Identification and Validation of the Hub Genes in Gastric Cancer Through Weighted Gene Co-Expression Network Analysis

Chunyang Li
Sichuan University
Haopeng Yu
Sichuan University
Yajing Sun
Sichuan University
Xiaoxi Zeng (zengxiaoxi@wchscu.cn)
Wei Zhang
Sichuan University

Research

Keywords: gastric cancer, weighted gene co-expression network analysis, WGCNA, LASSO regression, supervised machine learning

Posted Date: June 15th, 2020

DOI: https://doi.org/10.21203/rs.3.rs-34166/v1

Version of Record: A version of this preprint was published at PeerJ on March 5th, 2021. See the published version at https://doi.org/10.7717/peerj.10682.

Abstract

Background: Gastric cancer is one of the most lethal tumors and is characterized by poor prognosis and lack of effective diagnostic or therapeutic biomarkers. The aim of this study was to find hub genes serving as biomarkers in gastric cancer diagnosis and therapy.

Methods: GSE66229 from Gene Expression Omnibus (GEO) was used as training set. Genes bearing the top 25% standard deviations among all the samples in training set were performed to systematic weighted gene co-expression network analysis (WGCNA) to find candidate genes. Then, hub genes were further screened by using the “least absolute shrinkage and selection operator” (LASSO) logistic regression. Finally, hub genes were validated in GS54129 dataset from GEO by supervised learning methods logistic regression algorithms.

Results: 12 modules with strong preservation were identified by using WGCNA methods in training set. Of which, two modules significantly related to gastric cancer were selected as clinically significant modules, and 43 candidate genes were identified from these two modules. Then, *ACADL*, *ADIPOQ*, *ARHGAP39*, *ATAD3A*, *C1orf95*, *CCKBR*, *GRIK3*, *SCNN1G*, *SIGLEC11*, and *TXLNB* were screened as the hub genes. These hub genes successfully differentiated the tumor samples from the healthy tissues in an independent testing set through the logistic regression algorithm, with the area under the receiver operating characteristic curve at 0.882.

Conclusions: These hub genes bearing diagnostic and therapeutic values, and our results may provide a novel prospect for the diagnosis and treatment of gastric cancer in the future.

Background

Gastric carcinoma remains the fifth most frequently diagnosed cancer and the third leading cause of cancer-related deaths, with an estimated 1,033,701 new cases and 782,685 deaths worldwide in 2018 [1, 2]. Gastric cancer is also one of the most common malignancies and the third leading cause of death in China, where 427,100 cases with 301,200 deaths were observed in 2013 [3]. Despite the several existing treatments including chemo-, radio-, or targeted therapy, the overall 5-year survival rate of stomach cancer patients is still < 20% [4].

There are two types of gastric cancer, diffuse and intestinal types, which differ in their histological manifestations and epidemiological features (gender ratio and age at diagnosis) [5]. Histopathology is the gold standard approach for diagnosing gastric cancer; however, this approach is not suitable for everyone due to the invasive nature of the biopsy [6]. Although there are several commonly used serum biomarkers such as alpha-fetoprotein (AFP), carcinoembryonic antigen (CEA), cancer antigen 125 (CA125), and cancer antigen 19–9 (CA19-9) [7] for gastric cancer diagnosis, none of them are sensitive or gastric-cancer-specific [8]. Moreover, effective and specific targeted therapies for gastric cancer remain to be identified. Presently, the major treatment strategies for gastric cancer are anti-human epidermal growth factor

receptor 2 (HER2) and anti-vascular therapies [4]. However, resistance to the targeted agents is common in some gastric tumor types. Therefore, novel practical approaches are needed for specific diagnosis and effective treatment of gastric cancer. Accordingly, identification of the key genes and biomarkers that are involved in the pathogenesis of gastric cancer is of paramount significance.

With recent advancements in bioinformatics methods, comprehensive identification of potential biomarkers through large-scale screening of expression profiles has been proposed [9, 10]. A weighted gene co-expression network analysis (WGCNA) approach provides a systematic analysis to investigate the functional clustering of expression profiles, based on the theory that genes with similar expression profiles may have closely functional linkages and/or pathways [11–13]. This approach groups highly co-expressed genes into the same module. Modules bearing high correlation with certain clinical traits are identified as clinically significant modules[13]. The “least absolute shrinkage and selection operator” (LASSO) logistic regression was a useful method to select significant variables, especially suitable for high-dimensional gene data analysis [14, 15].

In order to find sensitive or specific biomarkers in gastric cancer, combination of WGCNA and LASSO regression were performed. These findings may provide potential diagnostic and therapeutic targets in future research and clinical intervention of gastric cancer.

Materials And Methods

Data collection and preprocessing

The workflow of this study is shown in Fig. 1. Raw expression datasets were downloaded from the Gene Expression Omnibus (GEO) database (http://www.ncbi.nlm.nih.gov/geo/) by using the keywords “stomach/gastric cancer/tumor/carcinoma,” “normal,” “GPL570,” and “*Homo sapiens*.” Our inclusion criteria for the datasets were that all the datasets 1) were included in the Affymetrix Human Genome U133 Plus 2.0 Array (Affymetrix, Santa Clara, CA, USA) and 2) had been derived from human case-control studies involving both gastric tumor and normal samples.

To avoid bias and batch effects among different cohorts, and to ensure a large sample size, the three datasets with the largest sample sizes as of March 25, 2020 [GSE66229 [16, 17], GSE13911[18] and GSE54129] were selected. GSE66229 contained 300 tumor and 100 normal samples, and was used as the training set to screen for the hub genes. GSE54129 (contained 111 tumor and 21 normal samples) served as the testing set to validate the hub genes. GSE13911, bearing 69 samples (38 tumor and 31 normal), was used to validate module preservation.

All the analyses in this study were conducted using R software (version 3.5.1). FitPLM weight, Relative Log Expression (RLE), Normalized Unscaled Standard Errors (NUSE), and RNA degradation images were analyzed to evaluate the quality of each dataset. Then, the “rma” function with the default parameters of the “affy” package was used to perform background correction and normalization [19]. Missing values in

each dataset were imputed by using the function “impute.knn” with the default parameters of the “impute” package [20]. Platform annotations were downloaded from the GEO database, and finally, the gene symbol expression matrices were acquired from each dataset for further analyses.

Weighted Gene Co-expression Network Construction

Weighted gene co-expression network in the training set was constructed using the “WGCNA” package [21, 22]. The genes with the top 25% SD among all the 400 samples in the expression matrix of the training set were selected as the input genes (5115 genes in total).

In brief, first, the appropriate soft-thresholding power (β) was selected by using the “pickSoftThreshold” function with the default parameters (herein, β = 4). Subsequently, the Pearson’s correlation matrix was calculated to evaluate the similarity among all the pair-wise genes by using the “cor” function with the default parameters. Then, the adjacency was calculated based on β and the Pearson’s correlation matrix by using the “TOMsimilarity” function with the default parameters, and the corresponding dissimilarity (dissTOM) was also calculated. Finally, average linkage hierarchical clustering was conducted according to the dissTOM value with a minimum size of 30 for each gene dendrogram.

Module eigengenes (MEs), considered the first principal component (PC) of gene expression patterns of a corresponding module, were obtained for each module. To further strengthen the reliability of the modules, a cut line was set at 0.25 so that modules bearing < 0.25 would be merged [23].

Module Preservation Analysis

GSE13911 was used to validate the module preservation of the training set. The raw expression data of GSE13911 were converted into a gene symbol expression matrix. As shown in Fig. 4A, 5115 genes in GSE13911 clustered into 11 colored modules (no genes were clustered in the grey module), as determined in the training set. All gene modules were found to bear strong conservation, as the *Z* summary scores were all > 10 (Fig. 4B).

Identification Of Clinically Significant Modules

Herein, we focused on the clinical traits of the tissue types (tumor or normal). The Pearson’s correlation between each ME and clinical traits were calculated to reveal the most significant modules. Modules with a Pearson’s correlation > 0.7 and *P*-value < 0.05 were selected as clinically significant modules.

Additionally, gene significance (GS), defined as the $\log_{10}$ transformation of the *P*-value in the linear regression between the gene expression and clinical traits, were also calculated. Modules with the highest absolute gene significance were considered to be closely correlated with clinical traits.

Functional Enrichment And Pathway Analyses Of Significant Modules

To determine whether the clinically significant modules were closely correlated with gastric cancer, GO functional annotation and KEGG pathway analyses were performed using the Database for Annotation, Visualization and Integrated Discovery (DAVID) (version 6.8) (https://david.ncifcrf.gov/home.jsp)[28, 29]. The visualization of the functional enrichment and pathway analyses was performed by the “GOplot” [30] and “ggplot2” [31] packages of R, respectively.

