Research Square

Preprints are preliminary reports that have not undergone peer review.
They should not be considered conclusive, used to inform clinical practice,
or referenced by the media as validated information.

Identification and Validation of Cysteine-Related Prognostic Signatures in Gastric Cancer

Xi Yang[1]
The First Affiliated Hospital of Guangxi Medical University
Mengda Wei[1]
The First Affiliated Hospital of Guangxi Medical University
Yanqing Huang[1]
The First Affiliated Hospital of Guangxi Medical University
Xiaoying Yang[1]
The First Affiliated Hospital of Guangxi Medical University
Qiuqi Ou[1]
The First Affiliated Hospital of Guangxi Medical University
Lei Tian
15277042675@163.com

The First Affiliated Hospital of Guangxi Medical University

Research Article

Keywords: Gastric carcinoma, Cysteine, Prognosis, Nomogram, Reverse transcription quantitative polymerase chain reaction

Posted Date: July 1st, 2026

DOI: https://doi.org/10.21203/rs.3.rs-9857128/v1

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

Additional Declarations: No competing interests reported.

Identification and Validation of Cysteine-Related Prognostic Signatures in Gastric Cancer

Authors: Xi Yang[1], Mengda Wei[1], Yanqing Huang[1], Xiaoying Yang[1], Qiuqi Ou[1], Lei Tian[1]*

Affiliations: [1]Department of Gastrointestine, Gastric Gland Surgery, The First Affiliated Hospital of Guangxi Medical University, Nanning, Guangxi, 530021, China.

Corresponding Author:
Lei Tian， MD
Address: Department of Gastric Gland Surgery, The First Affiliated Hospital of Guangxi Medical University, Nanning, Guangxi, 530021, China
Tel:15277042675
Email: 15277042675@163.com

Abstract

Background: The bioavailability of cysteine (Cys) within cancer cells can affect systemic metabolism and may contribute to the development of therapeutic resistance. Nevertheless, its prognostic significance and mechanisms of action in gastric carcinoma (GC) remain ambiguous. Consequently, this study aimed to identify key genes associated with cysteine in GC, to construct a robust prognostic model, and to clarify its potential molecular regulatory mechanisms.

Methods: Transcriptomic and clinical data pertaining to GC were retrieved from publicly accessible databases. The cys-differentially expressed genes (DEGs) were isolated by intersection of DEGs between GC and normal samples with cys-related key modular genes obtained from weighted gene co-expression network analysis (WGCNA). Subsequently, the proportional hazards (PH) assumption test, univariate and multivariate Cox regression analysis, Least absolute selection and shrinkage operator (LASSO) algorithm, and stepwise regression analysis were employed to acquired key genes. Upon constructing a risk model with these genes, further analysis yielded

independent prognostic factors, and a nomogram was developed based on these variables. Lastly, we engaged in enrichment analysis and the construction of a regulatory network and verified the expression of key genes to elucidate potential mechanisms in GC.

Results: The overlap of 2,599 DEGs with 3,895 key module genes yielded a total of 743 Cys-DEGs. Then, SERPINE1, MATN3, BAALC, SLC7A3, and RNASE1 were recognized as key genes for constructing a risk model with highly proficient in assessing the risk of GC patients through a series of analyses. Next, we constructed a nomogram that incorporated risk score, age and M-stage, demonstrating high predictive accuracy for survival outcomes in patients with GC. Groups at distinct levels of risk were found to display pathway enrichment in cancer development, progression, and cellular behavior regulation. Additionally, a lncRNA-miRNA-mRNA network was constructed featuring 79 lncRNA-miRNA interaction pairs. EZH2, GATAD1, KLF16, ZNF384, ARID4B, MLX, SIN3A, and RCOR2, were identified as pivotal transcription factors (TFs), with each having the ability to govern at least two of these critical genes concurrently. Conclusively, the expression validation results for BAALC, SLC7A3, and RNASE1, obtained through reverse transcription quantitative polymerase chain reaction (RT-qPCR), showed trends consistent with those observed in public datasets. Specifically, BAALC, SLC7A3, and RNASE1 all demonstrated downregulation in GC samples.

Conclusion: We identified five key genes associated with Cys in GC and integrated them into a risk prognostic model. These genes serve as signatures, providing a novel perspective for the prognosis of GC.

Keyword: Gastric carcinoma, Cysteine, Prognosis, Nomogram, Reverse transcription quantitative polymerase chain reaction

Abbreviations

GC	Gastric cancer
MATN3	Matrilin-3
TNM	Tumor node metastasis
ROC	Receiver operating characteristic

GEO	Gene Expression Omnibus
GO	Gene Ontology
BP	Biological pathways
AUC	Areas under the ROC curves
KEGG	Kyoto Encyclopedia of Genes and Genomes
OS	Overall survival
TCGA	The Cancer Genome Atlas
qPCR	quantitative Real-Time PCR
R-Nase	ribonuclease
LASSO	Least Absolute Shrinkage and Selection Operator
ROC	Receiver Operating Characteristic
DCA	Decision Curve Analysis
RNA	Ribonucleic Acid
m-RNA	Messenger RNA

1. Introduction

Gastric carcinoma (GC), predominantly characterized by stomach adenocarcinoma (STAD), accounts for more than 95% of cases and ranks as the fifth most diagnosed and the third most lethal cancer globally[1].Early-stage GC often presents with nonspecific symptoms akin to those of other upper gastrointestinal disorders, resulting in delayed diagnosis.[2] Consequently, by the time of presentation, many patients exhibit advanced-stage symptoms, including pain, fatigue, anorexia, and weight loss. Surgical resection is currently the main clinical treatment of GC, due to the late detection of GC, most of them belong to the advanced stage of GC, the surgical effect is not ideal, and the five-year survival rate after surgery is still less than 30%[3-4]. GC mainly originates from gastric epithelial tissues, including gastric mucosa and epithelial surface glands. Recognized as a heterogeneous malignancy, it exhibits heightened susceptibility to a range of environmental and genetic factors[5-6]. Infection by Helicobacter pylori (Hp) and exposure to its virulence

factors (CagA, VacA), smoking, and various benign gastric pathologies (e.g., gastric polyps, gastric ulcers, and chronic atrophic gastritis) have been identified as contributing etiological factors[7]. Therefore, GC research is increasingly focused on elucidating the molecular pathogenesis of GC. This includes identifying inactivating mutations in proto-oncogenes and tumor suppressor genes that drive its onset and progression. There is also a concerted effort to identify effective biomarkers and discover novel molecular targets with high specificity for therapeutic intervention.

Cysteine (Cys) is a sulfur-containing protein amino acid that has a free thiol group, which may confer functional site-specific properties to proteins containing this highly conserved residue. As a key component of carbon and sulfur metabolism, Cys plays a central role in cellular metabolism. Cys catabolism proceeds via two predominant pathways: one is through enzymatic decomposition to produce hydrogen sulfide (H2S) and organic intermediates will serve as carbon sources, and the second is through oxidative metabolism of Cys dioxygenase (CDO) [8]. Elevated plasma levels of Cys and homocysteine (Hcy) can function as pro-oxidants, inciting oxidative stress. In addition, Cys serves as a crucial source of energy and biomass, with its bioavailability in cancer cells affecting metabolic functioning and the emergence of treatment resistance [9]. Cys has been reported as a biomarker for cardiovascular disease, obesity, and insulin resistance due to its breakdown pathway that produces reactive oxygen species (ROS), including hydrogen peroxide [10]. As a precursor of many molecular compounds, Cys contributes to the survival and proliferation of cancer cells [11]. Notably, cys metabolism in pancreatic ductal adenocarcinoma has been observed to forestall ferroptosis in murine models, a process not yet widely reported in human malignancies, marking it as an avenue of substantial clinical and prognostic interest, particularly in GC. Current studies reveal that cysteine metabolism promotes cancer progression through multiple mechanisms [12]. demonstrated that glutamate-cysteine ligase catalytic (GCLC) and modifier

(GCLM) subunits are significantly overexpressed in gastric adenocarcinoma (77% and 80% positive staining, respectively), suggesting these enzymes as novel immunotargets. Cysteine availability regulates ferroptosis resistance via system Xc−/SLC7A11-mediated cystine import, which sustains glutathione synthesis and GPX4 activity to prevent lipid peroxidation[13]. Additionally, the KEAP1-NRF2 pathway plays a critical role, where cysteine modification of KEAP1 cysteine residues disrupts NRF2 degradation, activating antioxidant responses that promote cancer cell survival and chemoresistance[14-15].

Capitalizing on clinical insights and transcriptome sequencing data from public GC databases, the present study harnessed bioinformatics approaches to illuminate the prognostic capabilities and molecular regulatory processes of Cys in GC, offering novel insights for prognostication and clinical intervention

2. Materials and methods

2.1 Data source

Expression matrix and clinical information for STAD from The Cancer Genome Atlas (TCGA) were downloaded from the University of California Santa Cruz (UCSC) Xena database (https://xenabrowser.net/datapages/) (accessed on January 9, 2024). The dataset comprised 375 tumor tissue samples from patients diagnosed with GC and 32 adjacent normal gastric tissue samples, serving as the training cohort. An additional GC dataset. GSE84437, which included survival information, was sourced from the Gene Expression Omnibus (GEO) database (https://www.ncbi.nlm.nih.gov/gds). This dataset contained GC tissue samples from 433 patients based on the GPL6947 platform, as the validation cohort . Moreover, a set of 20 Cys-related genes (Cys-RGs) was identified and extracted from the "CYSTEINE-METABOLIC" gene set within the Molecular Signatures Database (MSigDB) (https://www.gsea-msigdb.org/gsea/msigdb).

2.2 Identification of differentially expressed genes (DEGs)

DEGs were identified from GC and normal samples [$|\log_2$Fold Change (FC)| >

1.5 and adj.*P* < 0.05] by DESeq2 package (v 1.38.0) [16] within training cohort. To visualize the DEGs, volcano plot and heatmap were generated through ggplot2 package (v 3.4.1) [17] and pheatmap package (v 1.0.12) [18], respectively.

2.3 Identification of key module genes associated with Cys

In order to elucidate genes associated with Cys, we initially calculated Cys-RGs that exhibited differential expression between GC and normal samples in training cohort (*P* < 0.05). Within training cohort samples, the expression scores of these differentially expressed Cys-RGs were then analyzed via GSVA package (v 1.42.0) [19]. Patients were dichotomized into high- and low-score groups using the median GSVA score of differentially expressed Cys-related genes in the training cohort as the cutoff. The median was selected to avoid arbitrary thresholding and to ensure balanced group sizes. Survival differences were examined between both score groups through Kaplan-Meier (K-M) survival curves using survminer package (v 0.4.9) [20]. Subsequent analysis involved weighted gene co-expression network analysis (WGCNA), which was performed using WGCNA package (v 1.70-3) [21], to gain key module genes related to Cys. A cluster analysis was employed to evaluate the need for excluding outlier samples to improve the accuracy of further analyses. Next, we determined an optimal soft-thresholding power (β) to construct a scale-free co-expression network, aiming for a scale-free fit index (R^2) above 0.85 and a mean connectivity approaching 0, which resulted in a topology indicative of a scale-free network. Gene similarity was assessed, producing a gene dendrogram, and modules were detected utilizing the dynamic tree cut algorithm with a minimum module size of 300 genes. Finally, correlation coefficients were computed between the modules and grouped traits, and modules demonstrating an absolute correlation value greater than 0.3 with the Cys-RGs scores were selected as key modules. Genes within these key modules were designated as key module genes associated with Cys.

2.4 Recognition and enrichment analysis of Cys-DEGs

The Cys-DEGs were identified through the intersection of DEGs and key

module genes associated with Cys, employing ggvenn package (v 0.1.9) [22]. Following this, we performed Gene Ontology (GO) and Kyoto Encyclopedia of Genes and Genomes (KEGG) analysis on Cys-DEGs to further uncover potential biological functions and pathways, facilitated by ClusterProfiler package (v 4.7.13) [23] (adj.*P* < 0.05). The GO analysis encompassed three fundamental categories: biological processes (BP), cellular components (CC) and molecular functions (MF).

2.5 Construction and validation of risk model

In the first place, proportional hazards (PH) assumption test were conducted on Cys-DEGs using CoxPH function from survival package (v 3.5.3) [20] leveraging GC patient samples from training cohort that included survival information [the clinical endpoint being overall survival (OS)]. The genes that satisfied PH assumption test (*P* > 0.05) were subjected to univariate Cox regression analysis [hazard ratio (HR) ≠ 1, and *P* < 0.01] to screen out survival-related genes. In that case, we further recognized key genes through a successive application of least absolute selection and shrinkage operator (LASSO) algorithm [Glmnet package (v 4.1.4) [24], PH assumption test, multivariate Cox regression analysis, and stepwise regression analysis, sequentially.

Within training cohort, the risk score for each GC patients was calculated according to relative expression of key genes and their coefficients obtained from the stepwise regression model. The formula used was $risk\ score = \sum_{i=1}^{n}(coefi * Xi)$, where Xi was the relative expression of key genes i, coefi was the stepwise regression model coeffcient of key genes i. Thereafter, GC patients were stratified into high and low risk groups via median value of risk score from the training cohort. KM analysis was conducted to assess differences in OS between these distinct risk groups. Subsequently, the distribution of risk scores, survival status of GC patients, and the expression of key genes in both groups were illustrated. To evaluate the predictive accuracy of the risk model, receiver operating characteristic (ROC) curve was generated

using pROC package (v 1.18.0) [25], facilitating the computation of area under curve (AUC) at 1, 2 and 3 year. Additionally, risk model underwent validation through validation cohort. The same formula and cutoff strategy were then applied in the validation cohort.

2.6 Independent prognostic analysis and construction of nomogram

Primarily, within training cohort, risk score and clinical characteristics (age, grade, gender, stage, T-stage, N-stage, M-stage, race) of GC patients were integrated to performed PH assumption test, followed by univariate and multivariate Cox regression analyses to identify independent prognostic factors. Next, the rms package (v 6.5-0) [26] was employed to develop a nomogram for predicting survival rate of GC patients at 1, 2, and 3 years, based on the identified independent prognostic factors. The predictive accuracy of the nomogram for survival was assessed through calibration curves. Additionally, decision curve analysis (DCA) was applied to determine if the nomogram offered substantial clinical benefits over the use of independent prognostic factors alone.

2.7 Correlation of risk score with clinical characteristics

A comparative analysis was conducted to examine differences in risk score and clinical characteristics, and visualization was achieved via ggplot2 package. Afterwards, differences in survival among differential clinical subtypes within distinct risk groups were compared via survminer package. Ultimately, the expression levels of key genes across various clinical subtypes were analyzed and compared.

2.8 Gene set enrichment analysis (GSEA)

In an effort to elucidate the pathways associated with both risk groups, an initial analysis of differential gene expression was performed utilizing the DESeq2 package, with genes ranked by their $\log_2$FC. Subsequently to this, GSEA was conducted by ClusterProfiler package (adj.P < 0.05) via rank results. The reference gene set was c2.kegg.v7.4.symbols downloaded from MSigDB.

2.9 Construction of regulatory network

With a view to elucidate the regulatory mechanisms governing the expression of key genes, we first predict the miRNAs that might regulate these genes in both miRDB (http://www.mirdb.org/) and TargetScan (https://www.targetscan.org/vert_80/). Then we gained miRNAs consistently predicted by both resources. In addition, the corresponding lncRNAs were predicted separately in both starBase database (https://rnasysu.com/encori/) and miRNet database (https://www.mirnet.ca). Afterwards, we determined the overlap in lncRNA predictions from these databases and established a lncRNA-miRNA-mRNA network. Besides, to pinpoint transcription factors (TFs) implicated in the regulation of key genes, we retrieved data from NetworkAnalyst database (https://www.networkanalyst.ca/) for retrieval and established a TF-mRNA network. We visualized all aforementioned results using Cytoscape software (v 3.7.2) [27].

2.10 Expression validation of key genes

We initially assessed the expression patterns of key genes in GC and corresponding normal samples across training and validation cohorts, employing ggplot2 package. Subsequent validation of gene expression levels was conducted via reverse transcription-polymerase chain reaction (RT-qPCR). Concretely, tumour tissue samples were gained from five GC patients in First Affiliated Hospital of Guangxi Medical University, while matching paracancerous tissue samples served as controls. Following sample acquisition, RT-qPCR was employed for analysis. This study was conducted in accordance with the principles of the Declaration of Helsinki. Approval was granted by the Ethics Review Committee of the First Affiliated Hospital of Guangxi Medical University (Approval No.: 2026-E0064). For the validation process, we extracted total RNA from approximately 50 mg of each tissue sample utilizing TRIzol (Ambion, Austin, USA) as per the manufacturer's protocol. RNA concentration and purity were assessed using a NanoPhotometer N50 instrument. We then synthesized cDNA from the total

RNA using the SweScript First-Strand cDNA Synthesis Kit (Servicebio, Wuhan, China), following the recommended procedure. Quantitative PCR was conducted on a CFX Connect Real-Time PCR System (Bio-Rad, USA) using Universal Blue SYBR Green qPCR Master Mix (Servicebio, Wuhan, China). Each reaction was performed in a 10 μL volume containing 3 μL cDNA, 5 μL 2× SYBR Green Master Mix, 1 μL forward primer (10 μM), and 1 μL reverse primer (10 μM). The thermal cycling conditions were as follows: 95°C for 1 min, followed by 40 cycles of 95°C for 20 s, 55°C for 20 s, and 72°C for 30 s. Melt-curve analysis was performed to confirm amplification specificity. GAPDH was used as the internal control. Primers used for the PCR are listed in **Table S1**, with GAPDH serving as the internal reference gene. Gene expression levels were quantified employing $2^{-\Delta\Delta Ct}$ method [28].

