Autophagy-related Long Non-coding RNA Signature Associated with Prognosis of Lung Adenocarcinoma

Guofei Zhang  
Second Affiliated Hospital of Zhejiang University

Jiayi Shen  
Zhejiang University-University of Edinburgh Institute, Zhejiang University School of Medicine

Zipu Yu  
Second Affiliated Hospital of Zhejiang University

Gang Shen  
Second Affiliated Hospital of Zhejiang University

Chengxiao Liang (✉ zycszju@yahoo.com)  
Zhejiang Hospital

Research Article

Keywords: lung adenocarcinoma, autophagy, IncRNA, overall survival, prognostic signature

DOI: https://doi.org/10.21203/rs.3.rs-136739/v1

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

Background

Evidence suggests that long non-coding RNAs (lncRNAs) are involved in various cancers. Here, we developed and evaluated an autophagy-related prognostic lncRNA signature for lung adenocarcinoma (LUAD).

Results

Using a publicly available microarray dataset from The Cancer Genome Atlas, we analyzed the lncRNA expression profile in a cohort of 439 LUAD patients. The lncRNA-mRNA co-expression network along with univariate and multivariate Cox regression analyses were used to determine 15 autophagy-related lncRNA signatures that were significantly correlated with patient overall survival. Autophagy-related lncRNA signatures stratified patients into high- and low-risk groups with significantly different survival (hazard ratio = 3.256, 95% confidence interval = 2.858–4.101, P < 0.001). The lncRNA signature was further confirmed in other independent datasets. Moreover, the lncRNA signature had prognostic value independent of routine clinical factors. Functional analysis indicated that autophagy-related lncRNA signatures may be involved in LUAD via known autophagy-related pathways.

Conclusions

This newly identified autophagy-related lncRNA signature is a more powerful prognostic tool than the clinicopathological factors routinely used to predict patient survival, and can provide further insights into the molecular mechanisms underlying LUAD.

Background

Lung adenocarcinoma (LUAD) is the most common form of lung cancer, causing high morbidity and mortality annually worldwide [1]. Although progress in the treatment of LUAD has improved the short-term survival rate, the impacts on long-term survival remain modest [2]. Thus, there is a need to elucidate the mechanism underlying LUAD and to accurately identify prognostic molecular markers for predicting clinical outcomes with clinical significance.

Autophagy is a catabolic process in which proteins and whole organelles in cells are guided to degradation by lysosomes. This process enables the cells to survive stress by providing it with a continuous supply of energy and biomolecules [3]. Autophagy can also inhibit cancer initiation as a tumor suppressor mechanism in the early stages of tumor formation [4, 5]. Karantza-Wadsworth et al. [6] reported that deletion of or defects in many autophagy-related genes cause DNA damage and genomic instability, which further promote the activation of proto-oncogenes, resulting in tumorigenesis. However, given the intricate and dual function that autophagy plays in the survival and apoptosis of cancer cells, the development of effective therapeutic approaches that target the link between autophagy and cancer
has been challenging. As such, the significance of autophagy as a novel and alternative approach in the development of LUAD therapies remains relatively unclear.

Long non-coding RNAs (lncRNAs) continue to attract substantial attention owing to increasing recognition of the roles they play in the pathogenesis of several diseases [7]. LncRNAs comprise a heterologous family of RNAs with lengths greater than 200 nucleotides that lack any detectable open reading frame [8]. Accumulating evidence indicates that lncRNAs play an important role in many basic biological processes such as transcription, cell cycle, imprinting, translation, and protein localization, and they have been highly implicated in cancer progression [9]. LncRNAs have been shown to modulate autophagy by interacting with numerous autophagy-related genes at various stages. For example, LINCO00515 has been reported to play different roles in several types of cancers, including inhibiting autophagy and chemoresistance in melphalan-resistant myeloma, as well as being involved in tumorigenesis in stage I LUAD [10]. The lncRNA lung cancer progression-association transcript 1 (LCPAT1) binds RCC2 to mediate autophagy and enhance the progression of lung cancer [11]. Despite these recent findings, information regarding the regulatory role of lncRNAs in mediating autophagy during lung cancer development is currently limited.

