Construction of a Prognostic Signature of 10 Autophagy-Related lncRNAs in Gastric Cancer

Wenwen Wang1,2, Qingshan Pei1, Lifen Wang1, Tong Mu1, Hua Feng1

1Department of Gastroenterology, Shandong Provincial Hospital Affiliated to Shandong First Medical University, Jinan, Shandong, People’s Republic of China; 2Shandong Provincial Engineering and Technological Research Center for Liver Diseases Prevention and Control, Shandong Provincial Hospital Affiliated to Shandong First Medical University, Jinan, Shandong, People’s Republic of China

Background: Autophagy plays a double-edged sword role in cancers. lncRNAs could regulate cancer initiation and development at various levels. However, the role of autophagy-related lncRNAs (ARlncs) in gastric cancer (GC) remains indistinct.

Methods: GC gene expression profile and clinical data were acquired from the Cancer Genome Atlas (TCGA). The prognostic signature composed of ARlncs was established via cox regression analysis. Kaplan–Meier (K-M) survival curve was adopted to show overall survival (OS). Independence and reliability of risk signature were visualized by cox regression analysis and ROC curve. A nomogram was constructed and the reliability was analyzed by ROC curve. Immune infiltrating cells and check points were also analyzed.

Results: A prognostic signature was constructed which stratified GC patients into high- and low-risk groups according to risk score calculated via the 10 ARlncs including LINC01094, AC068790.7, AC090772.1, AC005165.1, PVT1, LINC00106, AC026368.1, AC090912.3, AC013652.1, UICLM. Patients in high-risk group showed a poor prognosis (p<0.001). Cox regression analysis showed signature was an independent prognostic factor (p<0.001). Areas under curves (AUC) of ROC for risk signature for predicting OS outweighed age, gender, grade, T, M and N, which suggested the reliability of the signature. A nomogram was constructed with risk signature, T, M, N and age and its AUC of ROC for 1-, 3-, and 5-year was 0.700, 0.730, 0.757 respectively, which showed good reliability. Macrophage M2, T cell CD8+ and T cell CD4+ memory resting had greatest difference between the two risk groups according to CIBERSORE-ABS algorithm (p<0.001). CD274 (PD-L1), PDCD1 (PD-1) and PDCD1LG2 (PD-L2) were expressed higher in the high-risk group (p<0.05), which implied that immunotherapy may be a good choice for these patients.

Conclusion: The risk signature based on 10 ARlncs can serve as an efficacious prognostic predictor and guide the immunotherapies and precise treatment for GC patients.