Candidate Gene Selection

Candidate genes were selected according to the following criteria: 1) differentially expressed genes (DEGs) between gastric cancer samples and normal samples with $|\log_2 FC \text{ (Fold Change)}| > 1$ and adjusted *P*-value < 0.05 based on the “limma” package [32], 2) high module membership defined as |MM(Module Membership)| > 0.8 and a high significant correlation with a clinical trait as |GS| > 0.6.

Selection Of Hub Genes By Lasso Logistic Regression Analysis

Candidate genes were subjected to LASSO regression, which was performed using the “glmnet” package by setting alpha = 1, and ten-fold cross-validation for tuning parameter selection. Lambda was defined as the minimum partial likelihood deviance [33].

Validation Of Hub Genes

The machine learning method of logistic regression (by using the “glm” function with the default parameters) [34] was performed to determine whether the hub genes could correctly distinguish the tumor samples from the normal samples in a testing set.

Areas under the receiver operating characteristic (ROC) curves were calculated to compare the predictive effects among the different supervised machine learning models, and then the ROC curves were plotted using the “pROC” package [30]. An area under the curve (AUC) value between 0.8 and 0.9 is considered an excellent classification, while greater than 0.9 is considered as outstanding discrimination [35].

Results

Construction of co-expression networks

After the quality check of the input data, no sample was removed (**Supplementary Fig. 1**); herein, two clinical traits (tissue type and stage) are presented. According to different tissue types (tumor or normal), the 400 samples could be mainly divided into two clusters.

As shown in Fig. 2, the soft thresholding was set at 4, while the scale-free topology fit index reached 0.89, indicating approximate scale-free topology. Co-expressed gene modules were identified with the dynamic tree cut method (Fig. 3A) [23]. In total, 12 modules were found, and each color represented one module (Fig. 3B). The biggest module was the turquoise module, which contained 2120 genes, followed by the blue module, bearing 1503 genes. The grey module comprised 2 genes, which did not have a similar expression pattern and did not belong to any other module.

Selection Of Clinically Significant Modules

Herein, two factors (tissue type and stage) were included in module-trait relationship analysis. The most interesting clinical trait was the tissue type, which was designated as tumor or normal samples. After the assessment of the module-trait relationship (Fig. 5A), two modules that had > 0.7 Pearson's correlation with tissue types were found. The blue module had the highest correlation with gastric cancer (Pearson's correlation between gastric cancer and the blue module was 0.83, *P*-value = 9e-105), followed by the black module (Pearson's correlation = -0.72, *P*-value = 1e-65). Furthermore, the black and blue modules also had the highest gene significance (Fig. 5B), which denoted the association of the genes in the modules with gastric cancer.

In addition, as shown in **Supplementary Fig. 2**, the correlations between module membership and gene significance in the blue and black modules were 0.85 ($P < 0.001$) and 0.77 ($P < 0.001$), respectively. It was further demonstrated that these two modules were closely related to gastric cancer. Thus, they were selected as clinically significant modules for further analysis.

Functional Enrichment Of Clinically Significant Modules

Gene Ontology (GO) enrichment results showed that 1503 genes in the blue module mainly participated in 199 different biological processes (**Supplementary Table 1**). In the black module, 137 genes were mainly involved in 29 different biological processes (**Supplementary Table 2**). The top three most significantly enriched biological processes were cell division ($P = 2.29e-29$), G1/S transition of the mitotic cell cycle ($P = 6.32e-21$), and mitotic nuclear division ($P = 3.49e-19$) in the blue module (Fig. 6A), and potassium ion import ($P = 3.39e-05$), digestion ($P = 8.25e-04$), and multicellular organismal water homeostasis ($P = 0.0012$) in the black module (Fig. 6B).

KEGG (Kyoto Encyclopedia of Genes and Genomes) analysis showed that 1503 genes in the blue module were mainly enriched in 17 pathways (**Supplementary Table 3**). The top three most significantly enriched pathways were cell cycle ($P = 1.21e-15$), DNA replication ($P = 6.21e-15$), and the p53 signaling pathway (*P*

= 2.03e-06) (Fig. 7A). In the black module, 137 genes were involved in 10 pathways (**Supplementary Table 4**), and gastric acid secretion (*P* = 4.46e-05), protein digestion and absorption (*P* = 1.28e-04), as well as drug metabolism-cytochrome P450 (*P* = 0.0032) pathways were the top three most significantly enriched pathways (Fig. 7B).

Identification Of Hub Genes

Using the screening criteria of |GS| > 0.6 and |MM |> 0.8, 43 genes were identified. All these 43 genes were differentially expressed between the normal and tumor samples with |logFC| > 1 and adjusted *P*-value < 0.05 (Fig. 8, **Supplementary Table 5**). Finally, 10 genes [Acyl-CoA dehydrogenase long-chain (*ACADL*), Adiponectin (*ADIPOQ*); Rho GTPase activating protein 39 (*ARHGAP39*); ATPase family AAA-domain containing protein 3A (*ATAD3A*); *C1orf95* (also known as *STUM* gene), cholecystokinin B receptor (*CCKBR*); Glutamate receptor, ionotropic kainate 3 (*GRIK3*); sodium channel epithelial 1 subunit gamma (*SCNN1G*); Sialic acid-binding immunoglobulin-like lectin-11 (*SIGLEC11*); and Taxilin alpha (*TXLNB*)] were identified as the hub genes by using LASSO logistic regression (**Table 1**).

Validation Of The Hub Genes

GSE54129 was utilized as the testing set to validate the 10-gene model. The AUC value of this classifier upon using logistic regression was 0.882, indicating the excellent classification effects of the model (Fig. 9).

Discussion

Gastric cancer has the fifth highest incidence rate among all cancers, and it has a poor prognosis, with a low 5-year survival rate [36]. Patients with early-stage gastric cancer are asymptomatic and thus difficult to diagnose, and screening for significant genes that can effectively distinguish normal and tumor samples can provide very important auxiliary evidence in gastric cancer diagnosis[37].

In this study, the WGCNA approach was proposed to screen for the hub genes, which can effectively differentiate the normal gastric samples from tumor samples. WGCNA is an important approach to systematically describe the correlation patterns among genes and identifies modules of highly-correlated genes, followed by candidate biomarker screening. Nodes and edges in this co-expression network represent genes and the correlations between gene pairs, respectively. By using this systematic method, the connectivity in each module can be detected while also taking clinical traits into account [22].

Eventually, blue and black modules were identified as clinically significant modules. The GO and KEGG analyses revealed that the genes in these two modules were significantly enriched in the biological processes of the cell cycle, cell division, and stomach-related functions. All these biological functions are closely related to gastric cancer [38, 39]. *ACADL*, *ADIPOQ*, *ARHGAP39*, *ATAD3A*, *C1orf95*, *CCKBR*, *GRIK3*,

SCNN1G, *SIGLEC11*, and *TXLNB* were identified as the hub genes from the screening the black and blue modules.

ACADL is a mitochondrial enzyme catalyzing the initial step of fatty acid oxidation and has been reported to correlate with esophageal [40], breast [41], hepatocellular [42], and gastric [43] cancers. ACADL is upregulated in esophageal squamous cell carcinoma cell lines and specimens, and its upregulation is associated with disease progression [40]. However, another study reported the opposite result, in which ACADL was downregulated and significantly correlated with poor prognosis in hepatocellular carcinoma [42]. ACADL is also among the genes that are significantly differentially methylated between the ER + and ER- breast cancer tumors [44]. By comprehensive whole-genome and transcriptome sequencing analyses, a study has revealed that ACADL is one of the mutated genes in gastric cancer [43].

ADIPOQ is one of the most important adipocytokines secreted by adipocytes [45], and the polymorphisms of *ADIPOQ* have been reported to correlate with several types of cancer, including colorectal [46, 47] and breast [48] cancers. A study focusing on the molecular mechanisms ADIPOQ participated in has revealed that ADIPOQ induces cytotoxic autophagy in breast cancer cells through STK11/LKB1-mediated activation of the AMPK-ULK1 axis [49]. Another study has reported that miR-370 inhibits the proliferation, invasion, and epithelial-mesenchymal transition of gastric cells by directly downregulating receptor 4 of ADIPOQ [50]. Overexpression of another microRNA, miR-15b-5p, promotes the metastasis of gastric cancer by regulating ADIPOQ receptor 3 [51].

ATAD3A is a nuclear-encoded mitochondrial enzyme, involving in mitochondrial dynamics, cell death, and cholesterol metabolism [52]. It has been reported to correlate with hepatocellular carcinoma [53] and breast cancer [54], and it might be an effective therapeutic target in cancer treatment [55]. *ATAD3A* is differentially expressed between paclitaxel-resistant and -sensitive MCF7 breast cancer cells [54]. A study has revealed that ATAD3A is upregulated in hepatocellular carcinoma and ATAD3A upregulation is correlated with poor prognosis [53].