Herein, we aimed to identify lncRNAs that are functionally related to the process of autophagy in LUAD. Toward this end, we analyzed the expression profile of lncRNAs related to autophagy in The Cancer Genome Atlas (TCGA)-LUAD dataset and established a co-expression network of autophagy-related lncRNAs. These findings will offer new insights into the development of novel therapies against LUAD.

Methods

Data acquisition

We obtained RNA expression profiling data and related data of 497 LUAD patients from the TCGA database in February 2020. The lncRNA and mRNA expression data were downloaded in FPKM format, with a normalized count of log₂(x + 1) conversion. The criteria for screening data were as follows: (1) histopathological diagnosis of LUAD; (2) complete patient pathological stage and follow-up information available; and (3) absence of any other type of malignant tumor. As a result, we obtained RNA-sequencing data from 439 LUAD tissues (Supplementary Table S1). Data were processed as per the TCGA guidelines (https://www.cancer.gov/about-nci/organization/ccg/research/structural-genomics/tcga).

LncRNA and autophagy gene screening

Genes related to autophagy were retrieved from the human autophagy database (http://autophagy.lu/). The correlations between levels of autophagy-related genes and lncRNAs were determined using the Pearson correlation coefficient with R software (www.r-project.org). A correlation coefficient |R²| for lncRNA > 0.3 and P < 0.001 were used to define lncRNAs related to autophagy.
Construction of an lncRNA signature

To obtain more detailed information about the role of autophagy-related lncRNAs in LUAD, the differentially expressed lncRNAs related to overall survival (OS) were determined using a univariate Cox proportional hazards regression model (significance level set to 0.01). The obtained autophagy-related lncRNAs with differential expression were then used in multiple Cox regression analysis models, and a prognostic signature was constructed from the autophagy-related lncRNAs. A prognostic risk score that could predict the OS was determined as follows: risk score = \exp_1 \cdot \beta_1 + \exp_2 \cdot \beta_2 + \ldots + \exp_n \cdot \beta_n, where \exp represents the expression level, n is the total number of key lncRNAs, and \beta is the regression coefficient. LUAD patients were categorized into low-risk and high-risk groups on the basis of the median risk score, and the prognosis and predictive role of this risk score in LUAD were explored.

Independence of the prognostic lncRNA signature from clinical parameters in the TCGA cohort

Multivariate and univariate Cox analyses were used to predict the prognosis of clinical features and risk scores [including gender, age, pathological stage, and tumor lymph node metastasis (TNM) stage system] [12]. P < 0.05 was defined as a statistically significant difference. The Cox regression model was utilized to calculate the hazard ratios (HRs) and 95% confidence intervals (CIs). Only patients with complete clinical information were included in this analysis.

Nomogram construction and validation

A nomogram of the final multivariate model was constructed as a visual aid to manually obtain predicted values from the Cox model. Receiver operating characteristic (ROC) analysis with area under the curve (AUC) value calculations was also performed to predict the sensitivity and specificity of predicting OS through our proposed risk score and clinical characteristic comparison.

External validation of the lncRNA signature

We further validated the autophagy-related lncRNA signature using external datasets collected from the European Genomephenome Archive (EGA, http://www.ebi.ac.uk/ega/) under accession code EGAD00001004421 [13]. The external datasets utilized the same platform as the TCGA-LUAD dataset, Illumina HiSeq 2000. All clinical records were collected from OncoSG (https://src.gisapps.org/OncoSG/) (Supplementary Table S2).

Gene set enrichment analysis (GSEA)

GSEA analysis was performed to assess whether differentially expressed genes between the low-risk and high-risk groups were enriched during autophagy [14]. Gene expression data of the two groups in the TCGA-LUAD cohort and the dataset from Kyoto Encyclopedia of Genes and Genomes (KEGG) were imported into GSEA software and analyzed under the following settings: permutation = phenotype, gene ranking index = Signal2Noise, enrichment statistics = weighting, and permutation = 1000. Standardized
enrichment scores and P values were obtained, and a false discovery rate (FDR)-corrected P-value < 0.05 was set as the threshold for significance.

