Comprehensive analysis of multiomics data for the identification of a cuproptosis-related gene signature predicting prognostic outcomes and drug responses in gastric cancer

Hongxin Huang
The First Affiliated Hospital of Nanjing Medical University

Chuming Zhu
The First Affiliated Hospital of Nanjing Medical University

Mengpei Yan
The First Affiliated Hospital of Nanjing Medical University

Jihuan Wang
The First Affiliated Hospital of Nanjing Medical University

Jialun Lv
The First Affiliated Hospital of Nanjing Medical University

Lang Fang
The First Affiliated Hospital of Nanjing Medical University

Penghui Xu
The First Affiliated Hospital of Nanjing Medical University

Zetian Chen
The First Affiliated Hospital of Nanjing Medical University

Weizhi Wang
The First Affiliated Hospital of Nanjing Medical University

Zekuan Xu (✉ xuzekuan@njmu.edu.cn)
The First Affiliated Hospital of Nanjing Medical University

Research Article

Keywords: Cuproptosis, Gastric cancer, Prognosis, Drug response, Overall survival

Posted Date: August 28th, 2023

DOI: https://doi.org/10.21203/rs.3.rs-3283955/v1

License: This work is licensed under a Creative Commons Attribution 4.0 International License. 
Read Full License
Abstract

Background: Cuproptosis, a recently elucidated copper-dependent mechanism of cell death associated with the tricarboxylic acid cycle, lacks a comprehensive understanding of its relation to clinical prognosis and drug response in gastric cancer (GC). This study aims to discern potential prognostic signatures of cuproptosis-related genes (CRGs) and evaluate drug response.

Methods: Using publicly available datasets from TCGA and GEO, we initially obtained transcriptomic and clinical data of GC patients. We employed consensus clustering approach to delineate molecular subtypes based on the expression of CRGs. Utilizing least absolute shrinkage and selection operator (LASSO) regression analysis, we formulated a prognostic signature derived from the differentially expressed genes among these molecular subtypes. We constructed a nomogram that amalgamates both clinical characteristics and the prognostic model to provide a comprehensive prognosis prediction. Rigorous assessment of prognostic performance was carried out through Kaplan–Meier curve analysis, the log-rank test, univariate and multivariate Cox regression, and time-dependent ROC curve analysis. Tumor Immune Dysfunction and Exclusion (TIDE) and the pRRophetic package in R were used to assess the potential response to chemotherapy and immunotherapy. Seurat was utilized to analyze the general characterization of the single-cell dataset. Additionally, the validation of hub gene expression in both cells and clinical samples was undertaken via qRT–PCR.

Results: Upon conducting an exhaustive investigation into the distinct differential expression and prognostic implications of each CRG, we delineated two distinct cuproptosis-associated molecular subtypes. Following Lasso regression analyses, we formulated a prognostic model comprising six specific genes. Patients were effectively stratified into either high-risk or low-risk categories by utilizing this model. Patients classified as high-risk experienced poorer prognosis and were associated with higher TNM stages compared to those with low risk. Furthermore, patients belonging to the low-risk group exhibited enhanced benefits from chemotherapeutic drugs and demonstrated better susceptibility to immunotherapy. The validation of our prognostic model’s efficacy was established through ROC analysis, affirming its commendable sensitivity and specificity.

Conclusions: Our study illuminates the significance of cuproptosis in drug response and clinical prognosis in Asian GC patients, underscoring its clinical significance and providing a reliable tool for predicting overall survival in this patient population.

1. Background

Gastric cancer (GC) remains a prominent gastrointestinal malignancy with substantial global impact, resulting in an annual incidence of over one million new cases and approximately 0.76 million associated deaths worldwide[1]. Notably, regions including East Asia, Eastern Europe, and South America bear a disproportionate burden of this ailment, thereby engendering substantial challenges in the effective management of its incidence and corresponding mortality rates[2, 3]. It is imperative to acknowledge that
GC exhibits notable biological disparities between Asian and non-Asian populations[4], thereby complicating the establishment of a universally applicable unified predictive measure. Despite the widely used tumor node metastasis (TNM) classification for GC, its prognostic value is limited due to the inherent heterogeneity of the disease and the varying treatment responses observed among patients with similar classifications. Consequently, an exigent necessity emerges for the identification of steadfast prognostic markers and plausible therapeutic targets, with the ultimate aim of enhancing patient outcomes.

