Construction and Validation of an Immune-Related IncRNA Prognosis Model for Thyroid Cancer

Zheng Li  
Chongqing University Cancer Hospital

Hui Wang  
Chongqing University cancer hospital

Xia Deng  
Chongqing University Cancer Hospital

Jing Zhang  
Chongqing University Cancer Hospital

Wanyan Tang  
Chongqing University Cancer Hospital

Ling Wang  
Chongqing University Cancer Hospital

Wenxin You  
Chongqing University Cancer Hospital

Weiqi Nian (✉ oceanriverlake@163.com)  
Chongqing University cancer hospital  https://orcid.org/0000-0001-6596-9443

Research Article

Keywords: thyroid cancer, immune, IncRNA, prognosis, biomarker, TCGA

Posted Date: April 14th, 2022

DOI: https://doi.org/10.21203/rs.3.rs-722174/v3

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

Version of Record: A version of this preprint was published at Combinatorial Chemistry & High Throughput Screening on February 23rd, 2022. See the published version at https://doi.org/10.2174/1386207325666220223123917.
Abstract

Background: Immune-related long noncoding RNAs (lncRNAs) play an important role in the development of cancer. This study aimed to identify immune-related lncRNAs in thyroid cancer (THCA) and to develop a prognostic model for THCA.

Method: We downloaded immune-related gene sets from the Gene Set Enrichment Analysis (GSEA) website and obtained THCA gene expression and clinical data from The Cancer Genome Atlas (TCGA) database. Immune-related lncRNAs were then obtained by performing a correlation analysis on the expression of lncRNAs and immune-related genes. Prognostic model for THCA immune-related lncRNAs was developed through univariate Cox regression and multiple Cox regression analyses. We confirmed the results in clinical samples using quantitative real-time PCR.

Results: A total of 26 immune-related lncRNAs in THCA were obtained. Then we constructed a prognosis model composed of seven lncRNAs (LINC01614, AC017074.1, LINC01184, LINC00667, ACVR2B-AS1, AC090673.1 and LINC00900). Our model can be used as an independent prognostic factor. Principal component analysis displayed that the lncRNAs in the model can distinguish between high and low-risk groups. Clinical correlation analysis showed that the expression levels of AC090673.1 (P<0.05), LINC01184 (P<0.001), and LINC01614 (P<0.001) were related to disease stage, and LINC00900 (P<0.001) and LINC01614 (P<0.001) were related to T stage. We validated this model in cancer and paracancerous tissues from 24 THCA patients.

Conclusion: We identified and experimentally validated seven immune-related lncRNAs that can serve as potential biomarkers for THCA prognosis.

Introduction

Thyroid cancer (THCA) is a malignant tumour of the thyroid epithelium. Five-year survival rate for patients with invasive THCA is less than 50%. Most noninvasive patients have a good prognosis [1]. However, some patients with noninvasive THCA are prone to recur and distal metastases may occur [2]. Existing prognostic predictors for THCA do not accurately predict patients’ survival and recurrence [3,4]. Therefore, novel and reliable markers for the prognosis of patients with THCA are urgently needed.

1.1. Tumour immunotherapy in thyroid cancer

Tumor immunotherapy has developed into an important treatment approach. It is gradually becoming a hot topic in oncology therapeutics and brings hope to cancer patients. In advanced follicular carcinoma and anaplastic THCA, multiple preclinical studies of cancer immunotherapy showed prospect for anticancer applications. Anti PD-L1 treatment enhanced the effect of the BRAF inhibitor on THCA regression [5]. A variety of immune cells are present in the THCA microenvironment. These cells can fight tumors or promote them. In addition, proinflammatory cytokines and chemokines secreted by immune cells can promote or inhibit the proliferation of tumor cells [6]. The fact that cancer can survive in this
immune microenvironment illustrates the critical role of immune regulation in THCA. Chronic inflammation is associated with the presence and severity of THCA [7]. The expression of immune-related genes in cancer cells can regulate the abundance of infiltrating immune cells, Affecting tumor diagnosis, prognosis, and sensitivity to clinical treatment. These findings have prompted researchers to explore the expression of immune-related coding and noncoding genes in THCA and provide ideas for the clinical diagnosis, treatment, and prognosis of THCA from an immune perspective.