2.11 Statistical Analysis

All analyses were executed in R software (v 4.2.2). Differences between groups were analyzed by Wilcoxon test. $P < 0.05$ was considered statistically significant.

3. Results

3.1 Recognition of 2,599 DEGs and 3,895 key module genes

Primarily, in training cohort, we screened out 2,599 DEGs, comprising 1,207 up-regulated and 1,392 down-regulated genes in GC and compared to normal samples through differential expression analysis (**Fig. 1A, 1B**). Thereafter, we conducted the WGCNA within training cohort to acquire key modular genes associated with associated with Cys. Specifically, our findings revealed that 18 Cys-RGs were expressed in training cohort, with 11 Cys-RGs exhibiting differential expression between GC and normal groups ($P < 0.05$) (**Fig. 1C**). Furthermore, high score of Cys-RGs correlated with a higher survival rate in comparison to low score of Cys-RGs (**Fig. 1D**).Cluster analysis results revealed the absence of outlier samples within training cohort (**Fig. 1E**). Subsequently, after optimizing soft threshold (β) to four (R^2 close to 0.85 and mean connectivity value close to 0), it was determined that the network

exhibited characteristics of a scale-free network distribution (**Fig. 1F**). The construction of a co-expression network yielded a total of seven modules, each containing a minimum of 300 genes per gene module. (**Fig. 1G**). Lastly, the key modules, MEgreen (cor = 0.346) and MEblue (cor = 0.304), were identified, housing a combined total of 3,895 Cys-related key module genes (**Fig. 1H**).

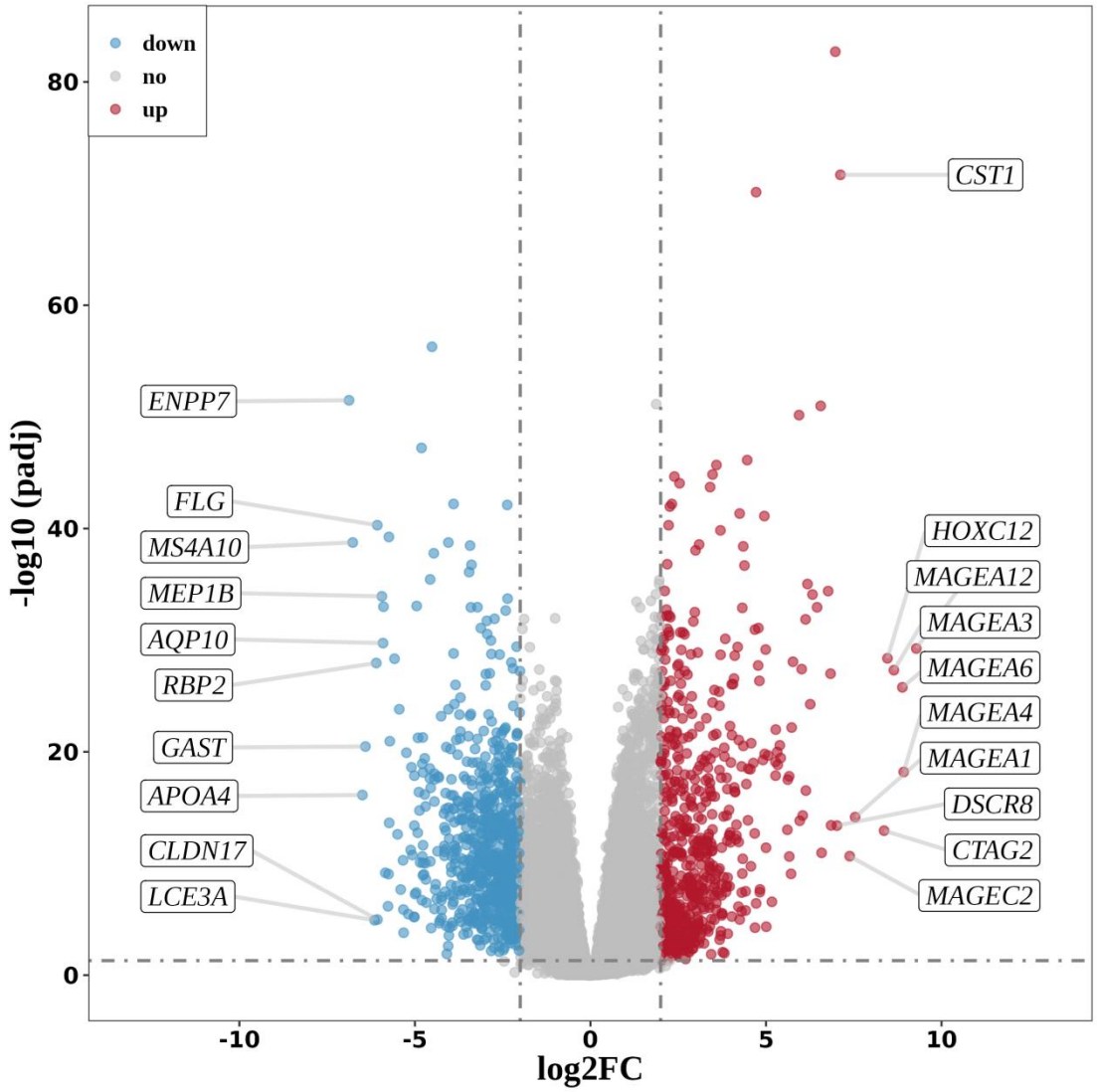


Fig. 1A

The top 10 most significantly up- and down-regulated genes are displayed with their gene names.

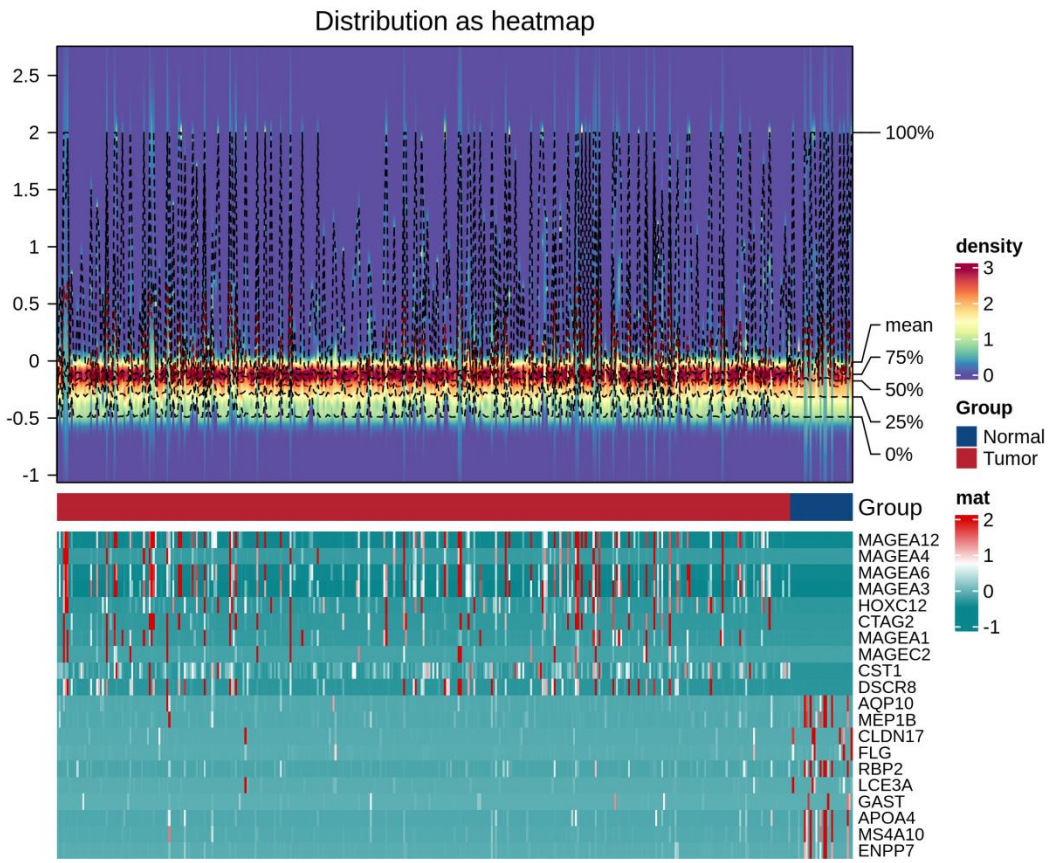


Fig. 1B

Gene expression heat map. The figure consists of two parts: the right panel displays the expression heatmap of the top 10 up- and down-regulated genes, with the x-axis representing samples and the y-axis representing genes. Red indicates high expression, whereas blue indicates low expression. The left panel shows the expression density heatmap of the same genes, displaying five quantile lines and the mean value.

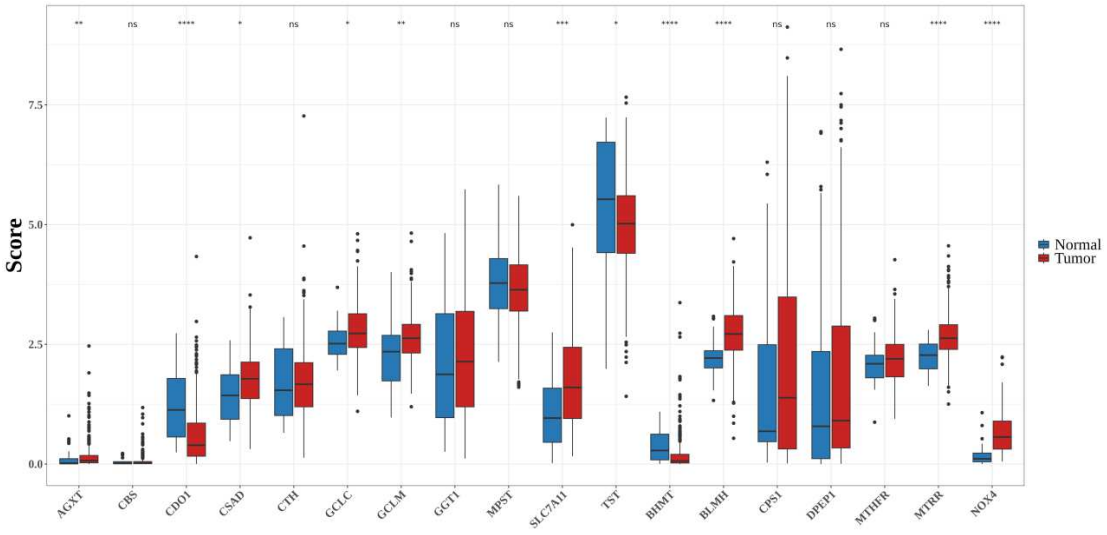


Fig. 1C

Differential analysis of Cys-related genes between disease and normal groups.

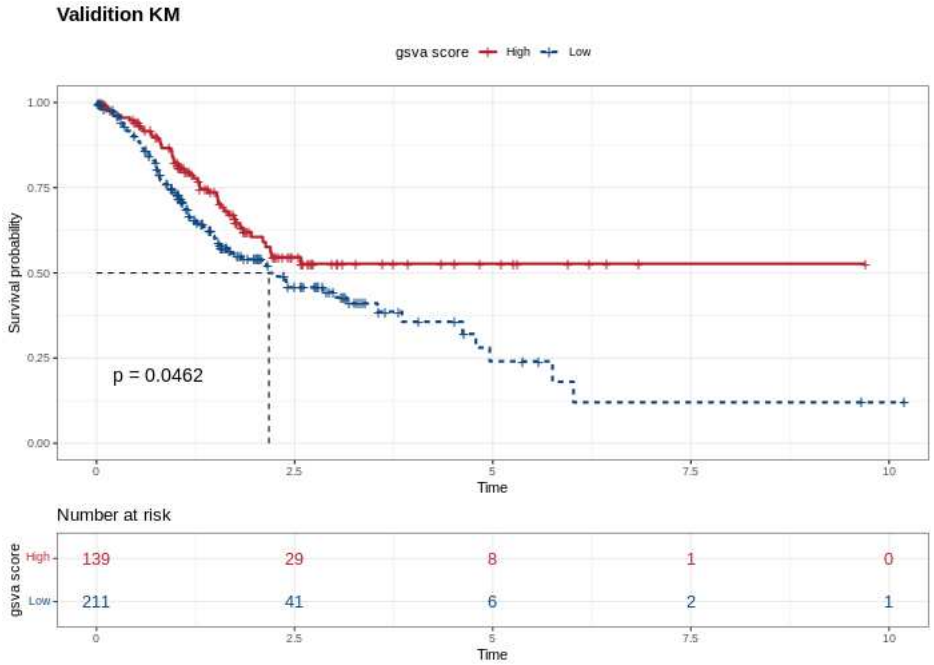


Fig. 1D

Survival analysis of high and low scoring groups of Cys-related differential genes

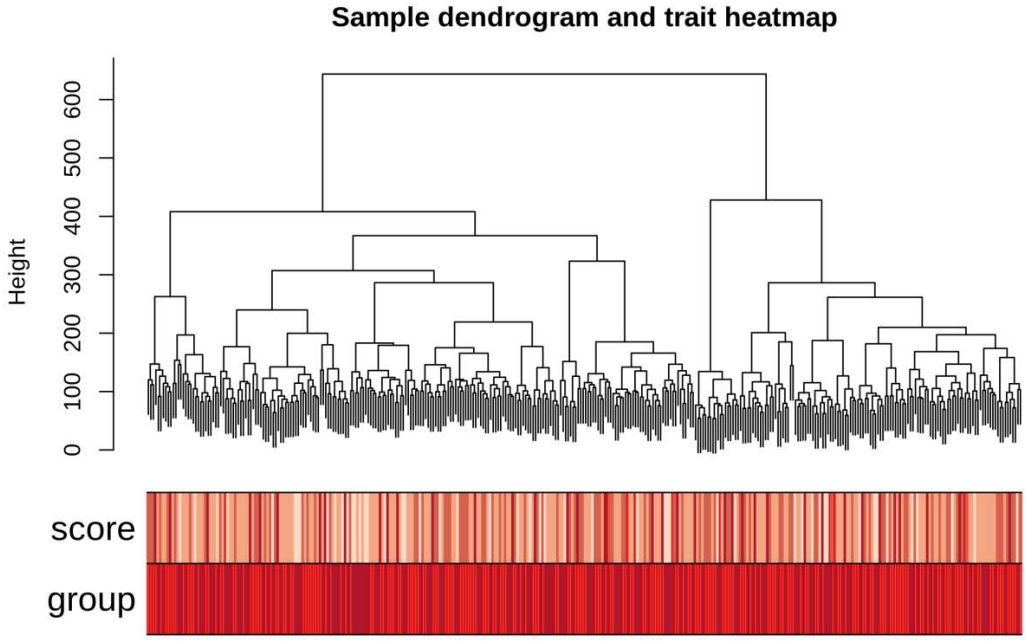


Fig. 1E

The branches represent samples, and the ordinate represents the height of hierarchical clustering. The scores and grouping of samples are shown below the graph

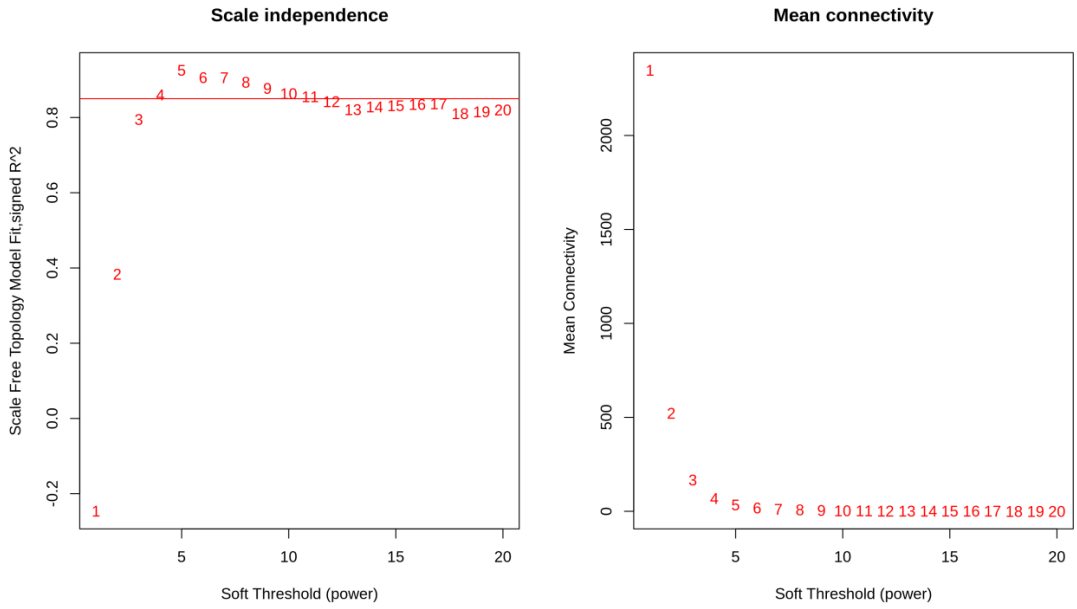


Fig. 1F

Soft threshold screening. The horizontal axis of the above graphs represents the weight parameter power value, and the vertical axis of the left graph has no

scale R^2. The higher the square of the correlation coefficient,the closer the network is to scale-free distribution.

