Functional Modules Analysis and Hub Gene Prognostic Values Evaluation Based on Co-Expression Network in Gastric Cancer

Yujia Liu

Zhejiang Provincial People's Hospital

Xiaoping Hu

Zhejiang Provincial People's Hospital

Zongfu Pan

Zhejiang Provincial People's Hospital

Yuchen Jiang

Zhejiang Provincial People's Hospital

Dandan Guo

Zhejiang Provincial People's Hospital

Tong Xu

Zhejiang Provincial People's Hospital

Jiao Sun

Zhejiang Cancer Hospital

Ping Huang

Zhejiang Provincial People's Hospital

Yiwen Zhang (✉ zjzyw2003@163.com)

Zhejiang Provincial People's Hospital

Research

Keywords: gastric cancer, differential expression genes, bioinformatics analysis.

Posted Date: June 17th, 2021

DOI: <https://doi.org/10.21203/rs.3.rs-600479/v1>

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

[Read Full License](#)

Abstract

Background: Gastric cancer is one of the most common fatal disease worldwide, but its mechanism and therapeutic targets are still unclear. In this study, we have analyzed the differences in gene modules and key pathways in gastric cancer patients, then elaborated the mechanism and effective treatment of gastric cancer with microarray data from the gene expression omnibus(GEO) database.

Methods: GEO2R tools were used to identify differential expression genes (DEGs), String database was employed to construct a protein-protein interaction (PPI) network. We imported the PPI network into the Cytoscape software to find key nodes, and employed statistical approach of MCODE to cluster genes. After that the ClueGO was used to enrich and annotate the pathways of key modules. To investigate the relationship between the upstream regulator and hub genes, the transcriptional regulatory network was built based on TFCAT database.

Results: 63 characteristic genes of gastric cancer are involved in regulation of ECM-receptor interaction, focal adhesion and protein digestion and absorption. SPARC, FN1, BGN and COL1A2 are four key nodes relating to tumor proliferation and metastasis, and their expression were strongly associated with poor survival ($p<0.05$). 13 transcription factors including PRRX1 have remarkable changes in gastric cancer, which may play a key role in hub gene regulation.

Conclusions: The present study defined the gene expression characteristics and transcriptional regulatory network that promote our understanding of the molecular mechanisms underlying the development of gastric cancer, and might provide new insights into targeted therapy and prognostic markers for the personalized treatment of gastric cancer.

1 Background

Gastric cancer, which is the fifth most common cancer worldwide and the second leading cause of cancer-related death in developed countries, affects about one million people per year ⁽¹⁾. Most cases, especially in developed countries, are diagnosed at a late stage, and curative treatment options are mostly restricted to surgery and often supplemented by neoadjuvant or adjuvant chemotherapy ⁽²⁾. However, despite significant advances in diagnosis and staging as well as the introduction of new treatment protocols, the rate of gastric cancer recurrences is still high. The long-term survival of patients affected by advanced gastric cancer following standard treatments is still far from satisfactory: 5-year survival rate ranges between 5 and 20% and median overall survival (OS) is less than 12 months⁽³⁾.

Thus, there is an urgent need to improve our understanding of the epidemiology, pathogenesis and molecular mechanism of gastric cancer and to identify more effective, less toxic therapeutic strategies. Gene expression profiles combined with bioinformatics analysis have shown great application prospects in explore diagnosis and prognosis markers for complex disease. The high-throughput platforms for analysis of gene expression, such as microarrays, has been used to detect the expression level of genes in cells and tissues which can help to screen the crucial genes and pathways associated with disease.

Microarray technology combined with bioinformatics analysis is able to analyze the pathways, biological processes, or molecular functions which the DEGs were involved in⁽⁴⁾.

In the current study, using microarray data from GEO database, we firstly identified DEGs between gastric cancer and normal tissues, detected the crucial genes and pathways associated with gastric cancer using bioinformatics methods. Then we investigated the relationship between the crucial genes and gastric cancer patient survival to improve our understanding gastric cancer and to identify more effective, less toxic therapeutic strategies.

2 Results

As mentioned earlier, we concentrated on the gastric cancer sample with the clinical data available in four training datasets. The detailed information for the patients, which were obtained from GEO matrix files, included in this study is listed in Table 1.

Table 1
Clinical characteristics of the study population

Clinicopathologic variables		n
Age	≤ 40	23
	> 40, ≤ 60	173
	> 60	145
	Unknown	56
Gender	Male	243
	Female	98
	Unknown	56
Tissue types	Tumor	351
	Normal	46
Lauren classification	Diffuse	73
	Intestinal	200
	Mixed	9
	Unknown	115
Ajcc Stage	1	60
	2	56
	3	90
	4	87
	Unknown	104
Adjuvant Chemotherapy	Yes	178
	No	95
	Unknown	124

2.1 identification of DEGs from the training sets

A total of 2636 DEGs were identified from training datasets including up-regulated genes and down-regulated genes in gastric carcinoma tissues (Fig. 1). Cluster analysis of validation dataset (GSE79973) shows that the expression of DEGs were significantly different between gastric carcinomas and normal mucosa (Fig. 2).

