Identification and Validation of an Individualized Autophagy-Clinical Prognostic Index in Gastric Cancer Patients

Qiu Jieping
Anhui Medical University
https://orcid.org/0000-0003-1887-4251

Jieping Qiu
Anhui Medical University

Mengyu Sun
Anhui Medical University

Yaoqun Wang
Anhui Medical University

Bo Chen (chenbo@ahmu.edu.cn)
https://orcid.org/0000-0001-5185-7432

Primary research

Keywords: gastric cancer, autophagy, prognosis, bioinformatics

Posted Date: March 5th, 2020

DOI: https://doi.org/10.21203/rs.3.rs-16081/v1

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

Background: The purpose of this study is to perform bioinformatics analysis of autophagy-related genes in gastric cancer, and to construct a multi-gene joint signature for predicting the prognosis of gastric cancer.

Methods: GO and KEGG analysis were applied for differentially expressed autophagy-related genes in gastric cancer, and PPI network was constructed in Cytoscape software. In order to optimize the prognosis evaluation system of gastric cancer, we established a prognosis model integrating autophagy-related genes. We used single factor Cox proportional risk regression analysis to screen genes related to prognosis from 222 autophagy-related genes in The Atlas Cancer Genome (TCGA) gastric cancer cohort. Then, the generated genes were applied to the Least Absolute Shrinkage and Selection Operator (LASSO). Finally, the selected genes were further included in the multivariate Cox proportional hazard regression analysis to establish the prognosis model. According to the median risk score, patients were divided into high-risk group and low-risk group, and survival analysis was conducted to evaluate the prognostic value of risk score. Finally, by combining clinical features and prognostic gene signatures, a nomogram was established to predict individual survival probability.

Results: GO analysis showed that the 38 differently expressed autophagy-related genes was enriched in cell growth, neuron death, and regulation of cell growth. KEGG analysis showed that the 38 differently expressed autophagy-related genes were related to platinum drug resistance, apoptosis and p53 signaling pathway. The risk score was constructed based on 4 genes (GRID2, ATG4D, GABARAPL2, CXCR4), and gastric cancer patients were significantly divided into high-risk and low-risk groups according to overall survival. In multivariate Cox regression analysis, risk score was still an independent prognostic factor (HR = 1.922, 95% CI = 1.573-2.349, P < 0.001). Cumulative curve showed that the survival time of patients with low-risk score was significantly longer than that of patients with high-risk score (P < 0.001). The external data GSE62254 proved that nomograph had a great ability to evaluate the prognosis of individual gastric cancer patients.

Conclusions: This study provides a potential prognostic marker for predicting the prognosis of GC patients and the molecular biology of GC autophagy.

Background

Gastric cancer (GC) is a common disease that threatens human health. It is composed of adenocarcinoma, squamous cell carcinoma, adenosquamous carcinoma, carcinoid, etc. most of them are gastric adenocarcinoma. In 2018, more than 1000000 people were diagnosed with gastric cancer, which caused about 783000 deaths (equivalent to 1 in every 12 deaths worldwide), making it the fifth most frequently diagnosed cancer and the third major cause of cancer death [1]. The prognosis of gastric cancer is related to pathological stage, location, tissue type, biological behavior and treatment. [2]. Up to now, histologic diagnosis and TNM staging are still the main methods to evaluate the prognosis of
gastric cancer [3]. However, the existing evaluation indicators can not cover all the disease information of patients, and can not be used to accurately predict the prognosis of GC patients. Therefore, it is necessary to dig out effective prognostic biomarkers to help optimize the prognosis evaluation system of gastric cancer. In the past few decades, people have learned more and more about the characteristics of tumors. One of the breakthroughs is the participation of autophagy process in the development of cancer [4–6].

