Bioinformatics Analysis and Identification of Potential Biomarkers in Gastric Cancer

Jingdi Yang

Nanchang University - Donghu Campus

Bo Peng

Nanchang University - Donghu Campus

Xianzheng Qin

Nanchang University Medical College: Medical College of Nanchang University

Tian Zhou (✉ zhoutian@ncu.edu.cn)

Nanchang University

Research

Keywords: gastric cancer, differentially expressed genes, microarray, Kaplan-Meier analysis, COL1A1, BGN

Posted Date: November 20th, 2020

DOI: <https://doi.org/10.21203/rs.3.rs-111288/v1>

License: © ⓘ This work is licensed under a Creative Commons Attribution 4.0 International License.

[Read Full License](#)

Abstract

Background: Although the morbidity and mortality of gastric cancer are declining, gastric cancer is still one of the most common causes of death. Early detection of gastric cancer is of great help to improve the survival rate, but the existing biomarkers are not sensitive to diagnose early gastric cancer. The aim of this study is to identify the novel biomarkers for gastric cancer.

Methods: Three gene expression profiles (GSE27342, GSE63089, GSE33335) were downloaded from Gene Expression Omnibus database to select differentially expressed genes. Then, Gene Ontology and Kyoto Encyclopedia of Genes and Genomes analysis were performed to explore the biological functions of differentially expressed genes. Cytoscape was utilized to construct protein-protein interaction network and hub genes were analyzed by plugin cytoHubba of Cytoscape. Furthermore, Gene Expression Profiling Interactive Analysis and Kaplan-Meier plotter were used to verify the identified hub genes.

Results: 35 overlapping differentially expressed genes were screened from gene expression datasets, which consisted of 11 up-regulated genes and 24 down-regulated genes. Gene Ontology functional enrichment analysis revealed that differentially expressed genes were significantly enriched in digestion, regulation of biological quality, response to hormone and steroid hormone, and homeostatic process. Kyoto Encyclopedia of Genes and Genomes pathway enrichment analysis showed differentially expressed genes were enriched in the secretion of gastric acid and collecting duct acid, leukocyte transendothelial migration and ECM-receptor interaction. According to protein-protein interaction network, 10 hub genes were identified by Maximal Clique Centrality method.

Conclusion: By using bioinformatics analysis, COL1A1, BGN, THY1, TFF2 and SST were identified as the potential biomarkers for early detection of gastric cancer.

Background

Gastric cancer (GC) is the third leading cause of cancer related deaths (8.2% of the total cancer deaths) and ranked as the fifth most common cancer (5.7% of the total cases), the number of new cases around the world is increasing every year (1, 2). There has been reported that an estimated 1,033,701 new cases of GC and estimated 782,685 deaths in 2018 (3). The incidence of GC has significant gender differences and male is twice as high as that in female (3). It also presents regional and ethnic variability. Eastern Asia with the highest rate around the world, but the rates in developed countries (eg, in Northern America and Northern Europe) are relatively low, particularly in American White populations (4). While African has the lowest rate. The main risk factor of GC is *Helicobacter pylori* and almost 90% non-cardia gastric cancer caused by *H. pylori* infection (5). Previous literature has shown that the eradication of *H. pylori* can obviously reduce the incidence of GC (6). However, other factors may also be the important risk factors. Dietary salt intake has been indicated as a risk factor of GC, and is associated with increased incidence and mortality (7). In addition, smoking, obesity, other dietary factors and some gastrointestinal disease like gastroesophageal reflux disease and chronic atrophic gastritis

have been identified as risk factors (8). Surgery like curative gastrectomy has been suggested as the only radical cure for early gastric cancer but it also involves the risk of recurrence (9). In contrast to advanced gastric cancer with 10% five-year survival rate, the five-year survival rate of early gastric cancer can be above 95% (10). However, the existing biomarkers can't be effectively diagnosed early gastric cancer and most patients with early gastric cancer present with nonspecific clinical symptoms until the advanced stage (11). It means that patients missed the best period of surgical radical treatment and recurrent GC also reduces the survival rate. Therefore, improving the detection methods of early GC and identifying the effective biomarkers for GC are crucial to enhance the survival rate of patients with gastric cancer.

Biomarker is defined as a biochemical indicator, and has been used widely in the field of clinical diagnosis. Although the existing biomarkers for GC could be used to evaluate the prognosis of GC, recurrence and its distant metastasis, they are not suitable for the detection of early GC because of low specificity and sensitivity, such as carcinoembryonic antigen (CEA) and carbohydrate antigen (CA-199) (12). Recently, there are some studies found that stanniocalcin-1 (STC1) and matrix metalloproteinase-9 (MMP9) have abnormal expression in GC (13, 14). The expression of STC1 was higher in GC, but decreased after surgical treatment. In addition, the sensitivity of STC is better than CEA and CA-199, and it's worth further exploring its clinical value (14). MMP9 expression was increased in GC and it had association with lymph node metastasis (13). Therefore, it suggested that MMP9 could be a potential biomarker for GC detection and could act as a possible molecular target for therapy. Despite these findings, the underlying molecular mechanisms of GC is not clear. The identification of new diagnostic biomarkers should still be carried out actively.

Microarray analysis is a high-throughput technology that has been used widely in the screening of abnormally expressed genes. It provides the possibility for the search for biomarkers and further study on the mechanism of diseases. In this study, three gene expression microarray datasets were analyzed to obtain differentially expressed genes (DEGs) in GC. Subsequently, integrated bioinformatics analysis was performed to help us identify gastric cancer related crucial genes and understand the likely underlying molecular mechanisms of carcinogenesis.