2.2 PPI network of DEGs

The PPI network of DEGs were constructed using String database (<https://string-db.org/>) (Fig. 3). Each node is a differentially expressed gene and the node degree is reflected by the size: larger degree value often corresponds to large size. We used MCODE plugin to enrich the module of DEGs which was shown in Fig. 4.

ClueGO was used to identify the key biological role of module. Genes involved in Platelet Degranulation, collagen fibril organization, endodermal cell differentiation, Nuclear Chromosome Segregation (Fig. 5).

2.3 The effect of the expression level of the crucial genes on patient survival

Survival analysis based on TCGA data showed that the high expression key nodes (SPARC, FN1, BGN, COL1A2) lead to a significantly decrease of the survival rate in gastric cancer patients ($p < 0.05$, Fig. 6).

2.4 The transcriptional regulatory network of DEGs

We built the transcriptional regulatory network of DEGs using data from TFCAT database, and detected 13 transcription factors of DEGs in gastric cancer, of which ASCL1, ESR2, ESRRB, GATA5, PAX8, MLLT1 are up-regulated transcription factors and PRRX1, HOXC10, NKX3-2, SALL4, HOXC6, HOXA10 are down-regulated transcription factors(Fig. 7).

2.5 GO and KEGG pathway enrichment analysis of DEGs

GO enrichment analysis showed that DEGs were significantly enriched in biological processes (BPs). As for up-regulated DEGs, the top three BPs are 'Extracellular matrix organization', 'Extracellular Structure Organization', and 'Collagen catabolic process' (Fig. 8a), whereas the main BPs of down-regulated DEGs include 'digestion', 'detoxification of copper ion' and 'stress response to copper ion (Fig. 8b).

KEGG pathway enrichment analysis indicated that DEGs were significantly enriched in at least ten pathways. The top three pathways of up-regulated DEGs are 'ECM-receptor interaction', 'focal adhesion'and 'Protein digestion and absorption' (Fig. 9a), while the top three pathways of down-regulated DEGs included 'chemical carcinogenesis', 'metabolism of xenobiotics by cytochrome P450', and 'Drug metabolism-cytochrome P450' (Fig. 9b).

3 Discussion

In the present study, we explored critical genes in gastric cancer based on bioinformatics methods. We integrated four original microarray datasets associated with gastric cancer (three training datasets, one validation dataset) and identified 2636 DEGs between gastric cancer tissues and normal tissues. PPI network analysis and microarray results ($p = 0.0095, 0.03, 0.031, 0.047$) indicated that SPARC, FN1, BGN and COL1A2 were key nodes of gastric cancer and significantly up-regulated in patients with less survival rate. These results showed that SPARC, FN1, BGN and COL1A2 had promoting effects in the happening and development of gastric cancer. SPARC could promote development of some tumors with highly metastatic characteristics, such as breast cancer and melanoma⁽⁵⁾. FN1 is involved in cell adhesion and

migration processes including embryogenesis, wound healing, blood coagulation, host defense, metastasis, as well as various biochemical processes⁽⁶⁾, which has been found to suppress apoptosis and promote viability, invasion, and migration in CRC by interacting with ITGA5. We also observed the potential role of COL1A2 in gastric cancer, collagens are important for cell adhesion and migration, angiogenesis, tissue morphogenesis, scaffolding and has been viewed as involving carcinogenesis^(7,8). However, little has been known about the role of collagens in gastric cancer.

Transcription factors (TF) control production rate of messenger RNA by binding to a transcription factor binding site on a DNA template to regulate the expression of genes in response to signals inside or outside the cell. Genes regulated by transcription factors are referred to as target genes (TGs). Transcription factors themselves are proteins encoded by genes, and thus are also regulated by other TFs. In this way, all the regulatory interactions which link transcription factors to their corresponding target genes are built into the transcriptional regulatory network (TRN)⁽⁹⁾. We used TFCAT database to conduct TFN in gastric cancer and identified that there were 13 TFs involved in Gastric cancer progression. Paired-related homeobox 1 (PRRX1) was down-regulated in gastric cancer which has been identified as a new EMT inducer⁽¹⁰⁾. Although the function of PRRX1 in gastric cancer was not well elucidated, PRRX1 expression levels were upregulated and positively correlated with metastasis in human gastric cancer. In addition, our GO and KEGG pathway analyses indicated that DEGs were involved in 'ECM-receptor interaction', 'focal adhesion', 'Protein digestion and absorption' (Fig. 9), and ECM-receptor interaction and focal adhesion pathway may be mediated by key nodes COL1A2 and key TF PRRX1.