Cholecystokinin is a well-known trophic factor for the gastrointestinal tract [56], and *CCKBR* is correlated with pancreatic [56–59], breast [60], and gastric [61] cancers. Gastrin upregulates CCKBR in gastric cancer cell lines [62], and thus it serves as a biomarker in gastric cancer treatment [63]. A study has shown that miR-148a has anti-cancer effects on gastric cancer through the inhibition of STAT3 and Akt activation by targeting *CCKBR* [64]. Another research has revealed that trastuzumab inhibits the growth of HER2-negative gastric cancer cells by regulating the CCKBR signaling pathway [61].

GRIK3 mainly participates in the neuroactive ligand-receptor interaction pathway, and GRIK3 upregulation is associated with poor survival in gastric cancer [65]. GRIK3 promotes epithelial-mesenchymal transition by regulating the SPDEF/CDH1 signaling in breast cancer cells [66].

There have been few studies focusing on the relationship between gastric cancer and *SCNN1G*, *ARHGAP39*, *Clorf95*, *SIGLEC11*, or *TXLNB*. *SCNN1G* is one of the genes significantly upregulated in Ewing's sarcoma and fibromatosis samples [67]. *ARHGAP39* mutations or variations in copy number or expression level are found in several types of tumor-like tissues from the central nervous system, skin, prostate, and gastrointestinal tract [68]. ARHGAP39 interacts with p53 and BAX, and decreased expression of ARHGAP39 increases cell proliferation, leading to tumorigenesis [69]. Clorf95 is one of the uncharacterized proteins correlated with diverse human cancers [70]. Another study focusing on scleroderma patients demonstrated the involvement of Clorf95 in cancer incidence [71]. Sialic acid-binding immunoglobulin-like lectin-11 (*SIGLEC11*) is a primate-lineage–specific receptor of human tissue macrophages, and it is also expressed in brain microglia [72, 73]. A missense mutation of *SIGLEC11* has been detected in pancreatic cancer patients [74], and SIGLEC11 is significantly upregulated in the poor prognostic group of pancreatic cancer patients [75]. Taxilin alpha (*TXLNA*), which is a binding partner of the syntaxin family, has been identified as a key factor in the coordination of intracellular vesicle trafficking, and it is upregulated in pancreatic adenocarcinoma patients [76].

In our study, several hub genes were implicated in the metabolic processes such as fatty acid oxidation and cholesterol metabolism. A previous study has demonstrated the association between metabolic syndrome and gastric cancer [77]. Low total serum cholesterol levels are correlated with an increased risk of gastric cancer in Chinese Han population [78]. A study has detected increased fatty acid oxidation in gastric cancer [79], and adipocytes fuel gastric cancer by mediating fatty acid metabolism [80].

All these results from the previous studies demonstrate that the hub genes identified in our study are closely correlated with gastric cancer and play important roles in cancer development, progression, or proliferation.

The significant module and hub genes identified in this study are biologically rational. First, the clinically significant module identified in our study bears strong preservation, implying that this clinically significant module is conservative and could also be reproduced in other datasets. Further, it suggests that that modules constructed by WGCNA are reliable. Second, most of the genes in the significant module were enriched for specific GO terms and KEGG pathways closely relating to stomach or cancer physiology. For instance, GO analysis demonstrated that most of the genes in the clinically significant modules were closely related to digestion, carbohydrate metabolic process, and gastric acid secretion, as well as cell division and cell cycle. KEGG enrichment analysis also indicated that most of the genes in the clinically significant module were implicated in gastric acid secretion, protein digestion and absorption, as well as glycerolipid metabolism and the p53 signaling pathway. Third, all the hub genes identified in our study had previously been reported to relate to cancer. Most of these hub genes are implicated in metabolic processes, influencing the development and progression of gastric cancer. It may thus be inferred that these genes are genuinely the hub genes in charge of the key processes in gastric cancer, and they deserve a deeper analysis and validation. Finally, by using machine learning methods, the hub genes were demonstrated to effectively discriminate the gastric tumor samples from normal samples. In

our study, the predictive effects of different machine learning methods were evaluated by AUC values [81]. Herein, the AUC values of both logistic regression was > 0.8, indicating the excellent predictive results. All the results indicated that the expression profiles of these ten hub genes have excellent predictive effects when discriminating gastric cancer samples from normal samples.

However, our study has limitations. First, all the hub genes were identified and validated only through bioinformatics, and further exploration of the biological functions and molecular mechanisms of these hub genes both *in vitro* and in *vivo* is required. Second, due to the limited availability of the data, we did not differentiate between intestinal-type and diffuse-type gastric cancers. More data are needed to analyze and identify the hub genes between these two types of gastric cancer and normal samples.

In summary, through WGCNA, we identified ten hub genes, which might serve as potential diagnostic and/or therapeutic biomarkers for gastric cancer. Profile data mining by bioinformatics analysis is an available method to find potential diagnostic or therapeutic biomarkers systematically. Nevertheless, further investigations about the molecular mechanisms in which these hub genes are involved are still needed to verify the involvement of these genes in gastric cancer. Our findings provide a better understanding of the molecular mechanisms and putative diagnostic or therapeutic biomarkers for gastric cancer.

Declarations

Acknowledgements

We sincerely thank Professor Qiu Li (Department of Cancer Center, West China Hospital) and Professor Yuan Yong (Department of Thoracic Surgery, West China Hospital) for their good suggestions on gastric cancer diagnosis, treatment, and management.

Funding

This work was supported by Science & Technology Department of Sichuan Province funding project (No. 2016FZ0108, 2017FZ0104), 1.3.5 project for disciplines of excellence, West China Hospital, Sichuan University (ZYJC18010).

Availability of data and materials

The datasets used and/or analyzed during the current study are downloaded

from the GEO database (http://www.ncbi.nlm.nih.gov/geo/), and the datasets

used in this study were GSE66229, GSE13911, and GSE54129.

Authors' contributions

CL participated in study design, data preprocessing and analysis, as well as manuscript writing. HY involved in data preprocessing, data analysis and results double check.YS participated in supervised learning analysis, and manuscript writing. XZ designed the study and help to write the manuscript. WZ supervised the study. All authors read and approved the final manuscript.

Ethics approval and consent to participate

Not applicable.

Consent for publication

Not applicable.

Competing interests

The authors declare that they have no competing interests.

References

1. Pormohammad A, Mohtavinejad N, Gholizadeh P, et al., Global estimate of gastric cancer in Helicobacter pylori-infected population: A systematic review and meta-analysis, *J Cell Physiol*, 2018.
2. Bray F, Ferlay J, Soerjomataram I, et al., Global Cancer Statistics 2018: GLOBOCAN Estimates of Incidence and Mortality Worldwide for 36 Cancers in 185 Countries, *CA Cancer J Clin*, 2018.
3. Chen W, Zheng R, Zhang S, et al. Cancer incidence and mortality in China, 2013. Cancer Lett. 2017;401:63–71.
4. Raimondi A, Nichetti F, Peverelli G, et al. Genomic markers of resistance to targeted treatments in gastric cancer: potential new treatment strategies. Pharmacogenomics. 2018;19(13):1047–68.
5. Karimi P, Islami F, Anandasabapathy S, et al. Gastric cancer: descriptive epidemiology, risk factors, screening, and prevention. Cancer Epidemiol Biomarkers Prev. 2014;23(5):700–13.
6. Yoon H, Kim N. Diagnosis and management of high risk group for gastric cancer. Gut Liver. 2015;9(1):5–17.
7. He CZ, Zhang KH, Li Q, et al. Combined use of AFP, CEA, CA125 and CAl9-9 improves the sensitivity for the diagnosis of gastric cancer. BMC Gastroenterol. 2013;13:87.
8. Smyth EC, Verheij M, Allum W, et al. Gastric cancer: ESMO Clinical Practice Guidelines for diagnosis, treatment and follow-up. Ann Oncol. 2016;27(suppl 5):v38–49.
9. Takeno A, Takemasa I, Doki Y, et al. Integrative approach for differentially overexpressed genes in gastric cancer by combining large-scale gene expression profiling and network analysis. Br J Cancer. 2008;99(8):1307–15.
10. Wang J, Ni Z, Duan Z, et al. Altered expression of hypoxia-inducible factor-1alpha (HIF-1alpha) and its regulatory genes in gastric cancer tissues. PLoS One. 2014;9(6):e99835.