Copper, as a pivotal participant in cellular signaling pathways, has emerged as a contributory factor in the advancement and progression of tumors, actively fostering tumor cell proliferation, angiogenesis, and metastasis[5]. Studies have illuminated the role of copper in promoting tumorigenesis through the activation of the phosphoinositide 3-kinase (PI3K)-protein kinase B oncogenic signaling cascade, facilitated by a copper transporter[6]. Recently, considerable attention has been directed toward cuproptosis, a novel modality of cell death. Cuproptosis occurs when copper interacts directly with lipoylated components of the tricarboxylic acid (TCA) cycle, culminating in the accumulation of lipoylated proteins and the ensuing degradation of iron-sulfur cluster proteins. This intricate sequence ultimately precipitates proteotoxic stress and culminates in cell death [7]. Remarkably, this phenomenon has been linked to the pathogenesis of diverse malignancies, spanning from liver and colorectal cancers to certain hematologic neoplasms such as acute myeloid leukemia [8–10], suggesting its potential involvement in a broader spectrum of malignancies. Notably, typical copper ionophores, such as disulfiram and elesclomol, have been utilized in cancer therapeutics due to their ability to induce cuproptosis by facilitating the intracellular delivery of copper ions[11, 12].

Delving deeper into the realm of cuproptosis-related genes (CRGs) holds the promise of furnishing invaluable elucidations concerning the intricate regulatory underpinnings of cuproptosis within the landscape of diseases[13]. Moreover, the predictive value of cell death patterns, including ferroptosis, pyroptosis, and disulfdptosis, has been increasingly recognized in assessing prognosis, characterizing the tumor immune microenvironment (TIME), and predicting immunotherapy responses in GC patients[14–16].

In this study, we conducted a comprehensive investigation into the expression patterns and functions of CRGs in gastric cancer. Building upon these findings, we introduced an innovative risk model to evaluate its prognostic value in GC patients. By juxtaposing the predictive efficacy of this risk model against other clinical characteristics, we aim to provide a more robust clinical predictor of therapeutic response and prognosis in GC. Additionally, we examined the association between risk scores and various clinical characteristics, immunotherapy scores, and drug sensitivity in GC patients. Our study endeavors to contribute to a better understanding of GC and potentially improve personalized treatment strategies for better patient outcomes.

2. Methods
2.1 Data acquisition

The data utilized in this study were sourced from reputable public databases. Specifically, the RNAseq data encompassing 100 normal tissue samples and 300 gastric cancer samples originating from the ACRG cohort (GSE66229) and data stemming from 200 gastric cancer patients (excluding 9) hailing from the Singapore Patient Cohort (GSE15459) constituted our training set[17]. The original files were downloaded and subjected to backdrop and quantile normalization. To mitigate nonbiological technological biases, the "ComBat" algorithm, an integral component of the "sva" package, was invoked to execute batch effect correction[18]. The testing set consisted of 340 patients from The Cancer Genome Atlas (TCGA) STAD, excluding 35 patients with incomplete clinical information. The TCGA cohort was subsequently partitioned into two distinct subgroups: an Asian cohort comprising 72 patients and a non-Asian cohort encompassing 268 patients, delineated on the basis of patient ethnicity. The clinical information of all included patients is provided in Table 1. Additionally, single-cell RNA-Seq data from GSE163558 were acquired for further analysis[19]. Furthermore, 13 cuproptosis-related genes were selected from a previous study[13], as listed in Supplementary Table S1. The schematic representation of our study's design is visually articulated in Figure 1.

Table 1: The clinical information of all included patients in the current study.
### Covariates

<table>
<thead>
<tr>
<th>Covariates</th>
<th>Training Set (n=491)</th>
<th>Testing Set#1 (n=72)</th>
<th>Testing Set#2 (n=268)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Database</td>
<td>GSE66229+GSE15459</td>
<td>TCGA-STAD-Asian</td>
<td>TCGA-STAD-nonAsian</td>
</tr>
<tr>
<td>Age, years</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>≤ 65</td>
<td>257</td>
<td>31</td>
<td>120</td>
</tr>
<tr>
<td>65</td>
<td>234</td>
<td>41</td>
<td>148</td>
</tr>
<tr>
<td>Gender</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Female</td>
<td>168</td>
<td>26</td>
<td>96</td>
</tr>
<tr>
<td>Male</td>
<td>323</td>
<td>46</td>
<td>172</td>
</tr>
<tr>
<td>Stage</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Stage I</td>
<td>61</td>
<td>7</td>
<td>40</td>
</tr>
<tr>
<td>Stage II</td>
<td>126</td>
<td>35</td>
<td>74</td>
</tr>
<tr>
<td>Stage III</td>
<td>168</td>
<td>28</td>
<td>120</td>
</tr>
<tr>
<td>Stage IV</td>
<td>136</td>
<td>2</td>
<td>33</td>
</tr>
<tr>
<td>Lauren's Classification</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Intestinal</td>
<td>244</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>Mixed</td>
<td>37</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>Diffuse</td>
<td>210</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>Censor</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Alive</td>
<td>244</td>
<td>52</td>
<td>150</td>
</tr>
<tr>
<td>Dead</td>
<td>247</td>
<td>20</td>
<td>118</td>
</tr>
</tbody>
</table>