Data analysis

All data analyses were conducted using R software (v 3.6.1; www.r-project.org). Mann–Whitney U tests and unpaired Student’s t-tests were employed to compare data among groups. The Kruskal-Wallis H test and one-way analysis of variance were employed to compare data among multiple groups. Fisher’s exact test and Pearson’s chi-square test were applied to assess the correlation between risk score and clinicopathological features. Visualization of the autophagy-related IncRNA/mRNA co-expression network was performed using Cytoscape software (v 4.0.3) [15]. A P value of < 0.05 defined a statistically significant difference. Kaplan–Meier survival analysis was executed using the R package “survival” [16].

Results

Co-expression network reconstruction for autophagy-related IncRNAs

In total, 232 genes identified to be related to autophagy were retrieved from the human autophagy database (HADb). We cross-referenced the 232 genes and obtained their levels of expression in the TCGA-LUAD dataset. A total of 14,142 IncRNAs were identified in the LUAD dataset. After excluding low-expression IncRNAs, we established an autophagy-IncRNA co-expression network to identify IncRNAs that are related to autophagy with common expression trends across different samples included in this study. Finally, 1651 IncRNAs were identified (|R^2| > 0.3 and P ≤ 0.001).

Identification of differentially expressed IncRNAs related to autophagy in LUAD

In total, 255 differentially expressed autophagy-related IncRNAs (61 downregulated and 194 upregulated) were identified between control and LUAD tissues under the threshold of |log_2 FC| ≥ 0.5 and an adjusted P-value < 0.05 (Supplementary Fig. S1).

Determination of a IncRNA signature related to autophagy for prognostic prediction of LUAD

First, we carried out univariate Cox regression analysis to determine IncRNAs with prognostic value based on the 255 IncRNAs identified to be related to autophagy. The prognostic IncRNAs related to autophagy were sorted in ascending order according to their P-values; P < 0.01 was set as the cut-off value and IncRNAs that met this requirement were used to develop the signature. In total, 21 IncRNAs were identified to be of significant prognostic value for LUAD patients (P < 0.01, Fig. 1). Further, lasso regression based on univariate analysis selected 15 autophagy-related IncRNAs, as shown in Table 1. Output data from the network analysis were imported into Cytoscape for network visualization (Fig. 2). Ultimately, nine
IncRNAs related to autophagy were identified as favorable prognostic factors for LUAD patients (MIR22HG, LINC01116, LINC00578, HSPC324, AL365181.3, AC125807.2, AC027031.2, AC021016.2, and AC010980.2; Fig. 3).

Table 1
The regression coefficients using lasso regression of 15 autophagy-related IncRNAs

<table>
<thead>
<tr>
<th>lncRNA</th>
<th>Coefficients</th>
</tr>
</thead>
<tbody>
<tr>
<td>MIR22HG</td>
<td>-0.0193</td>
</tr>
<tr>
<td>AC021016.2</td>
<td>-0.0706</td>
</tr>
<tr>
<td>AC027031.2</td>
<td>0.0483</td>
</tr>
<tr>
<td>LINC00578</td>
<td>-0.0689</td>
</tr>
<tr>
<td>AGAP1.IT1</td>
<td>0.0565</td>
</tr>
<tr>
<td>AC010980.2</td>
<td>0.0757</td>
</tr>
<tr>
<td>AL049555.1</td>
<td>0.0071</td>
</tr>
<tr>
<td>LINC01876</td>
<td>0.0430</td>
</tr>
<tr>
<td>AC125807.2</td>
<td>0.0538</td>
</tr>
<tr>
<td>LINC01116</td>
<td>0.0718</td>
</tr>
<tr>
<td>LINC01137</td>
<td>0.0445</td>
</tr>
<tr>
<td>AP001033.2</td>
<td>0.0779</td>
</tr>
<tr>
<td>HSPC324</td>
<td>-0.1358</td>
</tr>
<tr>
<td>LINC00857</td>
<td>0.0662</td>
</tr>
<tr>
<td>AL365181.3</td>
<td>0.0261</td>
</tr>
</tbody>
</table>