Keywords: gastric cancer, autophagy, TCGA, prognostic signature, immunotherapy

Introduction

Gastric cancer (GC) is the fifth most common cancer and fourth fatal cancer globally with over one million new cases and an estimated 769,000 deaths (equating to one in every 13 deaths globally) in 2020. It is notable that the incidence of stomach cancer (cardia and noncardia gastric cancers combined) among young adults (aged <50 years) has increased in recent years.1 With the development of gene-sequencing, the promises of early detection and molecularly precise diagnosis for many cancers have been possible, which could allow patients to accept targeted, personalized treatment with better outcomes.2 As TCGA is a database based on gene sequencing, the prognostic signature based on TCGA may provide proof for precise treatment of patients. Up to now, large amounts of prognostic signatures for GC have been developed, including competing endogenous RNA (ceRNA)-based prognostic biomarkers,3 Long non-coding RNAs (lncRNAs)-based risk signature,4 immune prognostic signature5 and others. However, autophagy-related lncRNAs (ARlncs)-based prognostic signature has not been studied.
Autophagy is an evolutionarily conserved, normal physiological catabolic degradation process in cells which occurs in response to different forms of stress, for example, nutrient deprivation, growth factor depletion, infection and hypoxia. Dysregulated autophagy has implications in health, leading to human diseases, such as cancer. In gastric carcinoma, autophagy plays a dual role, including a tumor-suppressor role and a tumor-promoter role. As to the chemoresistance of GC, autophagy can contribute to the formation of chemoresistance and meanwhile, it can reverse chemoresistance by promoting apoptosis and/or inhibiting epithelial–mesenchymal transition (EMT). These may because multiple proteins and/or noncoding RNAs (ncRNAs) that regulated autophagy play various roles in gastric cancer. Therefore, modulating autophagy for cancer treatment is an interesting therapeutic approach currently under intense investigation. A series of genes, which are named autophagy-related genes (ARGs) are involved in the process, such as ULK1 complexes, PIK3C3, Beclin 1 and others and a number of noncoding RNAs including IncRNAs, miRNAs are also responsible for regulating the process of autophagy. LncRNAs are RNA transcripts longer than 200 nucleotides which are not code for proteins. They participate in almost all cellular processes, including regulation of transcription, translation, protein modification, the formation of RNA-protein or protein-protein complexes and signaling pathways. Their regulatory functions in pathological and physiological conditions, especially in tumorigenesis and development have received widespread attention. A number of IncRNAs have been identified to be abnormally expressed in gastric cancer and some of them have been found to regulate proliferation, migration, invasion, and metastasis of gastric cancer cells. Recent studies have proved that autophagy-related IncRNAs also play important roles in various cancers via different mechanisms. Therefore, the functions of ARlncs in GC need further study.

In our study, autophagy-related IncRNAs were screened to construct a prognostic signature to stratify GC patients into low- and high-risk groups. A nomogram was constructed with risk signature and clinical features. Signaling pathways enrichment, Kaplan-Meier survival, relevance between prognostic signature and clinicopathological parameters, immune cells infiltration, immune function, immune checkpoint were also analyzed. The results will help clinicians predict survival and formulate individualized and precise treatments for gastric cancer patients.

Materials and Methods
Datasets
The RNA sequencing profiles with clinical information were downloaded from The Cancer Genome Atlas (TCGA) (https://portal.gdc.cancer.gov/). The cohort contains 375 tumor tissues and 32 normal tissues. And GC patients with clinical information were enrolled for later study (Table S1). The perl language was adopted to transform gene names from Ensemble IDs to a profile of gene symbols. Autophagy-related genes (ARGs) were downloaded from HAD (Human Autophagy Database) (http://www.autophagy.lu/index.html). There were 232 ARGs acquired from the database in total. The Infiltration Estimation of immune cells for all TCGA tumors was downloaded from TIMER2.0 (http://timer.compgenomics.org/). TCGA data are freely accessible and all above data acquired were fully complied with the access principles of the database. The present study was approved by the Biomedical Research Ethic Committee of Shandong Provincial Hospital (SWYX: NO. 2021-527).

Identification of Differentially Expressed IncRNAs (DElncs) in the TCGA Datasets
Based on the 232 ARGs sets obtained from HAD, we extracted 210 ARGs from the GC TCGA database and got 1215 ARlncs expression matrix on the base of co-expression analysis with corFilter=0.4 and p<0.001. Then the differentially expressed genes (DEGs) and IncRNAs between normal (n=32) and tumor (n=375) were obtained through R package to perform gene ontology (GO) and KEGG pathway enrichment analysis. The threshold value was set as |log fold change (FC)| >1 and fdr < 0.05.