Autophagy refers to the physiological and pathological process that relies on the lysosomes of cells to degrade or remove excess or damaged organelles, fold wrong proteins and invading microorganisms [7]. Autophagy widely exists in eukaryotic cells, which is one of the necessary ways of organelle renewal and metabolism dynamic balance in the organic body. It maintains the homeostasis of the intracellular environment at the basic level under the normal physiological state. When the body encounters hunger, growth factor deficiency, oxidative stress, injury and other events, the range of autophagy may raise dramatically, that is, megaautophagy, to provide nutrition and clean harmful substances [8]. Some signal transduction pathways are involved in autophagy regulation, such as mTOR signaling pathway, PI3K / Akt / (PKB) pathway, ras-raf-1-mek-er k1 / 2 signaling pathway [9].

Autophagy plays an intricate and contradictory role in all stages of tumor development. On the one hand, basic autophagy in normal cells can avoid the accumulation of some DNA mutation molecules and organelles, and prevent cell transformation[10]. On the other hand, the long-term upregulation of basal autophagy in tumor cells is conducive to the cancer cells being in a state of aging or dormancy, such as tumor stem cells, which will enhance the tolerance of tumor to chemotherapy or radiotherapy, and increase the recurrence rate of tumor. [11]. More and more studies have shown the significance of autophagy in gastric cancer. Beclin-1 is a key regulator of autophagy. Ahn et al. Studied the protein expression level of beclin-1 in 60 cases of gastric cancer, found that 83% of gastric cancer had beclin-1 expression, but almost no beclin-1 expression in normal gastric mucosa cells [12]. Kang et al. Also reached the same conclusion [13]. ATG5 is a key autophagy related protein. Ge et al found that the high expression of ATG5 was significantly related to poor overall survival(OS) and disease-free survival (DFS) in gastric cancer patients [14]. In addition, the high expression of L3 is related to the poor prognosis of gastric cancer [15]. These findings confirm the relationship between autophagy and gastric cancer, and indicate the great potential of autophagy related genes as a prognostic marker of gastric cancer.

In order to dig out autophagy-related genes in gastric cancer, we analyzed the differentially expressed autophagy-related genes in gastric cancer by GO and KEGG analysis, and constructed PPI network map. In order to optimize the prognosis evaluation system of gastric cancer, we established a prognosis model integrating autophagy related genes of gastric cancer. We used single factor Cox proportional risk regression analysis to screen genes related to prognosis from 222 autophagy related genes in The Atlas Cancer Genome (TCGA) gastric cancer dataset. Then, the generated genes were applied to the Least Absolute Shrinkage and Selection Operator (LASSO). Finally, the selected genes were further included in the multivariate Cox proportional risk regression analysis to establish the prognosis model. According to the median risk score, patients were divided into high-risk group and low-risk group, and survival analysis
was conducted to evaluate the prognostic value of risk score. Finally, by combining clinical features and prognostic gene signatures, a nomogram was established to predict individual survival probability.

**Methods**

**Data Acquisition**

Search for the word "autophagy" on the Human Autophagy-dedicated Database (HADb) (www.autophagy.lu) to retrieve autophagy related genes. The original RNA seq data set and clinical features of the TCGA gastric dataset can be downloaded from the TCGA website (https://portal.gdc.cancer.gov/). Use R (version 3.6.0) software to standardize and process data. GSE62254 data set was obtained from Gene Expression Omnibus (GEO, https://www.ncbi.nlm.nih.gov/geo/) for validation.

**Identification And Enrichment Analysis Of De-atgs**

The R package "clusterprofiler" was used to carry out Gene Ontology (GO) enrichment analysis including biological process (BP), cell components (CC) and molecular functions (MF) for the differentially expressed autophagy-related genes (DE-ATGs). The same tool is also used for the enrichment analysis of Kyoto Encyclopedia of Genes and Genomes (KEGG) enrichment analysis. Then, we use string database (https://string-db.org/) to construct protein-protein interaction (PPI) network of autophagy related genes, and import the data into the software of Cytoscape to visualize the interaction of PPI network.

