



OPEN

Integrated analysis identifies oxidative stress genes associated with progression and prognosis in gastric cancer

Zhengyuan Wu^{1,4}, Lin Wang^{2,4}, Zhenpei Wen² & Jun Yao^{2,3}✉

Oxidative stress (OS) reactions are reported to be associated with oncogenesis and tumor progression. However, little is known about the potential diagnostic value of OS in gastric cancer (GC). This study identified hub OS genes associated with the prognosis and progression of GC and illustrated the underlying mechanisms. The transcriptome data and corresponding GC clinical information were collected from The Cancer Genome Atlas (TCGA) database. Aberrantly expressed OS genes between tumors and adjacent normal tissues were screened, and 11 prognosis-associated genes were identified with a series of bioinformatic analyses and used to construct a prognostic model. These genes were validated in the Gene Expression Omnibus (GEO) database. Furthermore, weighted gene co-expression network analysis (WGCNA) was subsequently conducted to identify the most significant hub genes for the prediction of GC progression. Analysis revealed that a good prognostic model was constructed with a better diagnostic accuracy than other clinicopathological characteristics in both TCGA and GEO cohorts. The model was also significantly associated with the overall survival of patients with GC. Meanwhile, a nomogram based on the risk score was established, which displayed a favorable discriminating ability for GC. In the WGCNA analysis, 13 progression-associated hub OS genes were identified that were also significantly associated with the progression of GC. Furthermore, functional and gene ontology (GO) analyses were performed to reveal potential pathways enriched with these genes. These results provide novel insights into the potential applications of OS-associated genes in patients with GC.

Gastric cancer (GC), which was the third leading cause of cancer mortality until 2018, exhibits the fifth largest incidence rate worldwide, and remains a serious threat to human health^{1,2}. The occurrence and progression of GC is a complicated multi-step process involving various genetic and epigenetic risk factors³, where *Helicobacter pylori* infection is the most common⁴. Owing to the lack of specific symptoms, most patients with GC are diagnosed at an advanced stage, and thus, have a significantly poor 5-year survival rate⁵. Presently, the optimal management of patients with GC is surgical resection, although its overall 5-year survival rate is only 20–25%. Moreover, approximately half of the patients with GC who receive adjuvant therapy experience systemic or local tumor recurrence⁶. Unfortunately, GC diagnosis and treatment currently does not meet the needs for earlier diagnosis and longer survival time. Hence, exploration of novel biomarkers that provide increased predictive value is urgently required to improve prognostication for GC.

Currently, the unequivocal mechanism leading to GC carcinogenesis remains poorly understood; however, it is proposed that oxidative stress is an important factor in driving tumorigenesis and cancer progression through excessive production of reactive oxygen species (ROS)^{7–9}. As a characteristic of OS, ROS comprises free radicals or reactive nonradical species, including singlet oxygen, hydrogen peroxide (H₂O₂), and superoxide anion¹⁰. Additionally, ROS is dramatically elevated in patients with GC¹¹. In the absence of scavenging potential, ROS leads to genotoxicity and induces DNA damage^{12,13}. Moreover, accumulated DNA damage eventually induces various genomic mutations and initiates tumorigenesis^{14,15}. As a recognized risk factor of GC, increasing evidence

¹Department of Orthopedics Trauma and Hand Surgery, The First Affiliated Hospital of Guangxi Medical University, Nanning 530021, China. ²Department of Bone and Joint Surgery, The First Affiliated Hospital of Guangxi Medical University, Nanning 530021, China. ³Guangxi Collaborative Innovation Center for Biomedicine, Guangxi Medical University, Nanning 530021, China. ⁴These authors contributed equally: Zhengyuan Wu and Lin Wang. ✉email: junyao800524@163.com

indicates a positive association between *H. pylori* infection and gastric adenocarcinoma due to increased OS¹⁶. *Helicobacter pylori* infection generates ROS by activating various oxidant-producing enzymes, including inducible nitric oxide synthase and nicotinamide adenine dinucleotide phosphate (NADPH) oxidase¹⁷, and subsequently activates several pathways like Wnt, mTOR, and Ras, to initiate GC^{18–20}. These studies have clarified that OS is closely correlated with the progression of GC. Nevertheless, the prognostic value of these OS genes in GC prognosis prediction is largely unclarified, and the underlying mechanisms require further validation.

GC diagnostic methods mainly depend on imaging tests, molecular diagnostics, and histopathological examination. Only a small fraction of OS-related genes has been studied intensively and are known to play a key role in GC progression. Recently, large-scale tumor genome profiles have provided gene expression data, which provides an excellent chance to identify potential molecular markers^{21,22}. And bioinformatic analysis of OS genes in this study might also help to discover new GC diagnostic or prognostic markers to screen for innovative treatment targets. Herein, GC RNA-sequencing data and the corresponding clinical information were downloaded from The Cancer Genome Atlas (TCGA) and Gene Expression Omnibus (GEO) databases, and several prognosis-annotated OS genes were selected to construct a risk model. In addition, the relationship between the expression of OS genes and GC progression was examined using the weighted gene co-expression network analysis (WGCNA) method, which is a bioinformatics method that describes the relevance of gene sets and clinical traits between different samples^{23,24}. Gene ontology (GO) and Kyoto Encyclopedia of Genes and Genomes (KEGG) pathway enrichment analyses were also performed to probe the underlying mechanisms of OS genes in GC. Ultimately, a cluster of OS genes involved in the prognosis and progression of GC was identified, some of which might be developed as potential prognostic and diagnostic biomarkers in the future.

Material and methods

Data acquisition and differentially expressed OS genes (DEOSGs). The RNA-sequencing dataset comprised 375 GC samples and 32 normal gastric tissues with corresponding clinical information downloaded from TCGA database (<https://portal.gdc.cancer.gov/>) on May 3, 2020. To identify DEOSGs, 1399 OS protein domains were extracted from GeneCards (<https://www.genecards.org>) with a relevance score ≥ 7 (Supplement file 1), and further preprocessed with the limma package in view of a false discovery rate (FDR) < 0.05 and $|\log_2$ fold change (FC)| ≥ 1 , in accordance with previously reported methods²⁵. Meanwhile, genes with an average count value of < 1 were eliminated. This yielded 279 DEOSGs for further analysis. Additionally, gene profiles and clinical information of 433 patients with GC from the GSE84437 dataset (<https://www.ncbi.nlm.nih.gov/geo/>) were used as a validation cohort.

GO and KEGG pathway enrichment analysis. GO enrichment and KEGG pathway analysis^{26–28} were applied to systematically investigate the biological functions of selected DEOSGs using the Database for Annotation, Visualization, and Integrated Discovery (DAVID) version 6.8²⁹. GO analysis comprehensively comprised three terms: biological process (BP), cellular component (CC), and molecular function (MF). *P* and FDR values < 0.05 were considered significantly different.

Protein–protein interaction (PPI) network construction and module screening. The PPI information among all DEOSGs was identified using the search tool for the retrieval of interacting genes/proteins (STRING) online platform (<http://www.string-db.org/>)³⁰, and subsequently, their interactions were imported into the Cytoscape 3.7.0 software to construct and visualize a PPI network. The Molecular Complex Detection (MCODE) plug-in was also used to elect the virtual modules and hub genes in the PPI network with both MCODE scores and node counts > 5 ³¹. *P* < 0.05 was considered the significant threshold.

Prognostic model construction and efficacy evaluation. All hub DEOSGs in the key modules were subjected to univariate Cox regression analysis using the survival R package to explore the relationship between each gene and patients' overall survival; the genes with *P* < 0.05 were identified as prognosis related DEOSGs. Subsequently, these candidate genes were integrated into the least absolute shrinkage and selection operator (LASSO) regression³² to construct a potential risk signature of patients with GC. The formula for each sample's risk score was calculated as follows: $\text{risk score} = \sum \text{exp gene}_i * \beta_i$, where exp gene_i represents the relative expression value of OS genes, and β represents the regression coefficient.

Based on the LASSO prognostic model, patients were categorized into high- and low-risk groups, and the Kaplan–Meier method and log-rank test using R Bioconductor survival package were further conducted to compare the overall survival between the two subgroups. Additionally, the survivalROC and timeROC packages in R were calculated to validate the predictive accuracy and ability of the signature³³, and univariate and multivariate Cox regression analyses were also performed to evaluate the relationship between clinical characteristics and risk score. Finally, a nomogram incorporating calibration plots was constructed to forecast the clinical outcome of patients with GC using the RMS R package³⁴. All methods were also included in the GEO cohort to confirm the prognostic performance of this model.

Hub gene evaluation. To clarify the differential expression of 11 hub DEOSGs at a translational level, the Human Protein Atlas (HPA) online database (<http://www.proteinatlas.org/>) was used to discriminate between normal and GC tumor tissues³⁵. Furthermore, the expression of these DEOSGs in GC was also verified in TCGA dataset as described above.