Materials And Methods

Microarray data

The Gene Expression Omnibus (GEO, www.ncbi.nlm.nih.gov/geo) database is an international public repository and it stores the submitted high-throughput microarray datasets (15). Three gene expression profiles (GSE27342, GSE63089, GSE33335) were retrieved from GEO database (Affymetrix GPL5175 platform, Affymetrix Human Exon 1.0 ST Array). Based on the annotation information of the GEO platform, the probes were converted to the corresponding gene symbols. GSE27342 dataset contained 80 GC tissue samples and 80 paired adjacent noncancerous tissue samples (16). GSE63089 dataset contained 45 GC tissue samples and 45 paired adjacent normal tissue samples (17). GSE33335 dataset contained 25 GC tissue samples and 25 paired adjacent noncancerous tissue samples (18).

Screening for DEGs

The limma package (version 3.44.3, <http://bioconductor.org/packages/limma/>) in R programming language was applied to screen the DEGs between GC tissues and matched noncancerous tissues (19). The adjusted P-value (adj. P-value) which based on the Benjamini and Hochberg method was used to reduce false positive results from analysis. The cut-off criteria was set as the adj. P-value <0.05 and $|\log_2\text{FC (fold-change)}| \geq 1.5$, and genes that meet the above conditions can be regarded as DEGs (17). The R package venn (version 1.9, <https://CRAN.R-project.org/package=venn>) was utilized to visualize the overlap of DEGs in different datasets.

Visualization of DEGs

The free R package ggplot2 is a visualization package to construct statistical graphics and it could display the information of the datasets in an efficient and intuitive way. The volcano plot was constructed to show DEGs in different datasets by using the ggplot2 package (version 3.3.2, <https://CRAN.R-project.org/package=ggplot2>) (20).

Gene Ontology (GO) and pathway enrichment analysis of DEGs

The GO (<http://geneontology.org/>) knowledgebase is a public database that annotates genes and collects the information of genes functions (21). In order to explore the possible biological functions of DEGs, the GO knowledgebase was utilized for analysis. KOBAS (version 3.0, <http://kobas.cbi.pku.edu.cn/kobas3>) is an online bioinformatics tool to indicate the correlation among genes and putative pathways and diseases (22). The information of the Kyoto Encyclopedia of Genes and Genomes (KEGG) enrichment analysis was performed by KOBAS. P-value <0.05 was considered to reach statistical significance.

Construction of the protein-protein interaction (PPI) network of DEGs

Search Tool for the Retrieval of Interacting Genes (STRING, version 11.0, <https://string-db.org/>) is publicly available database to explore the information of protein-protein interaction (PPI) and further investigate the cellular mechanism (23). In order to ensure that the relationship among DEGs was statistically significant, the cut-off criteria was set as confidence score >0.400. Cytoscape (version 3.8.1) is an open software platform for biological research, which could integrate molecular interaction networks with gene annotations and other states data (24). Then, the PPI network was constructed by Cytoscape which visualizes DEGs subsets as an interactive network.

Identification and analysis of hub genes

The cytoHubba (version 0.1, <http://apps.cytoscape.org/apps/cytohubba>) is a plugin in Cytoscape and it was utilized to rank the PPI nodes according to the features of the network (25). Maximal Clique Centrality (MCC) method is one of the topological analysis methods in *cytoHubba plugin*, which has

better sensitivity and specificity, and it can help us to screen important nodes in PPI network as hub genes (25). Subsequently, KOBAS was used to analyze the enriched pathway of the identified hub genes.

Comparison of hub genes expression level

Gene Expression Profiling Interactive Analysis (GEPIA, <http://gepia.cancer-pku.cn/>) is a website based on TCGA and GTEx databases which analyzes the relationship of gene expression between cancer and noncancerous control (26). It was utilized to validate the expression level of hub genes between GC tissues and noncancerous tissues, and then the result of comparison was visualized in the form of box plot.

Kaplan–Meier survival analysis

The screened hub genes were processed for survival analysis. Kaplan–Meier plotter (<https://kmplot.com>) is an online web server which contains 10,188 cancer samples and integrates the relative gene expression and clinical trial data (27). The website calculates the hazard ratio (HR) with 95% confidence intervals and logrank P-value in its own R language environment, and makes the corresponding Kaplan–Meier survival curves. A logrank P-value <0.05 was considered to be statistically significant.

Results

Screening and visualization of DEGs in GC

In this study, we downloaded and processed three gene expression profiles (GSE27342, GSE63089, GSE33335). According to the above cut-off criteria (the adj. P-value <0.05 and $|\log_2FC$ (fold-change) ≥ 1.5), 540 DEGs in GSE27342 were selected, 183 DEGs in GSE63089 and 109 DEGs in GSE33335. Figure 1 visualizes DEGs from 3 datasets. The overlap of DEGs among three datasets included 35 genes, which consisted of 11 up-regulated genes and 24 down-regulated genes through comparing GC tissues to matched noncancerous tissues. The result is shown in the Venn diagram (Figure 2).