**Identification And Verification Of Prognostic Gene Signatures**

Single variable Cox proportional risk regression analysis was performed to screen autophagy-related genes (ATGs) significantly associated with overall survival (OS) in the TCGA gastric cancer dataset. The OS related genes identified were included in the LASSO regression analysis to screen the genes, and the multivariable Cox proportional risk regression analysis was carried out to establish the prognosis model of gastric cancer autophagy related genes. We used the following formula to calculate the risk score of each patient: risk score = Σ x J * coef J, where Xi is gene expression, coef J is coefficient, and X J is relative expression level of each ATG standardized by Z-score. The median risk score was determined as the critical value to divide the STAD dataset into high risk and low risk. In order to determine the role of risk score in predicting the clinical prognosis of GC patients, Kaplan-Meier Plotter was drawn to analyze the different survival time between high-risk group and low-risk group.

In order to study whether autophagy related risk index can be used as an independent predictor of OS in TCGA dataset of BC patients, single variable and multivariate Cox regression analysis were conducted.
Risk score, age, gender, tumor subtype, pathological stage and histological grade were used as covariates.

The Construction Of Nomogram

Age, gender, stage, grade, T, N, M, and risk score were used to construct the nomogram together using the “rms” and “survival” packages in R. Then, calibration curves were drawn to assess the consistency between actual and predicted survival.

Results

Identification and enrichment analysis of DE-ATGs

The TCGA gastric cancer cohort consisted of 443 patients with gastric cancer. We searched for ATGs in the HADB database. A total of 799 ATGs were selected, 222 of which were expressed in TCGA gastric dataset. In order to inquire about the potential signal pathways related to 222 autophagy related genes in gas cancer, we screened them and analyzed them with GO and KEGG. 222 autophagy related genes were screened by r-packet "limma", and the screening criteria were | lgfc | > 2, and adj. p < 0.05. Results showed 38 ATGs were differentially expressed in TCGA-STAD dataset (Fig. 1). GO analysis shows that these ATGs can be enriched in several basic biological processes(BP), including cell growth, positive regulation of cell protein localization, neuron death, regulation of cell growth (Fig. 2A). KEGG analysis showed that the 38 ATGs were mainly related to autophagy, platinum drug resistance, apoptosis, and p53 signaling pathway (Fig. 2B). These genes are linked to form a protein protein interaction(PPI) network, as shown in Fig. 2C (https://string-db.org/).

Construction Of Prognostic Markers In Tcga Gastric Cancer Dataset

222 ATGs were analyzed by single variable Cox regression. There are 10 genes associated with TCGA-STAD (Fig. 3A). These important genes entered the LASSO regression analysis, although including 8 genes, the model obtained the best performance (Fig. 3B, C, D). The 8 genes were further analyzed by multivariate Cox regression analysis, and finally 4 genes (GRID2, ATG4D, GABARAPL2 and CECR4) related to the prognosis of STAD were obtained. The coefficients of each gene are shown in Table 1.
Table 1
Genes included in prognostic gene signature.

<table>
<thead>
<tr>
<th>Gene symbol</th>
<th>Full name</th>
<th>coefficient</th>
<th>HR</th>
<th>pvalue</th>
</tr>
</thead>
<tbody>
<tr>
<td>GRID2</td>
<td>glutamate ionotropic receptor delta type subunit 2</td>
<td>1.865679589</td>
<td>6.460324759</td>
<td>0.001034234</td>
</tr>
<tr>
<td>ATG4D</td>
<td>Autophagy Related 4D Cysteine Peptidase</td>
<td>-0.378353964</td>
<td>0.684987997</td>
<td>0.021107318</td>
</tr>
<tr>
<td>GABARAPL2</td>
<td>GABA Type A Receptor Associated Protein Like 2</td>
<td>0.6654797</td>
<td>1.945423518</td>
<td>0.006096631</td>
</tr>
<tr>
<td>CXCR4</td>
<td>C-X-C Motif Chemokine Receptor 4</td>
<td>0.158333802</td>
<td>1.171557197</td>
<td>0.028902714</td>
</tr>
</tbody>
</table>