1.2. Immune-related LncRNA in thyroid cancer

The immune-related genes associated with the prognosis of THCA patients have not been fully revealed. Long noncoding RNAs (lncRNAs) are important components of noncoding RNAs and play an important role in pre-transcriptional, post-transcriptional, and posttranslational levels as well as in tumor development. Liyanarachi et al. found that the expression of lncRNAs XLOC051122 and XLOC006074 correlated significantly with lymph node metastasis and mutations in BRAF (V600E) in papillary THCA patients [8]. Lei et. al found that lncRNA TUG1 is related to the tumor progression of THCA [9]. Teng et al. showed that lncRNA PACERR and CYP4A22-AS1 are associated with the aggressiveness and disease-free survival of papillary THCA [10]. The clinical relevance and prognostic importance of immune-related lncRNAs need to be further investigated in THCA.

In this study, we provide insights into immune-related lncRNAs associated with THCA prognosis. A predictive model was developed and experimentally validated. The model served as an independent predictor. Our study provides potential immunotherapeutic targets for THCA.

Materials And Methods

2.1. Data collection

Expression profiles and clinical information of mRNA and lncRNA of patients with THCA were obtained from The Cancer Genome Atlas (TCGA: https://cancergenome.nih.gov/). It contains 58 paracancerous tissues and 509 THCA tissues. Patients with survival times of less than 30 days and unknown survival times were excluded.

2.2. Immune-related LncRNAs

We searched the Gene Set Enrichment Analysis (GSEA: https://www.gsea-msigdb.org/gsea/msigdb/index.jsp), using immune-related keywords “IMMUNE_RESPONSE” and “IMMUNE_SYSTEM_PROCESS”, and then downloaded the gene sets. The expression of immune-related genes was extracted from TCGA and analyzed with Perl. Immune-related lncRNAs were obtained by analyzing relevance between lncRNAs and immune-related genes by “limma”in R. The filter criteria were as follows: (1) all the genes in the simples had level of>0.5; (2) the levels of lncRNAs in all samples were in fluctuation.
2.3. Prognostic model construction

To screen immune-related lncRNAs that are significantly associated with prognosis, we analyzed the expression of lncRNAs in relation to patient survival time and status through univariate and multivariate COX regression analyses. The “survival” package in R was used (P<0.01). We used the function predicted to calculate the risk value for each patient. The calculation formula was: (coefficient of lncRNA1 × expression of lncRNA1) + (coefficient of lncRNA2 × expression of lncRNA2) + ... + (coefficient of lncRNA n × expression of lncRNA n).

2.4. Patients and samples collection

A total of 24 patients with THCA admitted to Chongqing University Cancer Hospital from January to June 2017 were randomly selected. The pathological diagnosis of each patient was assessed by three pathologists simultaneously according to the WHO classification criteria. Patients who received preoperative chemotherapy, radiotherapy, or any other adjuvant treatment were excluded from this study. The ethics has been approved by the Ethics Committee of Chongqing University Cancer Hospital. All participants signed informed consent forms.

2.5. RNA extraction and quantitative real-time PCR analysis

Tissues frozen in liquid nitrogen were collected, and RNA was extracted with on miRNeasy FFPE kit (Qiagen). MiRcute enhanced miRNA cDNA first strand synthesis Kit (TIANGEN) was used to reverse transcribe the RNA into complementary DNA. The qRT-PCR analysis was performed using the qPCR miRcute enhanced miRNA fluorescence quantification kit (TIANGEN). The results were normalised to glyceraldehyde-3-phosphate dehydrogenase (GAPDH) expression. We conducted three replicate experiments. The primers were synthesized by Thermofisher (Table S1).

2.6. Statistical analysis

All analyses were performed using R 3.4.3. The “survival” and “survminer” packages were used to plot survival curves for the high and low-risk groups. Heat maps were drawn using the “heatmap” package. The “ggpubr” package was used in correlation between lncRNAs and clinical features. Principal component analysis was performed using the “limma” and “scatterplot3d” packages. Univariate and multivariate Cox regression analysis were used to determine whether the constructed model can be used as an independent prognostic factor. SPSS 24.0 statistical software was used in analyzing the experimental data. The relative expression of lncRNA was calculated using the $2^{-\Delta\Delta Ct}$ method relative to GAPDH. The difference was considered significant when P<0.05.