Table 1. GO and KEGG pathway enrichment analysis	
Pathway ID	Pathway description
GO:0007586	digestion
GO:0065008	regulation of biological quality
GO:0009725	response to hormone
GO:0048545	response to steroid hormone
GO:0042592	homeostatic process
GO:1990169	stress response to copper ion
GO:0097501	stress response to metal ion
GO:0061687	detoxification of inorganic compound
GO:0010273	detoxification of copper ion
GO:0098771	inorganic ion homeostasis
hsa04971	Gastric acid secretion
hsa04966	Collecting duct acid secretion
hsa04670	Leukocyte transendothelial migration
hsa04512	ECM-receptor interaction
hsa04972	Pancreatic secretion
hsa04514	Cell adhesion molecules (CAMs)
hsa05160	Hepatitis C
hsa04530	Tight junction
hsa04510	Focal adhesion
hsa04978	Mineral absorption
Abbreviations: GO: Gene Ontology	

GO and KEGG enrichment analysis of DEGs

In order to further understand the biological functions of 35 overlapping DEGs, GO and KEGG pathway enrichment analysis were performed. GO functional enrichment analysis result showed the identified DEGs were significantly enriched in digestion, regulation of biological quality, response to hormone and steroid hormone, and homeostatic process. Among them, the most significantly enriched GO term was digestion. In addition, KEGG pathway analysis revealed that DEGs were mainly enriched in the secretion of gastric acid and collecting duct acid, leukocyte transendothelial migration, ECM-receptor interaction

and pancreatic secretion. Gastric acid secretion (P-value =1.02E-08) was the most relevant KEGG enriched analysis result. Based on P-value, the most prominent GO and KEGG terms are listed in Table 1. Figure 3 also displays the details of GO and KEGG terms.

Table 2. The top 10 proteins in PPI network ranked by MCC method

Rank	Gene symbol	Score	Expression in GC
1	<i>COL1A1</i>	16	up-regulated
2	<i>BGN</i>	14	up-regulated
3	<i>ATP4A</i>	13	down-regulated
4	<i>SPP1</i>	12	up-regulated
5	<i>TFF2</i>	10	down-regulated
6	<i>GIF</i>	10	down-regulated
7	<i>SST</i>	9	down-regulated
8	<i>GHRL</i>	9	down-regulated
9	<i>ATP4B</i>	7	down-regulated
10	<i>THY1</i>	7	up-regulated

Construction of the PPI network and selection of hub genes

The PPI network which contained 35 nodes and 42 edges was constructed by Cytoscape according to the information of DEGs from String. Figure 4 shows proteins with confidence score of interactions that was higher than 0.400. Moreover, the PPI network was further analyzed by *cytoHubba* plugin in Cytoscape. The top 10 genes of PPI network ranked by MCC method were regarded as hub genes (Figure 5). The hub genes were listed as the following order: COL1A1, BGN, ATP4A, SPP1, TFF2, GIF, SST, GHRL, ATP4B and THY1 (Table 2). Then, to have a better investigation of the hub genes, KOBAS was used to analyze the functional enriched pathways of these genes. The hub genes in the PPI network were significantly enriched for the secretion of gastric acid and collecting duct acid, ECM-receptor interaction and focal adhesion, oxidative phosphorylation and cAMP signaling pathway. Table 3 reveals the enriched pathways in detail.

Table 3. The enriched pathways of the hub genes			
Term	Count	P-value	Genes
Gastric acid secretion	3	9.32E-07	ATP4A, ATP4B, SST
Collecting duct acid secretion	2	2.42E-05	ATP4A, ATP4B
ECM-receptor interaction	2	2.27E-04	BGN, COL1A1
Oxidative phosphorylation	2	5.32E-04	ATP4A, ATP4B
Focal adhesion	2	1.17E-03	BGN, COL1A1
cAMP signaling pathway	2	1.35E-03	GHRL, SST

Comparison of hub genes expression level

In order to validate the expression level of the hub genes between GC tissues and normal tissues, GEPIA website was performed to compare the expression level of these genes. As the box plot in Figure 6 shown, all of these hub genes expression level were statistically significant because of P-value <0.05. The result indicated that COL1A1, BGN, SPP1, THY1 in GC tissues had an increased expression compared with noncancerous tissues. While ATP4A, TFF2, GIF, SST, GHRL and ATP4B in GC tissues revealed lower expression level.

Kaplan–Meier survival analysis of hub genes

The Kaplan–Meier survival analysis was processed by Kaplan–Meier plotter to predict the prognostic value of the hub genes. 881 GC samples from Kaplan–Meier plotter database revealed the result of survival analysis. Based on the result, there were 5 hub genes were identified and could be considered to be associated with overall survival (OS). Among these hub genes that associated with OS, three of them were up-regulated and two down-regulated. In GC patients, the high expression of COL1A1 [HR =1.48 (1.22~1.81), P-value =8.9E-5], BGN [HR =1.9 (1.55~2.31), P-value =1.5E-10] and THY1 [HR =1.8 (1.49~2.17), P-value =4.3E-10] and low expression of TFF2 [HR =0.78 (0.66~0.94), P-value =7.5E-3], SST [HR =0.82 (0.68~0.99), P-value =4.0E-2] had significant association with a poor OS. Besides, SPP1 which had a P-value greater than 0.05 and was not statistically significant. The prognostic survival analysis of the hub genes in GC patients are shown in Figure 7.

Discussion

Gastric cancer is one of the most common malignant tumors around the world and it has high incidence and mortality. It is estimated that the mortality of GC reaches 70% and is obviously higher than other epidemic diseases (28). With the improvement of standards of hygiene and the eradication of *H. pylori*, the incidence rate of GC has shown a decline trend and changed from the fourth most common cancer to

the fifth (29, 30). However, the mortality rate of GC has not been significantly decreased. Due to the application of radical surgery and chemotherapy, the five-year survival rate of early GC can increase up to 90%, but the prognosis of advanced GC is very poor (31). It's because that advanced GC missed the best therapeutic time window and has the potential for distant metastasis (32). Meanwhile, most patients with early GC do not display any symptom is also an important cause and it means that GC patients are difficult to detect at an early stage (33). Although surgical techniques are improving, novel chemotherapy and targeted therapy are developing, there is still no effective treatment for advanced gastric cancer (10). Therefore, the identification of the effective biomarkers for GC and further exploration of their correlative biological functions could help us to detect GC at early stage and understand the pathogenesis of GC, which in turns lead to the emergence of novel targeted therapy.