### 2.2 Differential Expression Analysis and Functional Enrichment

Differential expression analysis of genes was performed employing the "limma" R package[20], considering a stringent significance threshold of \( p < 0.05 \) and \(|\log_{2}FC| > 1.0\). To unveil intricate molecular interactions, we harnessed the capabilities of the STRING database to dissect protein–protein interaction networks. GO/KEGG enrichment analysis was conducted with the DAVID Bioinformatics Resources platform[21,22]. The results were visualized through the ggplot2 package. Gene set enrichment analysis (GSEA) was performed using GSEA software (version 4.3.2)[23,24].

### 2.3 Consensus clustering for cuproptosis-associated molecular subtypes

The prognostic relevance of the CRGs was evaluated through univariate Cox regression analysis. Consensus clustering analysis was conducted by the "ConsensusCluster Plus" R package to identify
cuproptosis-associated molecular subtypes in GC patients[25]. The optimal cluster count (k) was
determined by pinpointing the inflection point within the sum of squared error (SSE) curve while
considering k values from 2 to 6. The stability of the identified subtypes was confirmed by applying the t-
distributed stochastic neighbor embedding (tSNE) algorithm. Kaplan–Meier survival analysis was
conducted to assess the overall survival (OS) of each cuproptosis-associated subtype.

2.4 Construction and validation of the cuproptosis-related prognostic signature

A total of 160 differentially expressed genes (DEGs) expressed across the two cuproptosis-associated
subtypes were retained. Univariate Cox regression analysis was conducted on these DEGs to unearth
potential prognostic markers in the training set. A cuproptosis-related prognostic signature (CRPS) was
then constructed using the least absolute shrinkage and selection operator (LASSO) method[26]. For each
individual sample, a personalized risk score was computationally ascertained, employing the formula
Risk Score = Σ (Coef * Exp), where Coef and Exp correspond to the coefficients and expression levels
of each gene used in the analysis. With the median risk score as a demarcation, patients were judiciously
categorized into high- and low-risk groups. The predictive performance of the CRPS was subjected to a
comprehensive evaluation, incorporating both Kaplan–Meier survival analysis and time-dependent
receiver operating characteristic curves (ROC). The results were validated in the TCGA-Asian and TCGA-
non-Asian sets.

2.5 Clinical analysis and construction of the nomogram

Comparisons of risk scores across diverse clinical characteristics, including age, sex, pathological stage,
and Lauren's classification, were systematically performed. Stratified analyses were thoughtfully
executed, thereby rendering a meticulous evaluation of the prognostic value of the CRPS within distinct
subgroups delineated by these crucial clinical characteristics. Univariate and multivariate Cox regression
analyses were conducted to evaluate independent prognostic factors. A nomogram was developed based
on the identified independent prognostic factors, providing a visual tool for predicting patient survival
based on individual characteristics. The accuracy of the nomogram was assessed using calibration
curves, the C-index, ROC analysis, and decision curve analysis (DCA)[27].

2.6 Drug sensitivity analysis

Drug sensitivity analysis was conducted using several commonly used drugs for gastric cancer
treatment, including 5-fluorouracil, gemcitabine, cisplatin, doxorubicin, etoposide, docetaxel, and
paclitaxel[28]. The half-maximal inhibitory concentration (IC50) of these drugs was used to evaluate
therapy responses in the high- and low-risk groups. Additionally, the Tumor Immune Dysfunction and
Exclusion (TIDE) algorithm was employed to predict the efficacy of immune checkpoint blockade
therapy[29,30].

2.7 Single-cell RNA sequence analysis
We applied the Seurat R package (version 4.0.1) for single-cell RNA sequence analysis[31]. One Seurat object consisting of 3 primary gastric tumors was created. Genes that were expressed in fewer than three cells and cells displaying either fewer than 200 or over 5000 expressed features were both filtered out. The PercentageFeatureSet function was applied in the calculation of the proportion of mitochondrial genes within each cell. This metric facilitated the identification and exclusion of cells manifesting mitochondrial proportions exceeding 20%, a common indicator of low-quality or dying cells that are prone to mitochondrial contamination. Following these stringent filtration criteria, our dataset comprised 23949 features spanning 12014 cells, constituting a robust foundation for subsequent analysis. A total of 2000 distinct features were thereby extracted, the variation of which was evidently pronounced across the diverse cells. To achieve gene expression normalization, the ScaleData function within Seurat was judiciously invoked. Furthermore, linear dimensional reduction was judiciously pursued by invoking the RunPCA function, a technique that served as a pivotal stepping stone toward downstream analyses. Strategic determination of the optimal number of principal components (PCs) was achieved through the application of the JackStrawPlot and ElbowPlot functions, effectively facilitating an informed selection of PCs for subsequent analyses. Cells were clustered with the FindClusters function. Finally, we obtained 9 cell clusters and identified cell types with cell markers.