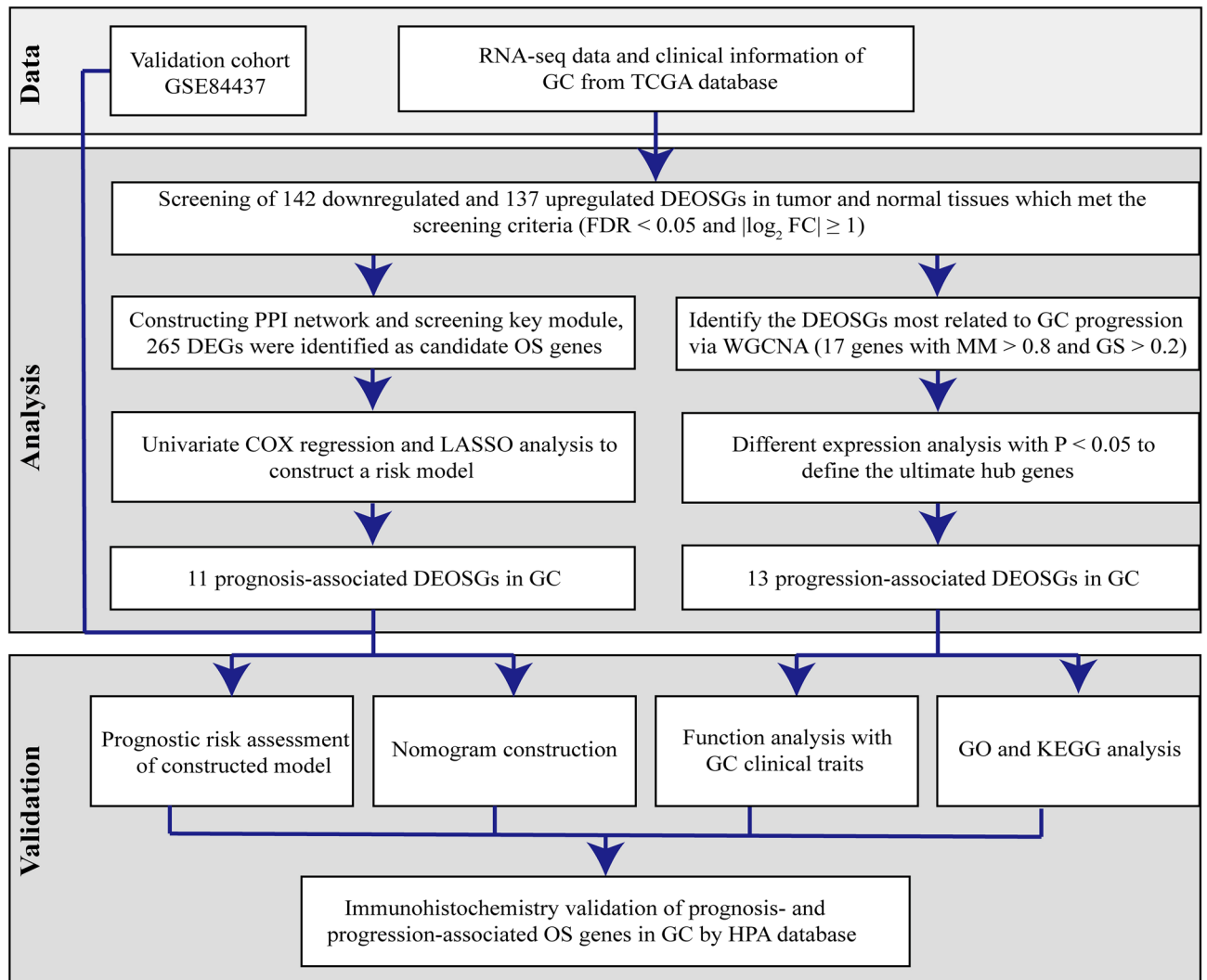


Figure 1. Flowchart describing the schematic overview of the study design.

WGCNA construction and identification of progression-annotated hub genes. A total of 279 DEOSGs in TCGA database were used to create a co-expression network using the WGCNA package in R^{36,37}. Briefly, a hierarchical clustering analysis of GC tissues with numerous clinicopathological features (OS days, OS state, age, gender, grade, and stage) was applied to remove outlier samples. Subsequently, Pearson's correlation coefficients for pairwise genes, and a weighted adjacency matrix was erected by the power function $a_{mn} = |c_{mn}|^\beta$ (c_{mn} = Pearson's correlation between gene m and gene n ; am_n = adjacency between genes m and n). After that, a suitable soft thresholding parameter β was screened to emphasize strong correlations and penalize weak correlations between genes. Then, the adjacencies were converted to a topological overlap matrix (TOM). In light of the TOM-based dissimilarity measure, the average linkage hierarchical clustering was conducted with a minimum module size of 50 for the DEOSG dendrogram, and the dissimilarity of module eigengenes was also calculated. Furthermore, two parameters [module eigengenes (MEs) and gene significance (GS)] revealed modules that were most relevant to GC progression. Hub genes comprising highly interconnected nodes within the module are regarded as functionally significant³⁸. Thus, in this study, after choosing a significant module, genes with high module membership ($MM > 0.8$) and $GS (> 0.2)$ were defined as candidate DEOSGs. Then, the transcriptional expression levels in normal tissues and GC samples were compared; those with significantly different expression levels were defined as the ultimate hub genes. To verify that hub genes were significantly associated with GC clinical traits, the relevance between hub genes and tumor grade was analyzed. The HPA database was used to verify the expression of the real hub genes. Meanwhile, GO and KEGG analyses using the R package were also used to identify the function and signaling pathways enriched with these hub genes.

Results

DEOSG identification and functional enrichment analysis. Bioinformatics analysis of publicly available datasets was performed according to the workflow illustrated in Fig. 1. A total of 1399 OS genes were included to compare differential expression between 32 normal stomach and 375 GC samples; 279 OS genes, comprising 142 downregulated and 137 upregulated genes, met the screening criteria ($P < 0.05$, $|\log_2 FC| > 1.0$)

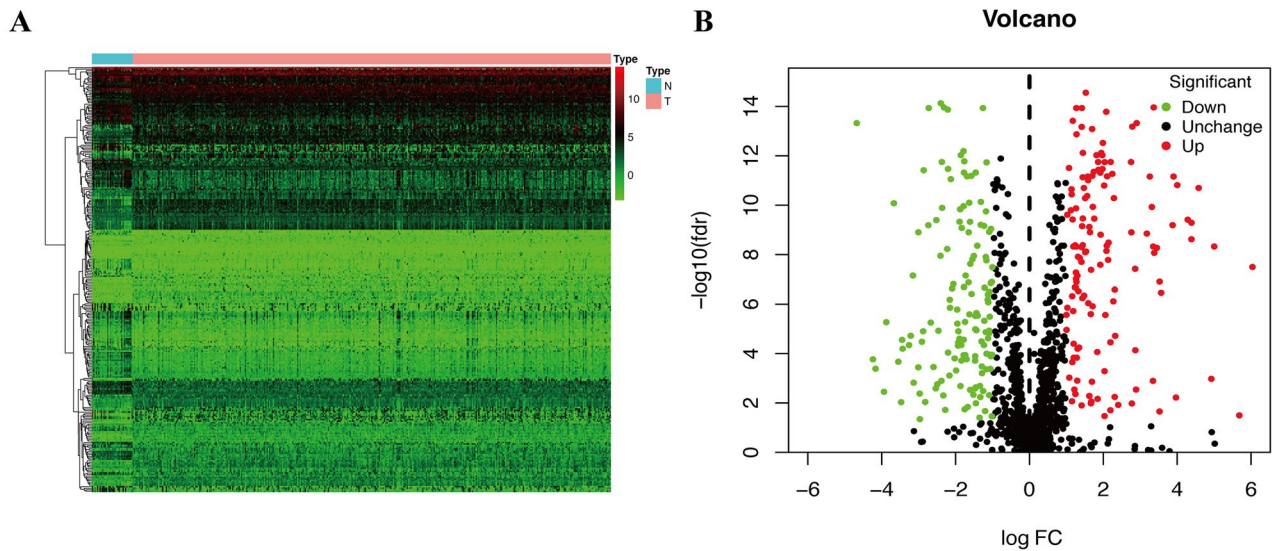


Figure 2. Identification of differentially expressed OS genes. (A) Volcano plot of DEOSGs between TCGA-GC and normal stomach samples. (B) Heatmap of DEOSGs. Dots in green represent down-regulated genes, dots in red represent up-regulated genes, and dots in black represent unchanged genes.

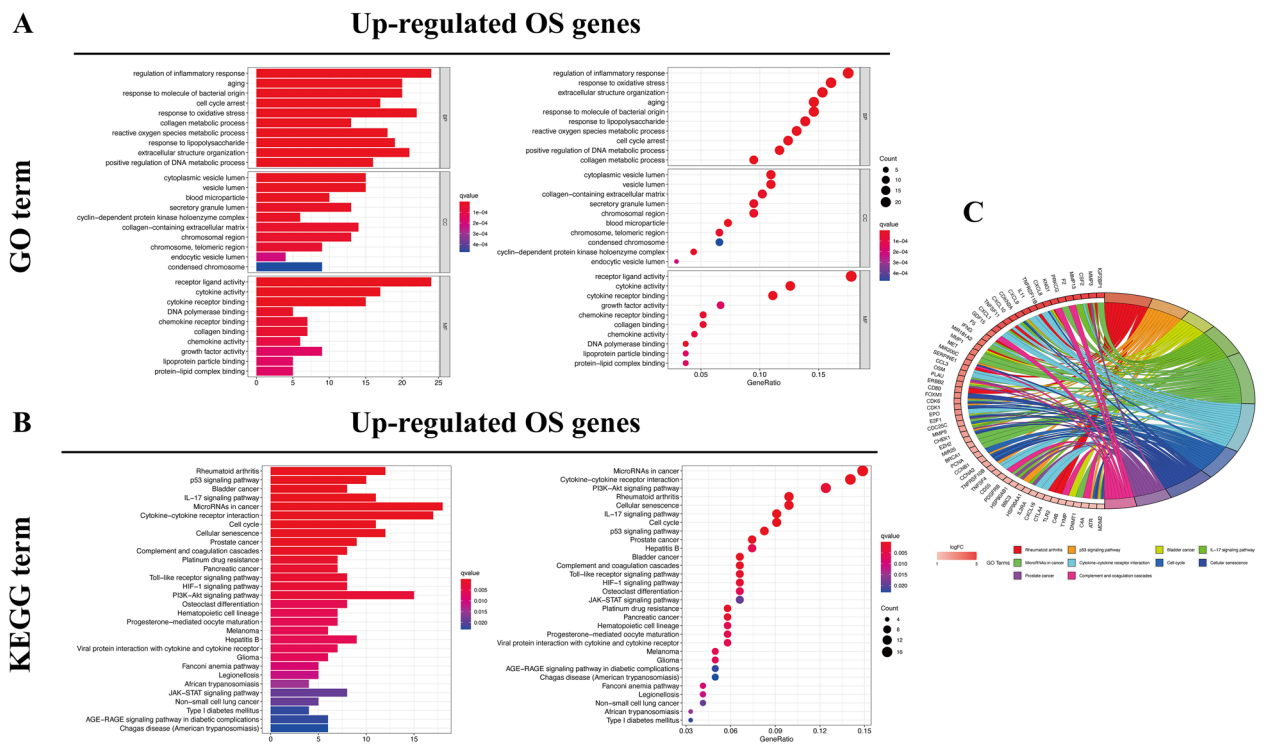


Figure 3. Functional enrichment analysis of up-regulated DEOSGs. (A) Top 10 classes of GO enrichment terms in biological process (BP), cellular component (CC), and molecular function (MF). (B) Top 30 classes of KEGG enrichment terms. (C) Circle diagram which enriched in KEGG analysis.

and were identified as DEOSGs. The expression distribution of DEOSGs in GC is displayed in Fig. 2A,B. To investigate the potential functional and molecular mechanisms of these identified DEOSGs in GC, GO and KEGG analyses were performed, and the pathways these genes enriched in are displayed in Figs. 3 and 4.