In this study, three gene expression profiles of GC (GSE27342, GSE63089, GSE33335) were utilized to select DEGs between GC tissues and noncancerous tissues. A total of 35 genes were identified, including 11 up-regulated genes and 24 down-regulated genes. GO and KEGG pathway enrichment analysis were executed to investigate the biological functions of DEGs. The results of GO functional enrichment analysis indicated that 35 genes were mainly enriched in the process of digestion, regulation of biological quality, response to hormone and steroid hormone, and homeostatic process. GC itself is a malignant tumor in the digestive system and most of risk factors are associated with digestion and diet. There are many studies have revealed that obesity acts as a risk factor of GC and could affect digestion (8, 34). While fat has metabolic capacity and could secret leptin and insulin-like growth factor (IGF), the change of leptin and IGF has been reported to be associated with gastrointestinal cancers (35). Moreover, Zhou et al. has proven that the abnormal expression of gastrointestinal hormones like hypergastrinemia have correlation with GC (36). Some previous studies have suggested that the production of steroid hormone could affect the development of GC (37, 38). Estrogen receptor (ER) acts as steroid hormone receptor and the dysregulation of ERs is associated with GC. ER α 36 is a kind of ER, and it could increase the expression of cyclin D1 to promote the proliferation of GC cells (39). In addition, the role of estrogen in GC is related to its concentration. Low concentration of estrogen can increase the expression of ER α 36 and promote GC cells growth, but high concentration of estrogen is on the contrary (40). Homeostatic process has also been demonstrated that is associated with GC (41). Autophagy is a vital homeostatic process and it plays both tumor-suppressor role and tumor-promoter functions in GC (42). One cause of GC development is autophagy could suppress GC cells at early stages and promote the growth of existing tumor (43).

The findings from KEGG pathway enrichment analysis showed DEGs were significantly enriched in the secretion of gastric acid and collecting duct acid, leukocyte transendothelial migration, ECM-receptor interaction and pancreatic secretion. It has been reported that gastric acid secretion is associated with GC and chronic [atrophic gastritis](#) which is regarded as a risk factor of GC (44, 45). The migration characteristics of leukocytes are essential to promote the systemic immune response (46). Enarsson et al. found that the transendothelial migration of T cells in GC patients was reduced and speculated that reduced transendothelial migration of T cells might help GC cells to evade immune responses and was beneficial for carcinogenesis (47). ECM, its full name is extracellular matrix, and the abnormal changes of

ECM contents are associated with the formation of cancer (48). The component of ECM like hyaluronan could play roles in promoting GC development (49). The relevant studies of the above GO and KEGG analysis have been confirmed to be associated with GC, which is also consistent with our results.

According to the PPI network and a total of 10 hub genes are obtained from it. Among these genes, COL1A1, BGN, SPP1 and THY1 were up-regulated and ATP4A, TFF2, GIF, SST, GHRL and ATP4B were down-regulated. We further explore these genes and analyze their possible biological functions. The result showed that ATP4A and ATP4B had correlation with most of enriched pathway. ECM-receptor interaction and focal adhesion were associated with COL1A1 and BGN, which two genes have the highest score in MCC method. Subsequently, we validated the expression level of these hub genes by using GEPIA website. The result as shown in Figure 6, COL1A1, BGN, SPP1, THY1 in GC tissues had an increased expression. While ATP4A, TFF2, GIF, SST, GHRL and ATP4B had a decreased expression level. The verification of expression level in the present study supports the screened hub gene and makes this study more credible. Finally, the Kaplan–Meier survival analysis indicated that up-regulated COL1A1, BGN and THY1, down-regulated TFF2 and SST were associated with a poor OS. It provides a possible research trend of GC, which is worth further exploring.

COL1A1, the full name is collagen type I α 1, and its protein is collagen type I which is a major component of ECM (50). There are many studies have been reported that the overexpression of COL1A1 is related to a variety of cancers, such as colorectal cancer, breast cancer and GC (51-53). Liu et al. found that the abnormal expression of COL1A1 in breast cancer cells promoted cancer metastasis and knockout of COL1A1 could inhibit this process (54). In colorectal cancer patients, overexpressed COL1A1 affected metastasis by regulating the WNT/PCP signaling pathway (51). As a component of ECM, high expression of COL1A1 could increase the stiffness of ECM and contribute to cancer development (55). Furthermore, miR-129-5p has been proved to inhibit the proliferation of GC and distant metastasis by suppressing the expression of COL1A1 (52). COL1A1 has also been recognized as a potential biomarker of GC for early detection at many times (56, 57).

BGN is also the component of ECM, and the upregulation of BGN has been reported as a biomarker for multiple cancers (58). A recent study demonstrated that BGN has a high expression in GC patients and BGN could interact with Toll like receptor 2 and 4 secreted by GC cells to activate the VEGF pathway and promote angiogenesis, which is beneficial to the progression of cancer (59). Therefore, BGN could be regarded as a biomarker of GC that is correlated with poor prognosis. THY-1 acts as a cell surface protein with immunoglobulin domain and it exhibits a high level in GC patients (60, 61). There are some studies has revealed that the formation of GC was associated with THY1 (62). The activation of Notch signaling pathway can promote cancer growth and THY1 stimulates Notch signaling pathway through the regulation of RhoA (63, 64). Furthermore, THY1 has been proven to be a target gene of miR-140-5p (65). Wu et al. has indicated that miR-140-5p can inhibit the development of gastric cancer by inhibiting the expression of THY1 and inactivating Notch signaling pathway (65). According to the above studies, THY1 can be used as a target for the treatment of GC, and it is worth further exploring its value in early detection.