2.8 Cell culture and patient sample collection

The normal gastric epithelial cell line GES-1 and human gastric cancer cell lines AGS, HGC-27 and MKN45 were purchased from the American Type Culture Collection (ATCC, Manassas, VA, USA). Cells were cultured in PRIM1640 medium (Gibco, Thermo Fisher Scientific, Inc., Waltham, MA, USA) supplemented with 10% fetal bovine serum (FBS, HyClone). Cells were routinely cultured in a humidified atmosphere containing 5% CO2 at 37 ºC. A total of 20 fresh tumor and paired adjacent normal tissues from patients with GC were collected at the First Affiliated Hospital of Nanjing Medical University. All patients provided written informed consent, and this study was approved by the ethics committee of Nanjing Medical University.

2.9 qRT–PCR

Total RNA was extracted by TRIzol reagent (Invitrogen, USA) and was then reverse transcribed into cDNA by the PrimeScript RT Master Mix Kit (Takara, Japan). The Universal SYBR Green Master Kit (Roche, Germany) was used for RT–qPCR, and the process was conducted by means of a 7500 Real-Time PCR System (Applied Biosystems, USA). The 2−ΔΔCT method was used to calculate the relative expression levels of samples, in which GAPDH was used as the internal reference of mRNAs. The primers are listed in Supplementary Table S2.

2.10 Statistical analysis

RStudio (version: 2022.12.0+353, version of R: 4.2.2) was used for all statistical analyses. Comparative analysis between groups was conducted using Student’s t test, Wilcoxon test, or one-way
3. Results

3.1 Expression and interactions of 13 CRGs in gastric cancer

We initiated our study by evaluating the expression profiles of the 13 CRGs in gastric cancer. As shown in Fig. 2A, ATP7B, DLAT, and SLC31A1 were upregulated, while ATP7A, DBT, DLD, DLST, FDX1, LIAS, LIPT1, and PDHB were downregulated. Additionally, we performed Pearson correlation analysis to uncover correlations among the CRGs (Fig. 2B), with FDX1 and DLAT showing the strongest correlation (R = 0.49). To further elucidate the interactions between these CRGs, we constructed a protein–protein interaction (PPI) network that retained all 13 CRGs, demonstrating intricate regulatory relationships with a robust confidence score (0.962) (Fig. 2C). Gene Ontology (GO) and Kyoto Encyclopedia of Genes and Genomes (KEGG) analysis revealed that the CRGs were primarily associated with mitochondrial processes, regulation of the tricarboxylic acid cycle, pyruvate metabolism, and copper ion homeostasis, consistent with previous findings (Fig. 2D-G).

3.2 Identification of cuproptosis-associated molecular subtypes in gastric cancer

Next, we assessed the prognostic value of the 13 CRGs through univariate Cox regression analysis. Remarkably, four genes (DLAT, FDX1, PDHA1, and SLC31A1) emerged as significantly linked to GC patient prognosis (Fig. 3A). Employing consensus clustering analysis based on all 13 CRGs, we discovered two cuproptosis-associated molecular subtypes (C1 and C2) with strong clustering stability at k = 2, as evidenced by high intragroup correlations and low intergroup correlations (Fig. 3B-E). C1 exhibited a more favorable overall survival (OS) than C2 (Fig. 3F-G). Further examination of gene expression patterns and clinical characteristics in the two cuproptosis-associated subtypes within the GSE66229 dataset revealed that all four CRGs associated with favorable prognostic effects were downregulated in C2 (Fig. 3H). The same conclusion was validated in GSE15459 (Figure S1). These results successfully identified two cuproptosis-associated molecular subtypes in gastric cancer.

3.3 Construction of a cuproptosis-related prognostic signature

To provide a practical tool for predicting GC patient prognosis, we constructed the CRPS. Utilizing the "limma" R package, a comprehensive exploration into the landscape of differentially expressed genes (DEGs) between the two cuproptosis-associated subtypes was initiated (Fig. 4A). Subsequently, univariate Cox regression analysis ultimately unearthed 80 cuproptosis-related DEGs bearing significant variance analysis, with p < 0.05 considered statistically significant. The significance levels were denoted as *p < 0.05, **p < 0.01, ***p < 0.001, ****p < 0.0001, and "ns" for nonsignificant.
associations with GC prognosis within the training set (Supplementary Table S3). A pivotal advancement was achieved through the application of the LASSO regression algorithm, revealing six genes (CRYAB, SYNM, TAC1, GPA33, SFRP4, and PPP1R1B) that are intrinsically predictive of overall survival (OS). This selection was meticulously guided by the optimal lambda value and the minimal partial likelihood of deviance (Fig. 4B-C). Pooling these six genes, the CRPS came to life, embracing the following risk score formulation: 

\[
\text{Risk Score} = (0.099349105 \times CRYAB) + (0.020473329 \times SYNM) + (0.041363499 \times TAC1) + (-0.025029339 \times GPA33) + (0.001439227 \times SFRP4) + (-0.017473611 \times PPP1R1B).
\]