Prognosis-related DEOSG screening and construction of a genetic risk score model for patients with GC. Using Cytoscape software and data from the STRING database, a PPI network with 265 nodes and 2736 edges was constructed (Fig. 5A). The MODE plugin in Cytoscape software was also used to identify the

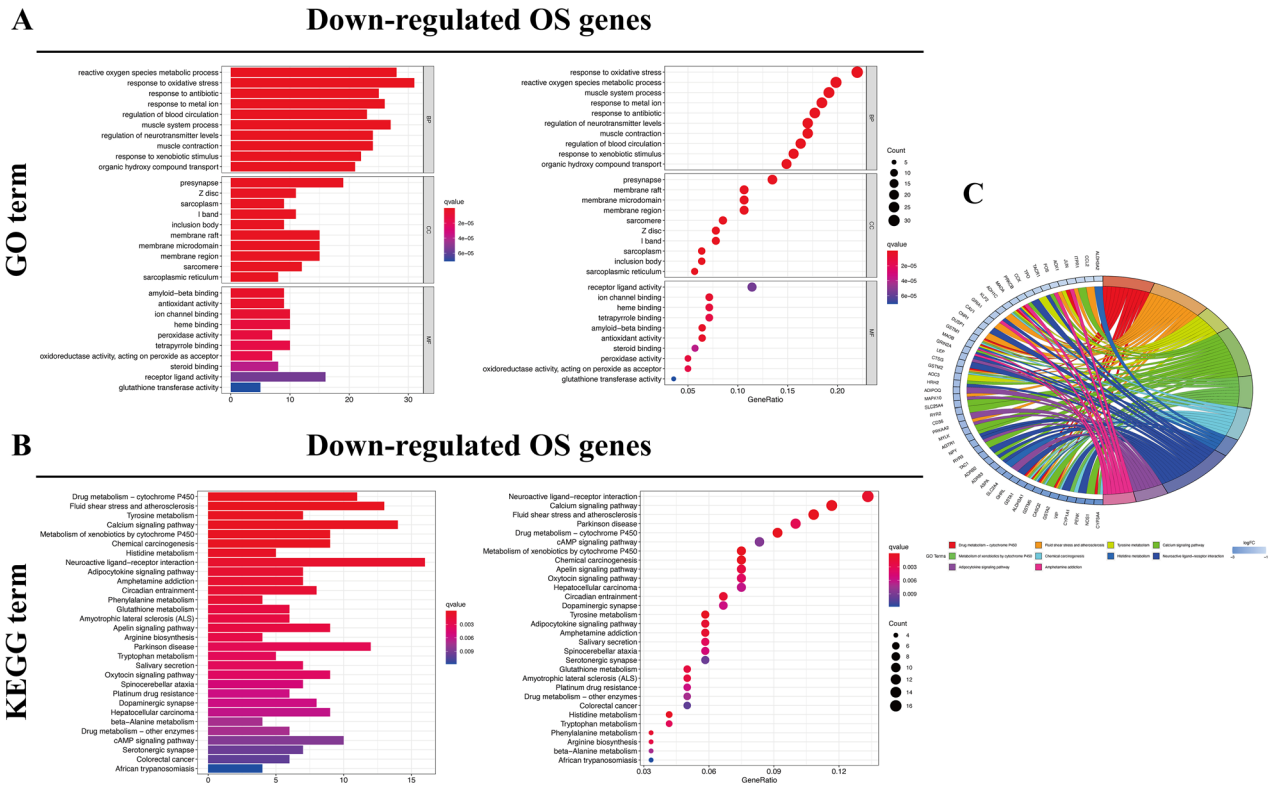


Figure 4. Functional enrichment analysis of down-regulated DEOSGs. **(A)** Top 10 classes of GO enrichment terms in biological process (BP), cellular component (CC), and molecular function (MF). **(B)** Top 30 classes of KEGG enrichment terms. **(C)** Circle diagram which enriched in KEGG analysis.

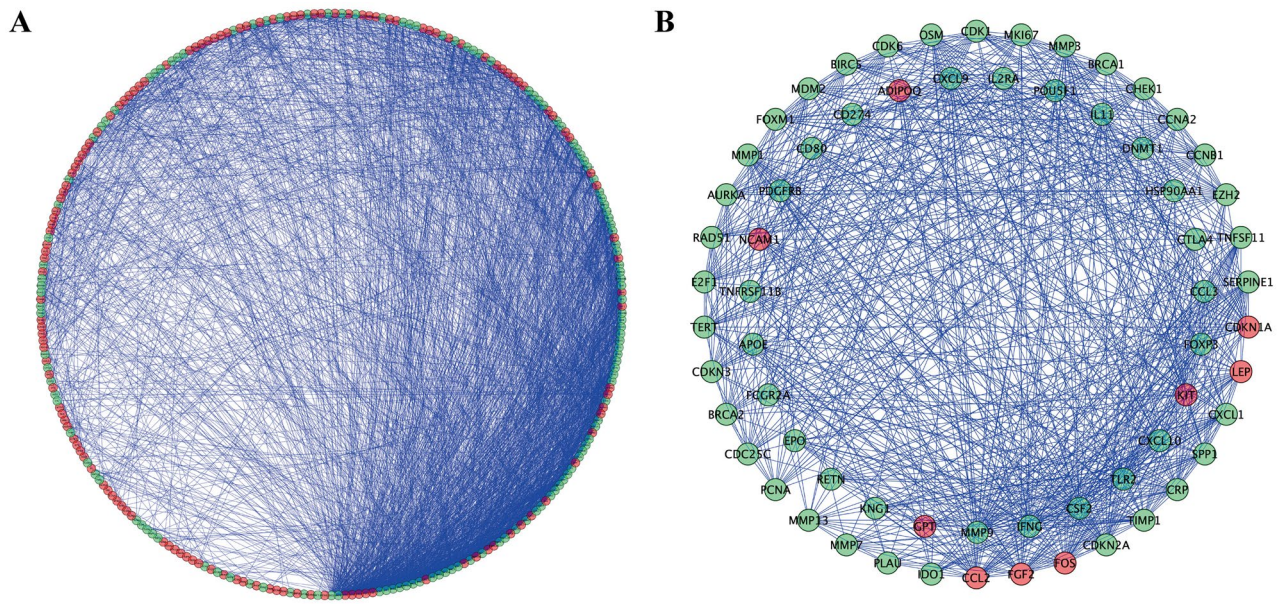


Figure 5. PPI network and modules screening. **(A)** PPI network of differentially expressed OS genes; **(B)** Top 2 critical modules from PPI network. Green circles represent down-regulated genes, and red circles represent up-regulated genes.

potential key modules in the network, and the top 2 significant modules were determined with 38 nodes and 329 edges, and 27 nodes and 158 edges, respectively (Fig. 5B).

To further identify prognostic-related DEOSGs, 265 differentially expressed genes (DEGs) were analyzed using univariate Cox regression analysis, and 24 DEOSGs were identified as GC prognostic-associated candidate OS genes with $P < 0.05$ (Fig. 6A). Subsequently, the LASSO algorithm was employed for specific OS gene range