In addition, TFF2 and SST have been reported as tumor suppressors, and low level of TFF2 and SST were found in GC patients (66-68). TFF2 is a short peptide that secreted by gastric mucous neck cells and is helpful for cellular protection of gastrointestinal tract (69). However, the exact mechanism between this protective effect and GC has not been elucidated. Recent studies have indicated that TFF2 could inhibit the development of GC and tumor invasion by interacting with Sp3, therefore down-regulated TFF2 might loss its anti-tumor effect (70). SST, an inhibitory gut peptide, it plays role in regulating cell proliferation and its analog octreotide has been used to inhibit the GC metastasis (71, 72). In GC patients, SST has a decreased expression and its promoter DNA methylation level is higher than normal patients (73). Low expression of SST promotes GC growth, while the inhibition of GC invasion and metastasis by overexpression of SST has been confirmed in vivo experiments (67). Accordingly, we speculate that promoter DNA methylation might regulate the expression of SST and further promotes GC growth through mechanisms that are not yet clear. Although the mechanism of TFF2 and STT in GC has not been elucidated, they are still worthy of being used as biomarkers for early diagnosis of GC.

Conclusions

In the present study, we used bioinformatics analysis to screen out potential biomarkers for early detection of GC. Although there is a lack of in vivo or in vitro experiments on gastric tissue samples, we still need to confirm these selected novel biomarkers. The findings of our study have been verified by multiple methods such as GEPIA and Kaplan–Meier plotter, and has been supported by previous researches. A total of 5 gastric cancer related crucial genes (COL1A1, BGN, THY1, TFF2 and SST) were identified. However, the biological functions of these genes and their exact mechanisms in GC are still worthwhile researching thoroughly.

Availability of data and materials

All data generated or analyzed during this study are included in the published article.

Abbreviations

GC: Gastric

GEO: Gene Expression Omnibus

DEGs: differentially expressed genes

GO: Gene Ontology

KEGG: Kyoto Encyclopedia of Genes and Genomes

GEPIA: Gene Expression Profiling Interactive Analysis

MCC: Maximal Clique Centrality

CEA: carcinoembryonic antigen

CA-199: carbohydrate antigen

STC1: stanniocalcin-1

MMP9: matrix metalloproteinase-9

OS: overall survival

IGF: insulin-like growth factor

ER: Estrogen receptor

Declarations

Acknowledgements

Not applicable.

Funding

No funding was received.

Author information

Jingdi Yang, Bo Peng and Xianzheng Qin contributed equally to this work.

Affiliations

Queen Mary School of Nanchang University, Nanchang, China

Jingdi Yang, Bo Peng and Xianzheng Qin

School of Basic Medical Sciences, Nanchang University, Nanchang, China

Tian Zhou

Corresponding authors

Correspondence to Tianzhou.

Authors' contributions

TZ made substantial contributions to conception and design. JY, XQ made substantial contributions to acquisition of data. BP, JY made substantial contributions to analysis and interpretation of data. BP, JY and XQ have been involved in drafting the manuscript and revising it critically for important intellectual

content. TZ has given final approval of the version to be published. 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. Ferro A, Peleteiro B, Malvezzi M, Bosetti C, Bertuccio P, Levi F, et al. Worldwide trends in gastric cancer mortality (1980-2011), with predictions to 2015, and incidence by subtype. *Eur J Cancer*. 2014;50(7):1330-44.
2. Venerito M, Link A, Rokkas T, Malfertheiner P. Gastric cancer - clinical and epidemiological aspects. *Helicobacter*. 2016;21 Suppl 1:39-44.
3. Bray F, Ferlay J, Soerjomataram I, Siegel RL, Torre LA, Jemal A. Global cancer statistics 2018: GLOBOCAN estimates of incidence and mortality worldwide for 36 cancers in 185 countries. *CA Cancer J Clin*. 2018;68(6):394-424.
4. Siegel R, Ma J, Zou Z, Jemal A. Cancer statistics, 2014. *CA Cancer J Clin*. 2014;64(1):9-29.
5. Plummer M, Franceschi S, Vignat J, Forman D, de Martel C. Global burden of gastric cancer attributable to *Helicobacter pylori*. *Int J Cancer*. 2015;136(2):487-90.
6. Lee YC, Chiang TH, Chou CK, Tu YK, Liao WC, Wu MS, et al. Association Between *Helicobacter pylori* Eradication and Gastric Cancer Incidence: A Systematic Review and Meta-analysis. *Gastroenterology*. 2016;150(5):1113-24 e5.
7. Fock KM, Talley N, Moayyedi P, Hunt R, Azuma T, Sugano K, et al. Asia-Pacific consensus guidelines on gastric cancer prevention. *J Gastroenterol Hepatol*. 2008;23(3):351-65.
8. Karimi P, Islami F, Anandasabapathy S, Freedman ND, Kamangar F. Gastric cancer: descriptive epidemiology, risk factors, screening, and prevention. *Cancer Epidemiol Biomarkers Prev*. 2014;23(5):700-13.
9. Youn HG, An JY, Choi MG, Noh JH, Sohn TS, Kim S. Recurrence after curative resection of early gastric cancer. *Ann Surg Oncol*. 2010;17(2):448-54.
10. Song Z, Wu Y, Yang J, Yang D, Fang X. Progress in the treatment of advanced gastric cancer. *Tumour Biol*. 2017;39(7):1010428317714626.

