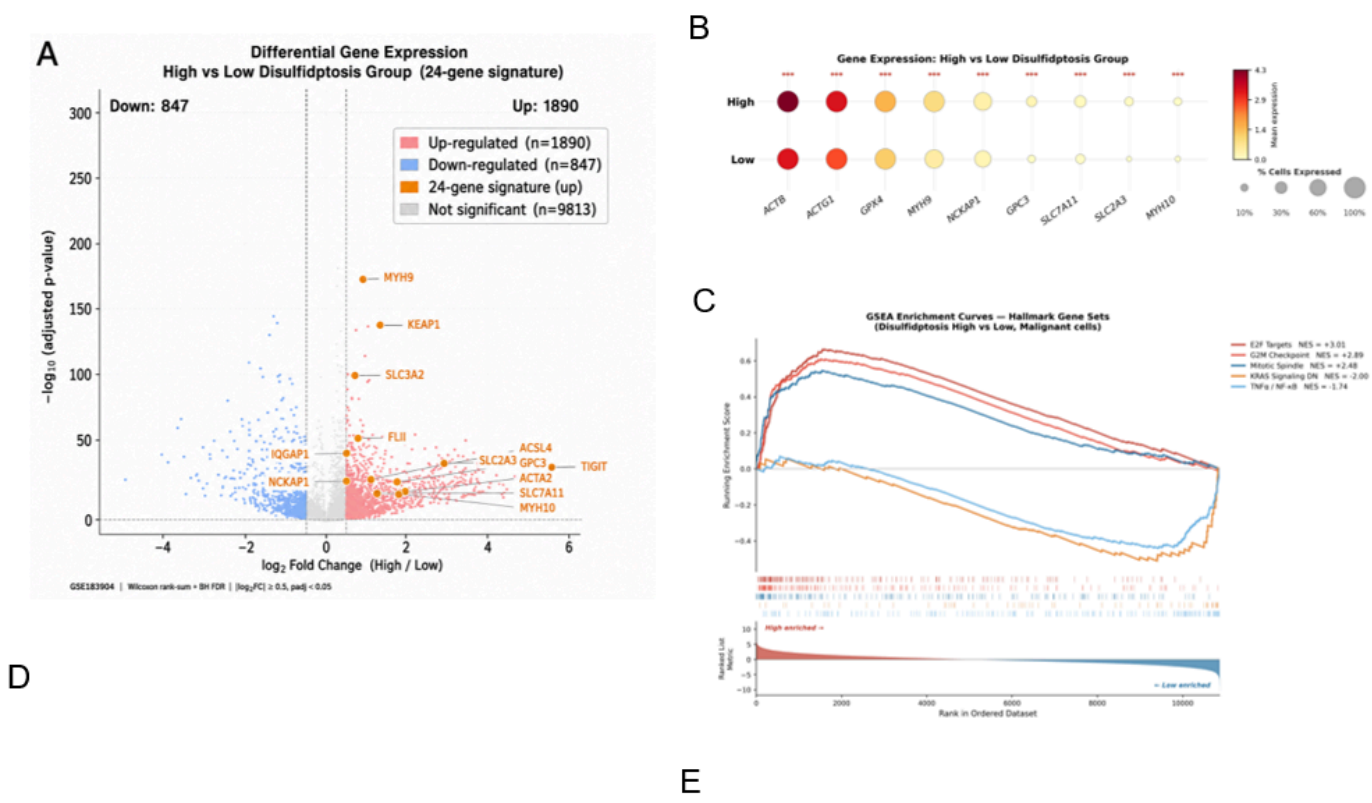


Figure 3

Transcriptional and pathway features of high- versus low-score malignant epithelial cells. (A) Volcano plot showing 1,890 upregulated genes and 847 downregulated genes in high-score cells relative to low-score cells. (B) Expression of representative cytoskeleton-, glucose-transport-, and cystine-transport-related genes. (C) Hallmark gene-set enrichment analysis. (D) Gene Ontology enrichment analysis. (E) KEGG pathway enrichment analysis. In enrichment plots, dot size represents gene count and color represents the adjusted P value. GSEA, gene-set enrichment analysis; GO, Gene Ontology; KEGG, Kyoto Encyclopedia of Genes and Genomes