The risk score of each patient was calculated on the basis of the relevant mRNA expression level and risk coefficient of each ATG. The risk score is used to forecast the prognosis of gastric cancer, and the median risk score is the critical value to divide patients into high-risk and low-risk groups. Heatmap was drawn to show gene expression profiles in high-risk and low-risk STAD groups (Fig. 4B). The genes with HR > 1 (GRID2, GABARAPL2, CXCR4) are considered to be dangerous genes, while those genes (ATG4D with HR < 1 as protective gene) (Fig. 4B). As shown in the Fig. 4B, patients in the high-risk group have more possibilities to express risk genes. In contrast, patients in the low-risk group have a disposition to express protective genes (Fig. 4B). Figure 4A shows the distribution of risk scores in patients with gastric cancer and the relationship between risk scores and survival time.

**Autophagy As An Independent Prognostic Factor**

We assessed the prognostic value of risk scores. For TCGA-STAD, the risk score in univariate analysis was significantly correlated with overall survival (OS) (HR = 1.648, 95% CI = 1.385–1.960, P < 0.001) (Fig. 5A). Multivariate analysis showed that the risk score was an independent prognostic indicator (HR = 1.922, 95% CI = 1.573–2.349, P < 0.001) (Fig. 5B). The Kaplan-Meier cumulative curve showed that the survival time of patients with low-risk score was significantly longer than that of patients with high-risk score (Fig. 5C). The AUC of risk score was significantly larger than that of other indicators, which proved that the Cox model had better ability to predict prognosis than other individual indicators.

**Construction And Verification Of Nomogram**

Nomogram is a powerful tool, which has been applied for quantifying individual risk in the clinical environment by integrating multiple risk factors. By combining four autophagy gene features, we
performed nomogram to predict the possibility of 3-year and 5-year OS. As shown in Fig. 6A, the score assigned to each factor is proportional to its risk contribution to survival. The indication of calibration curve matches well (Fig. 6B, C). Nomogram has been validated in the GSE62254 gastric cancer dataset, and the 3-year and 5-year calibration curves are respectively shown in Fig. 6D and Fig. 6E.

**Discussion**

Autophagy is a highly conserved evolutionary process in eukaryotic cells, which is involved in a series of cell homeostasis processes. There are three types of autophagy: megaautophagy, microautophagy and chaperone mediated autophagy. Megaautophagy is the only autophagy that can degrade organelles, which we usually call autophagy. The autophagy process can be summarized as two basic steps: first, the cytoplasmic material is wrapped by the autophagy body of the double membrane, and transported to the lysosome to form autophagy lysosome. Then, lysosomal enzymes are used to degrade the substances in the cells, so as to realize the metabolism of the cells and the renewal of some organelles. Autophagy related genes LC3, Beclin 1 and ATG5 are all biomarkers of autophagy, which are involved in autophagy regulation [16–18].

Many studies have shown that autophagy protein is closely related to the prognosis of GC patients. The high expression of ATG5 is closely related to the poor prognosis and drug resistance of gastric cancer [19]. It has been found that the disease-free survival rate and the overall survival rate of patients in the Beclin 1 high expression group are significantly increased [20]. However, some studies have come to the opposite conclusion [21]. These findings suggest that the abnormal expression of Beclin 1 is closely related to the prognosis of gastric cancer, but the divergence of the total and disease-free survival may be related to the follow-up degree of the researchers, the size of the sample size and the heterogeneity of the patients. Considering the importance of autophagy in gastric cancer, we can reasonably speculate that autophagy related genes have broad prospects in the prognosis evaluation of gastric cancer, and the multi gene signature generated by various algorithms will be better than a single molecule in the prediction of GC OS.