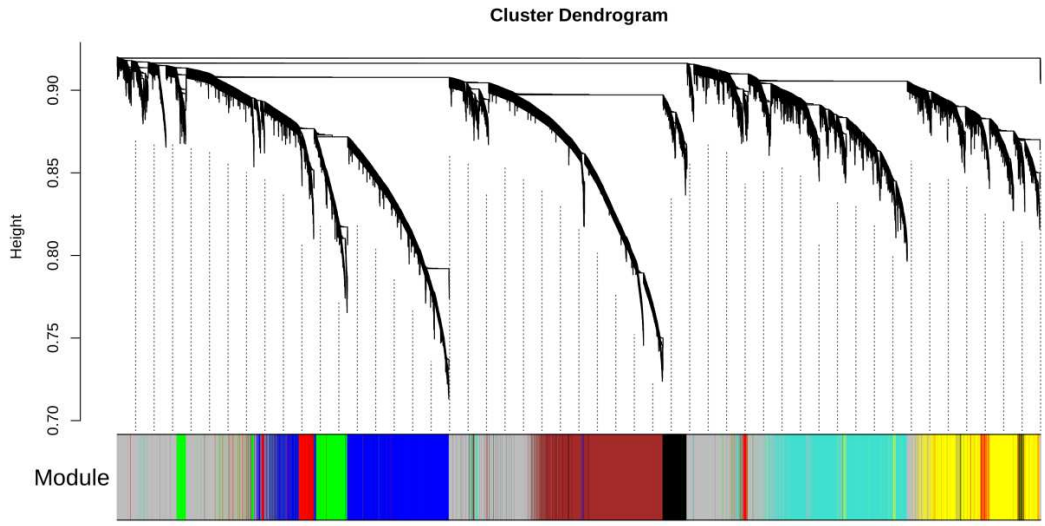


Fig. 1G

Note: The upper part is a hierarchical clustering tree diagram of genes, and the lower part is a gene module

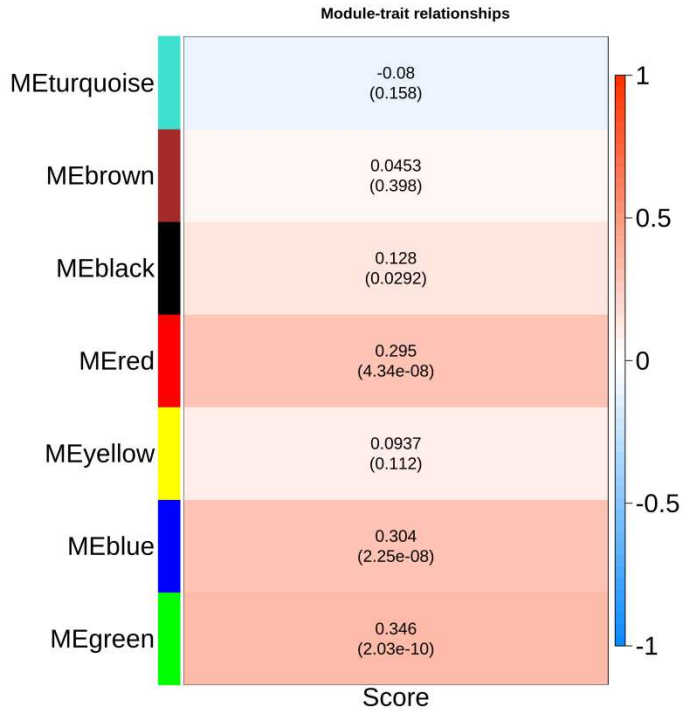


Fig. 1H

Module correlation heat map.

315 **3.2 A total of 743 Cys-DEGs were gained**

The overlap of 2,599 DEGs with 3,895 key module genes yielded a total of 743 Cys-DEGs (**Fig. 2A**). After that, these Cys-DEGs were subjected to GO and KEGG analysis. The GO analysis revealed 537 BP entries, including "muscle system process", "muscle contraction", and "muscle cell differentiation", and 108 CC entries were identified, like "collagen-containing extracellular matrix", "neuronal cell body", and "contractile fiber", as well as 53 MF entries were recognized, encompassing "actin binding", "receptor ligand activity", and "signaling receptor activator activity" (**Fig. 2B,2C,2D)**. The KEGG analysis demonstrated that Cys-DEGs were significantly enriched in 256 pathways such as “Neuroactive ligand-receptor interaction”, “Calcium signaling pathway”, “cAMP signaling pathway”, “Adrenergic signaling in cardiomyocytes”, “Vascular smooth muscle contraction” and other relevant pathways (**Fig. 2E)**. Overall, the Cys-DEGs exhibited a strong association with the processes governing muscle structure, function, and differentiation, and were implicated in critical biological pathways, including signal transduction, neuro-muscular interactions, and cardiovascular system functionality. The enrichment of these specific entries and pathways emphasized the potential involvement of Cys-DEGs in the regulation of muscular physiology and pathology. The data suggested that these genes might be pivotal in modulating muscle development, homeostasis, and disease etiologies.

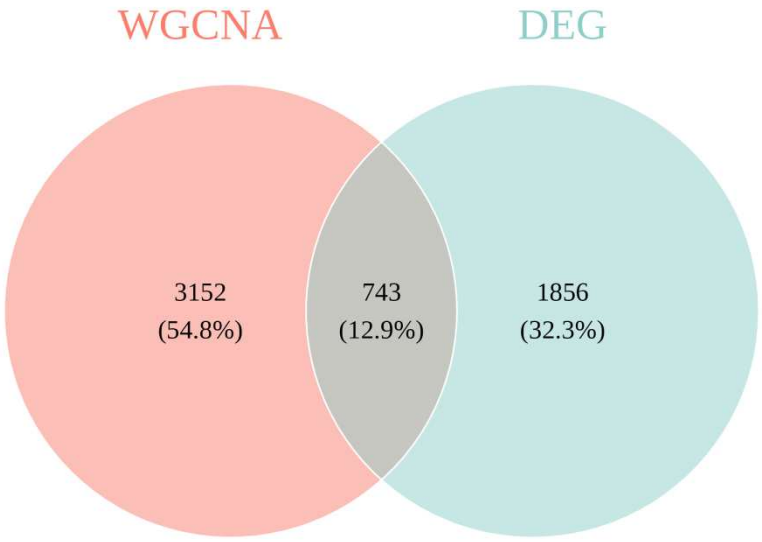

WGCNA
DEG
3152
(54.8%)
743
(12.9%)
1856
(32.3%)

Fig. 2A

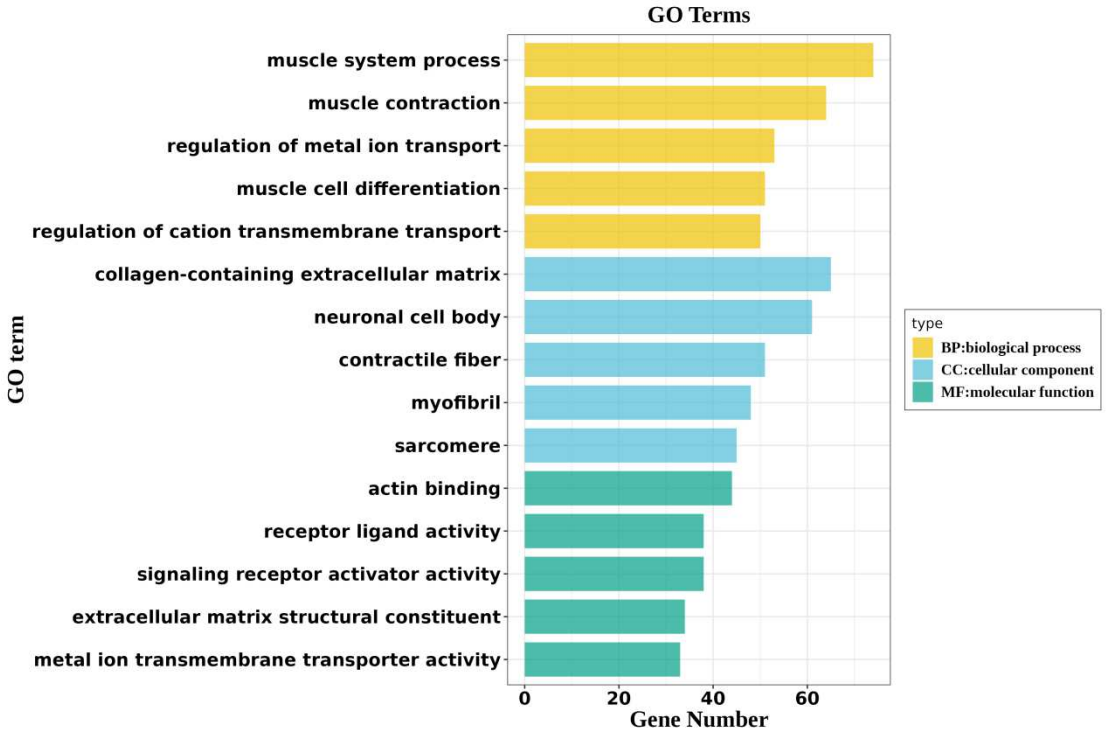

GO Terms
muscle system process
muscle contraction
regulation of metal ion transport
muscle cell differentiation
regulation of cation transmembrane transport
collagen-containing extracellular matrix
neuronal cell body
contractile fiber
myofibril
sarcomere
actin binding
receptor ligand activity
signaling receptor activator activity
extracellular matrix structural constituent
metal ion transmembrane transporter activity
GO term
0
20
40
60
Gene Number
type
BP:biological process
CC:cellular component
MF:molecular function

Fig. 2B

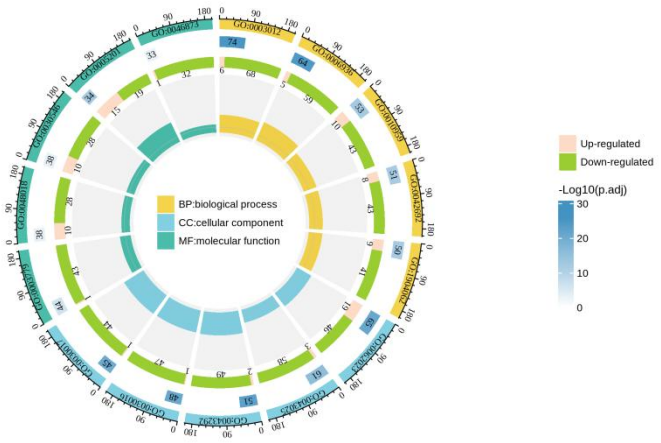


Fig. 2C

GO enrichment analysis of Cys-DEGs. Note: The first layer represents the IDs of GO functions; the second layer represents the depth of color indicates the significance, and the length and width and number represent the number of genes enriched in the function; the third layer represents the number of up- and down-regulated genes enriched in the function, and the color is used to distinguish between up- and down-regulation of genes; the color blocks in the innermost layer represent different functions, and the size represents the RichFactor of the pathway.

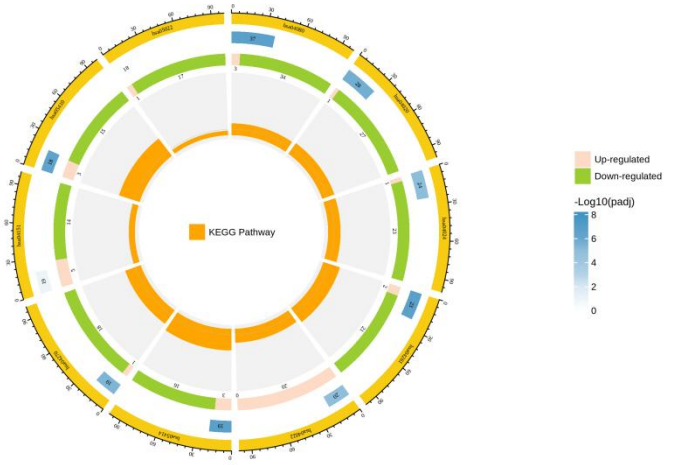


Fig. 2D

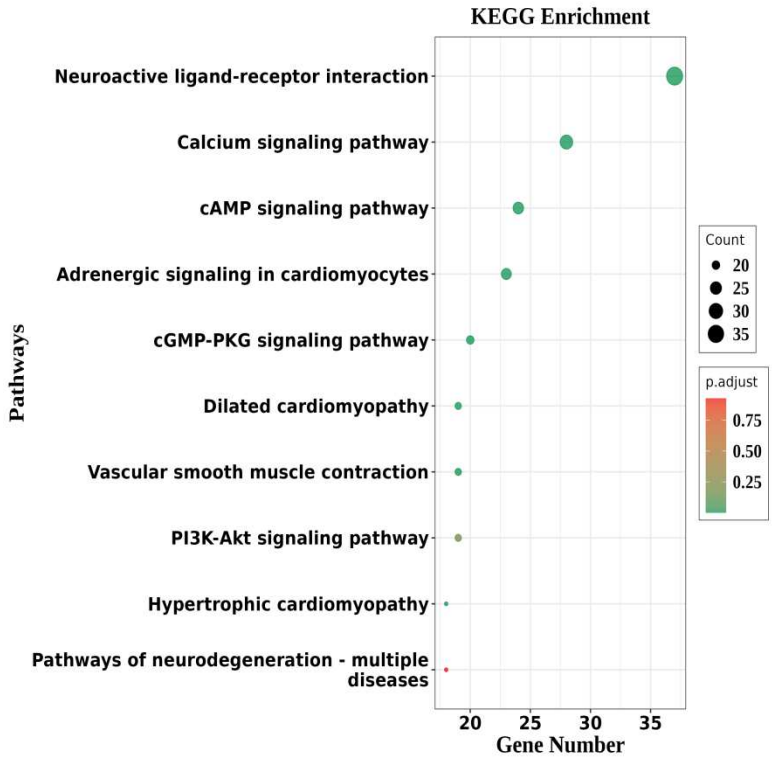


Fig. 2E

KEGG enrichment analysis of Cys-DEGs.

3.3 SERPINE1, MATN3, BAALC, GRP, SLC7A3, and RNASE1 were recognized as key genes

Using PH assumption test and univariate Cox regression analysis, we gained 44 survival-related genes with risk factors for GC characterized by HR greater than 1 (**Fig. 3A)**. Following that, ten genes were further screened out through LASSO algorithm (**Fig. 3B, 3C)**. These ten genes underwent further scrutiny through the PH assumption test and multivariate Cox regression analysis (**Fig. 3D)**. Of these, only two genes were found to be significantly associated with survival (P < 0.05), necessitating the application of stepwise regression analyses for further refinement and validation of the model. Through stepwise regression analysis, non-significant variables were eliminated, resulting in six key genes: SERPINE1, MATN3, BAALC, GRP, SLC7A3, and RNASE1, being identified for further analysis (**Fig. 3E)**.

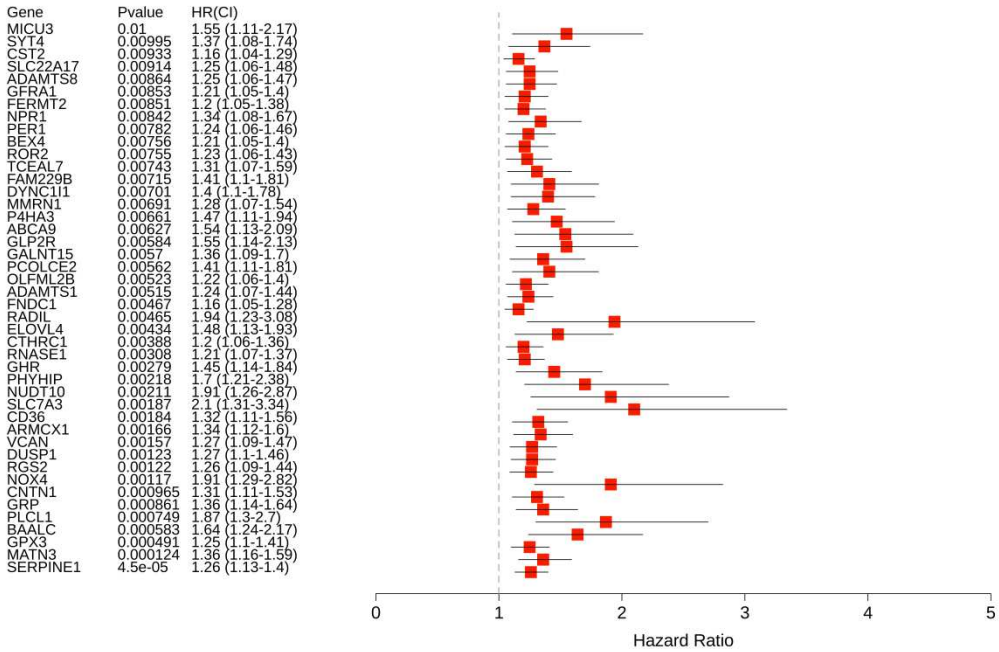


Fig. 3A

Forest plot of univariate Cox regression analysisunivariate Cox regression analysis of survival-related genes in the TCGA-STAD cohort; The red dots in the right figure indicate the risk ratios.>1 is a risk factor,and <l is a protective factor.

