Identification of Hub Genes And Pathways of Gastric Adenocarcinoma Using Bioinformatics Analysis

Yu Di (✉ diyu_daren@163.com)

Qilu hospital of Shandong University

Research

Keywords: gastric adenocarcinoma, hub genes, bioinformatics analysis

Posted Date: July 12th, 2021

DOI: <https://doi.org/10.21203/rs.3.rs-455851/v1>

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

[Read Full License](#)

Abstract

Object: Understanding hub genes associated with gastric adenocarcinoma (GC) development could lead to effective advances to diagnose and treat the diseases. In order to discover possible signal pathways and hub genes for the disease, we utilized the bioinformatics tools to analyze its mechanism.

Method: The gene chip of GSE7997 was downloaded from the GEO Datasets. Expression data of gastric cancer and its adjacent normal tissues were compared and the DEGs were acquired. The clusterProfiler and KEGG.db R packages were used for the analysis of its gene ontology process and KEGG pathways. What's more, the PPI network was constructed by the STRING website. The hub genes were acquired by the plugin of the Cytohubba in Cytoscape. Finally, these genes were examined by the TCGA datasets and potential drugs were explored by Connectivity map.

Results: The up regulated DEGs were mainly associated with the process of an extracellular matrix organization, an extracellular matrix organization, Collagen catabolic process, and multicellular organismal catabolic process. The down regulated DEGs were mainly associated with the process of digestion, cellular response to zinc ion, digestive system process, and organic hydroxy compound metabolic process. The up regulated DEGs were mainly located on PI3K-AKT signal pathways, human papillomavirus infection, Cytokine-cytokine receptor interaction, and focal adhesion process. The down regulated genes were mainly associated with protein digestion and absorption, mineral absorption, and the pancreatic secretion. Cytohubba had found hub genes of COL1A2, COL5A1, COL4A1, COL5A2, COL6A3, COL11A1, SERPINH1, FN1, and down-regulated COL2A1. These genes were associated with the process of transforming growth factor, extracellular matrix, cell adhesion, wound healing and so on. As examined by the TCGA datasets, these altered genes were associated with overall survival and no disease progress time ($P < 0.05$). Finally, we got the small molecule drugs of fenofibrate, benzbromarone, semustine, chloroquine, ondansetron, hydroxyachillin, megestrol, ciclopirox and monastrol for gastric cancer by Connectivity map ($P < 0.05$).

Conclusion: As mentioned above, we got 9 hub genes and tested by the TCGA datasets of gastric adenocarcinoma. They were associated with overall survival and disease free progressive time in gastric cancer patients. The bioinformatical analysis of the disease may enhance the understanding of the mechanism of disorders.

Introduction

Gastric cancer was one of the most common cancers throughout the world, and gastric adenocarcinoma was the most common type of the disorder. Despite its worldwide decline in incidence in recent years, gastric cancer remains a major killer in different countries[1]. Gastric cancer can be classified into diffuse and intestinal types based on its histological appearance, and into cardia and non-cardia cancer according to its location[2]. Patients with gastric cancer demonstrate “three high and three low” characteristics with a high rate of mortality and low rate of the early diagnosis, radical resection rate, and

5-year survival. Except for surgery for gastric cancer, its therapies include neoadjuvant chemotherapy, radiotherapy, molecular-targeted therapies and immunotherapy[3]. As with the development of computer science and bioinformatic technologies, genomics and model systems have enabled the discovery and validation of novel therapeutic strategies and effective objective indicators [4]. The digestive endoscopic technique and biomarkers of CEA and CA199 were the ordinary medical examinations for gastric carcinoma[5]. Most patients with gastric cancer can be diagnosed at an early stage. However, certain gastric tumors may be leaved out. The effect of molecular-targeted therapies of the disease including TKIs of EGFR(erlotinib and gefitinib) — as well as TKIs of EGFR and HER-2 — lapatinib were limited[6, 7]. So, A wholly perceptive of the molecular genetics and signaling pathways concerned with gastric cancer is a prerequisite for improvement of timely indicative and adjuvant therapeutic alternatives.