Signature Development and Correlations with Clinicopathological Characteristics
Univariate coxregression analysis was performed on the candidate ARlncs through survival package of R to filtrate IncRNAs with significant prognostic value (P < 0.001). Then the prognostic signature was established based on the multivariable cox regression analysis. Ten ARlncs were extracted from TCGA database to stratify GC patients into low- and high-risk groups.
based on different risk score of each patient in order to construct the prognostic signature ultimately. The following formula was employed to calculate the risk score for each patient: $\hat{\text{Risk Score}} = \hat{\text{Coef}} \times \hat{\text{Exp}}$, where $\hat{\text{Coef}}$ is the coefficient of the risk score for each patient and $\hat{\text{Exp}}$ is the expression of ARlncs. The Kaplan–Meier survival curve was visualized by the R package to verify the predictive ability of the risk score. Whether the risk score was an independent prognostic indicator was examined by univariate and multivariate cox regression analysis. The areas under the receiver operating characteristic (ROC) curve (AUCs) for 1-year, 3-year, 5-year for the prognostic signature were calculated. Chi square test was conducted to investigate whether relevant clinicopathological characteristics were associated with risk signature in GC. The Gene set enrichment analysis (GSEA) was implemented to survey the diverse signaling pathways of the different risk groups using GSEA (version 4.1.0) software. The co-expression network of genes and IncRNAs was established via Cytoscape.

**Construction of Nomogram**

A nomogram that included T, M, N, age and risk score was established to visualize the specific method and showed the ability to predict overall survival (OS) at 1-year, 3-year and 5-year. In addition, the corresponding calibration curve of the nomogram to value its predictive power was implemented.

**Results**

**Identification of ARlncs in GC**

There were 19,573 mRNAs and 14,056 IncRNAs extracted from GC TCGA database in total. Based on the 232 ARGs obtained from HAD, we screened 210 ARGs in GC and 1215 ARlncs according to the co-expression analysis further with corFilter > 0.4 and $p < 0.001$ (Table S2). Among them, a total of 451 differently expressed IncRNAs (DEIncRNAs) were recognized via $|\log FC| > 1$ and FDR < 0.05 compared to the normal tissue. The procedure was showed in Figure 1.

**Construction of Autophagy-Related IncRNA Prognostic Signature**

Based on the differently expressed ARlncs, there were 17 ARlncs with prognosis value preliminary screened via univariate cox regression analysis with $p<0.05$ (Figure 2A). Among them, 10 ARlncs were filtered to construct a prognostic IncRNAs signature based on multivariable cox regression analysis. The risk score for predicting OS was

![Figure 1 Flowchart of the analysis.](image-url)
Figure 2 Construction of prognostic signature. (A) 17 ARlncs with prognosis value were preliminary screened via univariate cox regression analysis. (B and C) The distributions of risk scores and survival status of patients showed that with the increase of risk score, the number of death increased. (D) The heatmapping of different expression levels of the 10 ARlncs between the high- and low-risk groups. (E) Kaplan–Meier (K–M) analysis of patients in the high- and low-risk groups which displayed patients in the high-risk group had a poor prognosis compared to the low-risk group (p<0.001). (F) Verification of the prognostic value of the ARlncs signature by ROC analysis for 1-, 3-, 5-year. (G) The prognostic value of the ARlncs signature and other index, such as age, gender, grade, stage, T, M and N and ARlncs signature showed a good reliability. (H) Decision Curve Analysis (DCA) to identify the reliability of ARlncs signature.
calculated as follows: risk score = \[\text{LINC01094} \times (0.4371)\] + \[\text{AC068790.7} \times (0.4167)\] + \[\text{AC090772.1} \times (0.7607)\] + \[\text{AC005165.1} \times (0.1492)\] + \[\text{PVT1} \times (−0.0881)\] + \[\text{LINC00106} \times (−0.2345)\] + \[\text{AC026368.1} \times (0.0583)\] + \[\text{AC090912.3} \times (0.2048)\] + \[\text{AC013652.1} \times (0.5587)\] + \[\text{UICLM} \times (−0.3886)\] (Table 1). Patients were divided into high-risk group (n = 185) and low-risk group (n = 186) on the basis of the optimal risk score cutoff value. The distributions of risk scores and survival status are shown in Figure 2B and C. The changing trends of expression levels of the 10 ARlncs as determined via heatmapting were consistent with their risk coefficient in the prognostic signature (Figure 2D). Kaplan–Meier curves suggested that patients in the low-risk group had a better OS than those in the high-risk group with p<0.001 (Figure 2E). In order to validate the reliability of the prognostic signature, the AUC of each ROC for each clinical characteristic and risk score were calculated. The AUC of the risk score for predicting 1, 3, and 5-year survival of GC was 0.707, 0.689, and 0.707, respectively (Figure 2F) and the AUC of the risk score outweighed age with 0.589, gender with 0.546, grade with 0.567, T with 0.564, M with 0.539 and N with 0.579 (Figure 2G). Decision Curve Analysis (DCA) showed that the distance between risk curve and all curve was the largest, suggesting that risk score was the best index to evaluate the prognosis (Figure 2H).