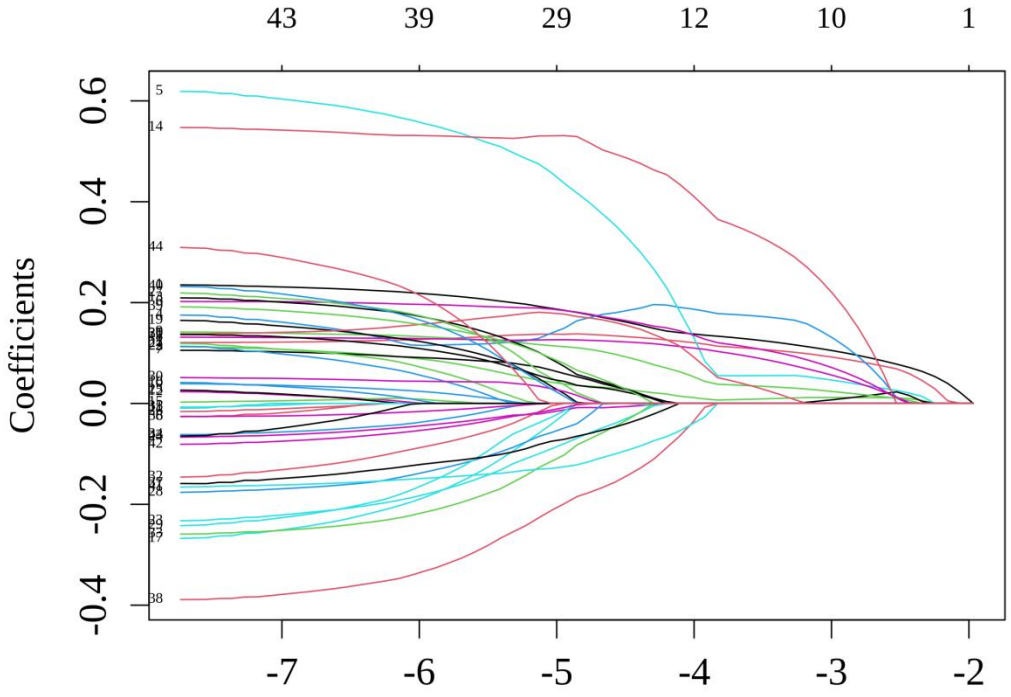


Fig. 3B

Ten cross validation of adjustment parameters in LASS0 analysis. The

horizontal axis represents the logarithm of lambdas, and the vertical axis represents the model error.

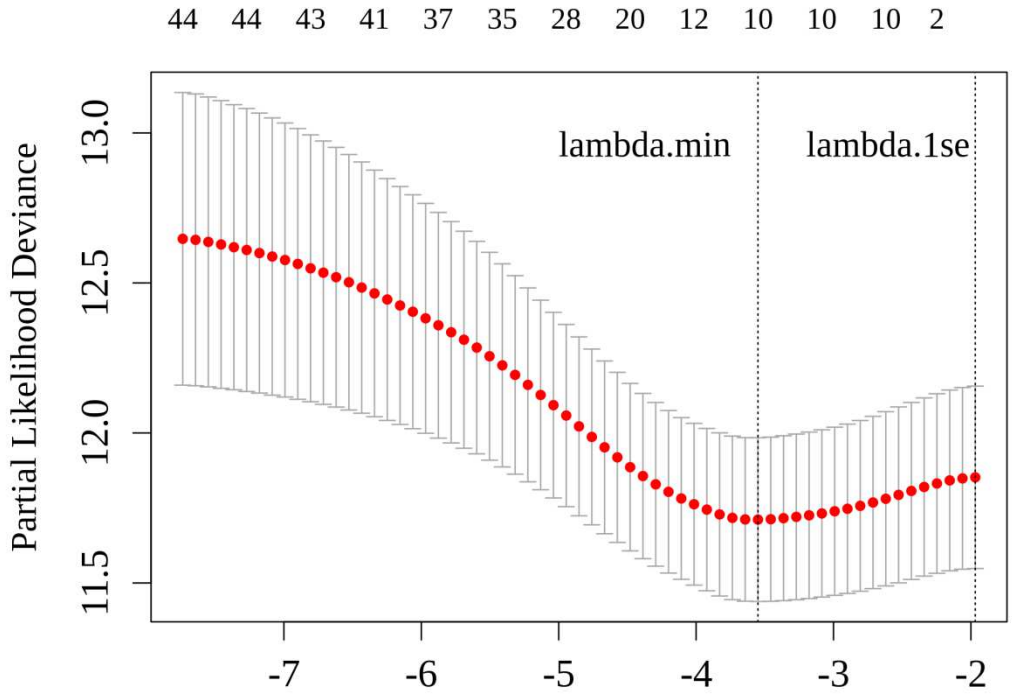


Fig. 3C

Elastic net regularization with partial likelihood deviance plot to determine optimal lambda;

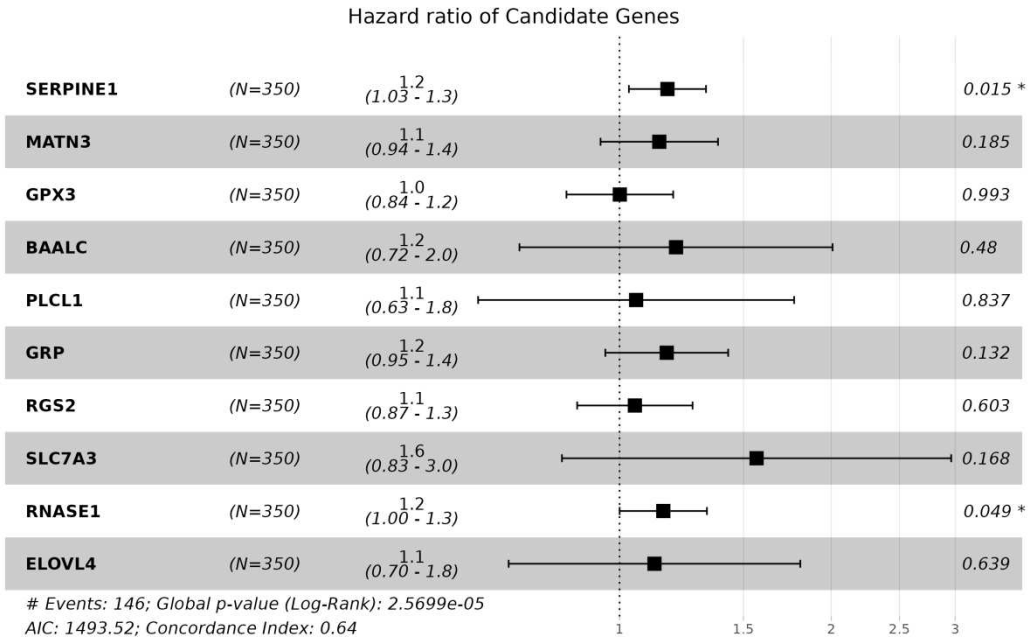


Fig. 3D

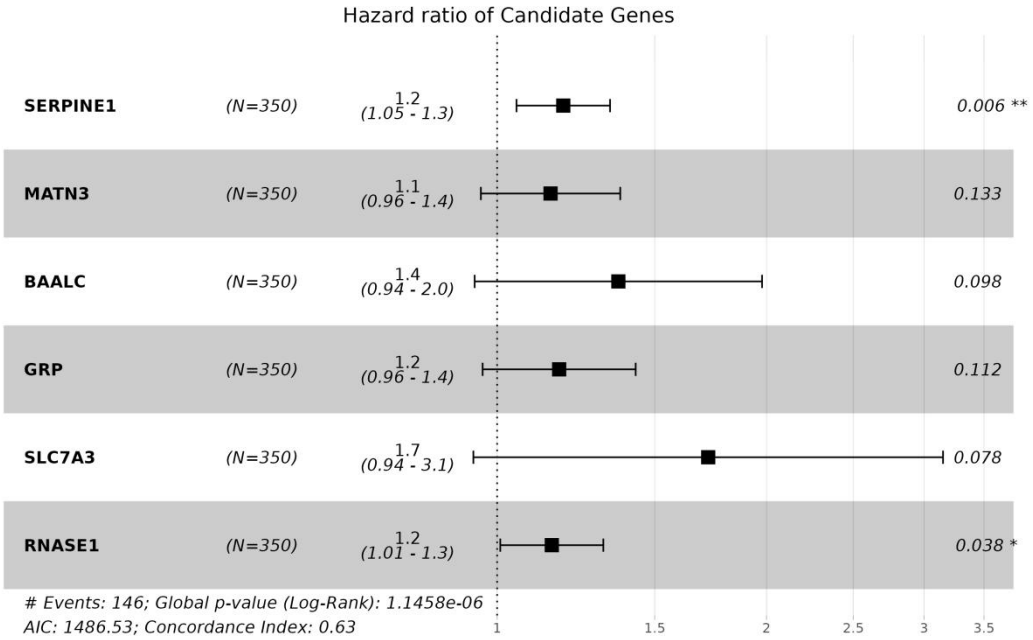


Fig. 3E

3.4 Risk model displayed strong proficiency in assessing the risk of GC patients

In this study, a risk model was developed according to key genes, with the risk score being derived from the following equation: RiskScore = SERPINE1 × (0.17016214860605) + MATN3 × (0.137437010604914) + BAALC × (0.311909330071195) + GRP × (0.159539249735826) + SLC7A3 × (0.543078157841349) + RNASE1 × (0.140643044299044). Within training cohort, GC patients were stratified into distinct risk groups (high risk individuals = 175 and low risk individuals = 175) via median risk score of 2.065561. The distribution of risk scores, survival status across distinct risk groups for GC patients were depicted in (**Fig. 4A, 4B**). As hypothesized patients classified as low risk exhibited longer survival compared to those classified as high risk (P < 0.0001) (**Fig. 4C**). The prognostic accuracy of the risk model was assessed through ROC analysis, yielding AUC of 0.65, 0.69 and 0.75 at 1, 2 and 3 years, respectively, indicating favorable predictive capabilities (**Fig. 4D**). We also presented the differential expression of key genes in the high- and low-risk stratifications among all disease samples, and the expression levels of all

genes were higher in the high-risk group (**Fig. 4E**). To ascertain the model's predictive reliability, external validation was conducted in a separate validation cohort. Importantly, results mirrored those from the training cohort, with AUC of 0.63, 0.61 and 0.60 for 1, 2 and 3 years (**Fig. 4F-J**). These outcomes affirmed the robustness of risk model in evaluating the prognostic risk of GC patients.

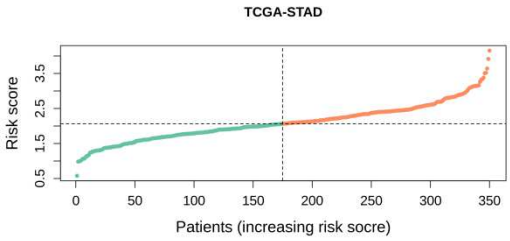


Fig. 4A

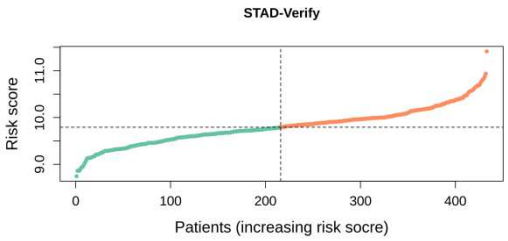


Fig. 4F

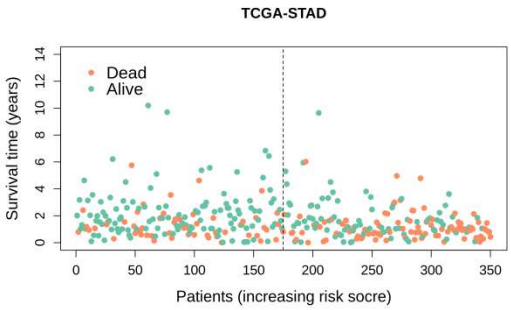


Fig. 4B

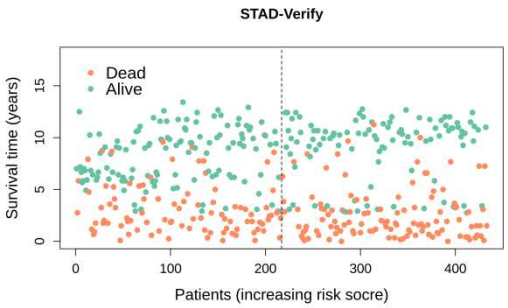


Fig. 4G

Fig. 4A, 4B

Fig. 4F, 4G

Risk score distribution in TCGA-STAD and validation cohort. (F) Survival time and status of patients in the TCGA-STAD and validation cohort;

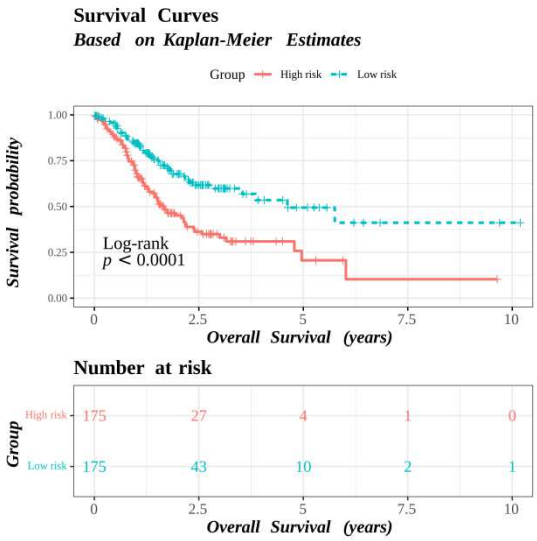


Fig. 4C

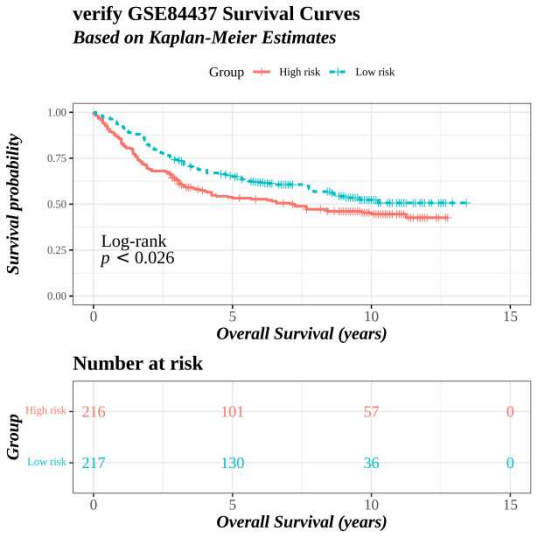


Fig. 4H

Note: The horizontal axis is the total survival time (days), and the vertical axis is the survival probability; red is the high-risk group, and blue is the low-risk group

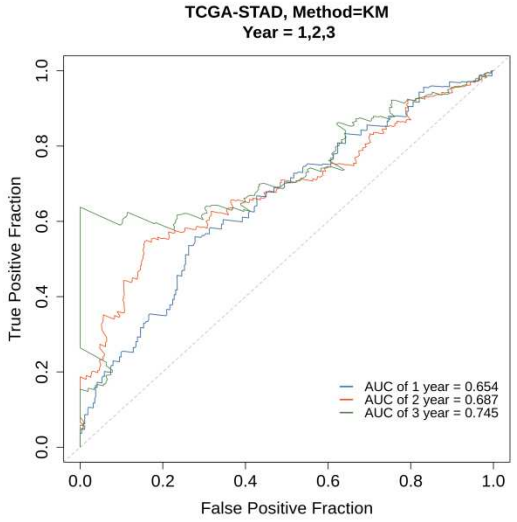


Fig. 4D

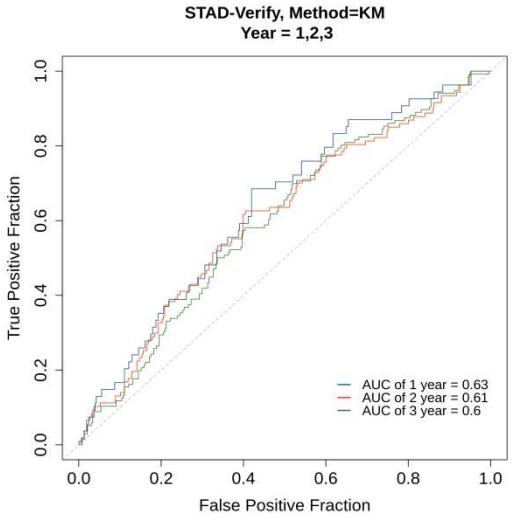
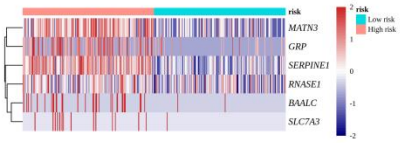


Fig. 4I

Note: The abscissa is the true positive rate, the ordinate is the false positive rate, and the area enclosed by the curve and the abscissa is called AUC. The closer the value of AUC is to 1, the better the diagnostic performance.