In this study, we analyzed the mRNA expression of 222 ATGs in the TCGA gastric cancer dataset. Single factor Cox regression analysis showed that 10 genes were related to the survival of STAD. We used LASSO regression to develop eight prognostic markers for the TCGA-STAD cohort. Finally, the signature of four genes was established by multivariate Cox regression. The risk score of each patient can be obtained by calculating the mRNA expression level and risk coefficient of the selected gene. In the TCGA-STAD cohort, risk scores significantly stratified patient outcomes. More importantly, in two independent geo gastric cancer datasets within the STAD, the prognostic power of the 4-gene signature was verified. Gene signature is often applied to forecast the prognosis of a variety of tumors in the past few years [22], which is even better than TNM staging and histopathological diagnosis in some extent [23]. Gene signatures based on ATGs have been reported in a variety of cancers, such as serous ovarian cancer, breast cancer, colon cancer and glioma [24–27]. For example, Liu and colleagues recently reported a 14
autophagy related signature (NRG1, itga3, map1lc3a) based on relapse free survival in patients with non-small cell lung cancer [28].

Bioinformatics enrichment analysis showed that 38 differentially expressed autophagy related genes (DE-ATGs) were mainly related to cell growth, positive regulation of cell protein localization, neuron death, regulation of cell growth, platinum drug resistance, apoptosis and p53 signaling pathway in STAD. Interestingly, Huang’s study found that autophagy plays a vital role in the platinum drug resistance of tumor cells [29]. In tumor treatment, apoptosis tolerance is an important mechanism for tumor drug resistance. Autophagy can prevent apoptosis induced by antitumor drugs and promote tumor drug resistance. However, autophagy cell death may be a death mode of apoptosis tolerant tumor cells, Autophagy has double effects on drug resistance of tumor cells [30]. There is also a lot of evidence implying the interaction between autophagy and apoptosis [31]. Autophagy may promote or hinder apoptosis.

Autophagy inhibited apoptosis when the environmental conditions were less affected. However, when autophagy causes excessive consumption of intracellular proteins and organelles, resulting in the inability of cells to survive, the cells will turn into apoptosis. In some cases, autophagy can also cause cell death. It is worth mentioning that autophagy and apoptosis involve many apoptosis related proteins, such as p53 and BH3 only proteins [32]. In the early stage of cancer cell formation, autophagy can inhibit the formation of cancer cells; after cancer cell formation, cancer cells use autophagy to promote the survival of cancer cells and inhibit cell apoptosis, which may lead to the resistance of cancer cells to chemotherapy drugs. Therefore, if we inhibit autophagy during chemotherapy, it will be beneficial to enhance the therapeutic effect.

Finally, we developed a nomogram to forecast individual clinical outcomes. Nomogram, which is based on multi factor regression analysis, uses multiple clinical indicators or biological attributes, and then uses line segments with scores, so as to achieve the setting purpose: based on the value of multiple variables to predict a certain clinical outcome or the probability of a certain type of event [33]. The nomogram transforms the complex regression equation into a simple and visual graph, which makes the prediction model more readable and valuable. This advantage makes the nomogram get more attention and application in medical research and clinical practice. Although traditional clinical and pathological features (e.g., TNM staging, tumor size, and histological subtypes) stemmed from gene signatures, risk scores can also be accepted into a predictive model nomogram to better forecast clinical outcomes [34]. Wang et al. Reported a nomogram to predict the recurrence free survival rate of malignant glioma, including the prognosis score calculated according to the characteristics of autophagy gene [35]. Compared with TNM staging, signature based on multi RNA showed a higher accuracy of prognosis. [36] in addition, as shown in the calibration curve, the 3-year and 5-year survival rates can be better predicted by using the nomogram integrating risk index and conventional prognostic factors [37]. Consistently, we show that the nomogram including four-ATG-signature can well predict the 3-year and 5-year survival probability of STAD patients. However, due to the lack of sufficient cases, we were unable to assess the predictive power of autophagy gene signatures in other independent gastric cancer data sets. In addition,
other potential prognostic variables related to OS in GC, such as neutrophil to lymphocyte ratio (NLR) should be investigated. Moreover, the expression of these 4 genes in gastric tissue and their prognostic effects need further study.