Methods

The gene chip of gastric adenocarcinoma of GSE7997

(<https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE79973>) [8] whose neoplasm staging varied in T1a ~ 2b was download from the GEO Datasets[9], including 10 RNA-Seq microchips. The platform of the chip was GPL570 [HG-U133_Plus_2] Affymetrix Human Genome U133 Plus 2.0 Array. Expression data of gastric adenocarcinoma and its adjacent normal tissues were compared and the DEGs were acquired by edgeR R packages[10]. The following condition of foldChange(FC) and false discovery rate(FDR) value was satisfied: $|\log FC| > 2$, $FDR < 0.01$, as showed in Fig. 1. The WEBGESTALT (<http://www.webgestalt.org/>) [11] tools and KEGG.db[12] R packages were used for the analysis of its gene ontology process and KEGG pathways. What's more, the PPI network was constructed by the STRING[13] website. The hub genes were acquired by the plugin of the Cytohubba[14] <http://apps.cytoscape.org/apps/cytohubba> in Cytoscape software. Then, these genes were annotated by GENCLIP2.0(<http://ci.smu.edu.cn/>)[15] and examined by the TCGA datasets[16], which was conducted by Cbioportal tools[17] <http://www.cbioportal.org/>. Finally, the DEGs were input into Connectivity map[18] <https://portals.broadinstitute.org/cmap/> to obtain the small molecule drugs for the disease.

Results

DEGs involved in gene ontology process

As mentioned above, about 187 up-regulated DEGs and 156 down-regulated DEGs were obtained by the filter of $|\log FC| > 2$, $FDR < 0.01$, as showed in Fig. 1. After gene ontology results of the DEGs by WEBGESTALT were showed in Fig. 2, the up regulated DEGs were mainly associated with the process of biological regulation, response to the stimulus, extracellular matrix organization, an extracellular matrix organization, Collagen catabolic process, multicellular organismal catabolic process and so on, as showed in Fig. 2a. The down regulated DEGs were mainly associated with the process of protein binding, digestion, cellular response to zinc ion, digestive system process, and organic hydroxy compound metabolic process and so on, as showed in Fig. 2b.

DEGs involved in KEGG pathways

The up regulated DEGs was mainly located on pathways of PI3K-AKTsignal pathways, human papillomavirus infection,Cytokine-cytokine receptor interaction, and focal adhesion process, as showed in Fig. 3a. The down regulated genes were mainly located on the pathways of protein digestion and absorption, mineral absorption, and the pancreatic secretion, as showed in Fig. 3b.

Cancer-associated hub genes haunted from the DEGs PPI network

The 343 DEGs were input into the STRING(<https://string-db.org/>)visualized by Cytoscape,the PPI network was showed in Fig. 4a, red indicated up-regulated and green indicated down regulated.TheCytohubba was used to determine the hub genes of the network.Nine hub genes of up-regulated COL1A2, COL5A1, COL4A1, COL5A2, COL6A3, COL11A1, SERPINH1, FN1, and down-regulated COL2A1, were discovered, as showed in Fig. 4b.These genes were associated with the process of transforming growth factor, extracellular matrix, cell adhesion, wound healing and so on, which was demonstrated by a literature analysis tool of GENCLIP2.0, as showed in Fig. 4c.

Validation of the hub genes by TCGA datasets

The cBioPortal is a web-based visualization and analysis engine that make complex cancer genomics data accessible to a wide range of cancer researchers and clinicians. The TCGA datasets was utilized to validate these genes. Nine hub genes of oncoplot of gastric cancer were shown in Fig. 5a: their genetic alteration, including mutation, amplification and deep depletion was displayed in the cases.These altered genes were connected with overall survival($P = 0.0416$)and no disease progress time($P = 0.0198$) of GC patients(Fig. 5b and 5c)($P < 0.05$).The mutual exclusivity of the nine hub genes was showed in table 1($P < 0.001$). Some of the genes were mutated exclusively. High gene expression of FN1, COL6A3, COL4A1, COL11A1 was associated with the low overall survival rate ($P < 0.05$), as showed in Fig. 6.