11. Tsai MM, Wang CS, Tsai CY, Huang HW, Chi HC, Lin YH, et al. Potential Diagnostic, Prognostic and Therapeutic Targets of MicroRNAs in Human Gastric Cancer. *Int J Mol Sci.* 2016;17(6).
12. Necula L, Matei L, Dragu D, Neagu AI, Mambet C, Nedeianu S, et al. Recent advances in gastric cancer early diagnosis. *World J Gastroenterol.* 2019;25(17):2029-44.
13. Chen SZ, Yao HQ, Zhu SZ, Li QY, Guo GH, Yu J. Expression levels of matrix metalloproteinase-9 in human gastric carcinoma. *Oncol Lett.* 2015;9(2):915-9.
14. Fang Z, Tian Z, Luo K, Song H, Yi J. Clinical significance of stanniocalcin expression in tissue and serum of gastric cancer patients. *Chin J Cancer Res.* 2014;26(5):602-10.
15. Barrett T, Wilhite SE, Ledoux P, Evangelista C, Kim IF, Tomashevsky M, et al. NCBI GEO: archive for functional genomics data sets–update. *Nucleic Acids Res.* 2013;41(Database issue):D991-5.
16. Cui J, Chen Y, Chou WC, Sun L, Chen L, Suo J, et al. An integrated transcriptomic and computational analysis for biomarker identification in gastric cancer. *Nucleic Acids Res.* 2011;39(4):1197-207.
17. Zhang X, Ni Z, Duan Z, Xin Z, Wang H, Tan J, et al. Overexpression of E2F mRNAs associated with gastric cancer progression identified by the transcription factor and miRNA co-regulatory network analysis. *PLoS One.* 2015;10(2):e0116979.
18. Cheng L, Wang P, Yang S, Yang Y, Zhang Q, Zhang W, et al. Identification of genes with a correlation between copy number and expression in gastric cancer. *BMC Med Genomics.* 2012;5:14.
19. Ritchie ME, Phipson B, Wu D, Hu Y, Law CW, Shi W, et al. limma powers differential expression analyses for RNA-sequencing and microarray studies. *Nucleic Acids Res.* 2015;43(7):e47.
20. Ito K, Murphy D. Application of ggplot2 to Pharmacometric Graphics. *CPT Pharmacometrics Syst Pharmacol.* 2013;2:e79.
21. Ashburner M, Ball CA, Blake JA, Botstein D, Butler H, Cherry JM, et al. Gene ontology: tool for the unification of biology. The Gene Ontology Consortium. *Nat Genet.* 2000;25(1):25-9.
22. Xie C, Mao X, Huang J, Ding Y, Wu J, Dong S, et al. KOBAS 2.0: a web server for annotation and identification of enriched pathways and diseases. *Nucleic Acids Res.* 2011;39(Web Server issue):W316-22.
23. Szklarczyk D, Gable AL, Lyon D, Junge A, Wyder S, Huerta-Cepas J, et al. STRING v11: protein-protein association networks with increased coverage, supporting functional discovery in genome-wide experimental datasets. *Nucleic Acids Res.* 2019;47(D1):D607-D13.
24. Shannon P, Markiel A, Ozier O, Baliga NS, Wang JT, Ramage D, et al. Cytoscape: a software environment for integrated models of biomolecular interaction networks. *Genome Res.* 2003;13(11):2498-504.
25. Chin CH, Chen SH, Wu HH, Ho CW, Ko MT, Lin CY. cytoHubba: identifying hub objects and sub-networks from complex interactome. *BMC Syst Biol.* 2014;8 Suppl 4:S11.
26. Tang Z, Li C, Kang B, Gao G, Li C, Zhang Z. GEPIA: a web server for cancer and normal gene expression profiling and interactive analyses. *Nucleic Acids Res.* 2017;45(W1):W98-W102.