11. Carlson MR, Zhang B, Fang Z, et al. Gene connectivity, function, and sequence conservation: predictions from modular yeast co-expression networks. BMC Genom. 2006;7:40.
12. Carter SL, Brechbuhler CM, Griffin M, et al. Gene co-expression network topology provides a framework for molecular characterization of cellular state. Bioinformatics. 2004;20(14):2242–50.
13. Zhou XG, Huang XL, Liang SY, et al. Identifying miRNA and gene modules of colon cancer associated with pathological stage by weighted gene co-expression network analysis. Onco Targets Ther. 2018;11:2815–30.
14. Zeng X, Li C, Li Y, et al., A network-based variable selection approach for identification of modules and biomarker genes associated with end-stage kidney disease, *Nephrology (Carlton)*, 2019.
15. Friedman J, Hastie T, Tibshirani R. Regularization Paths for Generalized Linear Models via Coordinate Descent. J Stat Softw. 2010;33(1):1–22.
16. Oh SC, Sohn BH, Cheong JH, et al. Clinical and genomic landscape of gastric cancer with a mesenchymal phenotype. Nat Commun. 2018;9(1):1777.
17. Cristescu R, Lee J, Nebozhyn M, et al. Molecular analysis of gastric cancer identifies subtypes associated with distinct clinical outcomes. Nat Med. 2015;21(5):449–56.
18. D'Errico M, de Rinaldis E, Blasi MF, et al. Genome-wide expression profile of sporadic gastric cancers with microsatellite instability. Eur J Cancer. 2009;45(3):461–9.
19. Gautier L, Cope L, Bolstad BM, et al. affy–analysis of Affymetrix GeneChip data at the probe level. Bioinformatics. 2004;20(3):307–15.
20. Hastie TTR, Narasimhan B. impute: Imputation for microarray data. Bioinformatics. 2001;17(6):6.
21. Zhang B, Horvath S. A general framework for weighted gene co-expression network analysis. Stat Appl Genet Mol Biol. 2005;4:Article17.
22. Langfelder P, Horvath S. WGCNA: an R package for weighted correlation network analysis. BMC Bioinformatics. 2008;9:559.
23. Chen L, Yuan L, Wang Y, et al. Co-expression network analysis identified FCER1G in association with progression and prognosis in human clear cell renal cell carcinoma. Int J Biol Sci. 2017;13(11):1361–72.
24. Chen L, Yuan L, Qian K, et al. Identification of Biomarkers Associated With Pathological Stage and Prognosis of Clear Cell Renal Cell Carcinoma by Co-expression Network Analysis. Front Physiol. 2018;9:399.
25. Obeidat M, Nie Y, Chen V, et al. Network-based analysis reveals novel gene signatures in peripheral blood of patients with chronic obstructive pulmonary disease. Respir Res. 2017;18(1):72.
26. Neidlin M, Dimitrakopoulou S, Alexopoulos LG. Multi-tissue network analysis for drug prioritization in knee osteoarthritis. Sci Rep. 2019;9(1):15176.
27. Langfelder P, Luo R, Oldham MC, et al., Is My Network Module Preserved and Reproducible?, *Plos Comput Biol*, 2011;7 (1).

28. Ashburner M, Ball CA, Blake JA, et al. Gene ontology: tool for the unification of biology. The Gene Ontology Consortium. Nat Genet. 2000;25(1):25–9.
29. Dennis G, Sherman BT, Hosack DA, et al., DAVID: Database for annotation, visualization, and integrated discovery, *Genome Biol*, 2003;4 (9).
30. Robin X, Turck N, Hainard A, et al. pROC: an open-source package for R and S + to analyze and compare ROC curves. BMC Bioinformatics. 2011;12:77.
31. H W, ggplot2: elegant graphics for data analysis, 2016.
32. Ritchie ME, Phipson B, Wu D, et al. limma powers differential expression analyses for RNA-sequencing and microarray studies. Nucleic Acids Res. 2015;43(7):e47.
33. Friedman J. Package 'glmnet', 2010.
34. Günther F, Fritsch S. neuralnet: Training of neural networks. The R journal. 2010;2(1):8.
35. Lemeshow DWHS. Applied Logistic Regression, 2nd Ed, 2000.
36. Sun F, Sun H, Mo X, et al. Increased survival rates in gastric cancer, with a narrowing gender gap and widening socioeconomic status gap: A period analysis from 1984 to 2013. J Gastroenterol Hepatol. 2018;33(4):837–46.
37. Tomasello G, Ghidini M, Liguigli W, et al. Targeted therapies in gastric cancer treatment: where we are and where we are going. Invest New Drugs. 2016;34(3):378–93.
38. Cao J, Wu N, Han Y, et al. DDX21 promotes gastric cancer proliferation by regulating cell cycle. Biochem Biophys Res Commun. 2018;505(4):1189–94.
39. Waldum HL, Sagatun L, Mjones P. Gastrin and Gastric Cancer. Front Endocrinol (Lausanne). 2017;8:1.
40. Yu DL, Li HW, Wang Y, et al. Acyl-CoA dehydrogenase long chain expression is associated with esophageal squamous cell carcinoma progression and poor prognosis. Onco Targets Ther. 2018;11:7643–53.
41. Li Z, Heng J, Yan J, et al. Integrated analysis of gene expression and methylation profiles of 48 candidate genes in breast cancer patients. Breast Cancer Res Treat. 2016;160(2):371–83.
42. Zhao X, Qin W, Jiang Y, et al. ACADL plays a tumor-suppressor role by targeting Hippo/YAP signaling in hepatocellular carcinoma. NPJ Precis Oncol. 2020;4:7.
43. Zhang J, Huang JY, Chen YN, et al. Whole genome and transcriptome sequencing of matched primary and peritoneal metastatic gastric carcinoma. Sci Rep. 2015;5:13750.
44. Li Z, Guo X, Wu Y, et al. Methylation profiling of 48 candidate genes in tumor and matched normal tissues from breast cancer patients. Breast Cancer Res Treat. 2015;149(3):767–79.
45. Parida S, Siddharth S, Sharma D. Adiponectin, Obesity, and Cancer: Clash of the Bigwigs in Health and Disease, *Int J Mol Sci*, 2019;20 (10).
46. Tan X, Wang GB, Tang Y, et al. Association of ADIPOQ and ADIPOR variants with risk of colorectal cancer: A meta-analysis. J Huazhong Univ Sci Technolog Med Sci. 2017;37(2):161–71.

47. Nimptsch K, Song M, Aleksandrova K, et al. Genetic variation in the ADIPOQ gene, adiponectin concentrations and risk of colorectal cancer: a Mendelian Randomization analysis using data from three large cohort studies. Eur J Epidemiol. 2017;32(5):419–30.
48. Mendez-Hernandez A, Gallegos-Arreola MP, Moreno-Macias H, et al. LEP rs7799039, LEPR rs1137101, and ADIPOQ rs2241766 and 1501299 Polymorphisms Are Associated With Obesity and Chemotherapy Response in Mexican Women With Breast Cancer. Clin Breast Cancer. 2017;17(6):453–62.
49. Chung SJ, Nagaraju GP, Nagalingam A, et al. ADIPOQ/adiponectin induces cytotoxic autophagy in breast cancer cells through STK11/LKB1-mediated activation of the AMPK-ULK1 axis. Autophagy. 2017;13(8):1386–403.
50. Feng Y, Sun T, Yu Y, et al. MicroRNA-370 inhibits the proliferation, invasion and EMT of gastric cancer cells by directly targeting PAQR4. J Pharmacol Sci. 2018;138(2):96–106.
51. Zhao C, Li Y, Chen G, et al. Overexpression of miR-15b-5p promotes gastric cancer metastasis by regulating PAQR3. Oncol Rep. 2017;38(1):352–8.
52. Teng Y, Lang L, Shay C. ATAD3A on the Path to Cancer. Adv Exp Med Biol. 2019;1134:259–69.
53. Liu X, Li G, Ai L, et al. Prognostic value of ATAD3 gene cluster expression in hepatocellular carcinoma. Oncol Lett. 2019;18(2):1304–10.
54. Daniel P, Halada P, Jelinek M, et al., Differentially Expressed Mitochondrial Proteins in Human MCF7 Breast Cancer Cells Resistant to Paclitaxel, *Int J Mol Sci*, 2019;20 (12).
55. Teng Y, Ren X, Li H, et al. Mitochondrial ATAD3A combines with GRP78 to regulate the WASF3 metastasis-promoting protein. Oncogene. 2016;35(3):333–43.
56. Rai R, Kim JJ, Tewari M, et al. Heterogeneous expression of cholecystokinin and gastrin receptor in stomach and pancreatic cancer: An immunohistochemical study. J Cancer Res Ther. 2016;12(1):411–6.
57. Mu G, Ding Q, Li H, et al. Gastrin stimulates pancreatic cancer cell directional migration by activating the Galpha12/13-RhoA-ROCK signaling pathway. Exp Mol Med. 2018;50(5):1–14.
58. Kumari S, Chowdhury J, Sikka M, et al. Identification of potent cholecystokinin-B receptor antagonists: synthesis, molecular modeling and anti-cancer activity against pancreatic cancer cells. Medchemcomm. 2017;8(7):1561–74.
59. Clawson GA, Abraham T, Pan W, et al, Cholecystokinin A, Receptor-Specific B. DNA Aptamer for Targeting Pancreatic Ductal Adenocarcinoma. Nucleic Acid Ther. 2017;27(1):23–35.
60. Meng LL, Wang JL, Xu SP, et al. Low serum gastrin associated with ER(+) breast cancer development via inactivation of CCKBR/ERK/P65 signaling. BMC Cancer. 2018;18(1):824.
61. Cui Y, Li SB, Peng XC, et al. Trastuzumab Inhibits Growth of HER2-Negative Gastric Cancer Cells Through Gastrin-Initialized CCKBR Signaling. Dig Dis Sci. 2015;60(12):3631–41.
62. Kim O, Yoon JH, Choi WS, et al. Gastrokine 1 inhibits gastrin-induced cell proliferation. Gastric Cancer. 2016;19(2):381–91.