Results

3.1. Acquired the immune-related lncRNAs in THCA
To make the research process clear, a flow chart was drawn to illustrate the research design and analysis methodology (Figure 1). We obtained THCA patients’ sequencing data and clinicopathological information from TCGA, composed of 58 paraneoplastic and 497 cancerous tissues (Table 1). Immune-related gene sets named “Immune Response” and “Immune System Processes” were downloaded from the GSEA website. 331 immune-related genes were obtained by Gene expression analysis. Then, 822 immune-related IncRNAs were obtained by conducting correlation analysis of immune-related genes with IncRNAs (|Cor| >0.4, P<0.001).

Table 1, The clinical characteristics of THCA patients in TCGA.

<table>
<thead>
<tr>
<th>Characteristics</th>
<th>Total N (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Age</td>
<td></td>
</tr>
<tr>
<td>&lt;60</td>
<td>387(77.9)</td>
</tr>
<tr>
<td>≥60</td>
<td>110(22.1)</td>
</tr>
<tr>
<td>Sex</td>
<td></td>
</tr>
<tr>
<td>Male</td>
<td>127(25.6)</td>
</tr>
<tr>
<td>Female</td>
<td>370(74.4)</td>
</tr>
<tr>
<td>TNM stage</td>
<td></td>
</tr>
<tr>
<td>I</td>
<td>285(57.3)</td>
</tr>
<tr>
<td>II</td>
<td>43(8.7)</td>
</tr>
<tr>
<td>III</td>
<td>112(22.5)</td>
</tr>
<tr>
<td>IV</td>
<td>57(11.5)</td>
</tr>
<tr>
<td>T stage</td>
<td></td>
</tr>
<tr>
<td>T1</td>
<td>100(20.1)</td>
</tr>
<tr>
<td>T2</td>
<td>167(33.6)</td>
</tr>
<tr>
<td>T3</td>
<td>170(34.2)</td>
</tr>
<tr>
<td>T4</td>
<td>58(11.8)</td>
</tr>
<tr>
<td>unknown</td>
<td>2(0.3)</td>
</tr>
</tbody>
</table>

3.2. Prognostic model construction

To obtain immune-related IncRNAs associated with prognosis, we analyzed the relationship between the expression of immune-related IncRNAs and survival status. Univariate Cox regression analysis revealed that 26 immune-related IncRNAs were significantly associated with patients’ survival (P<0.01, Table 2). Multivariate Cox regression analysis was performed on the above 26 immune-related IncRNAs for the screening of seven IncRNAs with high prognostic diagnostic values and establishment of a prognostic model for immune-related IncRNAs in THCA (Table 3). The formula for calculating each patient’s risk score was as follows:

Risk Score= (1.756 * LINC01614) + (2.649 * AC017074.1) + (2.323 * LINC01184) + (2.121 * LINC00667) + (3.816 * ACVR2B-AS1) + (-1.211 * AC090673.1) + (-1.486 * LINC00900).

Table 2, Univariate COX analysis of immune-related IncRNAs associated with THCA Prognosis.
Table 3, Multivariate Cox analysis of immune-related lncRNAs associated with THCA Prognosis.

<table>
<thead>
<tr>
<th>Gene symbol</th>
<th>Coefficients</th>
<th>P value</th>
</tr>
</thead>
<tbody>
<tr>
<td>LINC01614</td>
<td>1.756</td>
<td>0.000</td>
</tr>
<tr>
<td>LINC00900</td>
<td>-1.486</td>
<td>0.018</td>
</tr>
<tr>
<td>ACVR2B-AS1</td>
<td>3.816</td>
<td>0.000</td>
</tr>
<tr>
<td>AC017074.1</td>
<td>2.649</td>
<td>0.003</td>
</tr>
<tr>
<td>LINC01184</td>
<td>2.323</td>
<td>0.026</td>
</tr>
<tr>
<td>AC090673.1</td>
<td>-1.211</td>
<td>0.032</td>
</tr>
<tr>
<td>LINC00667</td>
<td>2.121</td>
<td>0.034</td>
</tr>
</tbody>
</table>