Patient stratification into high- and low-risk groups ensued, with the high-risk cohort exhibiting a trend toward earlier mortality and inferior outcomes compared to the low-risk counterpart (Fig. 4D). The CRPS demonstrated good predictive accuracy, with areas under the curve (AUCs) for the 5-year ROC of 0.699, 0.706, and 0.638 in the training, GSE66229, and GSE15459 sets, respectively (Fig. 4E). The distribution plot of the risk score alongside distinct expression levels of model genes within the two risk groups was meticulously showcased (Fig. 4F). This comprehensive synthesis of analysis underscored the robustness and potential clinical applicability of the CRPS.

3.4 Validation of the cuproptosis-related prognostic signature

Our prognostic model underwent rigorous validation within the expansive TCGA-STAD database. In the TCGA global cohort, patients were categorized into high-risk (n = 170) and low-risk (n = 170) groups. Similar to the training set, the low-risk group showed a more favorable prognosis than the high-risk group (p = 0.028) (Fig. 5A). However, the difference in OS was less pronounced compared to the training set. Considering the biological differences of GC in different races and that the training set was an Asian cohort, we further validated the prognostic model in the TCGA-Asian and TCGA-non-Asian cohorts. In TCGA-Asian, survival curves echoed the training set, with a significant distinction (p = 0.03) (Fig. 5C). However, in the TCGA-non-Asian cohort, the survival trend was similar but lacked statistical significance (p = 0.14) (Fig. 5E). Precision analysis through ROC confirmed the model's prowess, with remarkable AUCs in TCGA-Asian (AUC = 0.684), specifically for 1-year, 3-year, and 5-year survival (Fig. 5D). However, the model's performance was relatively limited in the TCGA global and TCGA-non-Asian cohorts (AUC = 0.547 and 0.523, respectively) (Fig. 5B and 5F). Collectively, this validation reinforced the model's robustness within specific racial contexts, underscoring its potential clinical value.

3.5 Evaluation of the clinical significance of CRPS and stratified analysis

Delving into the clinical implications of the CRPS-derived risk score was pivotal. An alluvial diagram artfully illustrated the intricate relationships among cuproptosis-associated subtypes, risk groups, and outcomes. C2 was linked to the high-risk group and subsequent mortality, while C1 was associated with the low-risk group and improved survival (Fig. 6A). Distinct risk scores further underscored this stratification, with C2 exhibiting higher risk scores than C1 (Fig. 6B). Analyzing correlations, Lauren's
classification of diffuse type and pathologic stages III and IV displayed alignment with high-risk scores. However, sex exhibited no substantial correlation with risk scores (Fig. 6C-D, 6L-M, 6F-G, 6I-J). Kaplan–Meier survival analysis, dissected by clinicopathological characteristics, offered illuminating insights. High-risk populations linked to age, sex, Lauren's classification, and pathologic stage exhibited significantly lower overall survival rates (Fig. 6E, 6H, 6K, 6N). This confluence of findings solidified the accuracy, independence, and wide applicability of our prognostic signature, affirming its potential as a powerful clinical tool.

3.6 Establishment and validation of the nomogram

The independence of prognostic factors was deduced via univariate and multivariate Cox regression analyses. Pathologic stage and risk score were validated as standalone prognostic determinants, with age showing potential significance alongside other clinical factors (Fig. 7A-B). ROC analysis highlighted the risk score's AUC of 0.699 for 5-year survival, although it was slightly lower than the pathologic stage's predictive capacity (Fig. 7C). Precision and convenience coalesced in our nomogram, seamlessly integrating age, pathologic stage, and risk score (Fig. 7D). Its accuracy was authenticated by calibration plots (Fig. 7E), with an elevated C-index of 0.747 (95% CI, 0.719 to 0.775), surpassing that of competing models (Supplementary Table S4). Ultimate validation was reached through decision curve analysis (DCA), where the nomogram emerged as a triumphant, asserting superior predictive power and accuracy (Fig. 7F).