63. Mjones P, Nordrum IS, Sordal O, et al. Expression of the Cholecystokinin-B Receptor in Neoplastic Gastric Cells. Horm Cancer. 2018;9(1):40–54.
64. Yu B, Lv X, Su L, et al. MiR-148a Functions as a Tumor Suppressor by Targeting CCK-BR via Inactivating STAT3 and Akt in Human Gastric Cancer. PLoS One. 2016;11(8):e0158961.
65. Gong B, Li Y, Cheng Z, et al. GRIK3: A novel oncogenic protein related to tumor TNM stage, lymph node metastasis, and poor prognosis of GC. Tumour Biol. 2017;39(6):1010428317704364.
66. Xiao B, Kuang Z, Zhang W, et al. Glutamate Ionotropic Receptor Kainate Type Subunit 3 (GRIK3) promotes epithelial-mesenchymal transition in breast cancer cells by regulating SPDEF/CDH1 signaling. Mol Carcinog. 2019;58(7):1314–23.
67. Sarver AE, Sarver AL, Thayanithy V, et al. Identification, by systematic RNA sequencing, of novel candidate biomarkers and therapeutic targets in human soft tissue tumors. Lab Invest. 2015;95(9):1077–88.
68. Nowak FV. Porf-2 = Arhgap39 = Vilse: A Pivotal Role in Neurodevelopment, Learning and Memory, *eNeuro*, 2018;5 (5).
69. Jones R, Identification of novel risk variants for sarcoma and other cancers by whole exome sequencing in cancer cluster families, *PhD Thesis, The University of Western Australia*, 2017.
70. Delgado AP, Brandao P, Chaoado MJ, Hamid S, Narayanan R. Open Reading Frames Associated with Cancer in the Dark Matter of the Human Genome. Cancer Genomics Proteomics. 2014;11(4):12.
71. Xu GJ, Shah AA, Li MZ, et al. Systematic autoantigen analysis identifies a distinct subtype of scleroderma with coincident cancer. Proc Natl Acad Sci U S A. 2016;113(47):E7526–34.
72. Angata T, Kerr SC, Greaves DR, et al. Cloning and characterization of human Siglec-11. A recently evolved signaling molecule that can interact with SHP-1 and SHP-2 and is expressed by tissue macrophages, including brain microglia. J Biol Chem. 2002;277(27):24466–74.
73. Shahraz A, Kopatz J, Mathy R, et al. Anti-inflammatory activity of low molecular weight polysialic acid on human macrophages. Sci Rep. 2015;5:16800.
74. Jones S, Zhang X, Parsons DW, et al. Core signaling pathways in human pancreatic cancers revealed by global genomic analyses. Science. 2008;321(5897):1801–6.
75. Stratford JK, Bentrem DJ, Anderson JM, et al. A six-gene signature predicts survival of patients with localized pancreatic ductal adenocarcinoma. PLoS Med. 2010;7(7):e1000307.
76. Lv S, Zhang G, Xie L, et al., High TXLNA Expression Predicts Favourable Outcome for Pancreatic Adenocarcinoma Patients, *Biomed Res Int*, 2020;2020 2585862.
77. Li F, Du H, Li S, et al. The Association Between Metabolic Syndrome and Gastric Cancer in Chinese. Front Oncol. 2018;8:326.
78. Kang R, Li P, Wang T, et al. Apolipoprotein E epsilon 2 allele and low serum cholesterol as risk factors for gastric cancer in a Chinese Han population. Sci Rep. 2016;6:19930.
79. Lee JS, Kim SH, Lee S, et al. Gastric cancer depends on aldehyde dehydrogenase 3A1 for fatty acid oxidation. Sci Rep. 2019;9(1):16313.

80. Tan Y, Lin K, Zhao Y, et al. Adipocytes fuel gastric cancer omental metastasis via PITPNC1-mediated fatty acid metabolic reprogramming. Theranostics. 2018;8(19):5452–68.
81. Huang J, Ling CX. Using AUC and accuracy in evaluating learning algorithms. Ieee T Knowl Data En. 2005;17(3):299–310.

Figures

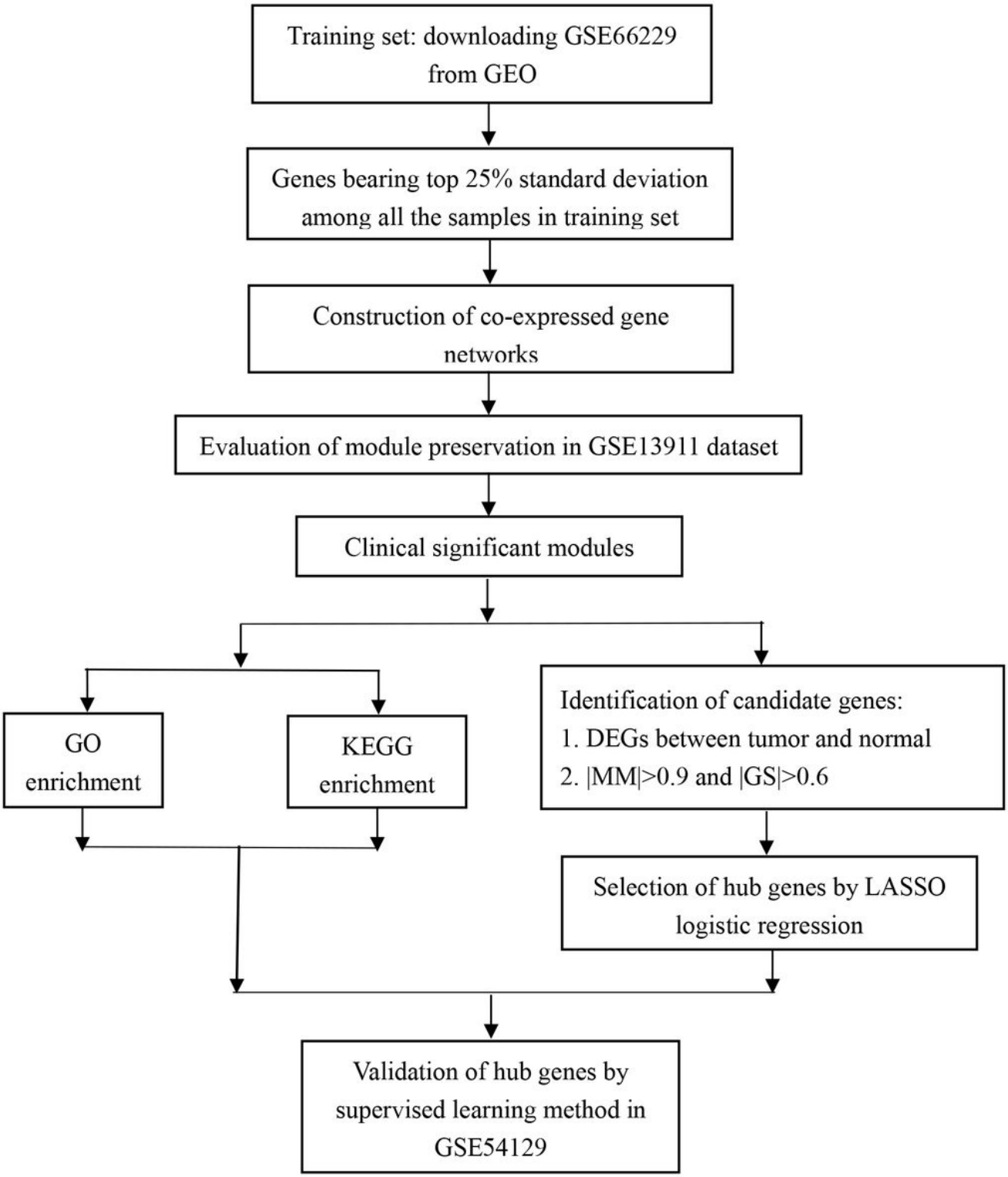

Figure 1

Flow diagram of the study: data collection, analysis, and hub gene selection and validation.