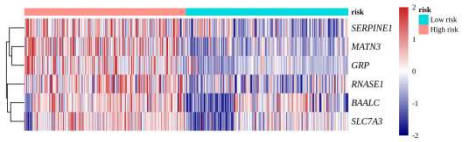


Fig. 4E Fig. 4J

Note: Row represents gene, column represents expression level, and the top is the grouping condition

3.5 Risk score, age and M-stage were recognized as independent prognosis factors

We further explored whether risk scores could serve as an independent predictor of prognosis in patients with GC. To this end, we integrated risk scores and clinical characteristics for analysis using PH assumption test, alongside univariate and multivariate Cox regression analysis in training cohort. Our findings revealed that only three factor: risk score, age and M-stage, emerged as independent prognostic indicators (**Fig. 5A, 5B**). Following that, we developed a nomogram to forecast survival rate of GC patients at 1, 2, and 3 years via risk score, age and M-stage (**Fig. 5C**). Calibration curve indicated that this nomogram achieved high predictive accuracy (**Fig. 5D**). DCA suggested that this nomogram provided significant clinical utility, surpassing that of the independent prognostic factors alone (**Fig. 5E**).

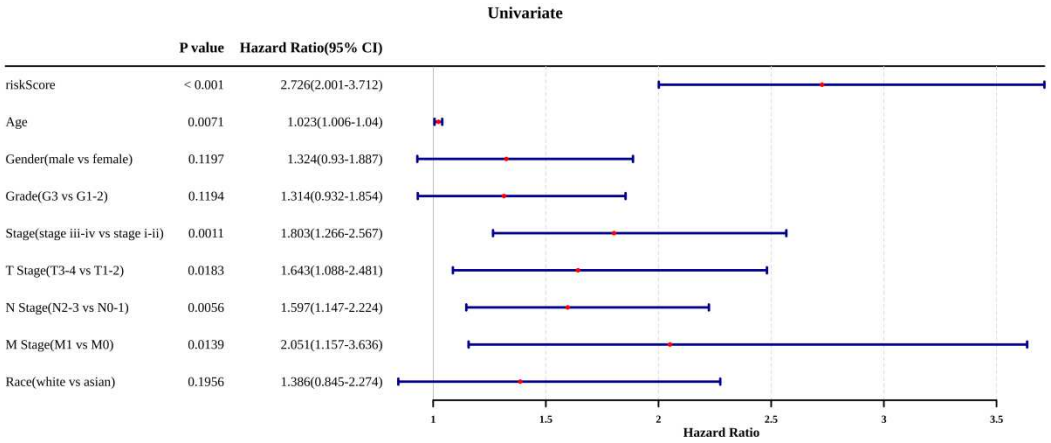


Fig. 5A

Results of forest map analysis using univariate cox regression

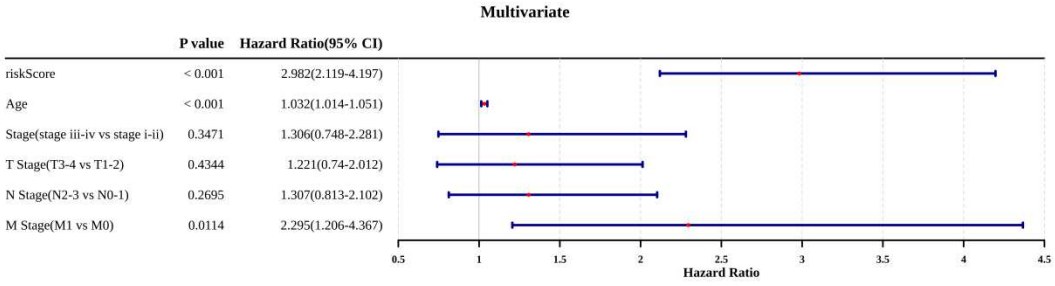
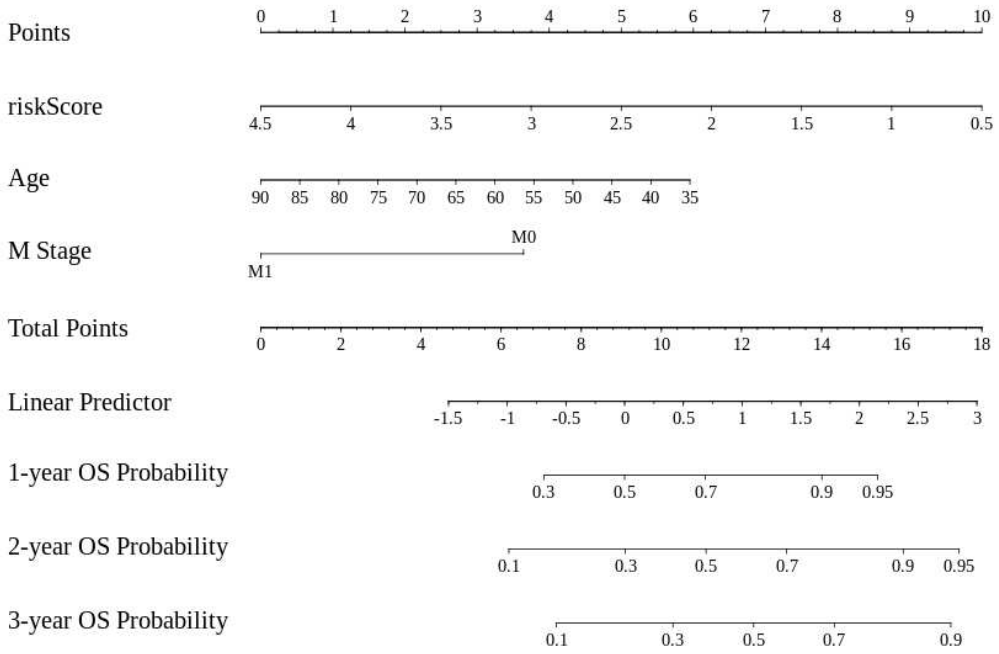


Fig. 5B

Display of forest map results from the multi-factor cox regression analysis



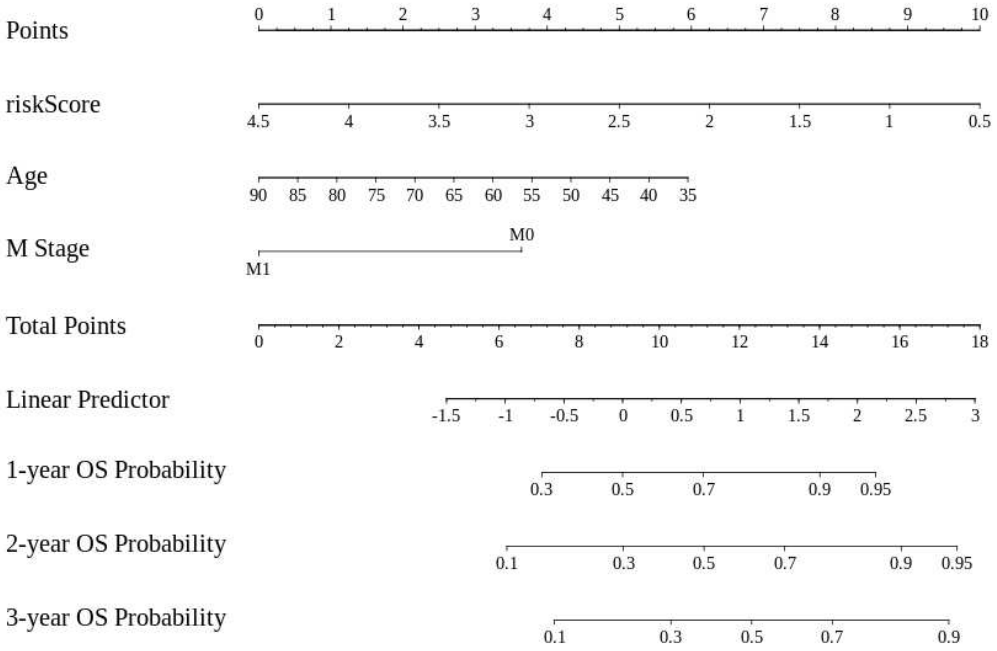


Fig. 5C

Development of a nomogram to predict 1-, 2-, and 3-year OS.

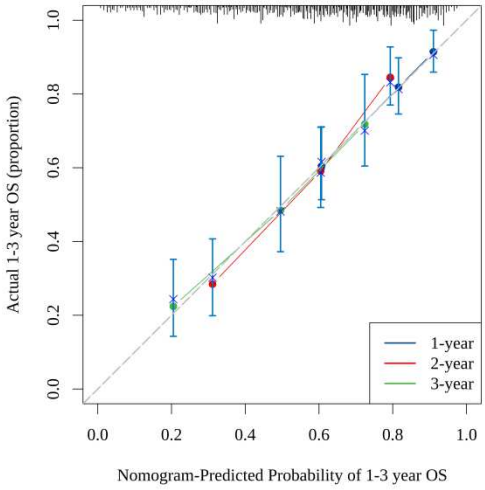


Fig. 5D

Note: The abscissa is the predicted event occurrence rate, and the ordinate is the observed actual event occurrence rate, both ranging from 0 to 1

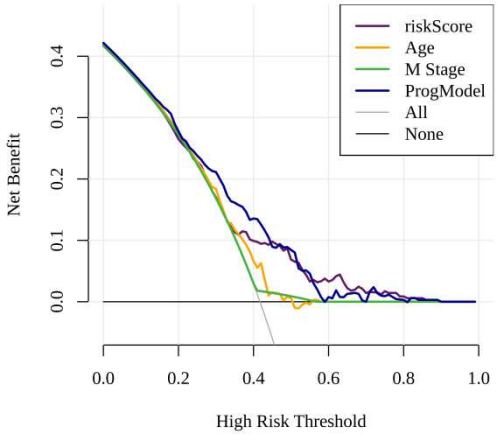


Fig. 5E

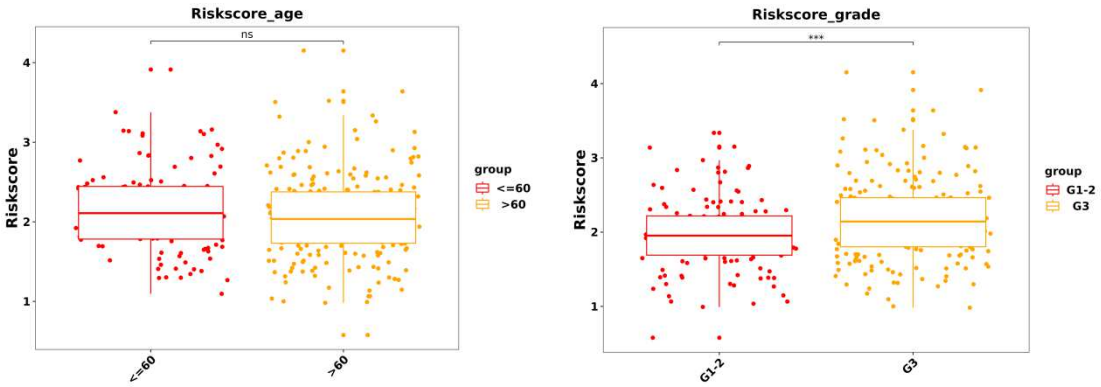
DCA decision curve

388

389 **3.6 Risk scores demonstrated strong correlation with the clinical**

characteristics of GC

We initially investigated the disparities in risk scores and clinical characteristics. Evidently, patients classified as Grade 3 (G3) and of white ethnicity exhibited higher risk scores compared to those identified as Grades 1 and 2 (G1-2) and of Asian ethnicity, respectively (**Fig. 6A**). This observation suggested that patients within the G3 category and those of white ethnicity might had a poorer prognosis for GC. Following this, we analyzed the survival differences across grade and race in distinct risk groups. It was apparent that within the G3 category, both white and Asian ethnic groups, patients classified in the low-risk category exhibited longer survival times compared to their high-risk counterparts (**Fig. 6B**). Moreover, our analysis revealed significant differential expression of key genes across various clinical characteristics. Specifically, SERPINE1 and RNASE1 exhibited significant variations in expression across different ages. SERPINE1's expression significantly varied with the grade of the disease, while MATN3's expression was markedly different across T-stages. GRP showed significant differential expression in relation to M-stage, and both BAALC and SLC7A3 exhibited significant variances in expression across different races (**Fig. 6C**).



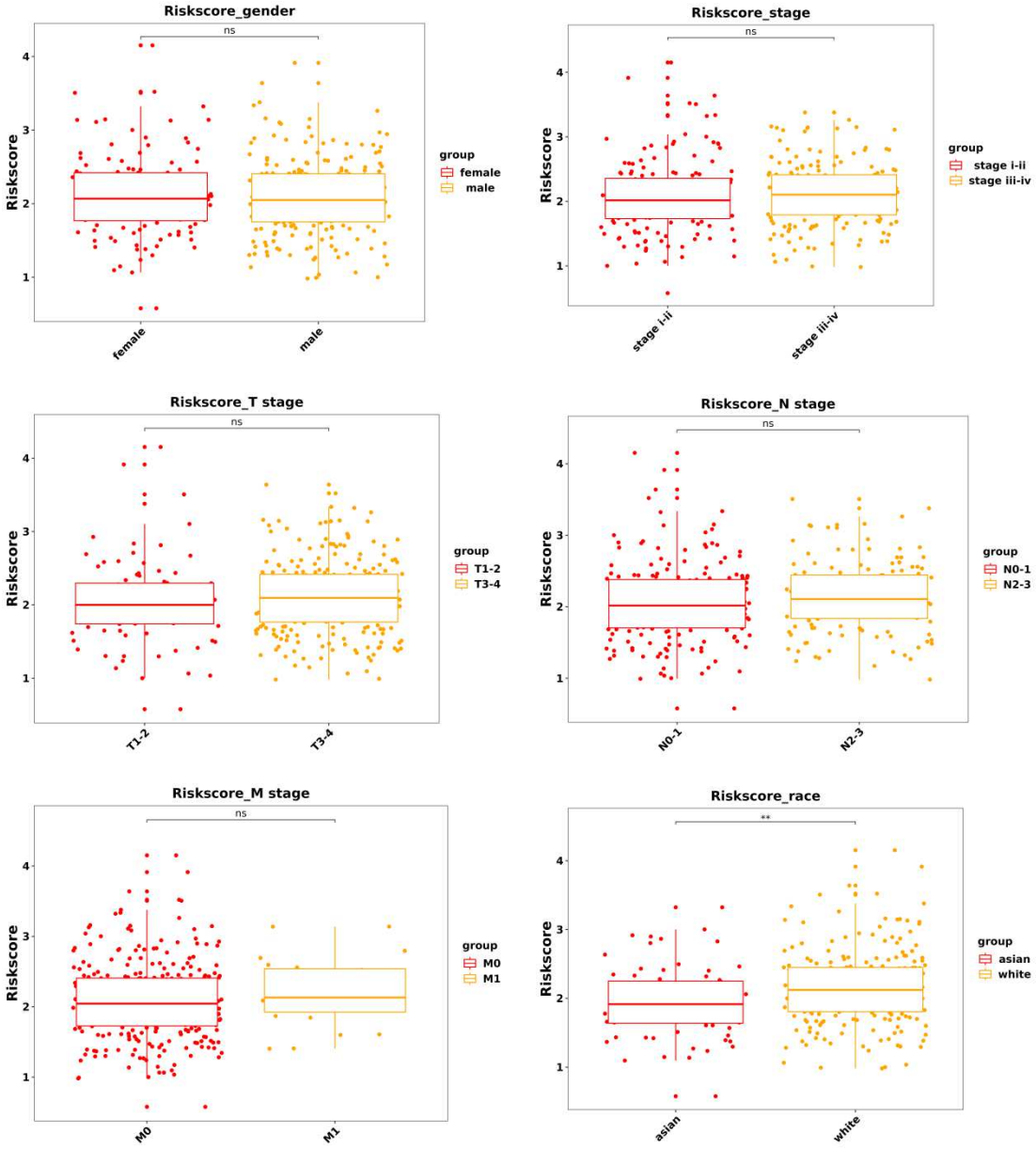
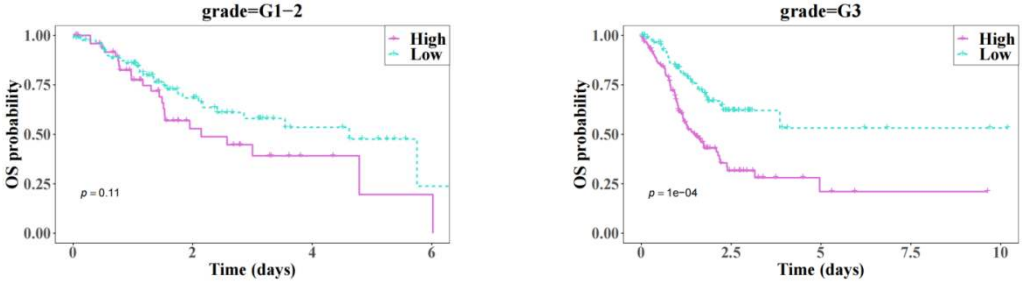


Fig. 6A

Differences in risk scores among patients with different clinical pathological characteristics