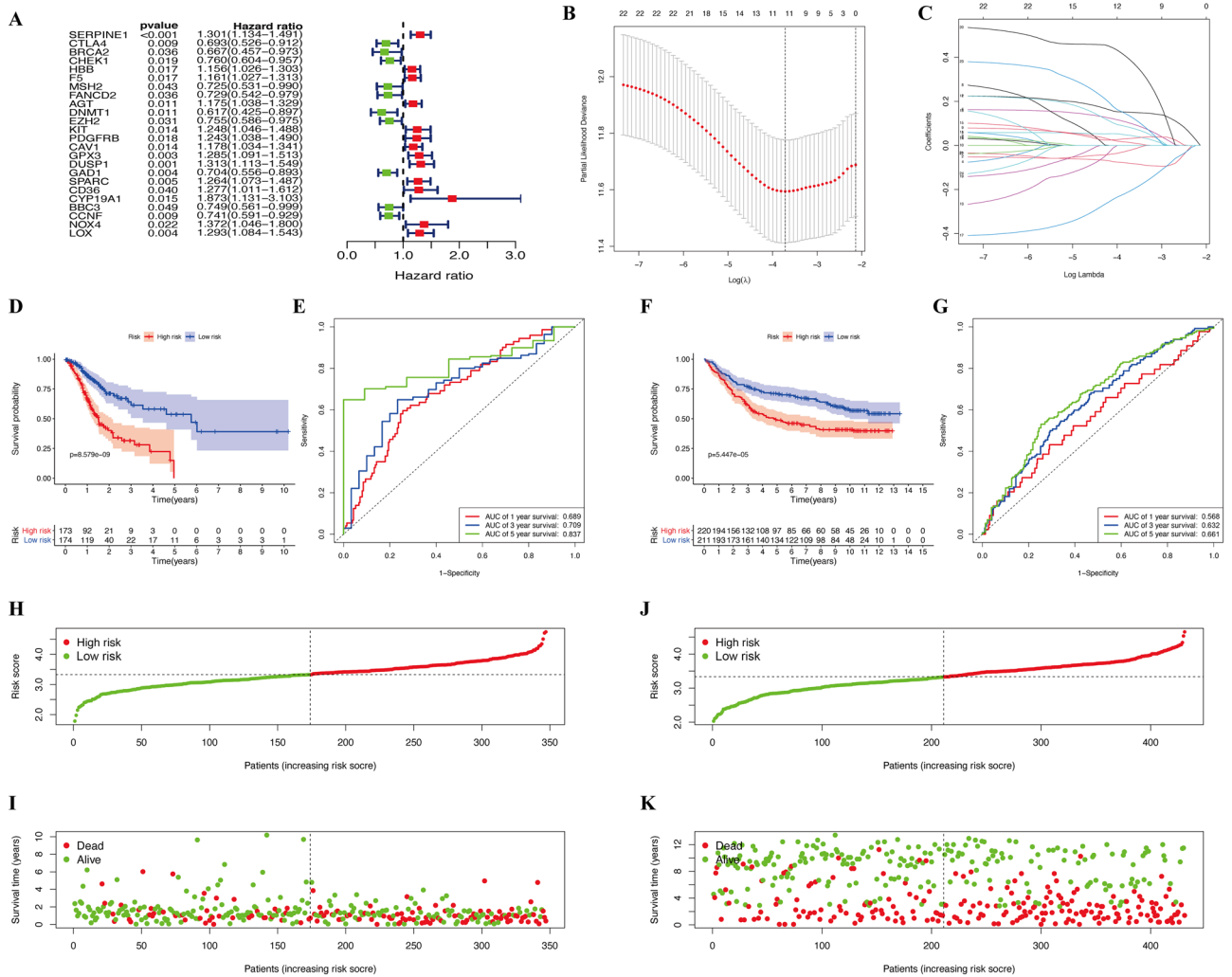


Figure 6. Construction of prognostic model in the TCGA and GSE84437 cohort. **(A)** Univariate Cox regression analysis for identification prognosis-associated OS genes. **(B–C)** LASSO analysis for determining the number of factors and constructing the prognosis prediction model. **(D)** Survival curve of TCGA cohort. **(E)** TimeROC curves for forecasting overall survival in TCGA cohort. **(F)** Survival curve of GSE cohort. **(G)** TimeROC curves for forecasting overall survival in GSE cohort. **(H–I)** Risk score distribution and survival status of TCGA cohort. **(J–K)** Risk score distribution and survival status of GSE cohort.

shrinkage (Fig. 6B,C), and 11 DEOSGs [serpin family E member 1 (*SERPINE1*), cytotoxic T-lymphocyte associated protein 4 (*CTLA4*), hemoglobin subunit beta (*HBB*), coagulation factor V (*F5*), angiotensinogen (*AGT*), proto-oncogene c-KIT (*KIT*), glutathione peroxidase 3 (*GPX3*), glutamate decarboxylase 1 (*GAD1*), cytochrome P450 family 19 subfamily A member 1 (*CYP19A1*), Bcl-2-binding component 3 (*BBC3*), and NADPH oxidase 4 (*NOX4*)] were selected to calculate the risk score; all patients with GC were separated into low- and high-risk groups according to the median risk score (Fig. 6H–K). The coefficients of the 11 DEOSGs are shown in Table 1. As indicated in Fig. 6D,F, the overall survival of patients with GC was significantly decreased with an increased risk score in both TCGA and GSE84437 databases. In addition, time-dependent ROC analysis indicated that the prediction model was quite credible, with the area under the ROC curve (AUC) reaching 0.837 at 5 years in TCGA database (Fig. 6E). Similar accuracy was also validated in the GSE84437 cohort, with the AUC reaching 0.661 at 5 years (Fig. 6G), which indicated that this prognostic model had moderate specificity and sensitivity. However, while GC patients were separated into the alive and dead subgroups, our increasing risk score no longer showed a clear correlation with GC patients’ survival time (Fig. 6I,K), which may suggest that our risk model can only predict the survival rate of the total population but cannot be used to predict the specific survival time of GC patients.

To determine whether the risk signature was an independent prognostic factor, univariate and multivariate Cox regression analyses were also performed. The risk score was an independent prognostic feature that was significantly connected with GC prognosis in both TCGA and GSE84437 databases (Fig. 7A–D). The ROC curve over 5 years showed that the prognostic model had a better predictive accuracy than other clinical features in the TCGA cohort (Fig. 7E). In the GSE84437 cohort, the prognostic model also showed better forecast performance than age, gender, and T stage (Fig. 7F). Moreover, the correlation between risk score and each clinicopathological

OS name	Univariate Cox regression analysis				LASSO coefficient
	HR	Lower 95% CI	Upper 95% CI	P value	
SERPINE1	1.3005	1.1340	1.4915	0.0002	0.1520
CTLA4	0.6929	0.5263	0.9121	0.0089	-0.0772
HBB	1.1564	1.0263	1.3029	0.0170	0.0506
F5	1.1611	1.0270	1.3126	0.0170	0.1119
AGT	1.1746	1.0383	1.3288	0.0106	0.0176
KIT	1.2478	1.0463	1.4880	0.0138	0.0907
GPX3	1.2850	1.0913	1.5131	0.0026	0.0283
GAD1	0.7044	0.5556	0.8931	0.0038	-0.2259
CYP19A1	1.8730	1.1306	3.1029	0.0148	0.4194
BBC3	0.7485	0.5610	0.9988	0.0491	-0.0313
NOX4	1.3723	1.0461	1.8001	0.0223	0.1537

Table 1. Eleven prognosis-associated OS genes with GC in the TCGA dataset were identified by LASSO analysis.

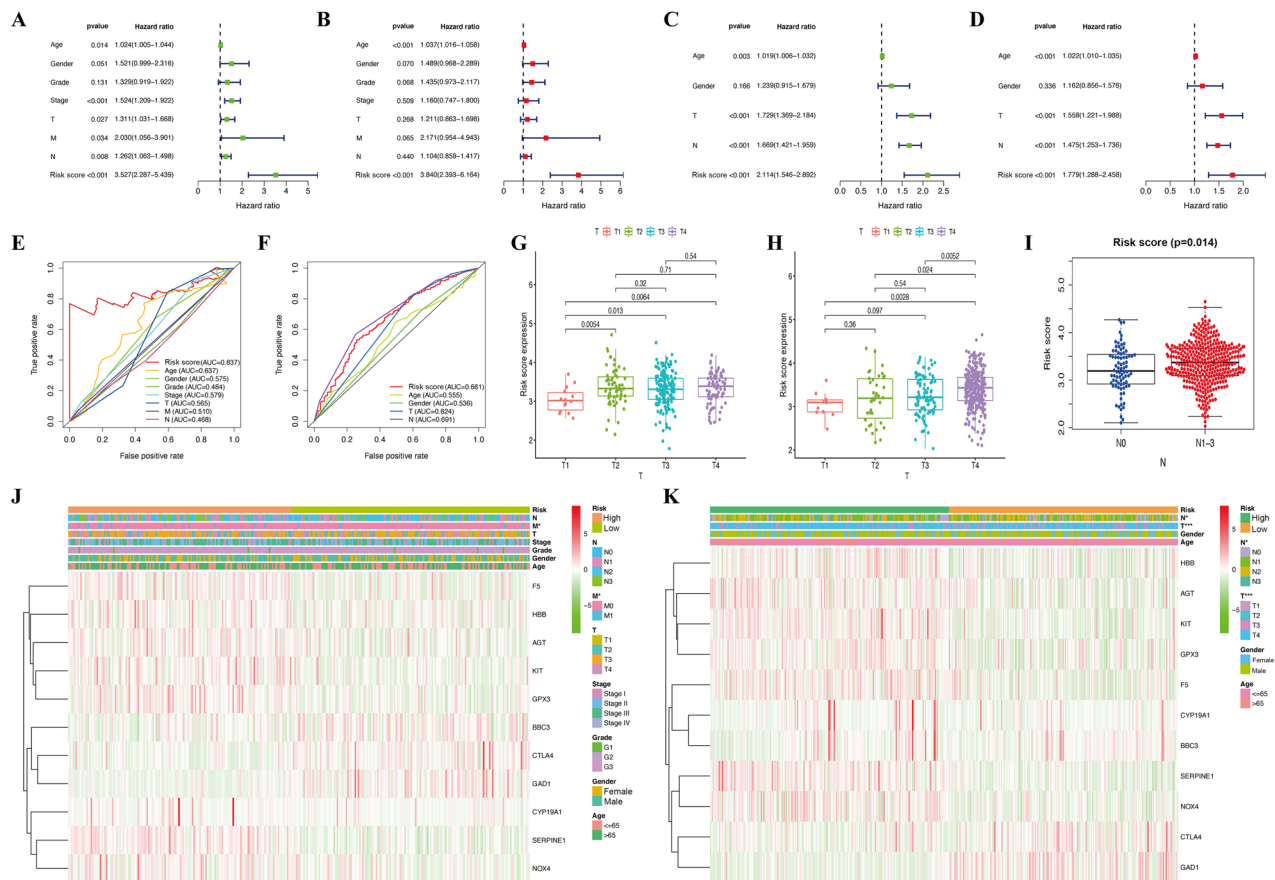


Figure 7. Efficacy evaluation of constructed prognostic model. Univariate (A) and multivariate (B) Cox regression analysis of the clinicopathological features in TCGA cohort. Univariate (C) and multivariate (D) Cox regression analysis of the clinicopathological features in GSE84437 cohort. ClinicalROC curves for forecasting overall survival in TCGA (E) and GSE84437 (F) cohort. (G) The relationship between the risk scores and T stage in TCGA cohort. The relationship between the risk scores and T stage (H) or N stage (I) in GSE84437 cohort. The heatmap shows the distribution of clinicopathological features and OS genes expression in TCGA (J) and GSE84437 (K) cohort. Columns in green represent down-regulated genes, columns in red represent up-regulated genes, and columns in white represent unchanged genes.