3.7 Different drug responses in the two risk groups

Subsequently, we observed that the high-risk cohort exhibited enrichment in tumor-associated pathways such as mTOR signaling, MAPK signaling, and TGF-BETA signaling, as unveiled by gene set enrichment analysis (GSEA). In contrast, the low-risk cohort displayed enriched drug metabolism, TCA cycle, and select repair pathways (Fig. 8A, Figure S1B). Furthermore, susceptibility assessment of anticancer drugs revealed that patients with low risk exhibited smaller IC50 values for most chemotherapeutic agents, including 5-fluorouracil, gemcitabine, paclitaxel, doxorubicin, and etoposide. Conversely, patients with high risk had notably lower IC50 values for cisplatin (Fig. 8B). Interestingly, the copper ionophore elesclomol showed potential benefit for patients with high risk. These findings imply a correlation between CRGs and medication susceptibility. Additionally, we focused on GSE66229, which provided additional information about mismatch repair (MMR) and microsatellites. As shown in Fig. 8C-D, patients with high risk had a larger proportion of pMMR and microsatellite-stable (MSS), indicating that they may receive less benefit from immunotherapy. Tumor Immune Dysfunction and Exclusion (TIDE) analysis validated that patients with high risk had a lower proportion of responders to immunotherapy (Fig. 8E). The lower response rate to immune checkpoint inhibitors may be attributed to an increase in cancer-associated fibroblasts (CAFs), promoting immune escape and carcinoma progression (Fig. 8F).
3.8 Experimental verification of hub gene expression and single-cell sequencing analysis

Translating our findings into the realm of experimentation, we embarked on verifying hub gene expression in GC samples. Utilizing qRT–PCR on 20 GC and paired adjacent normal tissue pairs, we unearthed notable differences. The expression levels of CRYAB, SYNM, TAC1, and SFRP4 were markedly higher in GC tissues, while GPA33 and PPP1R1B displayed lower expression levels (Fig. 9A). Extending our investigation, the expression of these six hub genes was scrutinized in the normal gastric epithelial cell line GES-1 and gastric cancer cell lines. The observations indicated significantly lower expression of GPA33 and PPP1R1B in GES-1 cells, while other genes exhibited no significant divergence between normal and malignant cell lines (Fig. 9B). Then, we analyzed the distribution of those six hub genes in different types of cells based on single-cell sequencing. We successfully identified 9 cell clusters, including epithelial cells (1720, markers: EPCAM, CLDN4), T cells (4766, markers: CD3D, CD3E), B cells (846, markers: CD79A, MS4A1), neutrophils (2601, markers: FCGR3B), macrophages (1004, markers: CD163), fibroblasts (474, markers: ACTA2, COL1A2), NK cells (281, markers: KLRC1, GNLY), endothelial cells (244, markers: RAMP2, VWF), and mast cells (78, markers: TPSAB1, TPSB2) (Fig. 9C-D). CRYAB and SFRP4 were primarily expressed in fibroblasts, SYNM was mainly expressed in fibroblasts and endothelial cells, and GPA33 and PPP1R1B were expressed in malignant epithelial tissues. TAC1 was weakly expressed (Fig. 9E). These results were consistent with the TIDE analysis and explained the different expression levels between tissues and cell lines.

4. Discussion

Gastric cancer (GC) persists as a formidable global health concern, marred by elevated mortality rates stemming from late detection and drug resistance[32]. Traditional survival prediction methods in GC have relied on pathological information and serum tumor markers, but the evolving era of precision medicine demands more accurate and personalized approaches[33, 34]. Additionally, GC showcases biological and epidemiological nuances in Asian versus non-Asian populations. This dichotomy underscores the urgency of crafting race-specific prognostic models to improve patient care and curtail mortality rates.

In our study, we utilized the GEO datasets GSE66229 and GSE15459 to select six cuproptosis-related genes (CRYAB, SYNM, TAC1, GPA33, SFRP4, and PPP1R1B) and established a prognostic model. The model identified CRYAB, SYNM, TAC1, and SFRP4 as high-risk factors in GC, while GPA33 and PPP1R1B were associated with a longer overall survival (OS). The model demonstrated good predictive performance with AUCs of 0.666, 0.694, and 0.699 for 1-year, 3-year, and 5-year survival, respectively. This risk score-based model effectively categorized patients into low-risk and high-risk groups, aiding precision treatment decisions and guiding the use of postoperative adjuvant chemotherapy, particularly cisplatin-based regimens for high-risk patients. Furthermore, the model's independence from other clinical data highlights its usefulness in predicting patient outcomes, enabling early detection of recurrent cases and the development of tailored surveillance strategies for high-risk groups.
The validation of the prognostic model using the TCGA-STAD global cohort confirmed its accuracy in predicting OS for Asian GC patients based on GEO datasets from Korea and Singapore. Nonetheless, the model's efficacy encountered constraints when applied to non-Asian populations. This phenomenon aligns with prior findings of similar ethnic variations in the United States, where Japanese immigrants to Hawaii exhibited higher GC incidence than local residents\[35\]. Such interethnic disparities underscore a pivotal role of ethnicity in shaping the trajectory of GC development and progression.