Validation of the Independent Prognostic Ability of the Prognostic Signature and Construction of the Predictive Nomogram with Risk Score

The univariate cox regression analysis was conducted which exhibited that age, T, M, N and risk score processed the prognostic value with p<0.05 (Figure 3A). The multivariate cox regression analysis was implemented on the basis of the former results which showed that age, M, N and risk score had the ability to be an independent prognostic factor with p<0.05 (Figure 3B). The relevance between risk score and clinical features was also assessed which displayed that T and grade were variable between low and high-risk groups (Figure 3C). A nomogram was then constructed for OS predication which consist of risk score and clinical features including T, M, N and age. It could be visualized to predict the survival probability of 1-, 3-, and 5-year (Figure 3D). In order to validate the reliability of the nomogram, the AUC of each ROC for 1-, 3-, and 5-year was calculated and the results were 0.700, 0.730, 0.757, respectively (Figure 3E). The calibration curve was also calculated which showed good accuracy of nomogram for prediction (Figure 3F).

Enrichment Analysis of DEGs, Gene Set Enrichment Analysis (GSEA) and Co-Expression Network

Based on the difference analysis, 26 genes were found differently expressed between normal and tumor tissue. Go and KEGG enrichment analysis were conducted with DEGs. As to biological process, the DEGs significantly enriched in cell growth, positive regulation of intracellular protein transport and neuron death. For cellular component, the DEGs significantly enriched in vacuolar membrane, mitochondrial outer membrane and organelle outer membrane. About molecular function, the DEGs significantly enriched in protein kinase regulator activity, kinase regulator activity, ubiquitin protein ligase binding and ubiquitin-like protein ligase binding (Figure 4A). In KEGG enrichment analysis,