Table 1.the mutual exclusivity of the nine hub genes($P<0.001$).

A	B	Neither	A Not B	B Not A	Both	Log2 Odds Ratio	p-Value	q-Value	Co-occurrence
COL6A3	FN1	470	43	22	10	2.313	<0.001	0.005	yes
COL4A1	COL6A3	454	38	40	13	1.957	<0.001	0.005	yes
COL4A1	COL5A2	474	42	20	9	2.344	<0.001	0.005	yes
COL5A2	FN1	491	22	25	7	2.644	<0.001	0.005	yes
COL1A2	COL5A1	455	48	30	12	1.923	<0.001	0.005	yes
COL5A2	COL6A3	472	20	44	9	2.271	<0.001	0.005	Yes

Table 2.the small molecular results of GC analyzed by Connectivity map.

cmap name	mean	enrichment	p	specificity
fenofibrate	-0.618	-0.874	0.00403	0.0194
benzbromarone	-0.475	-0.841	0.00803	0.0063
6-bromoindirubin-3'-oxime	-0.433	-0.585	0.00856	0.1122
semustine	-0.394	-0.739	0.00901	0.0519
chloroquine	-0.641	-0.729	0.0109	0.0114
ondansetron	-0.308	-0.713	0.01377	0.0075
hydroxyachillin	-0.346	-0.68	0.02312	0.012
megestrol	-0.427	-0.674	0.02528	0.0068
ciclopirox	-0.39	-0.655	0.03284	0.1238
monastrol	-0.294	-0.461	0.04559	0.1184

Mining small molecule drugs of GC by connectivity map

Genome-wide expression profiling of gene transcripts has been successfully applied in biomedical discovery. Based on the premises of this technology, Connectivity map provides a data-driven and systematic approach for discovering associations among genes, chemicals and biological conditions. After the ID transformation of DEGs by Netaffx(<http://www.affymetrix.com/analysis/index.affx>)[19], the connectivity map was used to explore the small potential drugs of the matched DEGs. Small molecule drugs of fenofibrate, benzbromarone, 6-bromoindirubin-3'-oxime, semustine, chloroquine, ondansetron, hydroxyachillin, megestrol, ciclopirox and monastrol had been discovered ($P < 0.05$).

Discussion

In our study, the gene chip of gastric adenocarcinoma of GSE7997 including 10 RNA-Seq microchips was download from the GEO Datasets. Expression data of gastric adenocarcinoma and its adjacent normal tissues were compared and the DEGs were acquired by edgeR R packages with the condition: $|\log FC| > 2$, $FDR < 0.01$. About 187 up-regulated DEGs and 156 down-regulated DEGs were obtained. The WEBGESTALT (<http://www.webgestalt.org/>) tools and KEGG.db R packages were used for the analysis of its gene ontology process and KEGG pathways. The up regulated DEGs were mainly associated with the process of biological regulation, response to stimulus, extracellular matrix organization, an extracellular matrix organization, collagen catabolic process, multicellular organismal catabolic process and so on. The down regulated DEGs were mainly associated with the process of protein binding, digestion, cellular response to zinc ion, digestive system process, and organic hydroxy compound metabolic process and so on. The up regulated DEGs was mainly located on pathways of PI3K-AKT signal pathways, human papillomavirus infection, Cytokine-cytokine receptor interaction, and focal adhesion process. The down regulated genes were mainly located on the pathways of protein digestion and absorption, mineral absorption, and the pancreatic secretion. What's more, the PPI network was constructed by the STRING website and visualized by Cytoscape. Nine hub genes were acquired by the plugin of the Cytohubba. They