While the TNM staging system has been valuable for risk assessment in GC, its ability to fully explain variability in prognosis among patients with the same stage is limited. The nomogram developed in our study, integrating genetic information with clinical data, demonstrated a 5.3% improvement in prediction accuracy compared to relying solely on pathologic staging (C-index: 0.75 vs. 0.71). Moreover, the six-gene expression assessment in the nomogram is simple and cost-effective compared to costly somatic mutation sequencing.

TIDE analysis and single-cell RNA sequencing have illuminated a potential dimension in the immunotherapy response among high-risk patients, attributed to cancer-associated fibroblasts (CAFs). In this intricate narrative, CRYAB and SFRP4 emerge as key protagonists. CAFs have a potent influence on diverse facets of tumor biology, including collagen deposition and immunosuppression, rendering them subjects of therapeutic exploration\[36\]. CRYAB, a notable member of the small heat shock protein family, occupies a prominent place. It deftly safeguards cells against unfavorable conditions by modulating processes integral to survival and stress recovery, such as protein degradation, cytoskeletal stabilization, and apoptosis\[37\]. Its contribution to gastric cancer cell migration and invasion through EMT, orchestrated by the NF-κB signaling pathway, has been previously documented\[38\]. Turning our gaze to SFRP4, an antagonist of Wnt ligands, we delve into its role as a glycoprotein. Operating as a counterforce to the canonical Wnt signaling pathway, which fuels cell proliferation and thwarts apoptosis, the significance of SFRP4 becomes clear\[39\]. Intriguingly, heightened intratumoral SFRP4 expression correlates with Wnt pathway activation, fostering tumor progression and predicting adverse survival outcomes in gastric cancer patients\[40\]. However, the synergy between CAFs and these genes requires further exploration, charting uncharted territory in this intricate web of interactions.

Despite these promising findings, our study has limitations. The lack of external validation in the TCGA database due to differences in the absolute value of the risk score warrants future research with larger sample sizes to enhance the reliability and generalizability of our prediction model.

5. Conclusions

Our research on CRGs revealed significant regulatory effects on clinical traits and prognosis in gastric cancer. We developed a six-gene-based prognostic model and a nomogram, showing reliable predictive ability for overall survival in Asian patients. However, the nomogram's predictive power was limited in the non-Asian population. Nonetheless, our nomogram improves the stratification of Asian gastric cancer
patients, aiding more accurate treatment decisions. Further validation and broader patient cohorts are needed to enhance the model's reliability and applicability for precision medicine in gastric cancer.

**Abbreviations**

GC  Gastric Cancer  
CRG  Cuproptosis-related Gene  
TNM  Tumor Node Metastasis  
TCA  Tricarboxylic Acid  
TIME  Tumor immune microenvironment  
TCGA  The Cancer Genome Atlas  
GSEA  Gene Set Enrichment Analysis  
OS  Overall Survival  
DEG  Differentially Expressed Gene  
LASSO  least absolute shrinkage and selection operator  
CRPS  Cuproptosis-related Prognostic Signature  
ROC  Receiver Operating Characteristic  
DCA  Decision Curve Analysis  
IC50  half-maximal inhibitory concentration  
GO  Gene Ontology  
KEGG  Kyoto Encyclopedia of Genes and Genomes  
AUC  Area Under the Curve  
MMR  Mismatch Repair  
MSS  Microsatellite-stable  
TIDE  Tumor Immune Dysfunction and Exclusion  
CAF  Cancer-associated Fibroblast
Declarations

Ethics approval and consent to participate

Tissues from patients with GC were collected at the First Affiliated Hospital of Nanjing Medical University. All patients provided written informed consent, and this study was approved by the ethics committee of Nanjing Medical University.

Consent for publication

The authors declare that they agree to submit the article for publication.

Availability of data and materials

The datasets used and analyzed during the current study are available online, and codes are available from the corresponding authors upon reasonable request.

Competing interests

The authors declare no potential conflicts of interest.

Fundings

This work was partially supported by the National Natural Science Foundation of China (82072708, 82273038); Special Foundation for National Science and Technology Basic Research Program of China (2019FY101104); Jiangsu Province Capability Improvement Project through Science, Technology and Education (Jiangsu Provincial Medical Key Discipline, ZDKK202222); the Program for Development of Innovative Research Team in the First Affiliated Hospital of NJMU, the Priority Academic Program Development of Jiangsu Higher Education Institutions (JX10231801); Jiangsu Key Lab of Cancer Biomarkers, Prevention and Treatment, Collaborative Innovation Center for Cancer Personalized Medicine, Nanjing Medical University.