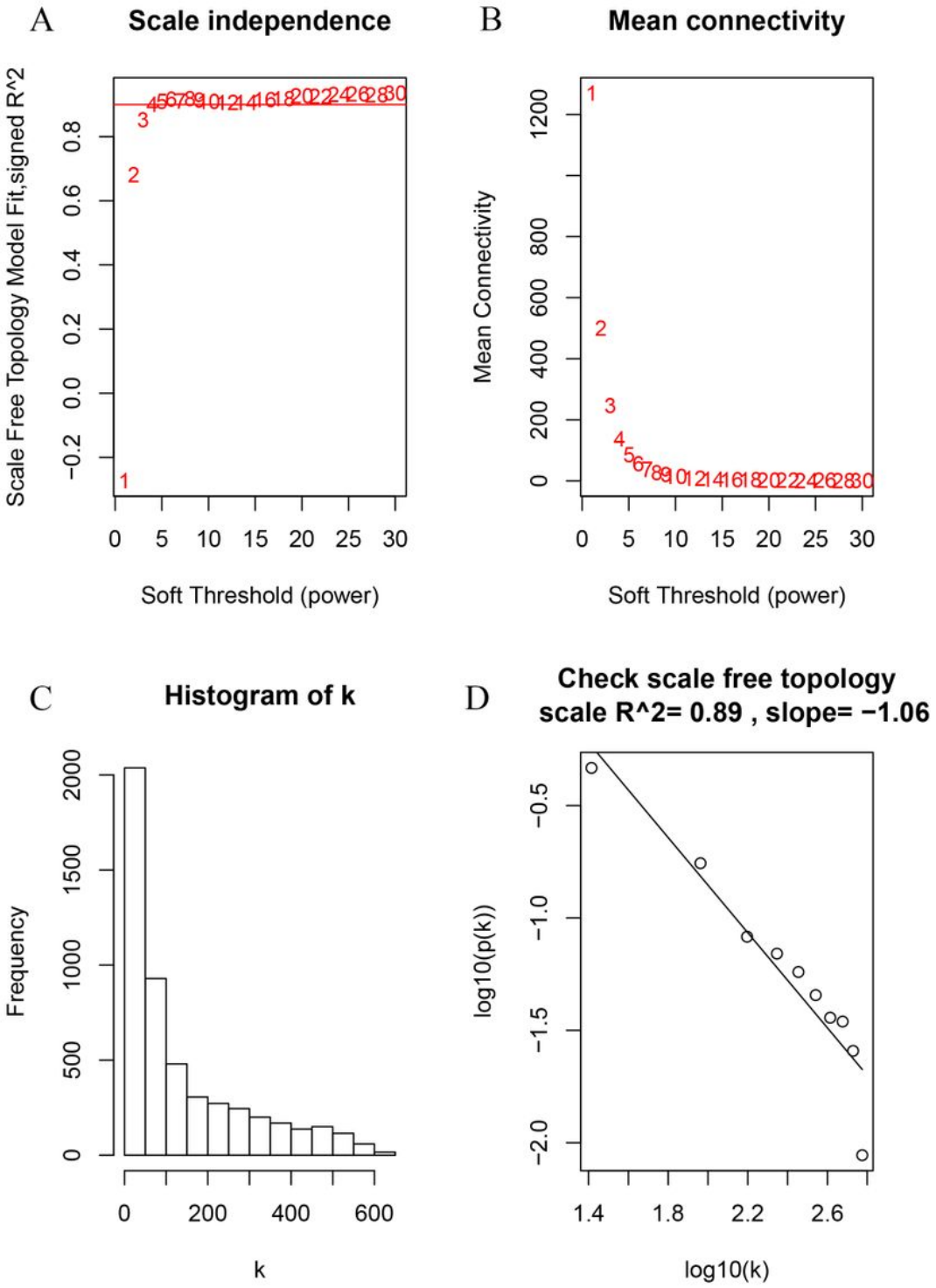

Figure 2

Determination of the soft-thresholding power in the weighted gene co-expression network analysis in the training set. (A) Screening soft-thresholding powers. (B) Analysis of the mean connectivity for various

soft-thresholding powers. (C) Histogram of the connectivity distribution with the soft-thresholding powers set at 4. (D) Checking the scale-free topology with the soft-thresholding powers set at 4.

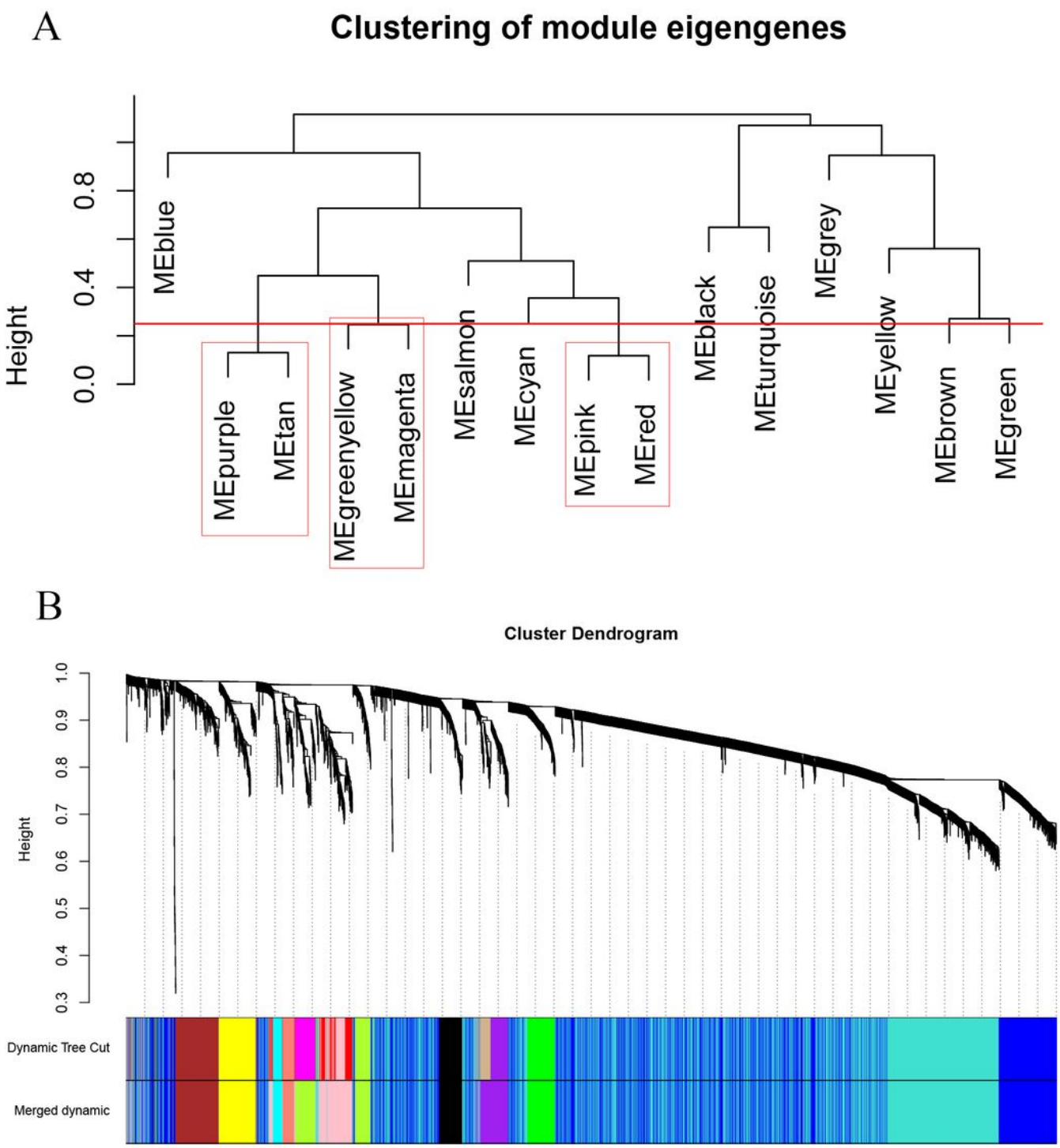

Figure 3

Clustering dendrograms of the 5115 genes in the training set. (A) Clustering of the module eigengenes to identify the merged modules. Upon setting the threshold at 0.25, 15 modules were merged into 12. modules (B) Co-expression module of the training set.