The autophagy-related IncRNA signature has better predictive ability for LUAD

We subsequently used risk scoring to design an autophagy-associated IncRNA signature. Patients with LUAD were sub-classified into low- and high-risk groups based on the median risk score (Fig. 4). The risk score strongly predicted the OS of LUAD patients, with the low-risk group having 57% OS at 5 years and the high-risk group having only a 22% OS (P < 0.001, Fig. 5). Moreover, Cox regression analyses showed that the risk score could predict the prognosis of LUAD patients (HR = 3.256, 95% CI: 2.858–4.101; P < 0.0001, Fig. 6A). Further, we used multivariate Cox regression analysis to explore if the risk score signature could independently predict the prognosis of patients with LUAD. The HR value was 2.919,
indicating that the risk score could be important in predicting the survival of LUAD patients, removing the effects of other factors (e.g. age, sex, stage, and TNM; Fig. 6B).

ROC analysis was then used to determine whether the risk score could improve the prediction of outcome as compared to commonly used clinical factor prediction models. The AUC for the risk score was 0.728 (Fig. 7). The nomogram was constructed based on six projecting factors, including risk score, age, sex, stage, T, and N. This suggested that in LUAD, the risk score of the autophagy-based IncRNA signature is appropriate to predict disease outcomes (Fig. 8).

**Application of the autophagy-based IncRNA signature to EGA cohorts**

The IncRNA signature was further verified in an additional dataset (EGA) using the same β value. A total of 152 patients were included in the verification of autophagy-related IncRNA signatures. The patients were classified into low-risk and high-risk groups based on the median risk score (Fig. S2). Similar to the findings from the TCGA dataset, low-risk patients in the EGA cohort had a better OS at 5 years than high-risk patients (82.4% vs. 55.5%; P < 0.001; Fig. 9). Cox regression analysis further confirmed this finding, wherein EGA patients in high-risk group had a shorter OS relative to those in the low-risk group (HR = 1.827, 95% CI: 1.167–2.861; P = 0.008; Fig. 10). Based on these findings, we verified the accuracy of the IncRNA signature in predicting the disease outcomes of LUAD patients.

**GSEA**

Enrichment map visualization of GSEA in LUAD patients was performed in the high- versus low-risk score groups based on the pathways provided by the KEGG database to examine the effects of the autophagy-related IncRNA signature on survival in the TCGA-LUAD and EGA cohorts. Sixty-seven gene sets were overtly enriched at FDR < 0.05 (Table S3). Amongst these gene sets, numerous pathways have a well-recognized role in cancers, which include cell cycle, DNA replication, p53 pathway, the cytokine-cytokine receptor interaction, and JAK/STAT signaling pathway. It is worth noting that GSEA indicated that these gene sets participate in the lysosome pathway, proteasome pathway, oxidative phosphorylation, and phosphatidylinositol signaling system, which are strongly linked to autophagy (Fig. 11). In summary, the selected autophagy-related IncRNAs participate in important cancer and autophagy pathways, and this revelation might aid in the development of an autophagy-targeted treatment for LUAD.

**Discussion**

LUAD is the most prevalent type of lung cancer and a prototype for precision oncology [17]. Accumulating evidence shows that IncRNAs regulate the pathological processes of lung cancer. Several studies have also identified some IncRNAs that affect the prognosis of non-small cell lung cancer; however, to date, there is no systematic method to identify IncRNA signature sets to predict survival for LUAD patients. Herein, we identified a novel and efficient autophagy-related IncRNA signature based on a TCGA dataset and validated its efficiency using an independent EGA dataset. Our signature was able to effectively
stratify patients according to OS. We further determined the efficacy of our signature in the training set and verification set, indicating that the signature has high prognostic value.