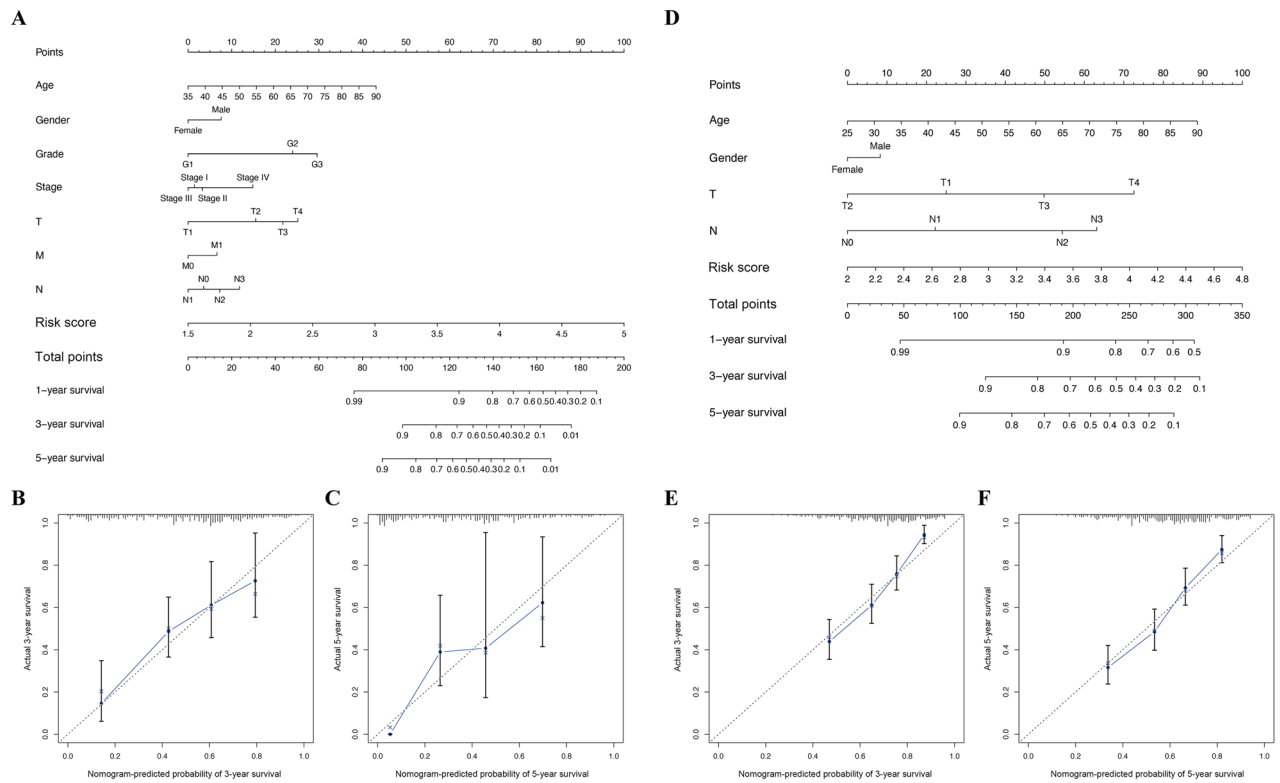


Figure 8. Nomogram plots construction. Nomogram of risk score and other clinical factors for predicting GC 1-, 3-, and 5-year overall survival in TCGA (A) and GSE84437 (D) cohort. (B–C) The calibration plot of the nomogram in TCGA cohort. (E–F) The calibration plot of the nomogram in GSE84437 cohort.

characteristic was evaluated; patients with GC in the T1 stage were significantly related to a lower risk score in the TCGA cohort (Fig. 7G). Meanwhile, the risk score was also associated with the different T and N stages of patients with GC in the GSE84437 cohort, and GC patients in the T4 or N1-3 stage were significantly related to a higher risk score (Fig. 7H–I). The heatmap revealed the expression of the 11 specific DEOSGs in each subgroup, and there were significant differences between the two risk subgroups with respect to the M stage in TCGA database (Fig. 7J). In the GSE84437 cohort, the expression trends of DEOSGs were quite similar to those in TCGA cohort, and there were significant differences between the two risk subgroups with respect to T and N stage (Fig. 7K).

A nomogram plot is another quantitative model to predict clinical outcomes of patients with GC. Thus, a nomogram plot was developed based on the risk score and other clinical characteristics, which allowed the calculation of the survival probabilities of each patient with GC at 1, 3, and 5 years (Fig. 8A,D). The calibration plots indicated good conformity between predicted and observed outcomes at 3 and 5 years in both TCGA and GSE84437 cohorts (Fig. 8B,C,E,F). These results indicated that the prognostic model showed great promise for predicting GC outcomes and clinical features.

Evaluation of the expression level of prognosis related DEOSGs in patients with GC. To further explore the transcriptional pattern of DEOSGs in patients with GC, the expression value of each key gene was extracted from TCGA database and a violin plot and heatmap was constructed. As shown in Supplement file 2A,B, the results indicated that *SERPINE1*, *CTLA4*, *F5*, *AGT*, *GAD1*, *CYP19A1*, *BBC3*, and *NOX4* were significantly overexpressed in GC samples, while the expression patterns of *HBB*, *KIT*, and *GPX3* were decreased compared to those in normal tissues. Similar results were obtained by analyzing the protein expression levels of the key DEOSGs in accordance with the immunohistochemistry results from the HPA database (Supplement file 2C).

Identification of hub DEOSGs for tumor grade by constructing a weighted co-expression network. Furthermore, WGCNA of 401 GC samples with complete clinical data from TCGA database was performed on 279 DEOSGs. Patients with GC with six types of clinical characteristics, including overall status, overall survival time, age, sex, tumor grade, and TNM stage were included for analysis, and probes with variances ranked in the top 25,000 were subjected to modules (Fig. 9A,B). To construct a scale-free network, the power of $\beta = 3$ (scale free $R^2 = 0.90$) was selected as the soft threshold (Fig. 9C); a total of six co-expressed modules were identified (Fig. 9D). Subsequently, to identify the module that was most related to GC progression, each module was assigned a different color. Among the modules, the brown module was specifically positively connected with tumor grade ($P < 0.05$), and genes in the turquoise module were negatively related to GC progression ($P < 0.05$, Fig. 9E). Thus, the data from these two modules were identified as candidate genes of interest in the training

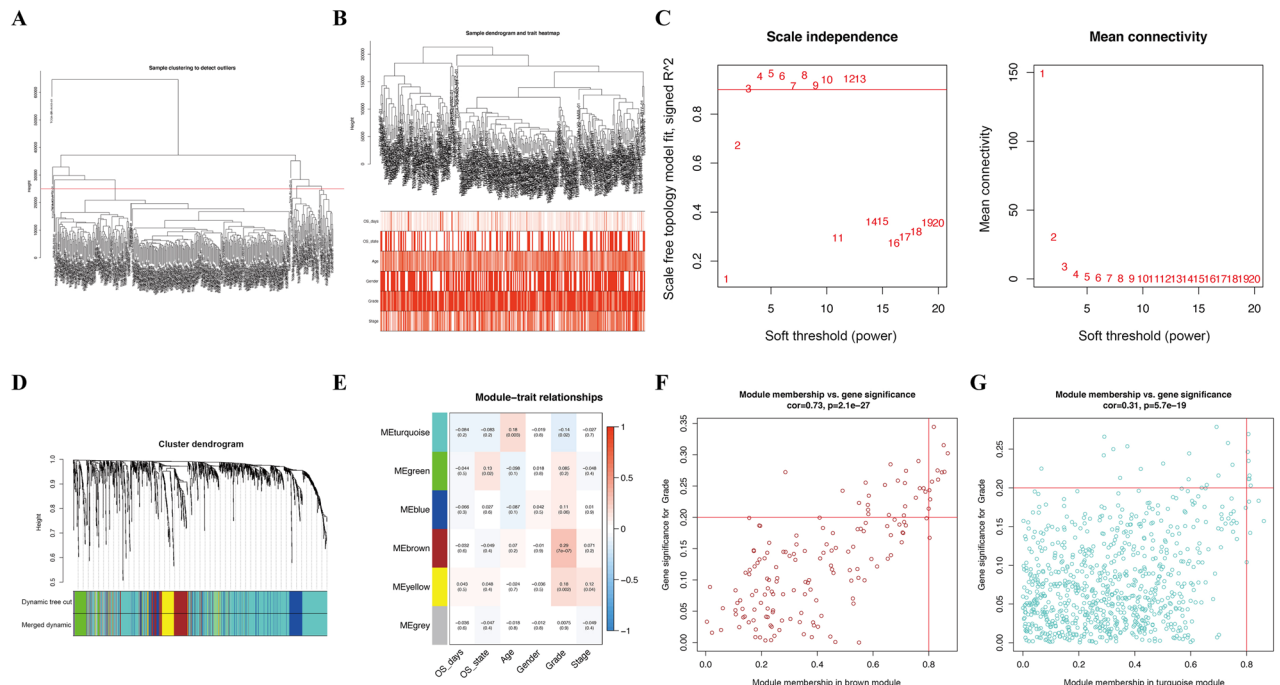


Figure 9. WGCNA analysis. (A) Samples clustering of OS genes from TCGA database to detect outliers. (B) Clustering dendrogram of GC samples and associated clinical traits. (C) The scale-free fit index for soft-thresholding powers. (D) A dendrogram of the differentially expressed genes clustered based on different metrics. (E) A heatmap showing the correlation between the gene module and clinical traits. Scatter plot of module eigengenes in brown (F) and turquoise (G) modules.

set. To further screen the most significant hub genes that were relevant to the tumor progression of GC, 17 genes with remarkable connectivity ($MM > 0.8$ and $GS > 0.2$) were identified in the brown and turquoise modules (Fig. 9F,G). Subsequently, the transcription of 17 candidate genes between GC samples and normal tissues were compared (Fig. 10A). Ultimately, 13 significant DEOSGs were identified as “real” GC progression-associated hub genes for further validation.