27. Gyorffy B, Lanczky A, Eklund AC, Denkert C, Budczies J, Li Q, et al. An online survival analysis tool to rapidly assess the effect of 22,277 genes on breast cancer prognosis using microarray data of 1,809 patients. *Breast Cancer Res Treat.* 2010;123(3):725-31.
28. Guggenheim DE, Shah MA. Gastric cancer epidemiology and risk factors. *J Surg Oncol.* 2013;107(3):230-6.
29. Sitarz R, Skierucha M, Mielko J, Offerhaus GJA, Maciejewski R, Polkowski WP. Gastric cancer: epidemiology, prevention, classification, and treatment. *Cancer Manag Res.* 2018;10:239-48.
30. Machlowska J, Baj J, Sitarz M, Maciejewski R, Sitarz R. Gastric Cancer: Epidemiology, Risk Factors, Classification, Genomic Characteristics and Treatment Strategies. *Int J Mol Sci.* 2020;21(11).
31. Tan Z. Recent Advances in the Surgical Treatment of Advanced Gastric Cancer: A Review. *Med Sci Monit.* 2019;25:3537-41.
32. Coccolini F, Cotte E, Glehen O, Lotti M, Poiasina E, Catena F, et al. Intraperitoneal chemotherapy in advanced gastric cancer. Meta-analysis of randomized trials. *Eur J Surg Oncol.* 2014;40(1):12-26.
33. Wagner AD, Unverzagt S, Grothe W, Kleber G, Grothey A, Haerting J, et al. Chemotherapy for advanced gastric cancer. *Cochrane Database Syst Rev.* 2010(3):CD004064.
34. Wolin KY, Carson K, Colditz GA. Obesity and cancer. *Oncologist.* 2010;15(6):556-65.
35. Aleman JO, Eusebi LH, Ricciardiello L, Patidar K, Sanyal AJ, Holt PR. Mechanisms of obesity-induced gastrointestinal neoplasia. *Gastroenterology.* 2014;146(2):357-73.
36. Zhou G, Yang J. Correlations of gastrointestinal hormones with inflammation and intestinal flora in patients with gastric cancer. *J BUON.* 2019;24(4):1595-600.
37. Frycz BA, Murawa D, Borejsza-Wysocki M, Wichtowski M, Spychala A, Marciniak R, et al. mRNA expression of steroidogenic enzymes, steroid hormone receptors and their coregulators in gastric cancer. *Oncol Lett.* 2017;13(5):3369-78.
38. Camargo MC, Goto Y, Zabaleta J, Morgan DR, Correa P, Rabkin CS. Sex hormones, hormonal interventions, and gastric cancer risk: a meta-analysis. *Cancer Epidemiol Biomarkers Prev.* 2012;21(1):20-38.
39. Deng H, Huang X, Fan J, Wang L, Xia Q, Yang X, et al. A variant of estrogen receptor-alpha, ER-alpha36 is expressed in human gastric cancer and is highly correlated with lymph node metastasis. *Oncol Rep.* 2010;24(1):171-6.
40. Ur Rahman MS, Cao J. Estrogen receptors in gastric cancer: Advances and perspectives. *World J Gastroenterol.* 2016;22(8):2475-82.
41. Cao Y, Luo Y, Zou J, Ouyang J, Cai Z, Zeng X, et al. Autophagy and its role in gastric cancer. *Clin Chim Acta.* 2019;489:10-20.
42. Qian HR, Yang Y. Functional role of autophagy in gastric cancer. *Oncotarget.* 2016;7(14):17641-51.
43. Maes H, Rubio N, Garg AD, Agostinis P. Autophagy: shaping the tumor microenvironment and therapeutic response. *Trends Mol Med.* 2013;19(7):428-46.

44. Saenz JB, Mills JC. Acid and the basis for cellular plasticity and reprogramming in gastric repair and cancer. *Nat Rev Gastroenterol Hepatol*. 2018;15(5):257-73.
45. Di Mario F, Goni E. Gastric acid secretion: changes during a century. *Best Pract Res Clin Gastroenterol*. 2014;28(6):953-65.
46. Schimmel L, Heemskerk N, van Buul JD. Leukocyte transendothelial migration: A local affair. *Small GTPases*. 2017;8(1):1-15.
47. Enarsson K, Lundin BS, Johnsson E, Brezicka T, Quiding-Jarbrink M. CD4⁺ CD25^{high} regulatory T cells reduce T cell transendothelial migration in cancer patients. *Eur J Immunol*. 2007;37(1):282-91.
48. Najafi M, Farhood B, Mortezaee K. Extracellular matrix (ECM) stiffness and degradation as cancer drivers. *J Cell Biochem*. 2019;120(3):2782-90.
49. Zhang H, Ren L, Ding Y, Li F, Chen X, Ouyang Y, et al. Hyaluronan-mediated motility receptor confers resistance to chemotherapy via TGFβ₂/Smad2-induced epithelial-mesenchymal transition in gastric cancer. *FASEB J*. 2019;33(5):6365-77.
50. Maasalu K, Nikopoulou T, Koks S, Noukas M, Kals M, Prans E, et al. Whole-exome sequencing identifies de novo mutation in the COL1A1 gene to underlie the severe osteogenesis imperfecta. *Hum Genomics*. 2015;9:6.
51. Zhang Z, Wang Y, Zhang J, Zhong J, Yang R. COL1A1 promotes metastasis in colorectal cancer by regulating the WNT/PCP pathway. *Mol Med Rep*. 2018;17(4):5037-42.
52. Wang Q, Yu J. MiR-129-5p suppresses gastric cancer cell invasion and proliferation by inhibiting COL1A1. *Biochem Cell Biol*. 2018;96(1):19-25.
53. Meng C, He Y, Wei Z, Lu Y, Du F, Ou G, et al. MRTF-A mediates the activation of COL1A1 expression stimulated by multiple signaling pathways in human breast cancer cells. *Biomed Pharmacother*. 2018;104:718-28.
54. Liu J, Shen JX, Wu HT, Li XL, Wen XF, Du CW, et al. Collagen 1A1 (COL1A1) promotes metastasis of breast cancer and is a potential therapeutic target. *Discov Med*. 2018;25(139):211-23.
55. Katara GK, Kulshrestha A, Mao L, Wang X, Sahoo M, Ibrahim S, et al. Mammary epithelium-specific inactivation of V-ATPase reduces stiffness of extracellular matrix and enhances metastasis of breast cancer. *Mol Oncol*. 2018;12(2):208-23.
56. Cao L, Chen Y, Zhang M, Xu DQ, Liu Y, Liu T, et al. Identification of hub genes and potential molecular mechanisms in gastric cancer by integrated bioinformatics analysis. *PeerJ*. 2018;6:e5180.
57. Li J, Ding Y, Li A. Identification of COL1A1 and COL1A2 as candidate prognostic factors in gastric cancer. *World J Surg Oncol*. 2016;14(1):297.
58. Hu L, Duan YT, Li JF, Su LP, Yan M, Zhu ZG, et al. Biglycan enhances gastric cancer invasion by activating FAK signaling pathway. *Oncotarget*. 2014;5(7):1885-96.
59. Hu L, Zang MD, Wang HX, Li JF, Su LP, Yan M, et al. Biglycan stimulates VEGF expression in endothelial cells by activating the TLR signaling pathway. *Mol Oncol*. 2016;10(9):1473-84.