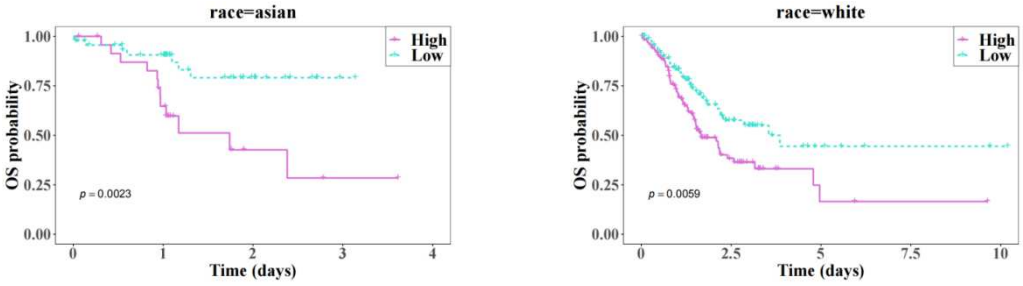
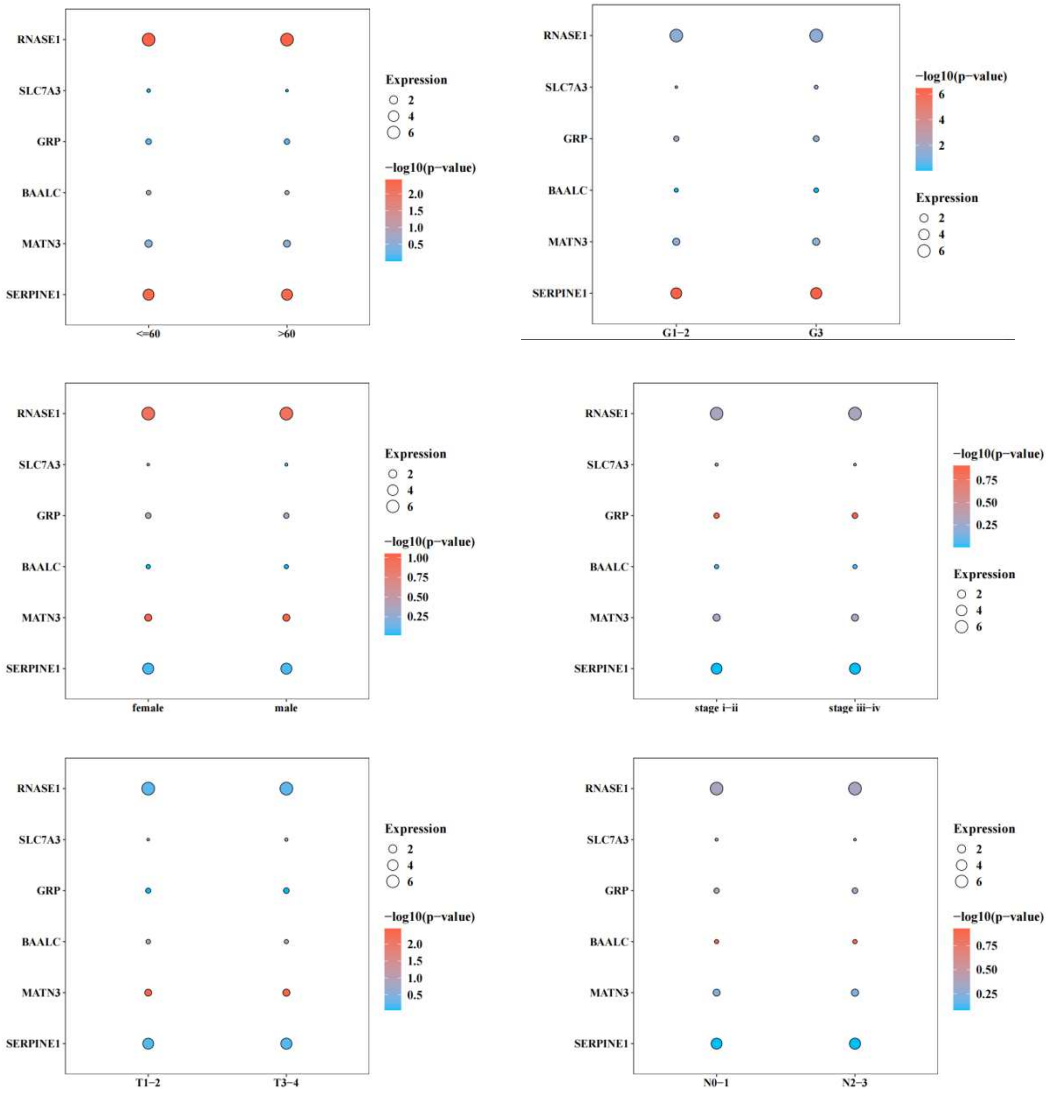


Fig. 6B

The OS display diagram of patients with different pathological characteristics in high-risk and low-risk groups



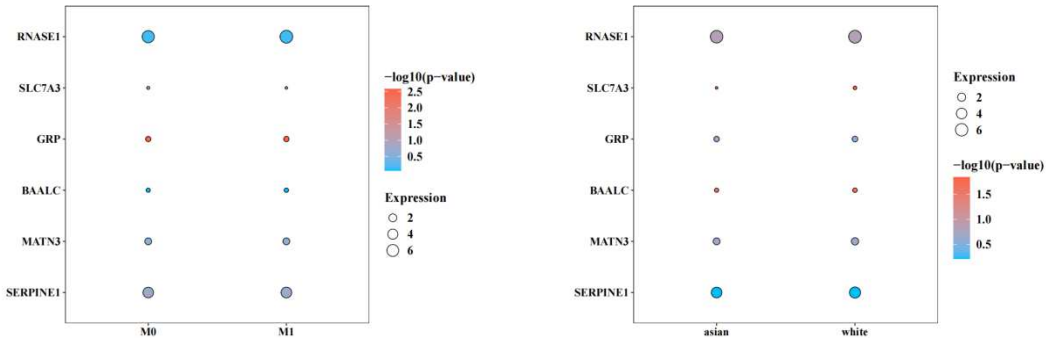


Fig. 6C

Prognostic gene expression analysis in patients with different clinical pathological characteristics

Note: Color indicates significance, and size indicates the expression level of prognostic genes

408

409 **3.7 Functional relevance of distinct risk groups in GC**

410 GSEA was employed to investigate the signaling pathways and potential
411 biological mechanisms correlated with distinct risk groups. To be specific, we
412 found that high risk groups predominantly exhibited enrichment in pathways
413 related to "ECM Receptor Interaction", "Focal Adhesion", and "Neuroactive
414 Ligand Receptor Interaction", while low risk groups were primarily enriched in
415 "Cell Cycle" and "Spliceosome" pathways (**Fig. 7A**).

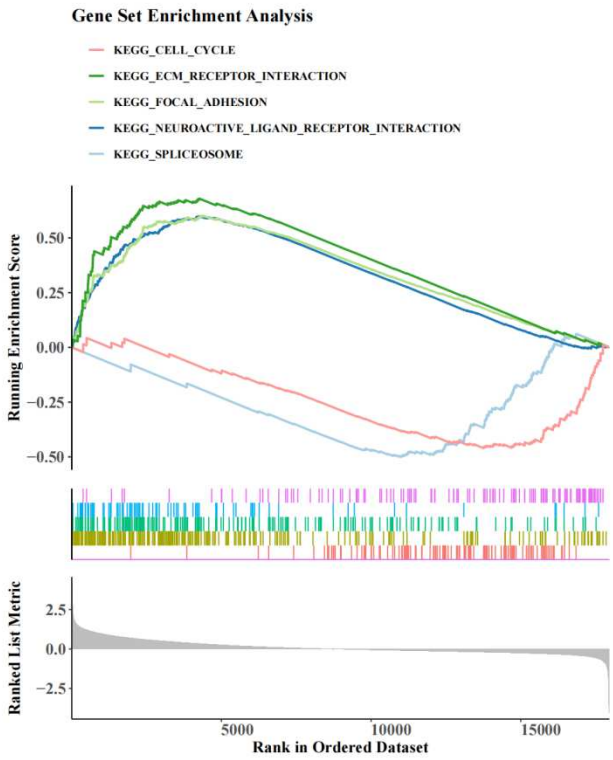


Fig. 7A
GSEA analysis of high-risk and low-risk groups

3.8 Potential molecular regulatory mechanisms of key genes

In this study, the miRDB identified 100 miRNAs corresponding to key genes, whereas TargetScan identified 48 miRNAs. The intersection of these predictions revealed a core set of 12 miRNAs. To elucidate the regulatory network further, we predicted lncRNAs that interact with these 12 miRNAs using starBase and miRNet. The starBase predicted 144 corresponding lncRNA-miRNA relationship pairs, and miRNet predicted 258 pairs. An intersection of these predicted sets revealed 79 significant lncRNA-miRNA interaction pairs, indicating a complex lncRNA-miRNA-mRNA network featuring intricate relationships, such as the SLC7A3-hsa-miR-136-5p-GAS5 interaction (**Fig. 7B**). Further, we extended our prediction analysis to TFs that potentially regulated the key genes, constructing a TF-mRNA network comprising 77 nodes and 82 edges. Notably, our analysis suggested that

EZH2, GATAD1, KLF16, ZNF384, ARID4B, MLX, SIN3A, and RCOR2 were pivotal TFs, as evidenced by their capacity to regulate at least two genes simultaneously (**Fig. 7C**).

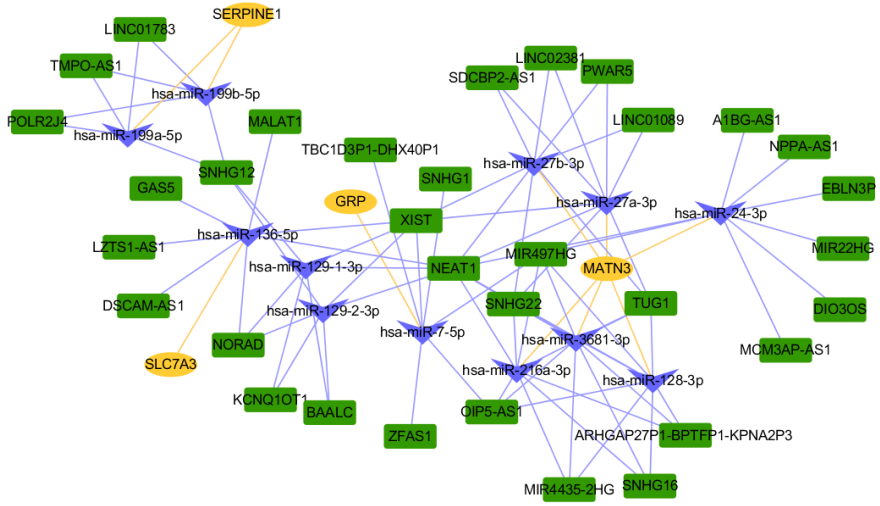


Fig. 7B
Construction of ceRNA network

Note: Yellow circles represent prognostic genes, blue arrows represent miRNAs targeting prognostic genes, and green squares represent lncRNAs targeting miRNAs

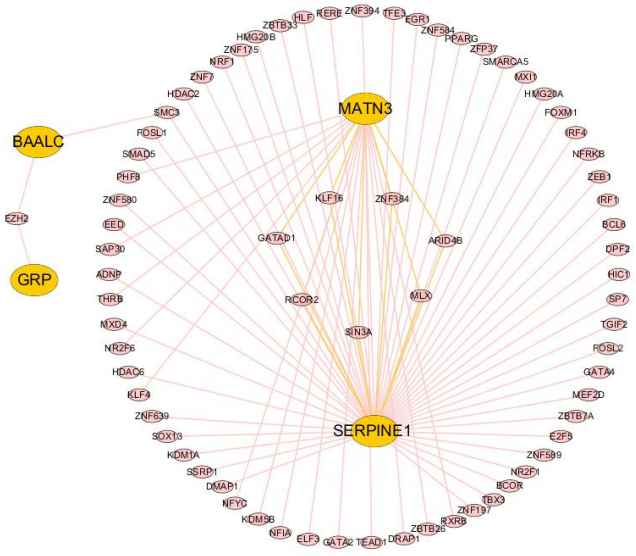


Fig. 7C
TF-mRNA regulatory network

3.9 Verification of key genes expression

In the training cohort, based on samples from the disease group and the normal group, the expression levels of SERPINE1 and MATN3 in gastric cancer samples of the disease group were higher than those in gastric cancer samples of the normal group ($P < 0.05$). In contrast, the expression levels of BAALC, GRP, SLC7A3, and RNASE1 in gastric cancer samples of the disease group were significantly lower than those in gastric cancer samples of the normal group ($P < 0.05$) (**Fig. 8A**). Concurrently, In validation cohort, SERPINE1 expression was significantly elevated in GC as opposed to normal samples ($P < 0.05$), while SLC7A3 and RNASE1 levels were markedly reduced in GC samples ($P < 0.05$) (**Fig. 8B**). The RT-qPCR results revealed that BAALC, SLC7A3, and RNASE1 exhibited a significant downregulation trend in GC samples compared to control samples ($P < 0.05$). This aligned with the trend observed in the aforementioned public dataset (**Fig. 8C-E**).

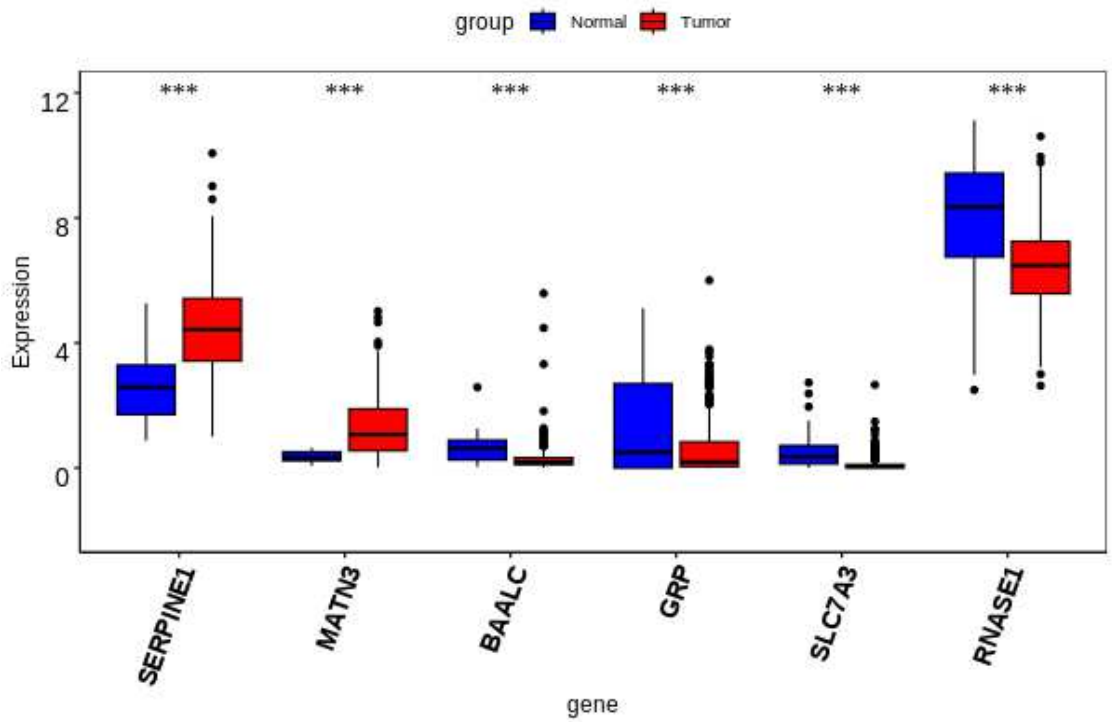


Fig. 8A
The differential level of prognostic genes between patients and normal controls in the training set

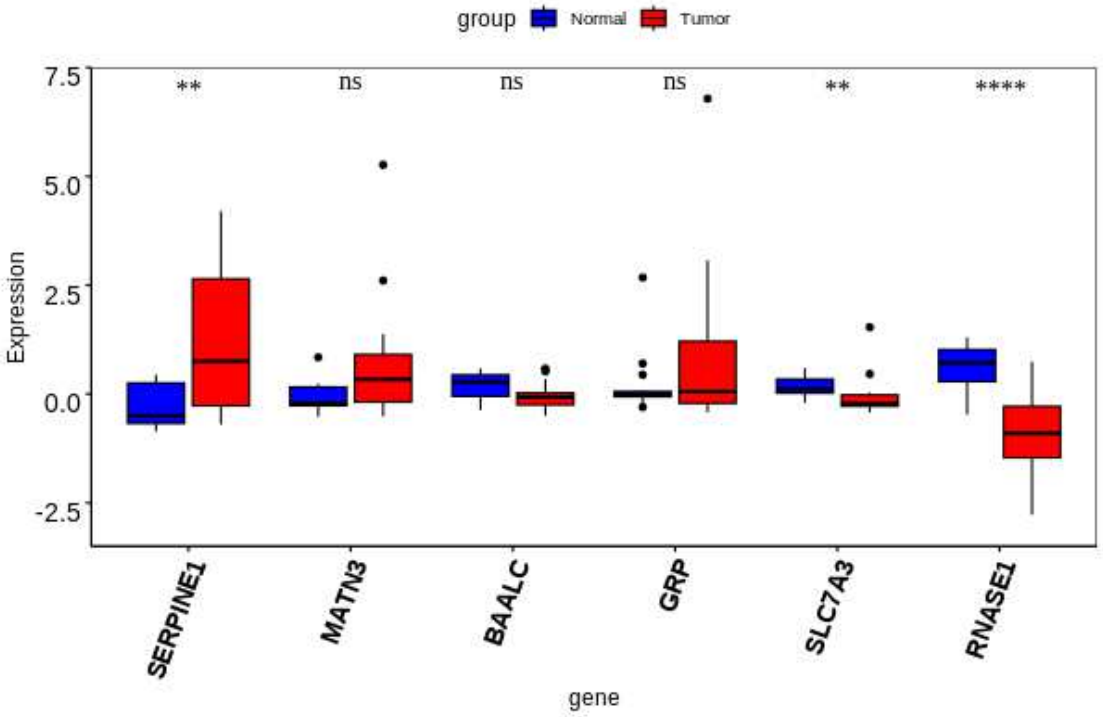


Fig. 8B
Differential levels of prognostic genes in the validation set of patients and normal controls

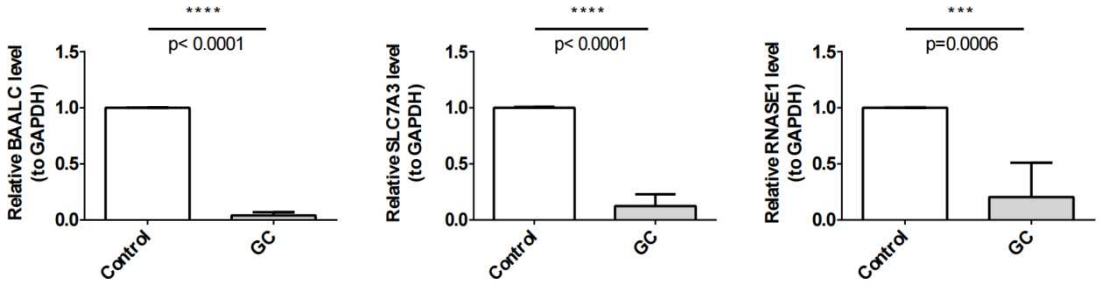


Fig. 8C
RT-qPCR to verify the expression level of prognostic genes

4. Discussion

GC remains the third leading cause of cancer death worldwide, with a lower survival rate and poorer prognosis [29]. The current treatment mode for GC is a comprehensive treatment mode that mainly involves surgical resection, combined with chemotherapy, immunotherapy, and targeted therapy [30-31]. Nevertheless, early gastric cancer usually causes no symptoms or only non-specific clinical manifestations, so many patients are already in the advanced stage when diagnosed.[32], culminating in a 5-year survival rate

below 30% for those with advanced GC[33]. Therefore, there is an urgent need for reliable biomarkers to evaluate the prognosis and treatment effectiveness of GC patients, thereby facilitating the development of personalized diagnostic and therapeutic strategies.