With a microarray data set from the GEO database and multiple bioinformatics methods, we explored the crucial genes and pathways in gastric cancer, which not only contributes to elucidating the pathogenesis of gastric cancer, but also provides prognostic markers and potential therapeutic targets for gastric cancer.

4 Conclusions

In summary, we screened the DEGs of microarray data from GEO database and constructed the co-expression network to illustrate the underlying mechanism of gastric cancer progression. Four key nodes relating to tumor proliferation and metastasis and their expression were strongly associated with poor survival. The transcriptional regulatory network analysis found that 13 transcription factors including PRRX1 may play a key role in hub gene regulation. Our study not only contributes to elucidating the pathogenesis of gastric cancer, but also provides prognostic markers and potential therapeutic targets for gastric cancer.

5 Materials And Methods

5.1 microarray data

Four Training datasets (GSE26899 (n = 108), GSE26942 (n = 217), GSE33651 (n = 52), GSE79973 (n = 20)) associated with gastric cancer were extracted from GEO database (<http://www.ncbi.nlm.nih.gov/geo/>; access time: Mar.26, 2018; species: homo). Three microarray datasets (GSE26899, GSE26942, and GSE33651) were utilized as training datasets in order to identify the DEGs between gastric cancer tissue and normal tissue. Gene expression profile (GSE79973) was added as validation datasets. If multiple probe sets correspond to one gene, the mean expression values of those probe sets were obtained.

5.2 identification of DEGs

GE02R was used to identify DEGs between gastric cancer tissues and non-cancerous tissues (<https://www.ncbi.nlm.nih.gov/geo/geo2r/>; B-statistic of zero corresponds to a 50–50 chance that the gene is differentially expressed. The B-statistic is automatically adjusted for multiple testing by assuming that 1% of the genes, or some other percentage specified by the user in the call to eBayes; P-value was adjusted by the BH approach). The DEGs were screened out according to adjusted P-value < 0.05 and $2^{\log_2(\text{Fold change})} > 1$.

5.3 analysis of the PPI network of DEGs

String database (www.string-db.org) was used to predict protein–protein association data and construct the PPI network of DEGs. Cytoscape software V3.6.0 (<http://www.cytoscape.org/>) was used to display the PPI network. Centiscape2.2 plugin was utilized to identify the degree centrality (DC) of gene nodes, and the DEGs were enriched in several modules by MCODE plugin^(11, 12).

5.4 the effect of the expression level of the crucial genes on patient survival

We used the UALCAN database (<http://ualcan.path.uab.edu/index.html>) to analysis cancer transcriptome and clinical data from The Cancer Genome Atlas (TCGA), which includes the effect of Key nodes expression level and clinic pathologic features on patient survival. We then estimated the relationship of the crucial genes and gastric cancer patient survival by Kaplan-Meier curve and Log-rank test.

5.5 constructing the transcriptional regulatory network of DEGs

The transcriptional regulatory network of DEGs was derived from the curated catalog of mouse and human transcription factors database⁽¹³⁾. Each TF in the collection is assigned to one or more functional taxa with a confidence judgement by an evaluating scientist based on literature review. The confirmed TFs were used in a sequence-oriented homology analysis to predict additional TFs.

5.6 gene ontology (GO) and pathway enrichment analysis of DEGs

Visualization and Integrated Discovery (DAVID) v6.8 was used for GO analysis (<https://david.ncifcrf.gov/>)⁽¹⁴⁾. The Kyoto Encyclopedia of Genes and Genomes (KEGG) knowledge database wastilized for

biochemistry pathway enrichment. We used R Profiler V3.6.0 to annotate and visualize the KEGG pathways of DEGs. We selected GO terms and pathways that the DEGs mainly enriched in with a cut-off criteria of $p < 0.05$.

Abbreviations

GEO: gene expression omnibus; DEGs: differential expression genes ; PPI: protein-protein interaction; OS: overall survival; DC: degree centrality; TCGA: The Cancer Genome Atlas; GO: Gene Ontology; DAVID: The Database for Annotation, Visualization and Integrated Discovery; BPs: biological processes; TF: Transcription factors; TGs: target genes; TRN: transcriptional regulatory network; PRRX1: Paired-related homeobox 1.

Declarations

Data Availability:

Datasets GSE26899 (n=108) was extracted from GEO database (<https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE26899>); Datasets GSE26942 (n=217) was extracted from GEO database (<https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE26942>); Datasets GSE33651 (n=52) was extracted from GEO database (<https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE33651>); Datasets GSE79973 (n=20) was extracted from GEO database (<https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE79973>).