Validation of the relationship between progression annotated DEOSGs and clinicopathological features of GC. Based on TCGA database, the expression of the genes Fc fragment of IgG receptor IIIa (*FCGR3A*), C-C motif chemokine receptor 5 (*CCR5*), interleukin-2 receptor alpha (*IL2RA*), cluster of differentiation 86 (*CD86*), cytochrome B-245 beta chain (*CYBB*), and cluster of differentiation 4 (*CD4*) were significantly elevated, while the expression of parkin RBR E3 ubiquitin protein ligase (*PRKN*), inositol 1,4,5-trisphosphate receptor type 1 (*ITPR1*), potassium calcium-activated channel subfamily M alpha 1 (*KCNMA1*), amine oxidase copper containing 3 (*AOC3*), fibroblast growth factor 2 (*FGF2*), ankyrin-2 (*ANK2*), and mitogen-activated protein kinase 10 (*MAPK10*) was significantly decreased in GC tissues (Fig. 10B). Immunohistochemistry images obtained from the HPA database also demonstrated the same expression trend of the 13 hub genes (Fig. 10C). Kaplan–Meier analysis was implemented to determine the capacity of hub genes to predict GC prognosis; only *MAPK10* was significantly correlated with GC overall survival (Supplement file 3). However, the regression analysis indicated that all 13 hub genes were strongly associated with GC grade (Fig. 10D), which indicated that these DEOSGs were mainly involved in GC progression, but with poor prognostic ability. Similar results were also confirmed in the relationship between the expression of the 13 hub DEOSGs and clinicopathological characteristics of GC. As shown in Table 2, the 13 hub genes were all significantly connected with GC grade (Supplement file 4A), while all 12 DEOSGs, except *PRKN*, were significantly related to the T stage of patients with GC (Supplement file 4C). Furthermore, the expression of *FCGR3A*, *CCR5*, *FGF2*, *KCNMA1*, *AOC3*, *ANK2*, and *MAPK10* was also significantly associated with patients’ TNM stage; only the expression of *KCNMA1* was associated with M stage (Supplement file 4B,D). Furthermore, GO and KEGG analyses were also used to identify the potential mechanisms of the 13 real hub genes in GC progression. GO enrichment results indicated that these genes were mainly enriched in calcium ion transport into the cytosol, external side of the plasma membrane, and coreceptor activity (Fig. 11A). Additionally, KEGG pathway analysis indicated that the 13 hub genes were mostly enriched in Kaposi sarcoma-associated herpesvirus infection, human immunodeficiency virus 1 infection, and Th1, Th2, and Th17 cell differentiation (Fig. 11B).

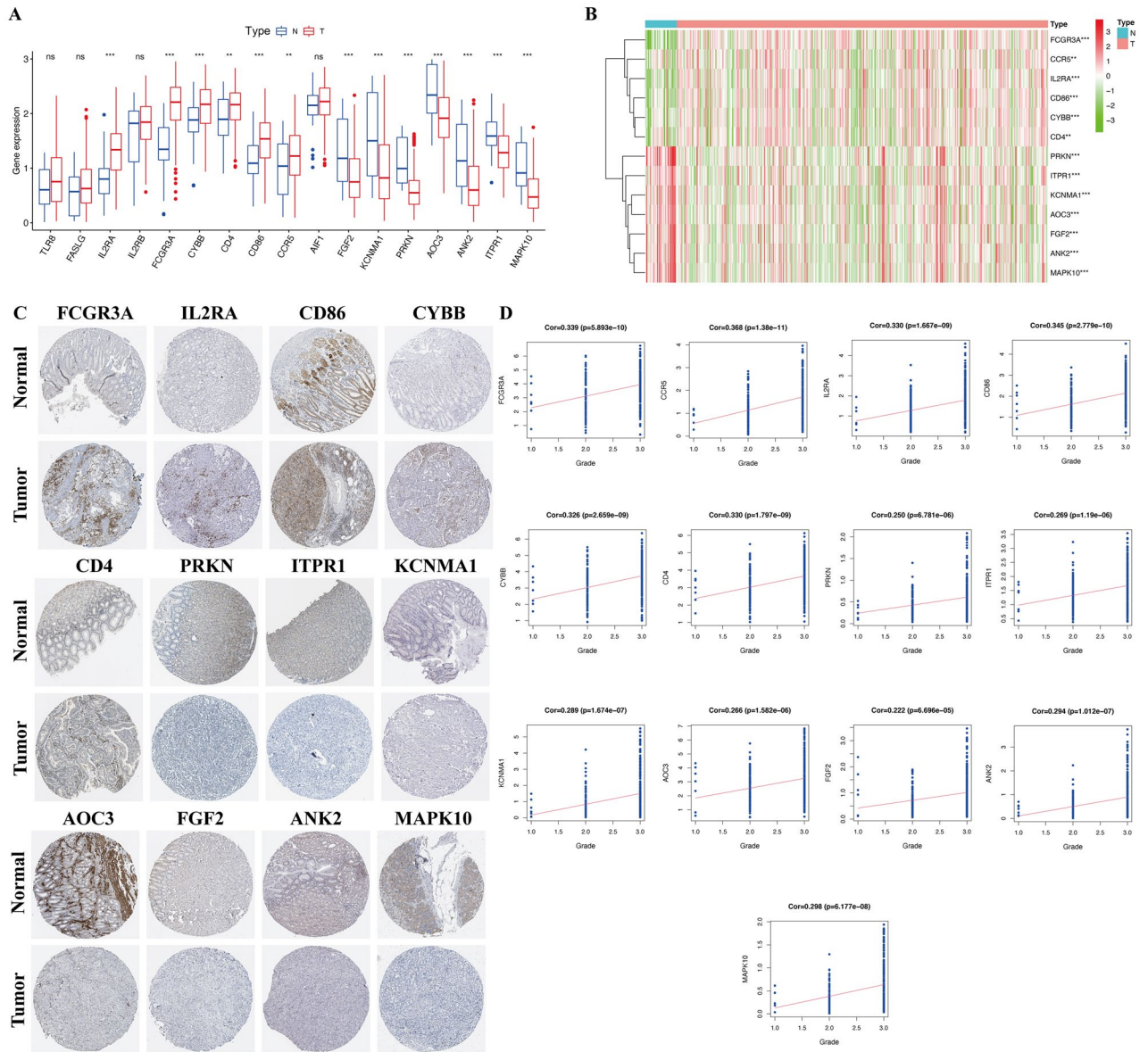


Figure 10. Connections between the expression of progression-associated OS genes and GC grade. (A) The mRNA expression pattern of progression-associated OS genes in TCGA cohort. (B) The heatmap reveals the transcription expression of progression-associated OS genes in TCGA cohort. (C) HPA database verifies the protein expression of hub progression-associated OS genes. (D) Correlation analysis between 13 progression-associated OS genes expression and tumor grade.

Clinical variables	P-value												
	IL2RA	FCGR3A	CYBB	CD4	CD86	CCR5	FGF2	KCNMA1	PRKN	AOC3	ANK2	ITPR1	MAPK10
Age	0.341	0.095	0.657	0.412	0.690	0.547	0.048	0.023	0.008	0.015	0.008	0.015	0.026
Gender	0.811	0.390	0.393	0.588	0.312	0.158	0.586	0.546	0.607	0.630	0.464	0.931	0.377
Tumor grade	2.115e-08	1.144e-08	2.426e-08	6.316e-09	2.159e-09	1.495e-10	6.117e-04	1.139e-08	3.142e-04	1.339e-05	7.204e-06	2.237e-05	5.315e-06
TNM stage	0.143	0.030	0.145	0.059	0.056	0.015	0.028	0.004	0.437	0.001	0.004	0.087	0.034
T stage	0.039	1.999e-04	0.005	0.010	0.002	0.003	7.24e-04	3.55e-04	0.529	2.205e-04	8.653e-05	0.027	0.019
M stage	0.411	0.586	0.818	0.620	0.201	0.351	0.500	0.020	0.105	0.456	0.316	0.873	0.770
N stage	0.154	0.277	0.258	0.258	0.181	0.326	0.367	0.946	0.267	0.982	0.406	0.101	0.311

Table 2. Correction between 13 hub progression-associated OS genes expression and clinicopathological characteristics of GC.

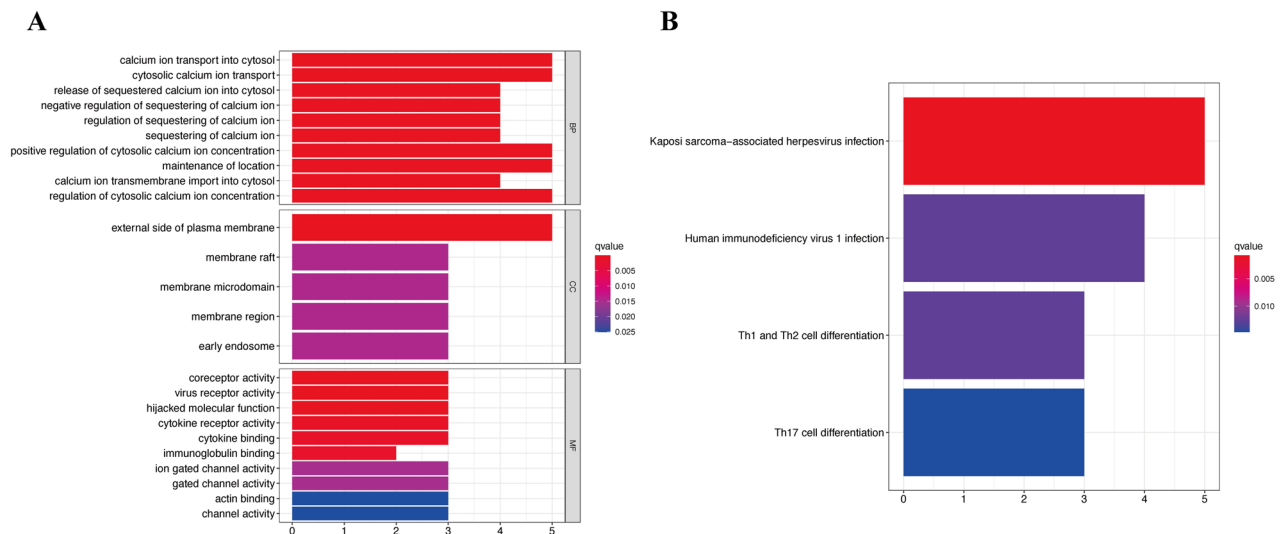


Figure 11. Function analysis of progression-associated OS genes. **(A)** Top 10 classes of GO enrichment terms about 13 hub genes in biological process (BP), cellular component (CC), and molecular function (MF). **(B)** KEGG enrichment terms about 13 hub progression-associated OS genes.