3.3. Predicted good and poor prognosis patients with THCA

According to their median risk scores, the patients were divided into high-risk group (n=248) and low-risk group (n=249). We ranked the risk score and marked the survival status of the patients in a scatter plot. Mortality was higher in the high-risk group, and mortality increased with risk-score (Figure 2A). The heat map revealed the relationship between the expression of lncRNAs in the model and the risk score in each patient (Figure 2B). As the expression risk score increased, five lncRNAs had elevated expression, which were defined as high-risk lncRNAs (i.e., LINC01614, AC017074.1, LINC01184, LINC00667 and ACVR2B-AS1), while AC090673.1 and LINC00900 had decreased expression, which were defined as low-risk lncRNAs. In addition, the results of the survival analysis showed a significant difference between the groups (P=1.41e-05, Figure 3).

3.4. Discrimination between high- and low-risk groups in different genomes
To verify the importance of lncRNAs as prognostic markers for THAC patients in our model, we conducted principal component analysis to determine whether lncRNAs can distinguish between high- and low-risk groups. Significant differences in the expression profiles of immune-related lncRNAs were found between the groups (Figure 4). The results indicated that the high- and low-risk groups can be discriminated by the lncRNAs in our model.

### 3.5. Evaluation of the prognostic models

We performed univariate and multivariate independent prognostic analyses to validate the accuracy and specificity of the prediction model and assess whether the model can be used as an independent prognostic factor. Univariate independent prognostic analysis revealed that the following factors were statistically significant, namely, age ($P<0.001$), stage ($P=0.003$), T stage ($P=0.030$) and the prognostic model ($P<0.001$) (Figure 5A). Multivariate independent prognostic analysis displayed that age ($P=0.002$) and the model ($P=0.025$) could predict the overall survival (OS) for patients with THCA (Figure 5B). This indicated that the model we constructed could predict the prognosis of THCA patients.

### 3.6. Analysis of the correlation between lncRNAs and clinical features

We performed correlation analysis to investigate the correlation between the expression of lncRNAs in the model and clinical characteristics. The results suggested that the levels of AC090673.1 ($P<0.05$), LINC01184 ($P<0.001$) and LINC01614 ($P<0.001$) were related to clinical stage. As the disease stage increased, the expression of AC090673.1 and LINC01614 fluctuated, and overexpression was observed in the late stage. By contrast, the expression of LINC01184 was relatively high and showed a trend of increasing before decreasing (Figure 6A). The levels of LINC00900 ($P<0.001$) and LINC01614 ($P<0.001$) were associated with the T stage. As T stage increased, the expression of LINC00900 increased slightly and then declined, whereas the expression of LINC01614 decreased and then increased (Figure 6B). These lncRNA prognostic markers accurately predicted the clinicopathological characteristics of patients.

### 3.7. Detected lncRNA expression levels in patients with THCA

To further evaluate the reliability of the constructed model, we examined the expression levels of seven lncRNAs in cancer and paracancer tissues of 24 patients with THCA (Table 4) through qRT-PCR (Table S2). The results showed that LINC01614, ACVR2B-AS1, AC017074.1, LINC01184, and LINC00667 were upregulated in THCA, whereas LINC00900 and AC090673.1 were downregulated (Figure 7). The experimental results were consistent with the bioinformatic analysis.

Table 4, The clinicopathological data of 24 THCA patients.
Characteristics | Total N (%)  
--- | ---  
Age  
<60 | 22(91.7)  
≥60 | 2(8.3)  
Sex  
Male | 12(50.0)  
Female | 12(50.0)  
TNM stage  
I | 11(45.8)  
II | 9(37.5)  
III | 3(12.5)  
IV | 1(4.2)  
Tumor (T)  
T1 | 17(70.9)  
T2 | 2(8.3)  
T3 | 2(8.3)  
T4 | 3(12.5)  
Nodes (N)  
N0 | 12(50.0)  
N1 | 9(37.5)  
N2 | 2(8.3)  
N3 | 1(4.2)  
Metastasis (M)  
M0 | 23(95.8)  
M1 | 1(4.2)  