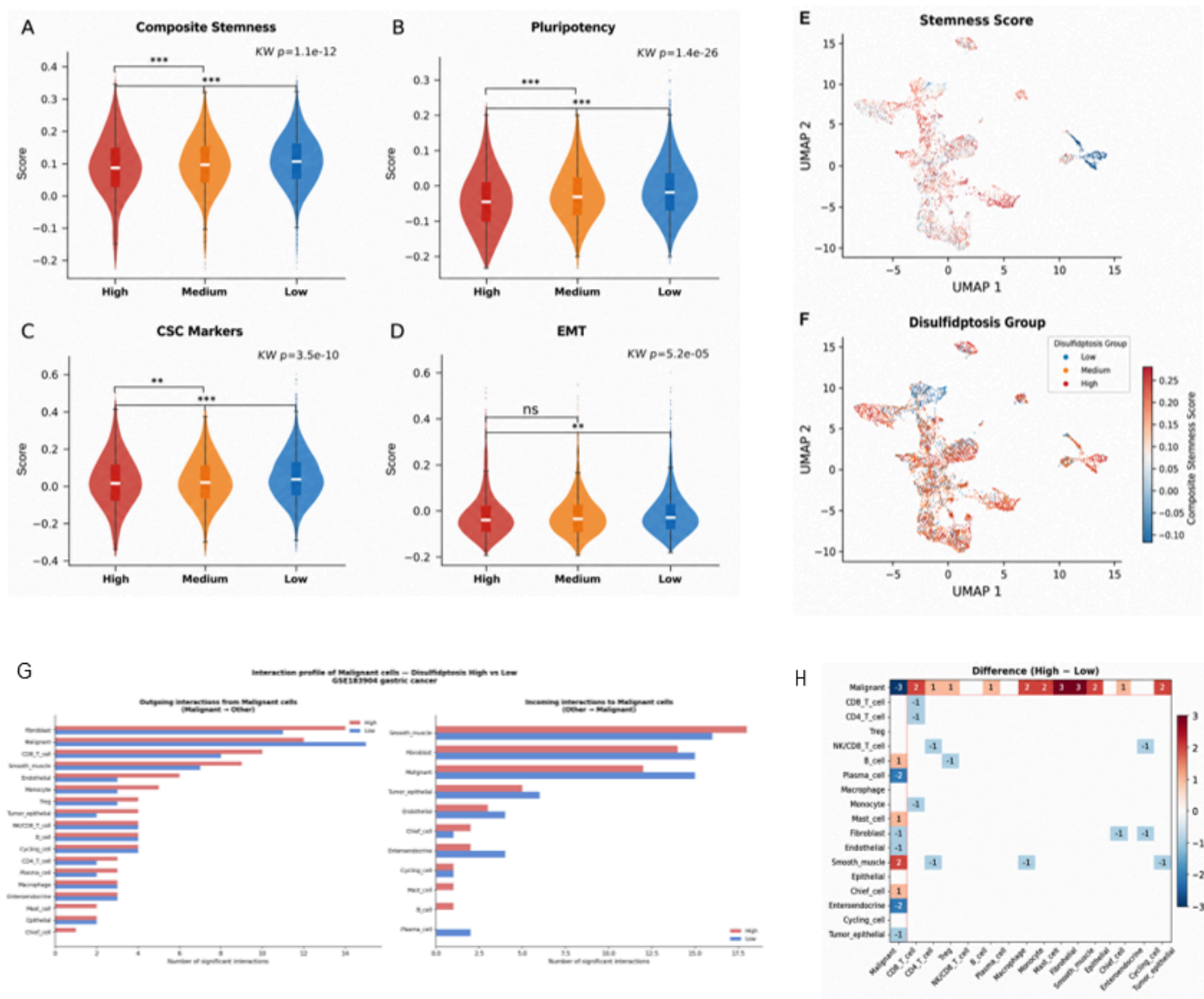


Figure 4

Stemness and cell-cell communication features across disulfidptosis-related malignant epithelial states. (A-D) Comparison of composite stemness, pluripotency, cancer stem-cell marker, and EMT scores among the low-, medium-, and high-score groups. (E) UMAP visualization of the composite stemness score. (F) UMAP distribution of the three disulfidptosis-related score groups. (G) Comparison of inferred outgoing and incoming interaction numbers between high- and low-score malignant cells across tumor-microenvironment populations. (H) Heatmap showing high-minus-low differences in predicted cell-cell communication. ns, $P \geq 0.05$; * $P < 0.05$; ** $P < 0.01$; *** $P < 0.001$. EMT, epithelial-mesenchymal transition