The metabolism of Cys plays a pivotal role in critical cellular processes including oxidative stress, energy production, and autophagy, providing valuable insights into cancer diagnostics and targeted therapeutics. For instance, glutamine has emerged as a potential biomarker for tumors[34-35], and alterations in arginine metabolism have been linked to the mechanisms of tumor development [36]. In cancer cells, Cys is obtained through transporter-mediated uptake and endogenous synthesis via sulfur transfer pathways. Once inside, it's converted to cystine, facilitating tumor growth, spread, and resistance to treatment [37]. Specific transport proteins, such as SLC7A5 and SLC3A1, are integral in this context, with increases in SLC3A1 expression correlating with enhanced progression of breast cancer, its levels are observed to be 1.5-fold higher in tumor tissues compared to adjacent non-tumorous tissues[38]. The sulfur transfer pathway, important for converting methionine to Cys, is associated with energy metabolism and the development of resistance in cancers like ovarian[39]and colon cancer[40], and has been implicated in inhibiting epithelial-mesenchymal transition (EMT) and cell migration in HepG2 cells [41]. Furthermore, the enzyme cystathionine gamma-lyase (CSE), a key player in sulfur metabolism, significantly influences tumor cell expansion and movement. Notably, breast cancer tissues exhibit elevated CSE mRNA levels compared to benign tissues[13], and a marked association has been documented between high CSE expression and decreased survival in prostate cancer, with a survival disparity of 20 months between the high and low expression groups[42]. These findings underscore the importance of Cys metabolism in various cancer types, highlighting its potential as a therapeutic target. Therefore, we boldly speculate that cys metabolism may be a potential target for treating GC. Furthermore, the

properties of Cys-RGs may offer prognostic value and predict treatment responses for GC patients, thereby guiding the exploration of cys metabolism's function within the context of GC.

By integrating rigorous statistical methods including differential expression analysis, WGCNA, PH assumption test, univariate and multivariate Cox regression analyses, LASSO algorithm, and stepwise regression analysis, we identified six cys-related key genes in GC: SERPINE1, MATN3, BAALC, GRP, SLC7A3, and RNASE1. Notably, SERPINE1 has been established as a prognostic biomarker for poor outcomes in GC and is strongly associated with EMT. Zhang et al[43].demonstrated that elevated SERPINE1 predicts poor survival and correlates with EMT molecular subtypes, with Tiplaxtinin inhibition suppressing EMT markers. Additionally, recent studies[44]revealed that the NFATC2/SERPINE1/JAK3/STAT3 axis drives gastric carcinogenesis, promoting proliferation, migration, invasion, and EMT, while contributing to immune evasion and anti-PD-1 resistance. These findings suggest SERPINE1 may serve as both a prognostic indicator and a therapeutic target to enhance immunotherapy efficacy in GC patients.MATN3 drives gastric cancer (GC) progression through epithelial-mesenchymal transition (EMT) pathways[45]. Recent studies demonstrate that MATN3 is significantly upregulated in GC tissues and correlates with poor prognosis. As an amino acid metabolism-related gene, MATN3 promotes GC cell proliferation, migration, and invasion while inhibiting apoptosis[46]. In our training cohort, MATN3 expression was markedly elevated in tumor samples and high-risk groups. Stepwise regression analysis identified MATN3 as an independent risk factor for GC with a hazard ratio (HR) of 1.1[47]. These findings align with recent evidence establishing MATN3 as a key prognostic biomarker in stomach adenocarcinoma. In breast cancer tissues, SLC7A3 expression is downregulated in breast cancer tissues compared to adjacent normal tissues[48]. Its knockdown promotes proliferation and invasion in MCF-7 and MDA-MB-231 cells, while elevated SLC7A3 predicts favorable prognosis,

indicating its tumor suppressor role[49]. BAALC (Brain and Acute Leukemia Cytoplasmic) is recognized as a prognostic gene associated with acute myeloid leukemia (AML), expressed at high levels in CD34+ hematopoietic stem and progenitor cells. Elevated BAALC expression serves as an independent adverse prognostic factor, predicting primary drug resistance, higher cumulative incidence of relapse, and inferior overall survival in AML patients[50]. Recent studies demonstrate that BAALC overexpression is associated with chemotherapy refractoriness across all cytogenetic risk groups, including normal karyotype AML[51]. High BAALC expression induces drug resistance genes (e.g., MDR1/ABCB1) and stem cell markers (CD133, CD34, KIT), contributing to leukemic cell survival[52]. RNASE1 drives macrophage polarization toward a tumor-promoting phenotype through ALK pathway activation [53]. Recent studies highlight its role in tumor microenvironment modulation[54]. Notably, we observed significant down-regulation of BAALC in gastric cancer (GC) samples across both training cohort and PCR validation. Leveraging key cuproptosis-related genes, we constructed a prognostic model stratifying patients into high- and low-risk groups based on median risk scores. Time-dependent ROC analysis demonstrated AUC values of 0.65, 0.69, and 0.75 at 1-, 2-, and 3-years, respectively, indicating superior predictive accuracy for 3-year survival compared to shorter intervals.

Comparatively, Yang et al. developed a lactose metabolism-related model showing mortality rates of 28%, 53.6%, and 66% at corresponding timepoints, with AUCs exceeding 0.65. Proportional hazards assumption testing and multivariate Cox regression identified risk score, age, and M-stage as independent prognostic factors (all $P<0.05$), maintaining significance after adjusting for covariates. These findings align with Chen et al. and Zhu et al., who similarly identified age and metastasis status as critical prognostic indicators in GC[55].

Wang et al. constructed a cuproptosis-related prognostic model achieving AUCs of 0.621, 0.654, and 0.670 at 1-, 3-, and 5-years[45]. Our model

demonstrates competitive performance, particularly for 3-year prediction. Similarly, cuproptosis-related models by Liu et al. and alternative prognostic signatures have shown comparable predictive efficacy in GC[56].

Integrating these three variables, we developed a nomogram demonstrating excellent calibration (slopes approximating 45°across all timepoints) and favorable decision curve analysis, confirming robust clinical utility for individualized survival prediction and treatment stratification in GC patients.

This research has also conducted a preliminary exploration of the association between risk scores and clinical characteristics, revealing that patients within the G3 category and of white ethnicity may have a less favorable prognosis. Within the G3 cohort, both Caucasian and Asian individuals categorized in the low-risk group demonstrated extended survival durations. This suggests a tight correlation between risk scores and patient clinical characteristics, which may well influence the survival prognosis for individuals with GC. Additionally, this study has revealed significant disparities in the expression of key genes in relation to clinical attributes, shedding light on the considerable correlation between key genes and these characteristics. Warsinggih's research on GC delineated the association between the disease's overall incidence and various factors including age, gender, tumor location, stage, type of surgical treatment, and pathological diagnostics[57]. Moreover, this study examined the correlation between clinical characteristics and gastric cancer within the context of developing a radiological model[58]. For instance, the expression of SERPINE1 is discovered to hold a notable link with the grading of the disease, positing it as a potential biomarker for the gravity of GC. Likewise, variations in the expression of MATN3 across different tumor dimensions (T-stages) indicate its influential role in tumor evolution, with its level of expression potentially serving as an indicator of tumor localization or as a predictor of tumor proliferation pace. A pronounced link between GRP and M-stages hints at underlying cellular mechanisms indicative of metastasis presence, assisting in prognosticating the aggressiveness of gastric cancer and its therapy

responses. Furthermore, significant expression variances of BAALC and SLC7A3 among distinct ethnic groups underscore the significance of ethnicity within the biology of gastric cancer, potentially revealing strategies for crafting bespoke treatments for patients from varying ethnic backgrounds. These discoveries could spark advancements in clinical decision-making, especially with the advent of personalized healthcare approaches.

In this study, GSEA was utilized to evaluate the functional pathways prevalent in distinct risk groups associated with GC. Results indicated that the high-risk group exhibited an enrichment of pathways linked to extracellular matrix (ECM) receptor interaction, focal adhesion, and neuroactive ligand-receptor interaction. Conversely, pathways related to the cell cycle and spliceosome were predominantly enriched in the low-risk group. ECM remodeling is critical in GC progression regulation, where fibrinogen 1 (FBN1) expression escalates in advanced GC and undergoes succinylation. This post-translational modification impedes FBN1 degradation by matrix metalloproteinases (MMPs), leading to its accumulation and contributing to tumor advancement through the activation of TGF-β1 and PI3K/Akt intracellular pathways[59]. ECM deregulation impacts tissue homeostasis and, by extension, cancer progression via the modulation of mechanical and biochemical signals. Specifically, ECM remodeling in gastric cancer fosters tumorigenesis through the activation of ECM receptors and pathways implicated in angiogenesis and extracellular metabolism[60]. These signaling pathways are intimately connected with cancer development, progression, and the modulation of cellular behavior. The analysis further revealed that the high-risk group's pathways were significantly associated with cell adhesion, migration, and ECM integration. For instance, copine 8 (CPNE8) enhances GC progression and invasion by regulating focal adhesion and the tumor microenvironment[61], which underscores the pathways' significance in tumor invasiveness and metastatic potential. In contrast, the low-risk group showed pathways related to cell proliferation and RNA splicing, suggesting their differential role in tumor growth. Highlighting the

therapeutic implications, Seiler M.'s research demonstrated that splicing regulation could serve as a potential target in spliceosome mutant cancers[62]. These findings considerably advance our comprehension of GC, enriching the breadth of our knowledge in this domain.

The construction of lncRNA-miRNA-mRNA network reveals intricate interactions, exemplified by the SLC7A3-hsa-miR-136-5p-GAS5 axis. Our analysis also highlights key TFs including EZH2, GATAD1, KLF16, ZNF384, ARID4B, MLX, SIN3A, and RCOR2 that warrant further investigation. The STAT3 signaling pathway is pivotal in tumorigenesis, primarily through the epigenetic modification of EZH2, which amplifies STAT3's activity by promoting its tyrosine phosphorylation. Research shows that STAT3-mediated upregulation of EZH2 correlates with advanced tumor node metastasis (TNM) stage and poor prognosis, suggesting that the concurrent inhibition of STAT3 and EZH2 may represent a novel therapeutic approach for GC[63]. In the context of Krüppel-like factors (KLFs), the KLF16 gene ablation significantly hampers the proliferation of GC cell lines, mediated by an increase in p21 expression and a decrease in CDK4 expression[64]. Concerning the glucose-responsive TF MLX and its partner MondoA, both are fundamental to the viability of male germ cells and human male reproductive cancer cells. Mlx deletion triggers metabolic changes in the testes, activates several stress response pathways, and induces GC apoptosis[65].

The RT-qPCR is a molecular technique widely used for the quantitative analysis of gene expression. It also remains an important tool for molecular monitoring, including the assessment of measurable residual disease in hematological malignancies. [66-67. To validate the robustness of our findings, RT-qPCR was employed to measure the expression of BAALC, SLC7A3, and RNASE1. The outcomes affirm the patterns depicted in the publicly available dataset, demonstrating a consistent downregulation of BAALC, SLC7A3, and RNASE1 in GC specimens.

Conclusion: Our investigation identified six Cys-ralated key genes as novel prognostic indicator for GC patients, and revealed their involvement in the initiation and progression of GC through bioinformatics and in vitro experimentation. However, this study has several limitations. Firstly, this is essentially an exploratory analysis mainly based on public transcriptome datasets, which may have inherent biases such as batch effects, platform heterogeneity and incomplete clinical information. Secondly, although RT-qPCR validation was performed, the sample size was relatively small. It only provided preliminary evidence for the expression levels, with no validation of related functions or mechanisms. Thirdly, the regulatory network and pathway analyses were conducted through computational inference without experimental verification. Therefore, the findings of this study should be interpreted with caution. Further validation using larger independent cohorts and more comprehensive in vitro and in vivo experiments are required to clarify the exact biological functions and clinical significance of these cysteine-related genes in gastric cancer..

Declarations

Ethics approval: This study was conducted in accordance with the principles of the Declaration of Helsinki. Approval was granted by the Ethics Review Committee of the First Affiliated Hospital of Guangxi Medical University (Approval No.: 2026-E0064).

Consent to participate: Informed consent was obtained from all individual participants included in the study.

Consent for publication: Not applicable.

Clinical Trial number:Not applicable.

Conflict of interest: The authors declare no confict of interest.References

Author’s contributions:

Xi Yang, Mengda Wei,Yanqing Huang, Xiaoying Yang, Qiuqi Ou Lei Tian contributed to the study's conception.Xi Yang, Yanqing Huang,Qiuqi Ou,were responsible for bioinformatics analysis and data processing; Mengda Wei and Xiaoying Yang were responsible for RT-qpcr experiments;Xi Yang were responsible for specimen collection,Lei Tian ,Xi Yang and Mengda Wei helped perform the analysis with constructive discussions. All authors read and approved the final manuscript.

Funding:

This research was supported by the Self-financed Scientific Research Project of the Health Commission of Guangxi Zhuang Autonomous Region (Project Contract No.: Z-A20220467). The funding body played no role in the design of the study, collection, analysis, and interpretation of data, or in writing the manuscript.

Availability of data and materials

The datasets generated and/or analysed during the current study are available in the TCGA repository (https://portal.gdc.cancer.gov/projects/TCGA-STAD), the GEO repository for validation cohort GSE84437 (platform GPL6947, https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE84437). The CYSTEINE-METABOLIC related gene set was retrieved from MSigDB (https://www.gsea-msigdb.org/gsea/msigdb/index.jsp). The miRNA-target databases used for regulatory network construction include miRDB (https://mirdb.org/), TargetScanHuman Release 8.0 (http://www.targetscan.org/), starBase (ENCORI platform, http://starbase.sysu.edu.cn/), miRNet (https://www.mirnet.ca/miRNet/home.xhtml), and NetworkAnalyst (https://www.networkanalyst.ca/NetworkAnalyst/uploads/ListUploadView.xhtml).

Acknowledgements

The authors wish to thank Professor Lei Tian from the Gastrointestinal Gland Surgery Department of the First Affiliated Hospital of Guangxi Medical University ,for his guidance.The authors also would like to thank TCGA, GEO databases, online website Kaplan-Meier Plotter, and Guangxi Medical University for providing platforms.

References：

[1] Chen J, Huang Q, Li YQ, et al. Comparative single-cell analysis reveals heterogeneous immune landscapes in adenocarcinoma of the

esophagogastric junction and gastric adenocarcinoma. Cell Death Dis. 2024;15(1):15. Published 2024 Jan 5. doi:10.1038/s41419-023-06388-6

[2] Taieb J, Bennouna J, Penault-Llorca F, Basile D, Samalin E, Zaanan A. Treatment of gastric adenocarcinoma: A rapidly evolving landscape. Eur J Cancer. 2023;195:113370. doi:10.1016/j.ejca.2023.113370.

[3] Manji GA, Lee S, Del Portillo A, et al. Chemotherapy and Immune Checkpoint Blockade for Gastric and Gastroesophageal Junction Adenocarcinoma. JAMA Oncol. 2023;9(12):1702-1707. doi:10.1001/jamaoncol.2023.4423.