Discussion

As a worldwide malignant tumor, GC is reported to have heterogeneous characteristics in the digestive system³⁹. Although many novel diagnostic techniques and surgical skills have been developed in recent years, they are not always sufficient. Therefore, it is imperative to identify GC prognosis-related molecules and determine the mechanism of tumor progression. In the present study, 279 DEOSGs were identified based on TCGA database, and relevant biological pathways and PPI networks were systematically identified for these genes. Pathway enrichment analysis revealed that the DEGs were significantly correlated to the progression of several types of tumors, such as melanoma, glioma, and bladder, prostate, and pancreatic cancers. Furthermore, these DEOSGs were significantly enriched in several biological processes, including inflammatory response, reactive oxygen species metabolic process, and response to oxidative stress. All of these biological processes have been reported to be significantly correlated with tumorigenesis and progression^{40–42}. Thus, these DEGs were significantly associated with GC progression and critical for the comprehensive evaluation of the mechanism of these DEOSGs.

In addition, the prognosis related DEOSGs were screened by univariate Cox and LASSO regression analysis, and a total of 11 DEOSGs: *SERPINE1*, *CTLA4*, *HBB*, *F5*, *AGT*, *KIT*, *GPX3*, *GAD1*, *CYP19A1*, *BBC3*, and *NOX4* were identified with a good prognosis in patients with GC. The expression patterns of these 11 DEOSGs on mRNA and protein levels using TCGA expression data and HPA database revealed that *SERPINE1*, *CTLA4*, *F5*, *AGT*, *GAD1*, *CYP19A1*, *BBC3*, and *NOX4* were overexpressed, while *HBB*, *KIT*, and *GPX3* were downregulated in GC tissues. These findings are mostly consistent those of previous studies showing that the expression of *SERPINE1* and *CTLA4* is elevated in gastric adenocarcinoma, and as a tumor carcinogenic gene, the overexpression of *SERPINE1* is significantly associated with GC aggressiveness and inferior overall survival^{43,44}. In addition, *GAD1* is overexpressed in lung adenocarcinoma and plays a virtual role in tumor progression⁴⁵; however, its role in GC outcome is unclear.

To further identify whether these specific DEOSGs could be used as prognostic factors, a novel prognostic prediction model was constructed based on these 11 hub genes. To our knowledge, this is the first OS-associated risk model for prognostication. Univariate and multivariate Cox regression analyses showed that the risk model was an independent prognostic factor with a robust prognostic value for GC. In addition, the survival analyses and ROC analyses also confirmed the major advantage of its biological implications for predicting GC prognosis. A similar scenario was also observed in the nomogram analysis that risk signature played a virtual role in predicting the overall survival of patients with GC, and its inspection efficiency was much better than that of other clinicopathological features. These explorations, for the first time, demonstrate the prognostic value of an OS gene-dependent risk model for patients with GC and provide a novel direction for further research.

To a great extent, OS plays a critical role in various stages of carcinogenesis and cancer progression^{46,47}. In recent years, many GC studies have identified that OS may be closely associated with progression owing to resulting DNA damage⁴⁸. However, in our previous study, the risk model, which was developed using the hub prognosis associated DEOSGs, was only associated with TNM stage of patients with GC, but there was no significant correlation with tumor grade. Therefore, in follow-up studies, we defined DEOSGs that were closely related to tumor progression. WGCNA is a widely used approach to identify potential biomarkers of interest^{49,50}. In the present study, 13 real hub genes (*FCGR3A*, *CCR5*, *IL2RA*, *CD86*, *CYBB*, *CD4*, *PRKN*, *ITPR1*, *KCNMA1*, *AOC3*, *FGF2*, *ANK2*, and *MAPK10*) tightly associated with GC progression were identified, and a series of bioinformatic analyses showed that these genes were both highly correlated with GC grade and may be potential biomarkers for predicting tumor stage. Moreover, considering the critical role of tumor progression in the overall survival of patients with GC, the prognostic value of these hub genes was assessed; *MAPK10* was significantly correlated with

patients' overall survival. Compared to previous studies, Gu's and Ying's group, who focused on the prognostic value of *MAPK10*, also confirm our conclusion that *MAPK10* is frequently downregulated in GC cell lines⁵¹ and predicts tumor progression and prognosis⁵². The functional and pathway enrichment analysis showed that these genes were mainly enriched in terms that were related to malignancy progression, including calcium ion transport into cytosol⁵³, coreceptor activity⁵⁴, and Kaposi sarcoma-associated herpesvirus infection⁵⁵, which may provide a perspective for exploring the role of prognosis-related DEOSGs in GC.

Nonetheless, this study has limitations. First, this study was designed as a retrospective analysis; therefore, more prospective studies should be performed to verify these results. Second, the results lack in vitro or in vivo exploration to confirm the reliability of the mechanistic analysis. Therefore, in the future, a number of experiments will be conducted to demonstrate the mechanistic connections between these genes and GC progression.

In conclusion, after a series of bioinformatic analyses and verifications, 11 prognosis-associated DEOSGs and 13 progression-associated DEOSGs were identified, which were related to the overall survival or tumor grade of patients with GC. A prognostic model was also constructed with powerful predictive effects. As far as we know, this is the first report of the construction of an OS-associated prognostic model for malignancies. This study provides novel research targets for studying the pathogenesis and progression of patients with GC.

Data availability

The data used to support the findings of this study are available from the corresponding author upon request.

Received: 6 October 2020; Accepted: 27 January 2021

Published online: 08 February 2021

References

1. Siegel, R. L., Miller, K. D. & Jemal, A. Cancer statistics, 2019. *CA Cancer J. Clin.* **69**, 7–34. <https://doi.org/10.3322/caac.21551> (2019).
2. Bray, F. *et al.* Global cancer statistics 2018: GLOBOCAN estimates of incidence and mortality worldwide for 36 cancers in 185 countries. *CA Cancer J. Clin.* **68**, 394–424. <https://doi.org/10.3322/caac.21492> (2018).
3. Necula, L. *et al.* Recent advances in gastric cancer early diagnosis. *World J. Gastroenterol.* **25**, 2029–2044. <https://doi.org/10.3748/wjg.v25.i17.2029> (2019).
4. Choi, Y. J. & Kim, N. Gastric cancer and family history. *Korean J. Intern. Med.* **31**, 1042–1053. <https://doi.org/10.3904/kjim.2016.147> (2016).
5. Hironaka, S. Anti-angiogenic therapies for gastric cancer. *Asia Pac. J. Clin. Oncol.* **15**, 208–217. <https://doi.org/10.1111/ajco.13174> (2019).
6. Niccolai, E., Taddei, A., Prisco, D. & Amedei, A. Gastric cancer and the epoch of immunotherapy approaches. *World J. Gastroenterol.* **21**, 5778–5793. <https://doi.org/10.3748/wjg.v21.i19.5778> (2015).
7. Zhou, F. *et al.* Jab1/Csn5-thioredoxin signaling in relapsed acute monocytic leukemia under oxidative stress. *Clin. Cancer Res.* **23**, 4450–4461. <https://doi.org/10.1158/1078-0432.Ccr-16-2426> (2017).
8. Brown, J. M. & Wilson, W. R. Exploiting tumour hypoxia in cancer treatment. *Nat. Rev. Cancer* **4**, 437–447. <https://doi.org/10.1038/nrc1367> (2004).
9. Kangari, P. *et al.* Enzymatic antioxidant and lipid peroxidation evaluation in the newly diagnosed breast cancer patients in Iran. *Asian Pac. J. Cancer Prev.* **19**, 3511–3515. <https://doi.org/10.31557/apjcp.2018.19.12.3511> (2018).
10. Lü, J. M., Lin, P. H., Yao, Q. & Chen, C. Chemical and molecular mechanisms of antioxidants: Experimental approaches and model systems. *J. Cell. Mol. Med.* **14**, 840–860. <https://doi.org/10.1111/j.1582-4934.2009.00897.x> (2010).
11. Wei, Y. C. *et al.* Oxidative stress in depressive patients with gastric adenocarcinoma. *Int. J. Neuropsychopharmacol.* **12**, 1089–1096. <https://doi.org/10.1017/s1461145709000091> (2009).
12. Wang, J. Y. *et al.* Skp2-mediated stabilization of MTH1 promotes survival of melanoma cells upon oxidative stress. *Cancer Res.* **77**, 6226–6239. <https://doi.org/10.1158/0008-5472.Can-17-1965> (2017).
13. Zhou, F., Shen, Q. & Claret, F. X. Novel roles of reactive oxygen species in the pathogenesis of acute myeloid leukemia. *J. Leukoc. Biol.* **94**, 423–429. <https://doi.org/10.1189/jlb.0113006> (2013).
14. Oates, J. C. & Gilkeson, G. S. The biology of nitric oxide and other reactive intermediates in systemic lupus erythematosus. *Clin. Immunol.* **121**, 243–250. <https://doi.org/10.1016/j.clim.2006.06.001> (2006).
15. Smith, J., Tho, L. M., Xu, N. & Gillespie, D. A. The ATM-Chk2 and ATR-Chk1 pathways in DNA damage signaling and cancer. *Adv. Cancer Res.* **108**, 73–112. <https://doi.org/10.1016/b978-0-12-380888-2.00003-0> (2010).
16. Yoon, J. H. *et al.* Gastrophilin 1 inhibits the carcinogenic potentials of *Helicobacter pylori* CagA. *Carcinogenesis* **35**, 2619–2629. <https://doi.org/10.1093/carcin/bgu199> (2014).
17. Park, S. *et al.* Amelioration of oxidative stress with ensuing inflammation contributes to chemoprevention of *H. pylori*-associated gastric carcinogenesis. *Antioxid. Redox Signal.* **6**, 549–560. <https://doi.org/10.1089/152308604773934305> (2004).
18. Jia, X. *et al.* Malignant transformation of human gastric epithelium cells via reactive oxygen species production and Wnt/ β -catenin pathway activation following 40-week exposure to ochratoxin A. *Cancer Lett.* **372**, 36–47. <https://doi.org/10.1016/j.canlet.2015.12.007> (2016).
19. Zhao, Z., Han, F., Yang, S., Wu, J. & Zhan, W. Oxamate-mediated inhibition of lactate dehydrogenase induces protective autophagy in gastric cancer cells: Involvement of the Akt-mTOR signaling pathway. *Cancer Lett.* **358**, 17–26. <https://doi.org/10.1016/j.canlet.2014.11.046> (2015).
20. Buti, L. *et al.* *Helicobacter pylori* cytotoxin-associated gene A (CagA) subverts the apoptosis-stimulating protein of p53 (ASPP2) tumor suppressor pathway of the host. *Proc. Natl. Acad. Sci. USA* **108**, 9238–9243. <https://doi.org/10.1073/pnas.1106200108> (2011).
21. Xia, L. *et al.* ANLN functions as a key candidate gene in cervical cancer as determined by integrated bioinformatic analysis. *Cancer Manag. Res.* **10**, 663–670. <https://doi.org/10.2147/cmar.S162813> (2018).
22. Yuan, L. *et al.* Identification of key genes and pathways in human clear cell renal cell carcinoma (ccRCC) by co-expression analysis. *Int. J. Biol. Sci.* **14**, 266–279. <https://doi.org/10.7150/ijbs.23574> (2018).
23. Wang, F. *et al.* Strong correlation between ASPM gene expression and HCV cirrhosis progression identified by co-expression analysis. *Dig. Liver Dis.* **49**, 70–76. <https://doi.org/10.1016/j.dld.2016.10.017> (2017).
24. Zhou, Z. *et al.* Overexpression of topoisomerase 2- α confers a poor prognosis in pancreatic adenocarcinoma identified by co-expression analysis. *Dig. Dis. Sci.* **62**, 2790–2800. <https://doi.org/10.1007/s10620-017-4718-4> (2017).
25. Li, W., Gao, L. N., Song, P. P. & You, C. G. Development and validation of a RNA binding protein-associated prognostic model for lung adenocarcinoma. *Aging (Albany NY)* **12**, 3558–3573. <https://doi.org/10.18632/aging.102828> (2020).