60. Rege TA, Hagood JS. Thy-1 as a regulator of cell-cell and cell-matrix interactions in axon regeneration, apoptosis, adhesion, migration, cancer, and fibrosis. *FASEB J.* 2006;20(8):1045-54.
61. Zhu GC, Gao L, He J, Long Y, Liao S, Wang H, et al. CD90 is upregulated in gastric cancer tissues and inhibits gastric cancer cell apoptosis by modulating the expression level of SPARC protein. *Oncol Rep.* 2015;34(5):2497-506.
62. Wang CJ, Zhang ZZ, Xu J, Wang M, Zhu CC, Zhuang C, et al. THY-1 (CD90) expression promotes the growth of gastric cancer cells. *Int J Clin Exp Pathol.* 2017;10(9):9878-88.
63. Du X, Cheng Z, Wang YH, Guo ZH, Zhang SQ, Hu JK, et al. Role of Notch signaling pathway in gastric cancer: a meta-analysis of the literature. *World J Gastroenterol.* 2014;20(27):9191-9.
64. Venkatesh D, Fredette N, Rostama B, Tang Y, Vary CP, Liaw L, et al. RhoA-mediated signaling in Notch-induced senescence-like growth arrest and endothelial barrier dysfunction. *Arterioscler Thromb Vasc Biol.* 2011;31(4):876-82.
65. Wu K, Zou J, Lin C, Jie ZG. MicroRNA-140-5p inhibits cell proliferation, migration and promotes cell apoptosis in gastric cancer through the negative regulation of THY1-mediated Notch signaling. *Biosci Rep.* 2019;39(7).
66. Katoh M. Trefoil factors and human gastric cancer (review). *Int J Mol Med.* 2003;12(1):3-9.
67. Wang GH, Zhou YM, Yu Z, Deng JP, Liu SF, Wei CZ, et al. Up-regulated ONECUT2 and down-regulated SST promote gastric cell migration, invasion, epithelial-mesenchymal transition and tumor growth in gastric cancer. *Eur Rev Med Pharmacol Sci.* 2020;24(18):9378-90.
68. Wu CS, Wei KL, Chou JL, Lu CK, Hsieh CC, Lin JM, et al. Aberrant JAK/STAT Signaling Suppresses TFF1 and TFF2 through Epigenetic Silencing of GATA6 in Gastric Cancer. *Int J Mol Sci.* 2016;17(9).
69. Hanisch FG, Bonar D, Schloerer N, Schroten H. Human trefoil factor 2 is a lectin that binds alpha-GlcNAc-capped mucin glycans with antibiotic activity against *Helicobacter pylori*. *J Biol Chem.* 2014;289(40):27363-75.
70. Cai Y, Yi M, Chen D, Liu J, Guleng B, Ren J, et al. Trefoil factor family 2 expression inhibits gastric cancer cell growth and invasion in vitro via interactions with the transcription factor Sp3. *Int J Mol Med.* 2016;38(5):1474-80.
71. Sun L, Coy DH. Somatostatin and its Analogs. *Curr Drug Targets.* 2016;17(5):529-37.
72. Huang MT, Chen ZX, Wei B, Zhang B, Wang CH, Huang MH, et al. [The inhibitive effects of celecoxib combined with octreotide on the metastasis of human gastric cancer in vivo]. *Zhonghua Yi Xue Za Zhi.* 2006;86(46):3255-9.
73. Jackson K, Soutto M, Peng D, Hu T, Marshal D, El-Rifai W. Epigenetic silencing of somatostatin in gastric cancer. *Dig Dis Sci.* 2011;56(1):125-30.

Figures

Figure 1

Visualization of DEGs. Volcano plot shows DEGs (the adj. P-value < 0.05 and $|\log_2\text{FC (fold-change)}| \geq 1.5$) between GC tissue samples and noncancerous tissue samples. The red dots represent statistical significance.

Figure 2

Venn diagram of DEGs. The overlap of DEGs among GSE27342, GSE63089, GSE33335 were displayed. The 3 datasets show an overlap of 35 DEGs.

Figure 3

GO and KEGG enrichment analysis of DEG. The most statistically significant GO and KEGG enrichment analysis of DEG in GC. (A) GO functional enrichment analysis of top 10; (B) KEGG pathway analysis of top 10.

Figure 4

PPI network construction. The PPI network of DEGs was visualized by Cytoscape. The size of Node is regulated by Degree, and the thickness of Edge is regulated by combined score.

Figure 5

The most significant module of PPI. The most significant module was selected as the hub gene by MCC method. The depth of the color represents the size of the score.

Figure 6

Expression levels of the hub genes. Increased expression of (A) COL1A1, (B) BGN, (D) SPP1 and (J) THY1 in GC. Decreased expression of (C) ATP4A, (E) TFF2, (F) GIF, (G) SST, (H) GHRL and (I) ATP4B in GC. *P-value < 0.05

Figure 7

Kaplan-Meier survival analysis of the hub genes. The prognostic survival analysis of the hub genes in GC patients. The red lines represent individuals with high expression of genes and black lines represent

individuals with low expression.