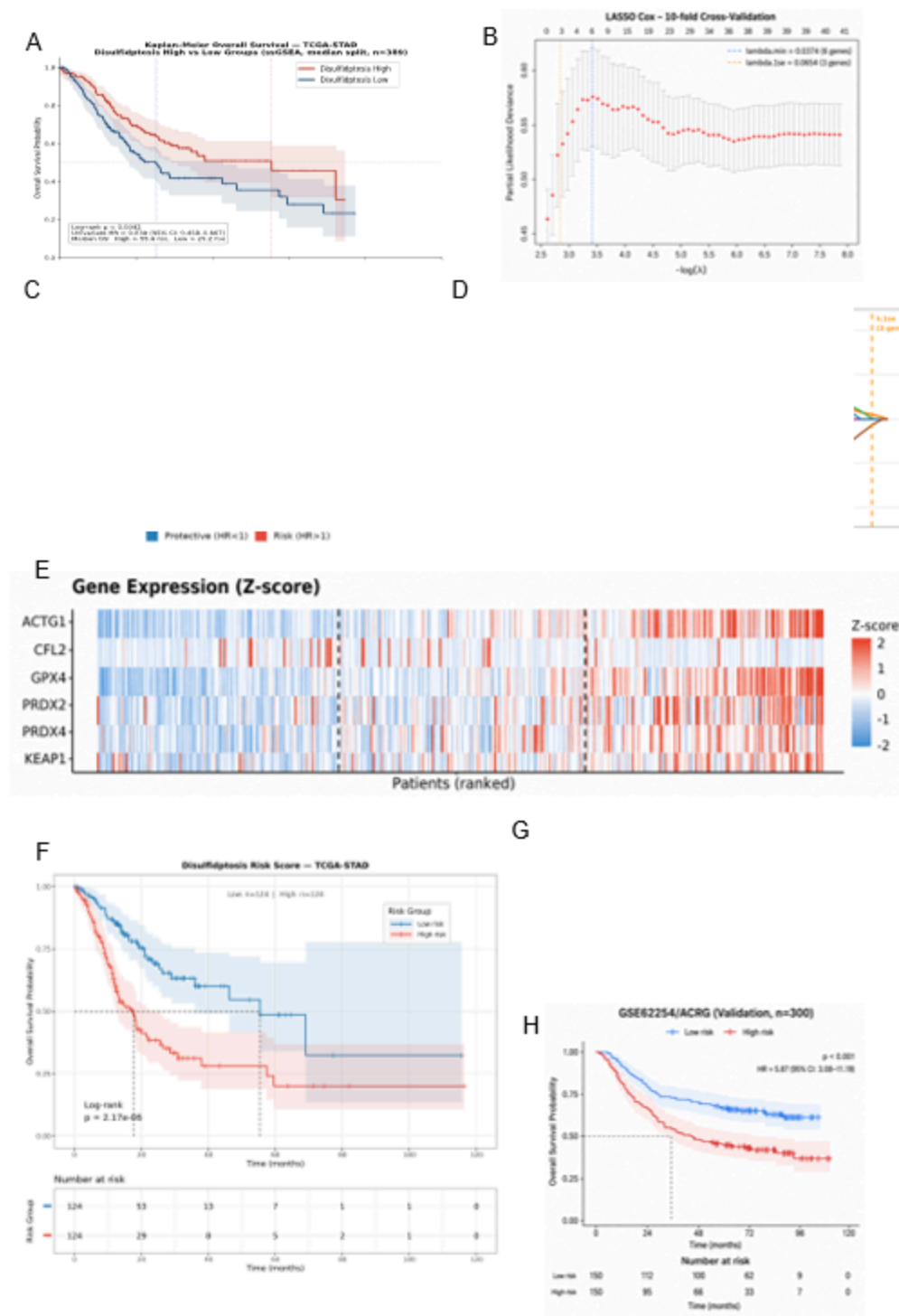


Figure 5

Construction and prognostic assessment of the six-gene signature across independent cohorts. (A) Overall survival according to the extended 60-gene bulk disulfidptosis-related score in TCGA-STAD. (B) Ten-fold cross-validation of the LASSO-Cox model using the 41-gene candidate subset. (C) Coefficients of the six genes retained in the final model. (D) LASSO coefficient paths for the 41 candidate genes. (E) Z-score expression heatmap of the six model genes in risk-ranked TCGA-STAD patients. (F-H) Kaplan-

Meier overall survival curves in TCGA-STAD, GSE84437, and GSE62254/ACRG, respectively. LASSO, least absolute shrinkage and selection operator; OS, overall survival