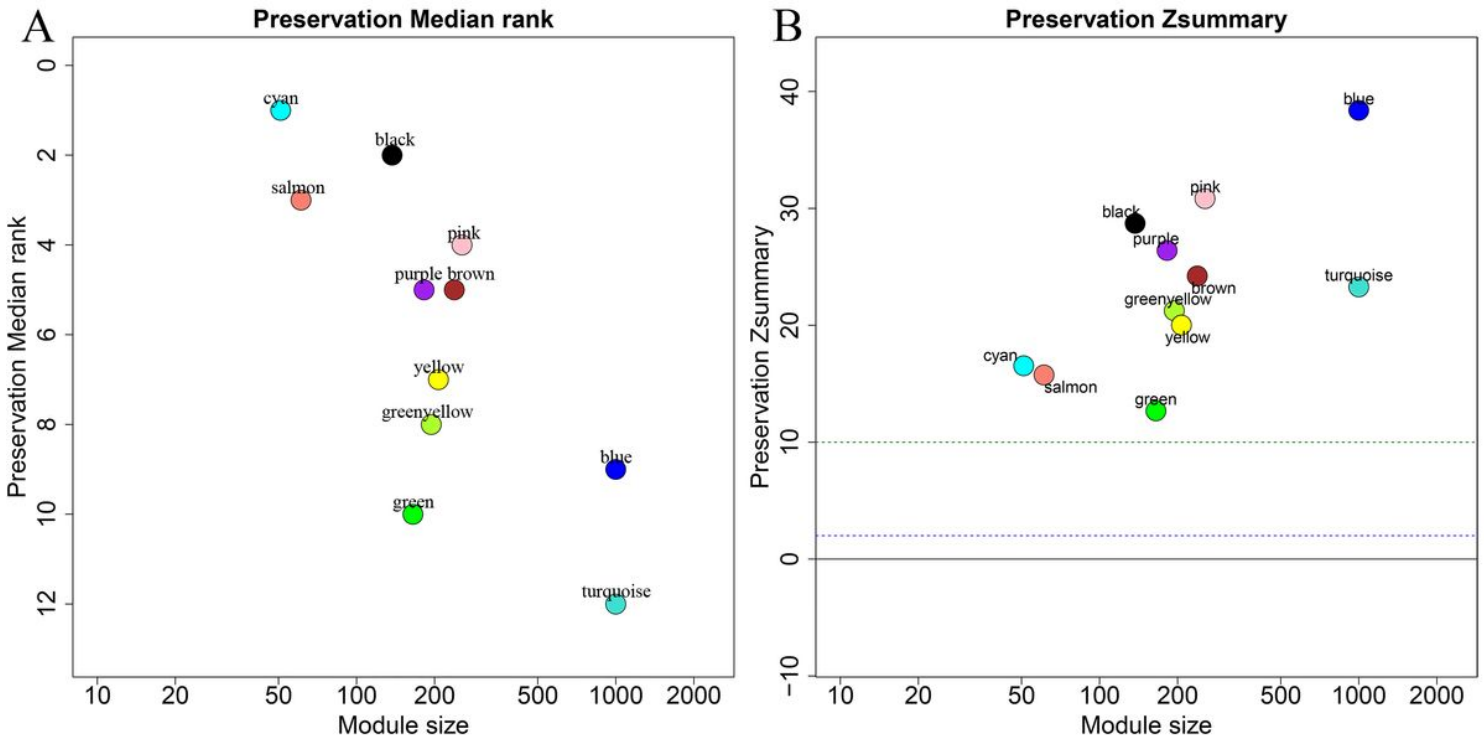

Figure 4

Evaluation of module preservation. The x- and y-axes present module size and preservation median rank (A) as well as preservation Z summary (B), respectively.

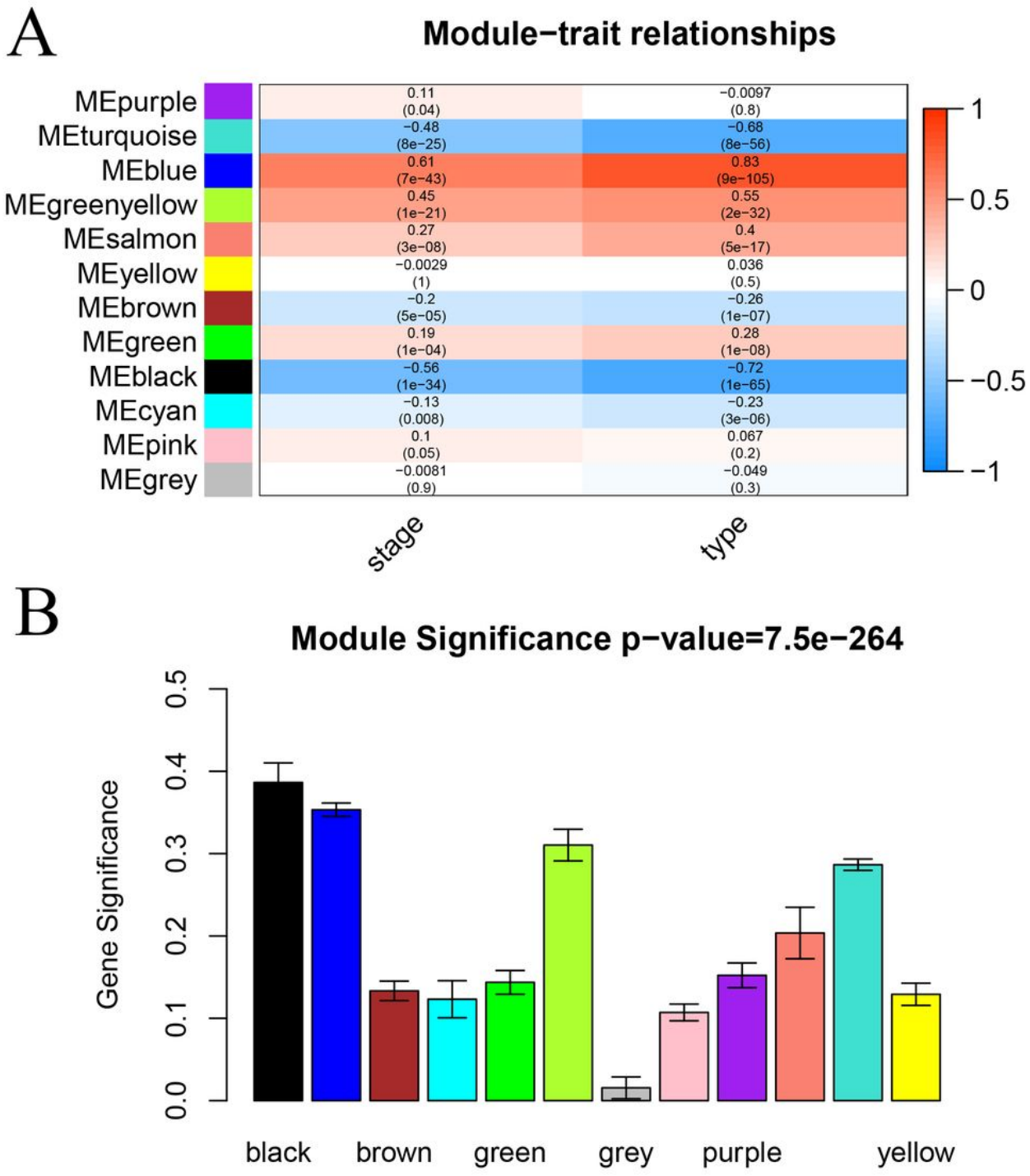

Figure 5

Identification of clinically significant modules. (A) Correlation matrix of ME and clinical traits. Rows and columns correspond to the eigengenes in the module and clinical traits, respectively. Pearson's correlation and P-values are presented in the cells. (B) Gene significance of gastric cancer across the modules.