were up-regulated COL1A2, COL5A1, COL4A1, COL5A2, COL6A3, COL11A1, SERPINH1, FN1, and down-regulated COL2A1. S. Tamizhalagan et al [20] has reported that COL1A2 and PCOLCE genes of this region were expressed in diffuse type gastric tumors, genome-wide expression neighbors of SHFM1 and COL1A2 also showed mutually exclusive expression pattern, and stratify intestinal and diffuse type gastric tumors. T. Li et al [21] suggested that COL4A1, VCAN, THBS2, TIMP1, COL1A2, SERPINH1, and COL6A3 were associated with the OS of gastric cancer. S. U. Yu et al [22] who studied MiR-33a reported that COL11A1 is overexpressed in gastric cancer tissues and regulates proliferation. After using a literature analysis tool of GENCLIP2.0, we found the obtained hub genes were associated with the process of transforming growth factor, extracellular matrix, cell adhesion, wound healing and so on. The cBioPortal and TCGA datasets were utilized to validate these genes. Nine hub genes were regarded as a whole factor for the patients with GC. These altered genes were associated with overall survival ($P = 0.0416$) and no disease progression time ($P = 0.0198$). The mutual exclusivity of the nine hub genes was displayed in our study, including COL6A3, COL4A1, COL5A2, COL1A2, FN1 and COL5A1 ($P < 0.001$). This indicated these genes were closely related to each other. High gene expression of FN1, COL6A3, COL4A1, COL11A1 was associated with the low overall survival rate ($P < 0.05$). Genome-wide expression profiling of gene transcripts has been successfully applied in biomedical discovery. Based on the premises of this technology, Connectivity Map provides a data-driven and systematic approach for discovering associations among genes, chemicals and biological conditions. After the ID transformation of DEGs by NetAffx [19] (<http://www.affymetrix.com/analysis/index.affx>), the connectivity map was used to explore the potential small drugs of the matched DEGs. Small molecule drugs of fenofibrate, benzbromarone, 6-bromoindirubin-3'-oxime, semustine, chloroquine, ondansetron, hydroxyachillin, megestrol, ciclopirox and monastrol had been revealed ($P < 0.05$). However, some drugs have been reported by literature for the treatment of gastric cancer, such as fenofibrate [23], semustine [24], chloroquine [25], ondansetron [26]. Other discovered drugs need more laboratory and clinical evidence for GCs.

Conclusions

As mentioned above, we got 9 hub genes by comparing normal and tumor tissue of GCs. Then the results were examined by the TCGA datasets of gastric adenocarcinoma. These genes were associated with overall survival and disease free progressive time in gastric cancer patients that demonstrated our results were pervasive and correct. Some small molecular drugs were mined by Connectivity map, which may provide a guidance for the clinical drug application for the GCs patients. In summary, our bioinformatical analysis of GCs may enhance the understanding of the mechanism of this kind of disorders and offer a feasible theory for clinical practice.

Abbreviations

DEGs: differentially expressed genes; GC: gastric cancer; GEO: gene expression omnibus; FC: fold change; TCGA: the cancer genome atlas; ID: identity document; OS: overall survival; COL1A2: recombinant

collagen type i alpha 2; FN1: fibronectin 1; PCOLCE: procollagen C-endopeptidase enhancer; VCAN: Versican core protein; SERPINH1: Serpin H1; TIMP1: Metalloproteinase inhibitor 1.

Declarations

Acknowledgements

None

Ethical Approval and Consent to participate

Not applicable.

Consent for publication

All the authors consent for publication of the study.

- Availability of supporting data

All the supporting data were available if required.

Competing interests

There is no competing interests among authors.

Funding

There is no funding for the article.

Authors' contributions

DIYu is responsible for the article.