To our knowledge, this is the first autophagy-related IncRNA predictive model developed based on RNA-sequencing data using a cohort of more than 100 LUAD patients. In the first step, we identified 1651 IncRNAs based on the IncRNA–autophagy gene co-expression network. The expression profiles of these IncRNAs in 439 LUAD patients were then analyzed by univariate and stepwise multiple Cox proportional hazards regression analyses. Fifteen IncRNA signatures related to autophagy were identified and used to establish a predictive model based on the linear combination of these IncRNAs. Using this predictive model, distinctive separation was observed in the survival curve between the high-risk and low-risk patient groups. The AUC obtained by ROC analysis was 0.728, demonstrating the high sensitivity and specificity of this IncRNA signature model. When considering other clinical factors together, multivariate Cox regression analysis indicated that the autophagy-associated IncRNAs could independently predict survival outcomes from these conventional clinical-pathological factors, including stage, sex, age, and TNM. Further stratified analysis showed that when predicting different survival rates of patients in the EGA cohort, the autophagy-related IncRNA signature has good discriminatory power.

To date, only a few IncRNAs have been well characterized with respect to their biological functions. Recent studies have found that the level of MIR22HG, one of our prognostic IncRNAs, is significantly reduced compared to that in normal lung tissue. Further, silencing MIR22HG triggers the signaling for cell survival and apoptosis via aberrant expression of the oncogenes p2, YBX1, and MET [18]. Other studies suggested that MET and p21 are important proteins involved in the autophagy pathway [19, 20]. In line with these previous results, our analysis based on the TCGA cohort showed that HSPC324 expression levels are distinctly higher in lung tumor tissues than those in normal tissues [21]. Further, abnormal expression of the IncRNA HSPC324 inhibits proliferation, and increases apoptosis and reactive oxygen species (ROS) accumulation in lung adenocarcinoma cells, as an important member of the ROS-dependent autophagy pathway [21, 22]. However, the biological functions of most of our prognostic IncRNAs are not clear.

LncRNAs participate in various biological activities by either promoting or suppressing gene expression at the transcriptional and post-transcriptional levels [23]. Thus, it is possible to deduce the biological role of an IncRNA from the functional perspective of co-expressed protein-coding genes (PCGs) [24]. Therefore, we studied the pattern of co-expression with regard to mRNAs and lncRNAs, and conducted functional enrichment analysis of the co-expressed PCGs to determine the biological function of prognosis-associated lncRNAs in the signature [25, 26]. Our results revealed that PCGs for which the expression value was negatively or positively related to the prognosis of IncRNA and also related to autophagy were enriched in several functional groups of recognized pathways related to autophagy, such as the lysosome pathway, proteasome pathway, oxidative phosphorylation, and phosphatidylinositol signaling system, which are also strongly related to the pathogenesis of LUAD. Therefore, it is reasonable to speculate that autophagy-related IncRNA markers might participate in LUAD by regulating PCGs in recognized autophagy-related biological pathways and activities.
Conclusion

In summary, this study identified IncRNA signatures associated with autophagy, which can predict the survival of LUAD patients. This signature shows good prognostic ability that is independent of routine clinical-pathological factors and can strongly predict the survival outcome of LUAD patients of the same TNM stage. This signature could thus be used to identify patients with high-risk scores who should receive more efficient and personalized treatment. In addition to providing a new potential biomarker to stratify disease risk in LUAD patients, this signature can also be used in further research to gain insights into the molecular mechanisms involved in the development of LUAD. However, further clinical searches are required to verify the predicted efficacy of these signatures, in addition to experimental studies to investigate the function of the prognostic IncRNAs.