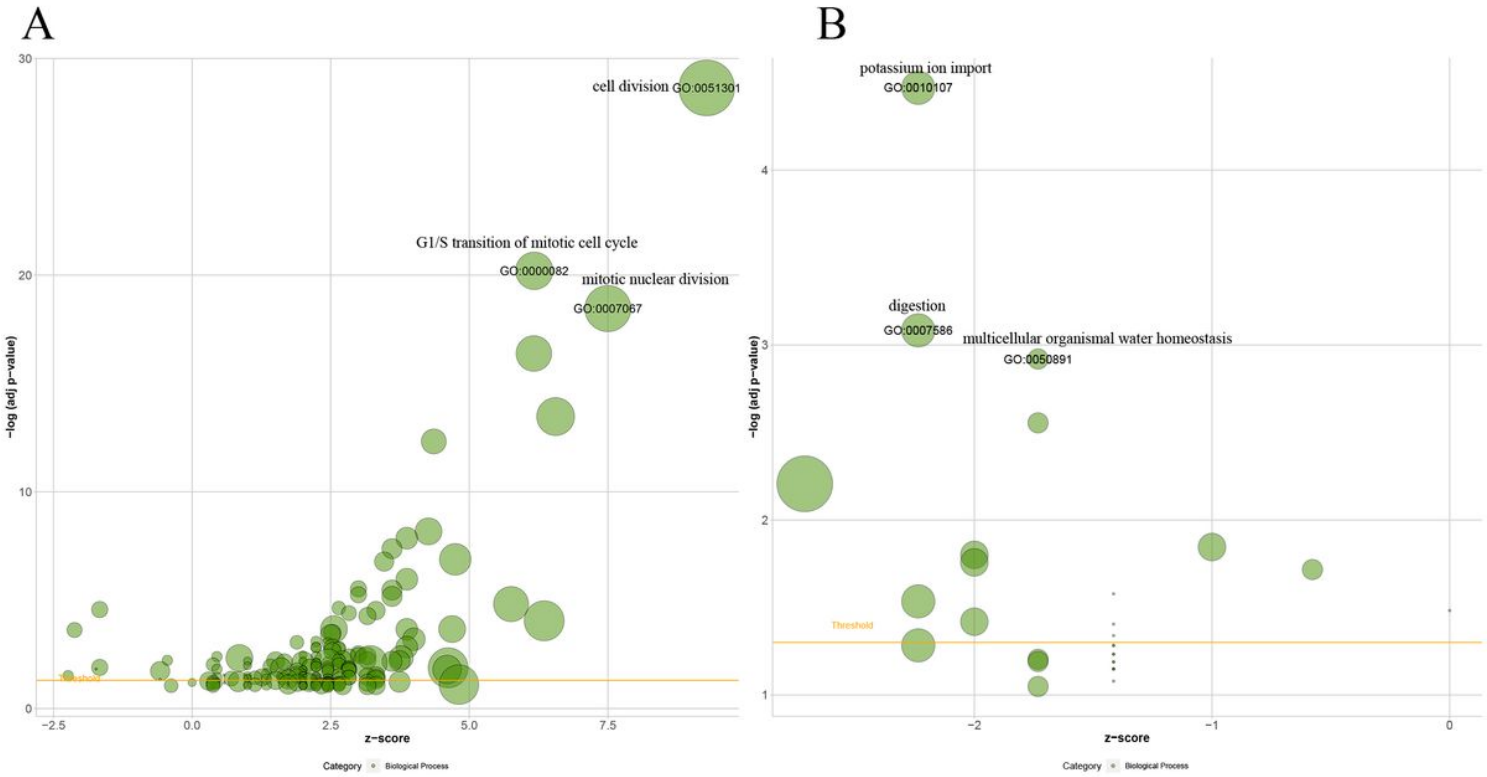

Figure 6

Functional enrichment analyses of biological process of the (A) blue and (B) black modules. The x- and y-axes present the z-score and log (adjusted P-value) of each GO term, respectively. The size of the bubble shows the numbers of the genes enriched in this GO term. Only top 3 GO terms with the most significant adjusted P-values in each module were presented.

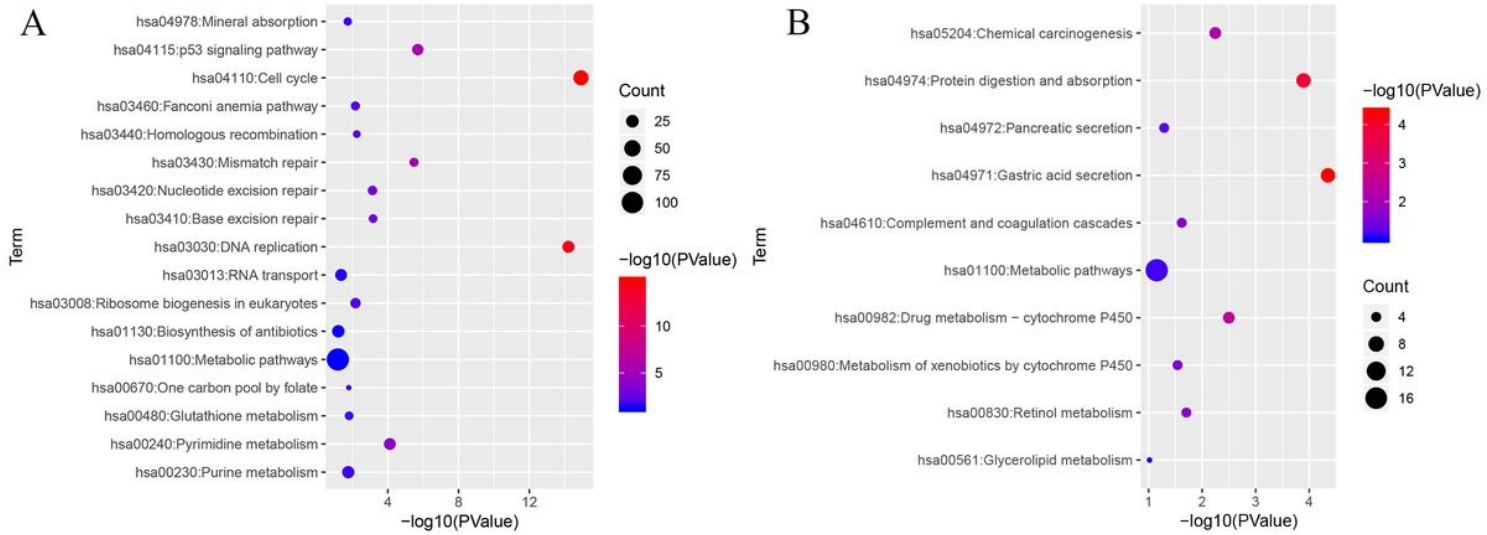

Figure 7

KEGG pathway enrichment analyses of the (A) blue and (B) black modules. The size of the bubble shows the numbers of the genes enriched in this pathway, while the colors indicate the enrichment significance.

Figure 8

Heatmaps showing the differential expression patterns of the candidate genes between the gastric cancer patients and normal controls in training set. The x- and y-axes present the samples and genes, respectively. In the x-axis, pink and green represent the gastric cancer and normal samples, respectively.

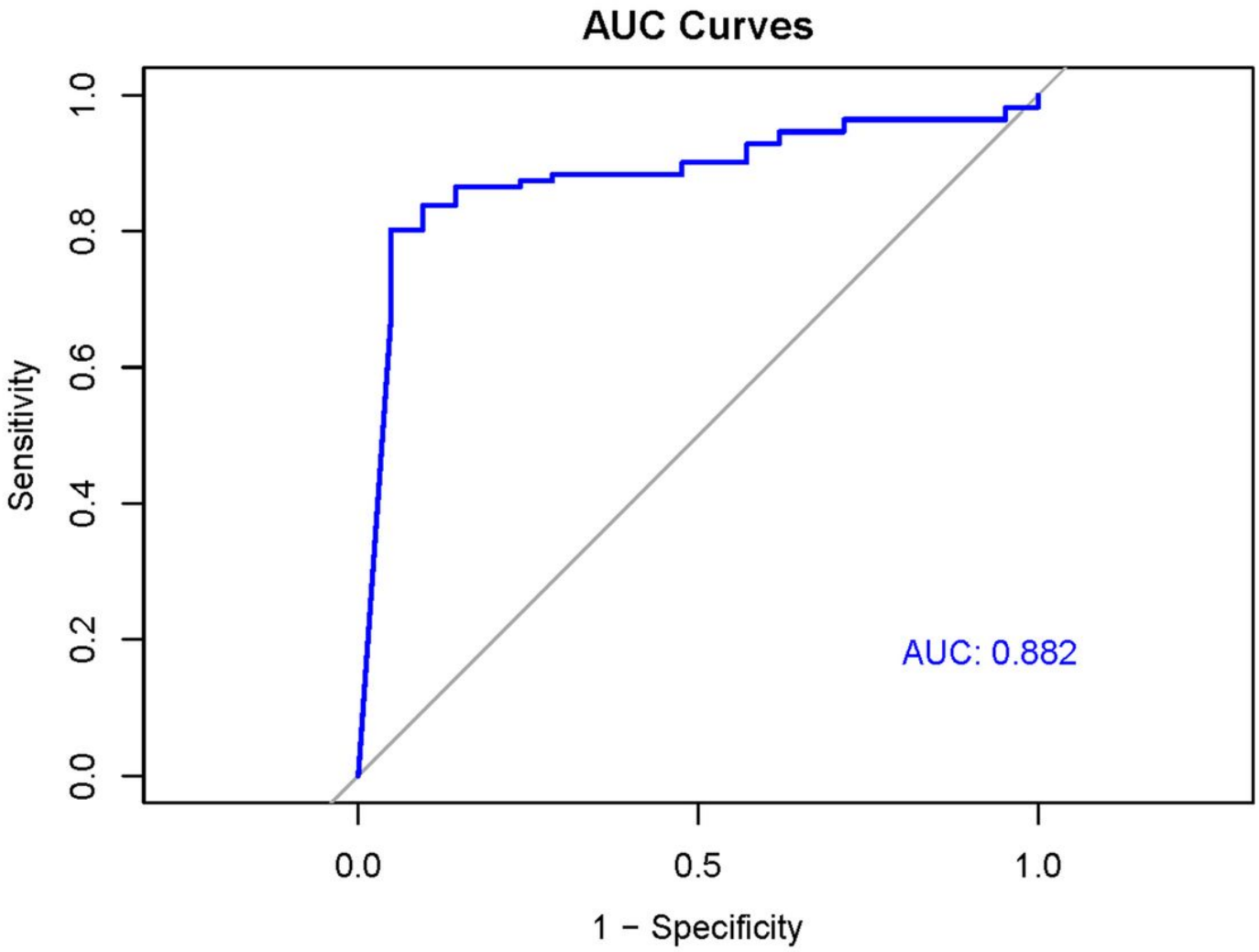

Figure 9

ROC curve of the classifier predicted by these 10 hub genes in the testing set.

Supplementary Files

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

- TableS5.xls
- TableS4.XLS
- TableS3.XLS
- TableS2.XLS
- TableS1.XLS
- ImageS2.TIF
- ImageS1.TIFF