Authors information

DIYu is the corresponding author of the article, who are from Qilu hospital of Shandong University, and Huangshi central hospital of Huangshi Institute of Technology: Hubei Polytechnic University.

References

1. Karimi, P., F. Islami, S. Anandasabapathy, et al., *Gastric cancer: descriptive epidemiology, risk factors, screening, and prevention*. Cancer Epidemiol Biomarkers Prev, 2014. **23**(5): p. 700-13.
2. Choi, Y.J. and N. Kim, *Gastric cancer and family history*. Korean J Intern Med, 2016. **31**(6): p. 1042-1053.
3. Song, Z., Y. Wu, J. Yang, et al., *Progress in the treatment of advanced gastric cancer*. Tumour Biol, 2017. **39**(7): p. 1010428317714626.
4. Tsimberidou, A.M., *Targeted therapy in cancer*. Cancer Chemother Pharmacol, 2015. **76**(6): p. 1113-32.
5. Shen, M., H. Wang, K. Wei, et al., *Five common tumor biomarkers and CEA for diagnosing early gastric cancer*. Medicine, 2018. **97**(19): p. e0577.
6. Sitarz, R., M. Skierucha, J. Mielko, et al., *Gastric cancer: epidemiology, prevention, classification, and treatment*. Cancer Management & Research, 2018. **10**: p. 239-248.
7. Potemski, R.C.A.J.P.-P.A.A.J.A.M.F.A.P., *Application of molecular targeted therapies for treatment of gastric cancer and gastroesophageal junction cancer*. Vol. 10. 2014. 244-252.
8. He, J., Y. Jin, Y. Chen, et al., *Downregulation of ALDOB is associated with poor prognosis of patients with gastric cancer*. Onco Targets Ther, 2016. **9**: p. 6099-6109.
9. Edgar, R. and A. Lash, 6. *The Gene Expression Omnibus (GEO): A Gene Expression and Hybridization Repository*. national center for biotechnology information, 2002.
10. Speth, K.A., *Comparison of Bioconductor Packages edgeR and DESeq for Analyzing Differential Expression*. 2015.
11. Zhang, B., S. Kirov, and J. Snoddy, *WebGestalt: an integrated system for exploring gene sets in various biological contexts*. Nucleic Acids Research, 2005. **33**(Web Server issue): p. 741-8.
12. Reid, J.F., *[BioC] KEGG.db in GoStats with Streptomyces coelicolor*.
13. Andrea, F., S. Damian, F. Sune, et al., *STRING v9.1: protein-protein interaction networks, with increased coverage and integration*. Nucleic Acids Research, 2012. **41**(D1).
14. Chin, C.H., S.H. Chen, H.H. Wu, et al., *cytoHubba: identifying hub objects and sub-networks from complex interactome*. BMC Systems Biology, 2014. **8**(S4): p. S11.
15. Wang, Jia-Hong, Zhao, et al., *GenCLiP 2.0: a web server for functional clustering of genes and construction of molecular networks based on free terms*. Bioinformatics, 2014.