<table>
<thead>
<tr>
<th>LncRNA</th>
<th>Coefficient</th>
<th>HR</th>
<th>HR.95CI (Lower)</th>
<th>HR.95CI (Upper)</th>
</tr>
</thead>
<tbody>
<tr>
<td>LINC01094</td>
<td>0.4371</td>
<td>1.5483</td>
<td>1.1787</td>
<td>2.0337</td>
</tr>
<tr>
<td>AC068790.7</td>
<td>0.4167</td>
<td>1.5169</td>
<td>0.8593</td>
<td>2.6777</td>
</tr>
<tr>
<td>AC090772.1</td>
<td>0.7607</td>
<td>2.1399</td>
<td>0.9854</td>
<td>4.6469</td>
</tr>
<tr>
<td>AC005165.1</td>
<td>0.1492</td>
<td>1.1609</td>
<td>0.9157</td>
<td>1.2032</td>
</tr>
<tr>
<td>PVT1</td>
<td>−0.0881</td>
<td>0.9157</td>
<td>0.6713</td>
<td>1.0180</td>
</tr>
<tr>
<td>LINC00106</td>
<td>−0.2345</td>
<td>0.7910</td>
<td>0.6780</td>
<td>1.0136</td>
</tr>
<tr>
<td>AC026368.1</td>
<td>0.0583</td>
<td>1.0600</td>
<td>0.8353</td>
<td>1.3170</td>
</tr>
<tr>
<td>AC090912.3</td>
<td>0.2048</td>
<td>1.2273</td>
<td>0.9888</td>
<td>1.5234</td>
</tr>
<tr>
<td>AC013652.1</td>
<td>0.5587</td>
<td>1.7483</td>
<td>1.1420</td>
<td>2.6767</td>
</tr>
<tr>
<td>UICLM</td>
<td>−0.3886</td>
<td>0.6780</td>
<td>0.4535</td>
<td>1.0136</td>
</tr>
</tbody>
</table>
Figure 3 Validation of the prognostic signature and construction of the Nomogram. (A) Univariate cox regression analysis exhibited that age, T, M, N and risk score processed the prognostic value (p<0.05). (B) The multivariate cox regression analysis showed that age, M, N and risk score had the ability to be an independent prognostic factor (p<0.05). (C) Relevance between risk score and clinical features. (D) A nomogram consisted of risk score and clinical features including T, M, N and age was constructed. (E) The AUC of each ROC for 1-, 3-, and 5-year to validate the reliability of the nomogram. (F) Calibration curve for the predictive accuracy of the nomogram.
the DEGs significantly enriched in platinum drug resistance, apoptosis, ErbB signaling pathway and p53 signaling pathway. It also showed that the DEGs were related to several viral diseases and cancers which suggested it may play a significant role in the immune (Figure 4B). Further functional annotation was conducted through GSEA, and the results showed differences in biological functions between the high-risk and low-risk groups. GSEA analysis results showed that dilated cardiomyopathy, arrhythmogenic right ventricular, ECM receptor interaction, hypertrophic cardiomyopathy and focal adhesion (Figure 5A–E) were enriched in the high-risk group. And the base excision repair, spliceosome, glyoxylate and dicarboxylate metabolism, oxidative phosphorylation and DNA replication (Figure 5F–J) were enriched in the low-risk group.

As to co-expression network, in order to present the connections and mechanisms linking prognosis-related ARlncs and related mRNAs better, the Cytoscape was used to visualize the results (Figure 5K).

**Immune Microenvironment, Immune Function and Checkpoint**

The difference of Stromal Score, Immune Score, microenvironment score and relative infiltration abundance of immune cells of the two risk groups were estimated based on the 7 algorithms such as TIMER, CIBERSORT, CIBERSORE-ABS, QUANTISEQ, MCPCOUNTER, XCELL and EPIC (Table S3). The results suggested that there were large amounts of microenvironment cells which had significant differences between the low- and high-risk group and most of them were enriched in high-risk group. According to the CIBERSORE-ABS algorithm, Macrophage M2, T cell CD8+ and T cell CD4+ memory resting were the types of cells with greatest difference between the two risk groups (Figure 6A). As to immune function, basically all the functions including APC co-inhibition, APC co-stimulation, CCR, checkpoint, cytolytic activity, HLA, inflammation promoting, parainflammation, T cell co-inhibition, T cell co-stimulation, Type-I IFN response and Type-II IFN response were enhanced in high-risk group except MHC-class I with no significant difference (Figure 6B). Given the importance of immune checkpoints in cancer treatment, the expression of checkpoint genes was compared between the two risk groups. We found TNFRSF25 had higher expressions in the low-risk group. Others including CD274 (PD-L1), PDCD1 (PD-1) and PDCD1LG2 (PD-L2) were expressed higher in the high-risk group (Figure 6C).

**Discussion**

GC is the common malignant tumor with great risks to human health. Even though plenty of new methods and drugs for diagnosis and treatments have emerged in recent years, prognosis of GC was still unsatisfactory. Therefore, individualized and precise diagnosis and treatments are desperately needed for GC patients.

Mounting evidences have indicated that lncRNAs as regulatory molecules in gene expression were directly implicated in a wide range of functions in the development of various human diseases, especially many types of cancers. 