Funding Statement

This study was supported by Zhejiang Provincial Natural Science Foundation of China, Zhejiang Province, People's Republic of China (grant number: LYQ20H310001, LY20H310001), the Medical and Health Research Program of Zhejiang Province (grant number: 2018KY297, 2019KY047, 2021KY491), the Zhejiang Provincial Program for the Cultivation of New Health Talents (to Yiwen Zhang).

Acknowledgements

We acknowledge TCGA and GEO database for providing their platforms and contributors for uploading their meaningful datasets.

Authors' contributions

Yiwen Zhang and Ping Huang conceived and designed the study. Yuchen Jiang, Dandan Guo and Zongfu Pan collected, processed and validated data. Xiaoping Hu, Tong Xu and Jiao Sun analyzed data and prepared figures. Yujia Liu drafted the manuscript.

Statements

Some preliminary data of this manuscript rather than the whole has been submitted as poster presentation in ESMO (EUROPEAN SOCIETY FOR MEDICAL ONCOLOGY) meetings, and then the journal of “ANNALS OF ONCOLOGY” summarized the poster presented in the ESMO as a SUPPLEMENT issue.

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. Ferlay J, Soerjomataram I, Dikshit R, Eser S, Mathers C, Rebelo M, et al. Cancer incidence and mortality worldwide: Sources, methods and major patterns in GLOBOCAN 2012. *International Journal of Cancer*. 2015.
2. Reddavid R, Sofia S, Chiaro P, Colli F, Trapani R, Esposito L, et al. Neoadjuvant chemotherapy for gastric cancer. Is it a must or a fake? *World J Gastroenterol*. 2018;24(2):274–89.
3. Mattiuzzi C, Lippi G. Current Cancer Epidemiology. *Journal of Epidemiology & Global Health*. 2019;9(4).
4. Kulasingam V, Diamandis EP. Strategies for discovering novel cancer biomarkers through utilization of emerging technologies. *Nat Clin Pract Oncol*. 2008;5(10):588–99.
5. Nagaraju GPC, Sharma D. Anti-cancer role of SPARC, an inhibitor of adipogenesis. *Cancer Treat Rev*. 2011;37(7):559–66.
6. Cai X, Liu C, Zhang T-N, Zhu Y-W, Dong X, Xue P. Down-regulation of FN1 inhibits colorectal carcinogenesis by suppressing proliferation, migration, and invasion. *J Cell Biochem*. 2018;119(6):4717–28.
7. Zhao Y, Zhou T, Li A, Yao H, He F, Wang L, et al. A potential role of collagens expression in distinguishing between premalignant and malignant lesions in stomach. *Anat Rec (Hoboken)*. 2009;292(5):692–700.
8. Jang M, Koh I, Lee JE, Lim JY, Cheong J-H, Kim P. Increased extracellular matrix density disrupts E-cadherin/ β -catenin complex in gastric cancer cells. *Biomater Sci*. 2018;6(10):2704–13.
9. Österlund T, Bordel S, Nielsen J. Controllability analysis of transcriptional regulatory networks reveals circular control patterns among transcription factors. *Integr Biol (Camb)*. 2015;7(5):560–8.

10. Guo J, Fu Z, Wei J, Lu W, Feng J, Zhang S. PRRX1 promotes epithelial-mesenchymal transition through the Wnt/ β -catenin pathway in gastric cancer. *Med Oncol*. 2015;32(1):393.
11. 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.
12. P ZT. Centrality Measures to Identify Influential Nodes in Complex Networks. *PLoS ONE*. 2016;035(014):209-10.
13. Fulton DL, Sundararajan S, Badis G, Hughes TR, Wasserman WW, Roach JC, et al. TFCat: the curated catalog of mouse and human transcription factors. *Genome Biol*. 2009;10(3):R29.
14. Lu P, Weaver VM, Werb Z. The extracellular matrix: a dynamic niche in cancer progression. *J Cell Biol*. 2012;196(4):395–406.

Figures

Figure 1

The Venn plot of DEGs

Figure 2

clustering heat map of DEGs

Figure 3

PPI network plot of DEGs

Figure 4

Key module of DEGs in PPI network

Figure 5

ClueGO network of pathways for DEGs

Figure 6

Kaplan-Meier survival curve analysis of key nodes including SPARC, FN1, BGN and COL1A2.

Figure 7

Transcriptional Regulatory Networks (TRN) for DEGs

Figure 8

The top 10 GO enriched by DEGs. a. Top 10 enriched GO pathways of up-regulated DEGs; b. Top ten enriched GO pathways of down-regulated DEGs

Figure 9

The top 10 KEGG pathways enriched by DEGs. a. Top ten enriched KEGG pathways of up-regulated DEGs; b. Top ten enriched KEGG pathways of down-regulated DEGs