**Conclusion**

In summary, our study established a novel 4-gene signature and nomogram to forecast the prognosis of GC patients, which may contribute to the clinical decision-making of individual therapy.

**Declarations**

**List of abbreviations**

GC  gastric cancer

ATG  autophagy-related gene

DE-ARG  differentially expressed autophagy-related genes

LASSO  the Least Absolute Shrinkage and Selection Operator (LASSO)

TCGA  The Atlas Cancer Genome

OS  overall survival

DFS  disease-free survival

GO  Gene Ontology

KEGG  Kyoto Encyclopedia of Genes and Genomes

PPI  protein-protein interaction

GEO  Gene Expression Omnibus

BP  biological process

CC  cell components

MF  molecular functions

STAD  stomach adenocarcinoma

**Ethics approval and consent to participate**

Not applicable.
Funding

This work was supported by the National Natural Science Foundation (NO.81602425) and Anhui Provincial Teaching Research Project (NO.2016jyxm0529) and National Innovation and Entrepreneurship Project for College Students (NO.201710366009).

Acknowledgements

Not applicable.

Author information

Affiliations

Jieping Qiu\(^1\), Mengyu Sun\(^1\), Yaoqun Wang\(^1\), Bo Chen\(^2\)

\(^1\)Department of Clinical Medicine, The First Clinical College, Anhui Medical University, Hefei, China

\(^2\)Department of Gastrointestinal Surgery Center, The First Affiliated Hospital of Anhui Medical University, Hefei, China

Contributions

Jieping Qiu conceived, designed, analyzed the data, and write the manuscript. Mengyu Sun helped to search for some relevant papers for this research. Yaoqun Wang analyzed the data and generated the figures and tables. Bo Chen guided the research process. All authors read and approved the final manuscript.

Corresponding authors

Correspondence to Bo Chen.

Ethics declarations

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


**Figures**

Figure 1

Heatmap of the expression levels of 38 DE-ATRs in TCGA-STAD.
Figure 2

GO, KEGG enrichment analysis and PPI network. (A) GO analysis of 38 differentially expressed autophagy-related genes. (B) KEGG analysis of 38 differentially expressed autophagy-related genes. (C) PPI network diagram of 38 differentially expressed autophagy-related genes.
Figure 3

Regression analysis to select autophagy genes related to prognosis of gastric cancer. (A) Forest map of autophagy genes related to STAD survival, analyzed by univariate Cox regression. (B) Boxplot of autophagy genes associated with STAD survival, analyzed by LASSO regression. (C) LASSO coefficient spectrum of 10 genes in STAD. Generate a coefficient distribution map for a logarithmic (λ) sequence. (D) Selecting the best parameters for STAD in the LASSO model (λ)
**Figure 4**

Characteristics of prognostic gene signatures. (A) Distribution of risk score and patient survival time, and status of STAD. The black dotted line is the optimal cut-off value for dividing patients into low-risk and high-risk groups. (B) Heat map of autophagy-related gene expression profiles in the prognostic signature of STAD.
Figure 5

Autophagy-related gene signatures are significantly associated with gastric cancer survival. (A) Univariate Cox regression analysis. Forest plot of associations between risk factors and the survival of STAD. (B) Multiple Cox regression analysis. The autophagy-associated gene signature is an independent predictor of TCGA-STAD. (C) Kaplan-Meier analysis of TCGA gastric cancer patients was stratified by median risk. High risk scores are associated with general poor survival of TCGA-STAD. (D) Multi-index ROC curve of risk score and other indicators.
Figure 6

The nomogram can predict the prognosis probability in STAD. (A) A nomogram of the STAD cohort (training set) used to predict the OS. (B-C) Calibration maps used to predict the 3-year (B) and 5-year survival (C) in the training set. Calibration plots for 3-year (D) and 5-year survival (E) in the GSE62254 gastric cancer cohort (test group). The x-axis and y-axis represent the predicted and actual survival rates.
of the nomogram, respectively. The solid line represents the predicted nomogram, and the vertical line represents the 95% confidence interval.