![Figure 4 Go (A) and KEGG (B) enrichment analysis of DEGs.](https://doi.org/10.2147/IJGM.S348943)
Figure 5 Distinct pathways enriched in the high-risk and low-risk groups and co-expression network of ARlncs. (A–E) Pathways enriched in the high-risk group. (F–J) Pathways enriched in the low-risk group. (K) Co-expression network of ARlncs.
have found that some lncRNAs were upregulated in GC and have an oncogenic effect by inducing proliferation and migration of GC cells. Meanwhile, some of them had tumor-suppressive role by inhibiting GC cells migration and invasion. As a type II programmed cell death, autophagy also plays crucial roles in cancers. In GC, it displayed tumor-suppressor and a tumor-promoter roles via related signaling pathways, genes and ncRNAs. Importantly, numerous researches have indicated the vital roles of lncRNAs in autophagy-inducing progression or inhibition of GC. Therefore, constructing a prognosis signature composed with lncRNAs associated with autophagy will be of great practical value.
In our study, 17 ARlncs with prognostic value were extracted by the initial screen via univariate cox regression analysis and 10 of them were filtered to construct a prognostic IncRNAs signature based on multivariable cox regression analysis. The signature stratified patients into low- and high-risk groups via the cutoff value of risk score and patients in different groups displayed different OS, which implicated that the patients in the low-risk group had a longer survival time. According to the univariate and multivariate cox regression analysis, the risk signature was proved to be an independent prognostic factor. A nomogram was established ultimately for predicting 1-, 3-, and 5-year overall survival probabilities in GC patients with good accuracy. The roles of immune cells infiltrating the tumor microenvironment and immune checkpoint genes in GC patients were also investigated. The results highlighted potential therapeutic targets in GC patients.

As to the 10 ARlncs which constructed the prognosis signature, four of them have been proved to play vital roles in multiple cancers. Xu et al found that the expression of LINC01094 was elevated in ovarian cancer (OC) tissues and cell lines and high LINC01094 expression was associated with higher FIGO stage, lymph node metastasis and the shorter overall survival rate in patients with OC. In addition, in other cancers, such as glioblastoma, glioma, clear cell renal cell carcinoma, LINC01094 also exhibited the effect of accelerating the cancer development. In our study, the HR of LINC01094 was 1.640 refer to the univariate cox regression analysis, which meant LINC01094 was an effector enhancing the risk of GC. LINC00106 is lowly expressed in Thyroid cancer (TC) specimens, which attenuates migratory and invasive abilities in TC by inhibiting EMT as a tumor suppressor. In GC, it also proved to be highly expressed in low-risk group compared to the high-risk group and acted as a negative regulator. PVT1 has been discovered processing oncogenic properties and regulating proliferation and growth of many cancers and the upregulation was involved poor prognosis in esophageal adenocarcinoma. But in our study, the PVT1 was found to be a low risk factor. Hence it needs further research in the future study. The IncRNA UICLM was significantly up-regulated in cases of colorectal cancer (CRC) with liver metastasis and UICLM expression was associated with poor patient survival. But contrary to what was found in CRC, UICLM was discovered to be a negative regulator with higher expression in the low-risk group in GC. At present, there is no relevant study about AC005165.1, AC068790.7, AC090772.1, AC026368.1, AC090912.3, AC013652.1 and more in-depth study will be carried out in the future.

Equally important, our research adopted the TIMER database to reveal connections between the risk signature and immune infiltration levels in gastric cancer. The results showed the associations of the risk signature and numerous immune cells. With respect to check point, most related genes were highly expressed in the high-risk group such as PD-L1, PD-1 and PD-L2, which implied that immune therapy seemed to be a good choice for these patients.

Conclusions
In conclusion, our study constructed an ARlncs-based prognostic signature. And immune checkpoints such as PD-L1, PD-1 and PD-L2 tended to be highly expressed in the high-risk group, which suggested that immunotherapy may be a promising therapy for these patients. A nomogram was constructed based on risk signature and clinical characteristics and displayed good accuracy to predict the prognosis of GC patients.

Funding
The study was funded by the Natural Science Foundation of Shandong Province (ZR2020QH035).

Disclosure
The authors report no conflicts of interest in this work.

References