Authors’ contributions

WZ W and HX H conceived and designed the research. PH X and ZT C prepared and processed the dataset. HX H and CM Z analyzed and interpreted the data. MP Y and JH W prepared figures and tables. JL L and FL L wrote the manuscript. ZK X supervised the study and acquired funding support. All authors contributed to the article and approved the final version of the manuscript.

Acknowledgments

We would like to thank the Core Facility of the First Affiliated Hospital of Nanjing Medical University for its help in the detection of experimental samples. In addition, we would like to acknowledge the GEO (GSE66229, GSE15459, and GSE163558) network and TCGA for providing data.
References


**Figures**
Figure 1

Schema of the current study.
Figure 2

Expression variations, correlation analysis, and functional analyses of CRGs in GC. (A) The expression of CRGs between GC and normal samples. (B) Correlations between the expression of CRGs. (C) PPI network of the CRG-encoded proteins. (D-G) The GO and KEGG analysis of CRGs in the DAVID database, D: KEGG, E: biological process, F: cellular component, G: molecular function.
Figure 3

Characterization of two cuproptosis-associated molecular subtypes based on CRGs in GC. (A) Univariate Cox regression analysis of CRGs. (B) Consensus matrix when k = 2. Consensus CDF (C), deltaarea (D), tracking plot (E), and tSNE plots (F) for validation of the clustering results. (G) Kaplan–Meier OS curves for GC patients between C1 and C2. (H) Unsupervised clustering of CRGs in the GSE66229 cohort. The
cuproptosis-associated molecular subtypes, stage, age, TNM stage, Lauren's classification and survival status were used as patient annotations.

Figure 4

Construction of the CRPS. (A) Identification of 160 DEGs between the two cuproptosis-associated molecular subtypes. (B) LASSO regression analysis of 80 prognostic DEGs. (C) Cross-validation method.
to select optimal genes. (D) Kaplan–Meier analyses of OS between the two risk groups in the training, GSE66229, and GSE15459 sets. (E) The 1-, 3- and 5-year ROC curves of the CRPS in the training, GSE66229, and GSE15459 sets. (F) Ranked dot, scatter plots and heatmap of the model gene expression in the training, GSE66229, and GSE15459 sets.
Validation of the CRPS. (A) Kaplan–Meier analyses and (B) the 1-, 3- and 5-year ROC curves of the CRPS in TCGA-STAD-global. (C) Kaplan–Meier analyses and (D) the 1-, 3- and 5-year ROC curves of the CRPS in TCGA-Asian. (E) Kaplan–Meier analyses and (F) the 1-, 3- and 5-year ROC curves of the CRPS in the whole TCGA-non-Asian cohort.
Evaluation of the clinical significance of CRPS. (A) Alluvial diagram of cuproptosis-associated molecular subtypes, risk groups and outcomes. (B) Comparisons of the risk score in the two cuproptosis-associated molecular subtypes. Relationships of the risk score and Lauren’s classification (C–D), age (F–G), sex (I–J) and p-stage (L–M). K–M survival analysis for clinicopathological subgroups divided by Lauren’s classification (E), age (H), sex (K) and p-stage (N).

Figure 7
Establishment and validation of the nomogram. Univariate Cox regression analysis (A) and multivariate Cox regression analysis (B) of clinical characteristics and risk score. (C) ROC curve of clinical characteristics and risk score. (D) The nomogram for predicting GC patients’ 1-, 3-, and 5-year OS probability. (E) The calibration plots of the nomogram in the training, GSE66229 and GSE15459 sets. (F) Decision curve analysis (DCA) curves for the nomogram.

Figure 8
Different drug responses in the two risk groups. (A) Significantly enriched pathways in the two groups according to GSEA. (B) Estimated drug sensitivity in patients with high and low risk scores. (C) Comparisons of MMR in the two risk groups. (D) Comparisons of microsatellites in the two risk groups. (E) Comparisons of the proportions of nonresponders and responders to immunotherapy between the two risk groups. (F) Comparisons of CAFs, MDSCs and M2 TAMs in the two risk groups.

Figure 9
Experimental validation and single-cell sequencing analysis. (A) qRT–PCR of the six hub genes in clinical samples. (B) qRT–PCR of the six hub genes in several gastric cells. (C) Identification of 9 cell clusters by tSNE based on single-cell sequencing data. (D) Markers of 9 cell clusters. (E) The expression and distribution of 6 hub genes at the single-cell level.

**Supplementary Files**

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

- FigureS1.tif
- FigureS2.tif
- TableS14.xlsx