16. Tomczak, K., P. Czerwinska, and M. Wiznerowicz, *The Cancer Genome Atlas (TCGA): An immeasurable source of knowledge*. Contemporary Oncology / Współczesna Onkologia, 2015. **19**(1A): p. A68-77.
17. Lukasse, P., F. Schaeffer, O.P. Casals, et al., *Gene set variation analysis in cBioPortal*. 2018.
18. Qu, X.A. and D.K. Rajpal, *Applications of Connectivity Map in drug discovery and development*. Drug Discovery Today, 2012. **17**(23-24): p. 1289-1298.
19. Liu, G., A.E. Loraine, S. Ron, et al., *NetAffx: Affymetrix probesets and annotations*. Nucleic Acids Research, 2003. **31**(1): p. 82-86.
20. Tamilzhalagan, S., M. Muthuswami, and K. Ganesan, *Juxtaposed genes in 7q21-22 amplicon contribute for two major gastric cancer sub-Types by mutual exclusive expression*. Molecular Carcinogenesis, 2016. **56**(4).
21. Li, T., X. Gao, L. Han, et al., *Identification of hub genes with prognostic values in gastric cancer by bioinformatics analysis*. World Journal of Surgical Oncology, 2018. **16**(1): p. 114.
22. Yu, S.U., G. Song, X.N. Zhang, et al., *MiR-33a is poorly expressed in gastric cancer tissues and inhibits the proliferation, migration and invasion of HGC-27 cells*. Progress of Anatomical ences, 2018.
23. Hong, C., Y.R. Song, Y. Bai, et al., *Anti-proliferative mechanism of fenofibrate against human gastric tumor MGC803 cells*. Chinese Journal of Hospital Pharmacy, 2015.
24. *Combination chemotherapy of advanced colorectal cancer utilizing 5-fluorouracil, semustine, dacarbazine, vincristine, and hydroxyurea: A phase III trial by the eastern cooperative oncology group (EST: 4275)*. 1982. **49**(8): p. 1555-1560.
25. Jing, X., Y. Xu, W. Cheng, et al., *Tanshinone I induces apoptosis and pro-survival autophagy in gastric cancers*. Cancer Chemotherapy & Pharmacology, 2016. **77**(6): p. 1171-1181.
26. Cascinu, S., R. Labianca, F. Graziano, et al., *Intensive weekly chemotherapy for locally advanced gastric cancer using 5-fluorouracil, cisplatin, epidoxorubicin, 6S-leucovorin, glutathione and filgrastim: a report from the Italian Group for the Study of Digestive Tract Cancer (GISCAD)*. British Journal of Cancer.

Figures

Volcano

Figure 1

Expression data of gastric adenocarcinoma and its adjacent normal tissues were compared and the DEGs were acquired by edgeR R packages[10]. The following condition of foldChange(FC) and false discovery rate(FDR) value was satisfied: $|\log FC| > 2 \times FDR < 0.01$, as showed in Fig.

Figure 2

After gene ontology results of the DEGs by WEBGESTALT wereshowed in Fig 2, the up regulated DEGs were mainly associated with the process of biological regulation, response to the stimulus, extracellular matrix organization, an extracellular matrix organization, Collagen catabolic process, multicellular organismal catabolic processand so on, as showed in Fig a. The down regulated DEGs were mainly associated with the process of protein binding, digestion, cellular response to zinc ion, digestive system process, and organic hydroxy compound metabolic process and so on, as showed in Fig b.

The 343 DEGs were input into the STRING<https://string-db.org/> visualized by Cytoscape the PPI network was showed in Fig a, red indicated up-regulated and green indicated down regulated. The Cytohubba was used to determine the hub genes of the network. Nine hub genes of up-regulated COL1A2, COL5A1, COL4A1, COL5A2, COL6A3, COL11A1, SERPINH1, FN1, and down-regulated COL2A1, were discovered, as showed in Fig b. These genes were associated with the process of transforming growth factor, extracellular matrix, cell adhesion, wound healing and so on, which was demonstrated by a literature analysis tool of GENCLIP2.0, as showed in Fig c.

Figure 5

The cBioPortal is a web-based visualization and analysis engine that make complex cancer genomics data accessible to a wide range of cancer researchers and clinicians. The TCGA datasets was utilized to validate these genes. Nine hub genes of oncoplot of gastric cancer were shown in Fig 5a: their genetic alteration, including mutation, amplification and deep depletion was displayed in the cases. These altered genes were connected with overall survival (P=0.0416) and no disease progress time (P=0.0198) of GC patients (Fig b and c) (P<0.05).

Figure 6

Some of the genes were mutated exclusively. High gene expression of FN1, COL6A3, COL4A1, COL11A1 was associated with the low overall survival rate ($P<0.05$), as showed in Fig.