[4] Saito Y, Kusuhara M, Ohno A, et al. Endoscopic submucosal dissection for adenoma in gastric adenocarcinoma and proximal polyposis of the stomach. Endoscopy. 2023;55(S 01):E928-E929. doi:10.1055/a-2119-0677

[5] Wheless MC, Comer M, Gibson MK. Evolving Treatment Landscape for Advanced Esophageal and Gastroesophageal Junction Adenocarcinoma. Curr Oncol Rep. 2024;26(11):1469-1488. doi:10.1007/s11912-024-01607-5

[6] Masetti M, Al-Batran SE, Goetze TO, et al. Efficacy of ramucirumab combination chemotherapy as second-line treatment in patients with advanced adenocarcinoma of the stomach or gastroesophageal junction after exposure to checkpoint inhibitors and chemotherapy as first-line therapy. Int J Cancer. 2024;154(12):2142-2150. doi:10.1002/ijc.34894

[7] Zhou H, Sun H, Liu X, et al. Combined effect between WT1 methylation and Helicobacter pylori infection, smoking, and alcohol consumption on the risk of gastric cancer. Helicobacter. 2019;24(5):e12650. doi:10.1111/hel.12650

[8] Sharma A, Singh LR. An insight into the pharmacology of cysteine/methionine containing peptide drugs. Eur J Med Chem. 2024;271:116456. doi:10.1016/j.ejmech.2024.116456

[9] Lee SJ, Zhang X, Wu E, et al. Structural changes in NOTCH3 induced by CADASIL mutations: Role of cysteine and non-cysteine alterations. J Biol Chem. 2023;299(6):104838. doi:10.1016/j.jbc.2023.104838

[10] Li H, Ma T, Remsberg JR, et al. Assigning functionality to cysteines by base editing of cancer dependency genes. Nat Chem Biol. 2023;19(11):1320-1330. doi:10.1038/s41589-023-01428-w

[11] Ghosh R, Pradhan D, Debnath S, Mansingh A, Nagesh N, Chatterjee PB. A Hydrogen Bonded Non-Porous Organic-Inorganic Framework for Measuring Cysteine in Blood Plasma and Endogenous Cancer Cell. Chemistry. 2024;30(50):e202401255. doi:10.1002/chem.202401255

[12] Song M, Pan W, Yu X, et al. Minimal Residual Disease Detection: Implications for Clinical Diagnosis and Cancer Patient Treatment. MedComm (2020). 2025;6(6):e70193. Published 2025 May 15. doi:10.1002/mco2.70193

[13] Huang D, Li J, He Z, et al. Pan-cancer and experimental analyses reveal the immunotherapeutic significance of CST2 and its association with stomach adenocarcinoma proliferation and metastasis. Front Immunol. 2025;15:1466806. Published 2025 Jan 24. doi:10.3389/fimmu.2024.1466806

[14] Zheng H, Chen H, Cai Y, et al. Hydrogen sulfide-mediated persulfidation regulates homocysteine metabolism and enhances ferroptosis in non-small cell lung cancer. Mol Cell. 2024;84(20):4016-4030.e6. doi:10.1016/j.molcel.2024.08.035

[15] Patra S, Patra D, Dey J. Camptothecin-loaded Surface-modified Nanostructured Lipid Carriers Consisting of L-Cysteine- and L-Cystine-derived Lipids: Physicochemical Characterization and In Vitro Evaluation. Chem Asian J. 2025;20(17):e00180. doi:10.1002/asia.202500180

[16] 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. doi:10.1186/s13059-014-0550-8

[17] Gustavsson EK, Zhang D, Reynolds RH, Garcia-Ruiz S, Ryten M. ggtranscript: an R package for the visualization and interpretation of transcript isoforms using ggplot2. Bioinformatics. 2022;38(15):3844-3846. doi:10.1093/bioinformatics/btac409

[18] Gu Z, Hübschmann D. Make Interactive Complex Heatmaps in R. Bioinformatics. 2022;38(5):1460-1462. doi:10.1093/bioinformatics/btab806

[19] Hänzelmann S, Castelo R, Guinney J. GSVA: gene set variation analysis for microarray and RNA-seq data. BMC Bioinformatics. 2013;14:7. Published 2013 Jan 16. doi:10.1186/1471-2105-14-7

[20] Ramsay IS, Ma S, Fisher M, et al. Model selection and prediction of outcomes in recent onset schizophrenia patients who undergo cognitive training. Schizophr Res Cogn. 2017;11:1-5. Published 2017 Nov 8. doi:10.1016/j.scog.2017.10.001

[21] Langfelder P, Horvath S. WGCNA: an R package for weighted correlation network analysis. BMC Bioinformatics. 2008;9:559. Published 2008 Dec 29. doi:10.1186/1471-2105-9-559

[22] Zheng Y, Gao W, Zhang Q, et al. Ferroptosis and Autophagy-Related Genes in the Pathogenesis of Ischemic Cardiomyopathy. Front Cardiovasc Med. 2022;9:906753. Published 2022 Jun 30. doi:10.3389/fcvm.2022.906753

[23] Yu G, Wang LG, Han Y, He QY. clusterProfiler: an R package for comparing biological themes among gene clusters. OMICS. 2012;16(5):284-287. doi:10.1089/omi.2011.0118

[24] Ramsay IS, Ma S, Fisher M, et al. Model selection and prediction of outcomes in recent onset schizophrenia patients who undergo cognitive training. Schizophr Res Cogn. 2017;11:1-5. Published 2017 Nov 8. doi:10.1016/j.scog.2017.10.001

[25] Li Y, Lu F, Yin Y. Applying logistic LASSO regression for the diagnosis of atypical Crohn's disease. Sci Rep. 2022;12(1):11340. Published 2022 Jul 5. doi:10.1038/s41598-022-15609-5

[26] Sachs MC. plotROC: A Tool for Plotting ROC Curves. J Stat Softw. 2017;79:2. doi:10.18637/jss.v079.c02

[27] Liu P, Xu H, Shi Y, Deng L, Chen X. Potential Molecular Mechanisms of Plantain in the Treatment of Gout and Hyperuricemia Based on Network Pharmacology. Evid Based Complement Alternat Med. 2020;2020:3023127. Published 2020 Oct 22. doi:10.1155/2020/3023127

[28] Livak KJ, Schmittgen TD. Analysis of relative gene expression data using real-time quantitative PCR and the 2(-Delta Delta C(T)) Method. Methods. 2001;25(4):402-408. doi:10.1006/meth.2001.1262

[29] Zhu Y, Hu Y, Wang P, et al. Comprehensive bioinformatics and experimental analysis of SH3PXD2B reveals its carcinogenic effect in gastric carcinoma. Life Sci. 2023;326:121792. doi:10.1016/j.lfs.2023.121792

[30] Zhao W, Lin J, Cheng S, Li H, Shu Y, Xu C. Comprehensive analysis of COMMD10 as a novel prognostic biomarker for gastric cancer. PeerJ. 2023;11:e14645. Published 2023 Mar 9. doi:10.7717/peerj.14645

[31] Yin X, Xing W, Yi N, et al. Comprehensive analysis of lactylation-related gene sets and mitochondrial functions in gastric adenocarcinoma: implications for prognosis and therapeutic strategies. Front Immunol. 2024;15:1451725. Published 2024 Oct 16. doi:10.3389/fimmu.2024.1451725

[32] Niu X, Ma F, Li F, et al. Integration of bioinformatics and cellular experiments unveils the role of SYT12 in gastric cancer [retracted in: BMC Cancer. 2026 Feb 21;26(1):291. doi: 10.1186/s12885-026-15764-2.]. BMC Cancer. 2024;24(1):1331. Published 2024 Oct 29. doi:10.1186/s12885-024-13077-w

[33] Tian S, Liu Y, Liu P, Nomura S, Wei Y, Huang T. Development and Validation of a Comprehensive Prognostic and Depression Risk Index for Gastric Adenocarcinoma. Int J Mol Sci. 2024;25(19):10776. Published 2024 Oct 7. doi:10.3390/ijms251910776

[34] Ekici S, Risk BB, Neill SG, Shu HK, Fleischer CC. Characterization of dysregulated glutamine metabolism in human glioma tissue with 1H NMR. Sci Rep. 2020;10(1):20435. Published 2020 Nov 24. doi:10.1038/s41598-020-76982-7

[35] Ramachandran R, Grantham T, Parvataneni S, et al. Gastric Cancer: Clinical Features, Screening, Diagnosis, Treatment, and Prevention. J Community Hosp Intern Med Perspect. 2024;14(2):49-57. Published 2024 Mar 4. doi:10.55729/2000-9666.1304

[36] Fultang L, Gamble LD, Gneo L, et al. Macrophage-Derived IL1β and TNFα Regulate Arginine Metabolism in Neuroblastoma. Cancer Res. 2019;79(3):611-624. doi:10.1158/0008-5472.CAN-18-2139

[37] Bonifácio VDB, Pereira SA, Serpa J, Vicente JB. Cysteine metabolic circuitries: druggable targets in cancer. Br J Cancer. 2021;124(5):862-879. doi:10.1038/s41416-020-01156-1

[38] Jiang Y, Cao Y, Wang Y, et al. Cysteine transporter SLC3A1 promotes breast cancer tumorigenesis. Theranostics. 2017;7(4):1036-1046. Published 2017 Feb 26. doi:10.7150/thno.18005

[39] Bhattacharyya S, Saha S, Giri K, et al. Cystathionine beta-synthase (CBS) contributes to advanced ovarian cancer progression and drug resistance. PLoS One. 2013;8(11):e79167. Published 2013 Nov 13. doi:10.1371/journal.pone.0079167

[40] Da D, Pan Z, Zeng L, et al. Glutamate-cysteine ligase catalytic and its modifier function as novel immunotargets in gastric adenocarcinoma. Asian J Surg. 2023;46(1):143-149. doi:10.1016/j.asjsur.2022.02.005

[41] Phoo NLL, Dejkriengkraikul P, Khaw-On P, Yodkeeree S. Transcriptomic Profiling Reveals AKR1C1 and AKR1C3 Mediate Cisplatin Resistance in

Signet Ring Cell Gastric Carcinoma via Autophagic Cell Death. Int J Mol Sci. 2021;22(22):12512. Published 2021 Nov 19. doi:10.3390/ijms222212512

[42] Huang D, Li J, He Z, et al. Pan-cancer and experimental analyses reveal the immunotherapeutic significance of CST2 and its association with stomach adenocarcinoma proliferation and metastasis. Front Immunol. 2025;15:1466806. Published 2025 Jan 24. doi:10.3389/fimmu.2024.1466806

[43] Patra S, Patra D, Dey J. Camptothecin-loaded Surface-modified Nanostructured Lipid Carriers Consisting of L-Cysteine- and L-Cystine-derived Lipids: Physicochemical Characterization and In Vitro Evaluation. Chem Asian J. 2025;20(17):e00180. doi:10.1002/asia.202500180

[44] Lin J, Xu N, An W. Screening the Biomarkers and Related Medicines for Gastric Adenocarcinoma. J Coll Physicians Surg Pak. 2024;34(3):290-295. doi:10.29271/jcpsp.2024.03.290

[45] Li J, Xie B, Wang H, Wang Q, Wu Y. Investigating MATN3 and ASPN as novel drivers of gastric cancer progression via EMT pathways. Hum Mol Genet. 2024;33(23):2035-2050. doi:10.1093/hmg/ddae129

[46] Zhu W, Fu M, Li Q, et al. Amino acid metabolism-related genes as potential biomarkers and the role of MATN3 in stomach adenocarcinoma: A bioinformatics, mendelian randomization and experimental validation study. Int Immunopharmacol. 2024;143(Pt 1):113253. doi:10.1016/j.intimp.2024.113253

[47] Li X, Lei Y. Construction of a prognostic risk model for Stomach adenocarcinoma based on endoplasmic reticulum stress genes. Wien Klin Wochenschr. 2024;136(11-12):319-330. doi:10.1007/s00508-023-02306-0

[48] Li D, Xu J, Dong X, et al. Diagnostic and prognostic value of MATN3 expression in gastric carcinoma: TCGA database mining. J Gastrointest Oncol. 2021;12(4):1374-1383. doi:10.21037/jgo-21-267

[49] Wang P, Song Y, Li H, et al. SIRPA enhances osteosarcoma metastasis by stabilizing SP1 and promoting SLC7A3-mediated arginine uptake. Cancer Lett. 2023;576:216412. doi:10.1016/j.canlet.2023.216412

[50] Jaque-Cabrera L, Buggiani J, Bignon J, Daira P, Bernoud-Hubac N, Martin JR. Knockdown of the snoRNA-Jouvence Blocks the Proliferation and Leads to the Death of Human Primary Glioblastoma Cells. Noncoding RNA. 2025;11(4):54. Published 2025 Jul 18. doi:10.3390/ncrna11040054

[51] Li S, Liu Z, Chen Q, Chen Y, Ji S. A novel fatty acid metabolism-related signature identifies MUC4 as a novel therapy target for esophageal squamous cell carcinoma. Sci Rep. 2024;14(1):12476. Published 2024 May 30. doi:10.1038/s41598-024-62917-z

[52] Olkhovskiy IA, Gorbenko AS, Stolyar MA, et al. Study of mRNA of WT1, BAALC, EVI1, PRAME and HMGA2 genes in whole blood samples. Исследование мРНК генов WT1, BAALC, EVI1, PRAME и HMGA2 в образцах цельной крови. Klin Lab Diagn. 2022;67(10):613-620. doi:10.51620/0869-2084-2022-67-10-613-620

[53] Liu C, Zhou C, Xia W, et al. Targeting ALK averts ribonuclease 1-induced immunosuppression and enhances antitumor immunity in hepatocellular carcinoma. Nat Commun. 2024;15(1):1009. Published 2024 Feb 2. doi:10.1038/s41467-024-45215-0

[54] Baldus CD, Tanner SM, Ruppert AS, et al. BAALC expression predicts clinical outcome of de novo acute myeloid leukemia patients with normal cytogenetics: a Cancer and Leukemia Group B Study. Blood. 2003;102(5):1613-1618. doi:10.1182/blood-2003-02-0359

[55] Li W, Zhong Q, Deng N, et al. Identification of a novel prognostic model for gastric cancer utilizing glutamine-related genes. Heliyon. 2024;10(19):e37985. Published 2024 Sep 19. doi:10.1016/j.heliyon.2024.e37985

[56] Langer C, Radmacher MD, Ruppert AS, et al. High BAALC expression associates with other molecular prognostic markers, poor outcome, and a

distinct gene-expression signature in cytogenetically normal patients younger than 60 years with acute myeloid leukemia: a Cancer and Leukemia Group B (CALGB) study. Blood. 2008;111(11):5371-5379. doi:10.1182/blood-2007-11-124958

[57] Zhu W, Fu M, Li Q, et al. Amino acid metabolism-related genes as potential biomarkers and the role of MATN3 in stomach adenocarcinoma: A bioinformatics, mendelian randomization and experimental validation study. Int Immunopharmacol. 2024;143(Pt 1):113253. doi:10.1016/j.intimp.2024.113253

[58] Zhu X, Wang J, Jin X, Chen Y, Hu L, Zhao J. Construction and evaluation of a prognostic risk assessment model of gastric cancer by using hypoxia features. Mutat Res. 2022;825:111795. doi:10.1016/j.mrfmmm.2022.111795

[59] Wu Z, Li G, Wang W, et al. Immune checkpoints signature-based risk stratification for prognosis of patients with gastric cancer. Cell Signal. 2024;113:110976. doi:10.1016/j.cellsig.2023.110976

[60] Zhong T, Zhang W, Guo H, et al. The regulatory and modulatory roles of TRP family channels in malignant tumors and relevant therapeutic strategies. Acta Pharm Sin B. 2022;12(4):1761-1780. doi:10.1016/j.apsb.2021.11.001

[61] Zhang P, Cao X, Guan M, et al. CPNE8 Promotes Gastric Cancer Metastasis by Modulating Focal Adhesion Pathway and Tumor Microenvironment. Int J Biol Sci. 2022;18(13):4932-4949. Published 2022 Jul 25. doi:10.7150/ijbs.76425

[62] Seiler M, Yoshimi A, Darman R, et al. H3B-8800, an orally available small-molecule splicing modulator, induces lethality in spliceosome-mutant cancers. Nat Med. 2018;24(4):497-504. doi:10.1038/nm.4493

[63] Pan YM, Wang CG, Zhu M, et al. STAT3 signaling drives EZH2 transcriptional activation and mediates poor prognosis in gastric cancer.

Mol Cancer. 2016;15(1):79. Published 2016 Dec 9. doi:10.1186/s12943-016-0561-z

[64] Ma P, Sun CQ, Wang YF, et al. KLF16 promotes proliferation in gastric cancer cells via regulating p21 and CDK4. Am J Transl Res. 2017;9(6):3027-3036. Published 2017 Jun 15.

[65] Carroll PA, Freie BW, Cheng PF, et al. The glucose-sensing transcription factor MLX balances metabolism and stress to suppress apoptosis and maintain spermatogenesis. PLoS Biol. 2021;19(10):e3001085. Published 2021 Oct 20. doi:10.1371/journal.pbio.3001085

[66] Chea M, Rigolot L, Canali A, Vergez F. Minimal Residual Disease in Acute Myeloid Leukemia: Old and New Concepts. Int J Mol Sci. 2024;25(4):2150. Published 2024 Feb 10. doi:10.3390/ijms25042150

[67] Vonk CM, Al Hinai ASA, Hanekamp D, Valk PJM. Molecular Minimal Residual Disease Detection in Acute Myeloid Leukemia. Cancers (Basel). 2021;13(21):5431. Published 2021 Oct 29. doi:10.3390/cancers13215431

Supplementary Files

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

- AuthorStatement.docx