26. Kanehisa, M. & Goto, S. KEGG: Kyoto encyclopedia of genes and genomes. *Nucleic Acids Res.* **28**, 27–30. <https://doi.org/10.1093/nar/28.1.27> (2000).
27. Kanehisa, M. Toward understanding the origin and evolution of cellular organisms. *Protein Sci.* **28**, 1947–1951. <https://doi.org/10.1002/pro.3715> (2019).
28. Kanehisa, M., Furumichi, M., Sato, Y., Ishiguro-Watanabe, M. & Tanabe, M. KEGG: Integrating viruses and cellular organisms. *Nucleic Acids Res.* <https://doi.org/10.1093/nar/gkaa970> (2020).
29. da Huang, W., Sherman, B. T. & Lempicki, R. A. Systematic and integrative analysis of large gene lists using DAVID bioinformatics resources. *Nat. Protoc.* **4**, 44–57. <https://doi.org/10.1038/nprot.2008.211> (2009).
30. Szklarczyk, D. *et al.* STRING v11: Protein–protein association networks with increased coverage, supporting functional discovery in genome-wide experimental datasets. *Nucleic Acids Res.* **47**, D607–d613. <https://doi.org/10.1093/nar/gky1131> (2019).
31. Bader, G. D. & Hogue, C. W. An automated method for finding molecular complexes in large protein interaction networks. *BMC Bioinform.* **4**, 2. <https://doi.org/10.1186/1471-2105-4-2> (2003).
32. Jiang, Y. *et al.* ImmunoScore signature: A prognostic and predictive tool in gastric cancer. *Ann. Surg.* **267**, 504–513. <https://doi.org/10.1097/sla.0000000000002116> (2018).
33. Heagerty, P. J. & Zheng, Y. Survival model predictive accuracy and ROC curves. *Biometrics* **61**, 92–105. <https://doi.org/10.1111/j.0006-341X.2005.030814.x> (2005).
34. Gu, H. Y. *et al.* Risk score based on expression of five novel genes predicts survival in soft tissue sarcoma. *Aging (Albany NY)* **12**, 3807–3827. <https://doi.org/10.18632/aging.102847> (2020).
35. Thul, P. J. *et al.* A subcellular map of the human proteome. *Science* <https://doi.org/10.1126/science.aal3321> (2017).
36. Mason, M. J., Fan, G., Plath, K., Zhou, Q. & Horvath, S. Signed weighted gene co-expression network analysis of transcriptional regulation in murine embryonic stem cells. *BMC Genom.* **10**, 327. <https://doi.org/10.1186/1471-2164-10-327> (2009).
37. Horvath, S. & Dong, J. Geometric interpretation of gene coexpression network analysis. *PLoS Comput. Biol.* **4**, e1000117. <https://doi.org/10.1371/journal.pcbi.1000117> (2008).
38. Albert, R. Scale-free networks in cell biology. *J. Cell. Sci.* **118**, 4947–4957. <https://doi.org/10.1242/jcs.02714> (2005).
39. Chi, Y., Wang, H., Wang, F. & Ding, M. PHTF2 regulates lipids metabolism in gastric cancer. *Aging (Albany NY)* **12**, 6600–6610. <https://doi.org/10.18632/aging.102995> (2020).
40. Lin, Y., Jiang, M., Chen, W., Zhao, T. & Wei, Y. Cancer and ER stress: Mutual crosstalk between autophagy, oxidative stress and inflammatory response. *Biomed. Pharmacother.* **118**, 109249. <https://doi.org/10.1016/j.biopha.2019.109249> (2019).
41. Landskron, G., De la Fuente, M., Thuwajit, P., Thuwajit, C. & Hermoso, M. A. Chronic inflammation and cytokines in the tumor microenvironment. *J. Immunol. Res.* **2014**, 149185. <https://doi.org/10.1155/2014/149185> (2014).
42. Bhattacharyya, A., Chattopadhyay, R., Mitra, S. & Crowe, S. E. Oxidative stress: An essential factor in the pathogenesis of gastrointestinal mucosal diseases. *Physiol. Rev.* **94**, 329–354. <https://doi.org/10.1152/physrev.00040.2012> (2014).
43. Yang, J. D., Ma, L. & Zhu, Z. SERPINE1 as a cancer-promoting gene in gastric adenocarcinoma: Facilitates tumour cell proliferation, migration, and invasion by regulating EMT. *J. Chemother.* **31**, 408–418. <https://doi.org/10.1080/1120009x.2019.1687996> (2019).
44. Schlößer, H. A. *et al.* Immune checkpoints programmed death 1 ligand 1 and cytotoxic T lymphocyte associated molecule 4 in gastric adenocarcinoma. *Oncoimmunology* **5**, e1100789. <https://doi.org/10.1080/2162402x.2015.1100789> (2016).
45. Tsuboi, M. *et al.* Prognostic significance of GAD1 overexpression in patients with resected lung adenocarcinoma. *Cancer Med.* **8**, 4189–4199. <https://doi.org/10.1002/cam4.2345> (2019).
46. Reuter, S., Gupta, S. C., Chaturvedi, M. M. & Aggarwal, B. B. Oxidative stress, inflammation, and cancer: How are they linked?. *Free Radic. Biol. Med.* **49**, 1603–1616. <https://doi.org/10.1016/j.freeradbiomed.2010.09.006> (2010).
47. Hecht, F. *et al.* The role of oxidative stress on breast cancer development and therapy. *Tumour Biol.* **37**, 4281–4291. <https://doi.org/10.1007/s13277-016-4873-9> (2016).
48. Butcher, L. D., den Hartog, G., Ernst, P. B. & Crowe, S. E. Oxidative stress resulting from helicobacter pylori infection contributes to gastric carcinogenesis. *Cell. Mol. Gastroenterol. Hepatol.* **3**, 316–322. <https://doi.org/10.1016/j.jcmgh.2017.02.002> (2017).
49. Wan, Q., Tang, J., Han, Y. & Wang, D. Co-expression modules construction by WGCNA and identify potential prognostic markers of uveal melanoma. *Exp. Eye Res.* **166**, 13–20. <https://doi.org/10.1016/j.exer.2017.10.007> (2018).
50. Yin, L., Cai, Z., Zhu, B. & Xu, C. Identification of key pathways and genes in the dynamic progression of HCC based on WGCNA. *Genes (Basel)* <https://doi.org/10.3390/genes9020092> (2018).
51. Ying, J. *et al.* Epigenetic disruption of two proapoptotic genes MAPK10/JNK3 and PTPN13/FAP-1 in multiple lymphomas and carcinomas through hypermethylation of a common bidirectional promoter. *Leukemia* **20**, 1173–1175. <https://doi.org/10.1038/sj.leu.2404193> (2006).
52. Li, L. & Luo, Z. Dysregulated miR-27a-3p promotes nasopharyngeal carcinoma cell proliferation and migration by targeting Mapk10. *Oncol Rep* **37**, 2679–2687. <https://doi.org/10.3892/or.2017.5544> (2017).
53. Papp, B. *et al.* Endoplasmic reticulum calcium pumps and tumor cell differentiation. *Int. J. Mol. Sci.* <https://doi.org/10.3390/ijms21093351> (2020).
54. Howlett, M. *et al.* Differential regulation of gastric tumor growth by cytokines that signal exclusively through the coreceptor gp130. *Gastroenterology* **129**, 1005–1018. <https://doi.org/10.1053/j.gastro.2005.06.068> (2005).
55. Ganem, D. KSHV and the pathogenesis of Kaposi sarcoma: Listening to human biology and medicine. *J. Clin. Invest.* **120**, 939–949. <https://doi.org/10.1172/jci40567> (2010).

Acknowledgements

This work has been financially supported by Innovation Project of Guangxi Graduate Education (Grant No. YCSW2019117).

Author contributions

Zhengyuan Wu and Jun Yao conceived and designed the study. Zhenpei Wen and Lin Wang analyzed data. Zhengyuan Wu wrote this manuscript. All authors reviewed and approved the final manuscript.

Competing interests

The authors declare no competing interests.

Additional information

Supplementary Information The online version contains supplementary material available at (<https://doi.org/10.1038/s41598-021-82976-w>).

Correspondence and requests for materials should be addressed to J.Y.

Reprints and permissions information is available at www.nature.com/reprints.

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.



Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.

© The Author(s) 2021