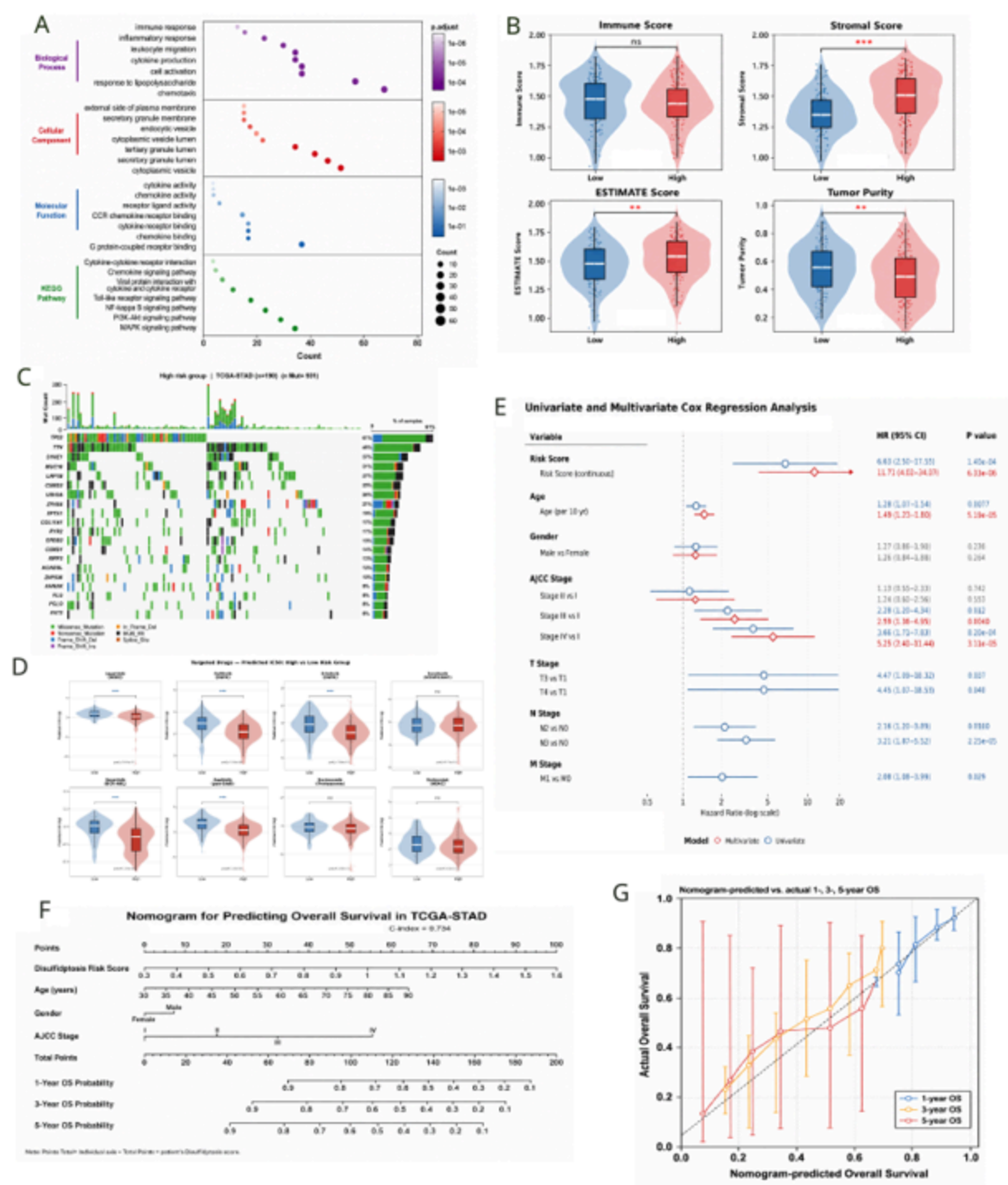


Figure 6

Biological characteristics, predicted drug sensitivity, and clinical value of the disulfidptosis-related risk model. (A) GO and KEGG enrichment analyses of genes associated with the high-risk group, highlighting immune/inflammatory, cytokine/chemokine, and related signaling programs; dot size represents gene count and color represents the adjusted P value. (B) Comparison of ImmuneScore, StromalScore,

ESTIMATEScore, and tumor purity between the low- and high-risk groups. (C) Somatic mutation landscape of the high-risk TCGA-STAD group; each column represents one patient and colors indicate mutation types. (D) Comparison of predicted log-transformed IC50 values for lapatinib, gefitinib, erlotinib, sorafenib, dasatinib, sapitinib, bortezomib, and vorinostat between the low- and high-risk groups; adjusted P values and significance levels are shown. (E) Univariable and multivariable Cox regression analyses of the risk score and clinical variables; hazard ratios and 95% confidence intervals are shown. (F) Nomogram integrating the risk score, age, sex, and AJCC stage to predict 1-, 3-, and 5-year overall survival; the C-index was 0.734. (G) Calibration curves comparing predicted and observed 1-, 3-, and 5-year overall survival probabilities; the diagonal dashed line indicates perfect agreement. GO, Gene Ontology; KEGG, Kyoto Encyclopedia of Genes and Genomes; IC50, half-maximal inhibitory concentration; HR, hazard ratio; CI, confidence interval; OS, overall survival

Supplementary Files

This is a list of supplementary files associated with this preprint. Click to download.

- [ESM1SupplementaryTableS1.xlsx](#)