Abbreviations

IncRNA, long noncoding RNA; LUAD, lung adenocarcinoma; TCGA, The Cancer Genome Atlas; OS, overall survival; TNM, tumor lymph node metastasis; HR, hazard ratio; CI, confidence interval; ROC, receiver operating characteristic; AUC, area under the curve; EGA, European Genomephenome Archive; GSEA, gene set enrichment analysis; HADb, Human Autophagy Database

Declarations

Ethics approval and consent to participate

Not applicable

Consent for publication

Not applicable

Availability of data and materials

The study was based upon data generated by the TCGA Research Network (http://cancergenome.nih.gov/) and EGA (https://www.ebi.ac.uk/ega/).

Competing interests

The authors declare that they have no competing interests.

Funding
This work was supported by supported by Scientific Research Fund of Zhejiang Provincial Education Department, No. Y201534188.

Authors’ contributions

ZG and LC conceived and designed the study, and drafted the manuscript. SJ, YZ and SG performed the data acquisition, analyzed the data and interpreted the results. ZG and YZ wrote and edited the manuscript. LC contributed to critical revisions of the text. All authors read and approved the final manuscript.

Acknowledgments

Not applicable.

References


Figures
Figure 1

Forest plot showing the results of univariate Cox proportional hazard regression analysis for 21 autophagy-associated lncRNAs.
Figure 2

Network of IncRNAs with prognostic value that were co-expressed with autophagy-related genes in the TCGA-LUAD dataset. In the centric positions, red nodes represent IncRNAs and blue nodes represent autophagy genes.
Figure 3

Prognostic impact of the nine lncRNAs in patients with lung adenocarcinoma by the Kaplan–Meier method.
Figure 4

Risk scores of the autophagy-associated IncRNAs in LUAD patients in the TCGA dataset. (a) Low- and high-risk score groups of autophagy-related IncRNA signatures in LUAD patients. (b) Survival duration and status of LUAD patients. (c) Heatmap of 15 key IncRNA expression profiles in patients with LUAD. The colors from blue to red show the growth trend from low to high levels.
Figure 5

Risk scores of autophagy-associated IncRNAs of LUAD patients in the TCGA dataset as determined by the Kaplan–Meier method. Low-risk samples had good survival outcomes.
<table>
<thead>
<tr>
<th>p value</th>
<th>Hazard ratio</th>
</tr>
</thead>
<tbody>
<tr>
<td>age</td>
<td>0.228</td>
</tr>
<tr>
<td>gender</td>
<td>0.385</td>
</tr>
<tr>
<td>stage</td>
<td>&lt;0.001</td>
</tr>
<tr>
<td>T</td>
<td>&lt;0.001</td>
</tr>
<tr>
<td>N</td>
<td>&lt;0.001</td>
</tr>
<tr>
<td>M</td>
<td>0.019</td>
</tr>
<tr>
<td>riskScore</td>
<td>&lt;0.001</td>
</tr>
</tbody>
</table>

Figure 6

Forest plot depicting hazard ratios (HRs) and 95% confidence intervals (95% CI) of the autophagy-associated lncRNAs in LUAD patients for overall survival. (a) Univariate analysis; HR = 3.256, 95% CI: 2.858–4.101. (b) Multivariate analysis; HR = 2.919, 95% CI: 2.277–3.742.
Figure 7

Receiver operating curve showing the risk scores of the autophagy-associated IncRNAs, with an area under the curve of 0.728.
Figure 8

Nomogram constructed to predict the 1-, 3-, and 5-year survival rates of patients with lung adenocarcinoma based on the risk signature with autophagy-related lncRNAs.
Figure 9

Kaplan–Meier curves for LUAD patients in the EGA dataset based on the autophagy-related lncRNA risk score model.
Figure 10

Forest plot depicting hazard ratios (HRs) and 95% confidence intervals (95% CIs) of the autophagy-related lncRNA signature for overall survival of patients with lung adenocarcinoma in the EGA dataset. (a) Univariate analysis. (b) Multivariate analysis.
Figure 11

Significantly enriched KEGG pathways in the TCGA cohort based on gene set enrichment analysis (GSEA). Five representative KEGG pathways in high- and low-risk patients are shown.

Supplementary Files

This is a list of supplementary files associated with this preprint. Click to download.

- SupplementaryMaterial.docx