**Discussion**

IncRNAs play critical roles in different stages of cancer immunity, including antigen release, antigen presentation, immune activation, immune cell migration, infiltration of cancer and killing of cancer cells. IncRNAs may be key regulators in reshaping the tumor immune microenvironment [11,12]. We analyzed the correlation between IncRNAs and immune-related genes in THCA and developed an immune-related IncRNA prediction model composed of seven IncRNAs (LINC01614, AC017074.1, LINC01184, LINC00667, ACVR2B-AS1, AC090673.1 and LINC00900). The high-risk IncRNA LINC01614 in this model has been studied in breast cancer (BC). Vishnubalaji et al. found increased expression of LINC01614 in primary BC tissues and BC cell lines and concluded that it is an unfavorable prognostic marker for BC [13]. Wang et al. confirmed this result [14]. Sun and Ling found that high expression of LINC01614 indicated low OS rate in patients with non-small cell lung cancer (NSCLC) [15]. In our model, another high-risk IncRNA INC00667 was found to be a tumor promoter in NSCLC. Involved in the proliferation, migration, and angiogenesis of NSCLC cells, it is significantly associated with the OS of patients with NSCLC [16]. INC00667 correlated with the prognosis of BC and ovarian and hepatocellular carcinomas [17-20]. Zhou et al. found that LINC01184 is overexpressed in tumor-infiltrating B lymphocytes and involved in the immunosuppression of cancer [21]. ACVR2B-AS1 was also one of the high-risk IncRNAs in our model. The high expression of ACVR2B-AS1 is an independent poor prognostic factor for OS and relapse-free survival in patients with liver cancer [22].

The prognostic role of IncRNAs in THCA has been investigated from different perspectives, such as methylation and competing endogenous RNA. Li et al. analyzed the methylation and transcriptome of THCA, identified 483 epigenetically regulated IncRNAs associated with the development of THCA, and
found that the methylation-driven 5-lncRNA-based signature is an independent prognostic factor [23]. Li et al. identified Lnc5m4 as an independent prognostic factor in a competing endogenous RNA network consisting of IncRNA, miRNA, and mRNA [24]. In this study, we constructed an immune-related prognostic model of THCA, which can be used as an independent prognostic factor.

In addition, we analyzed the correlation between the clinical characteristics of patients with THCA and the seven IncRNAs in our model. The results showed that AC090673.1, LINC01184, and LINC01614 were associated with clinical stage. LINC00900 and LINC01614 were associated with the T stage. LINC01614 is associated with clinical stage in patients with glioma and NSCLC [15,25]. LINC01184 is associated with the several clinicopathological profiles of rectal cancer, including clinical stage, tumor size, lymph nodes, and distal metastasis [26]. Regarding LINC00900, Wang et al. showed that the expression of LINC00900 increased with tumor grade in patients with primary glioblastoma [27]. LINC00900 is not differentially expressed in the neurogliomas of different clinical stages [28]. The relationship between these IncRNAs and the clinical characteristics of THCA patients needs to be further investigated.

The present study has some limitations. The model needs sample validation before it can be used in clinical settings. Recently, an immune-related IncRNA prognosis model based on the Gene Expression omnibus database was constructed [29]. The immune-associated gene sets in our study were derived from the GSEA database, and the prognostic model constructed was validated in multiple clinical samples, which made our conclusions more convincing. The prognostic model constructed in that study crossed over with our model but was not fully consistent. The possible reasons are the different sources of data and different sets of immune-related genes selected.

**Conclusion**

In summary, we constructed and experimentally validated a prognostic model for immune-related IncRNAs in THCA. The IncRNAs were significantly correlated with the clinical stage and T stage of THCA. Our results provide new targets for the prognosis and treatment of THCA.

**List Of Abbreviations**
<table>
<thead>
<tr>
<th>ABBREVIATIONS</th>
<th>Remarks</th>
</tr>
</thead>
<tbody>
<tr>
<td>IncRNAs</td>
<td>long non-coding RNAs</td>
</tr>
<tr>
<td>THCA</td>
<td>thyroid cancer</td>
</tr>
<tr>
<td>GSEA</td>
<td>Gene Set Enrichment Analysis</td>
</tr>
<tr>
<td>TCGA</td>
<td>The Cancer Genome Atlas</td>
</tr>
<tr>
<td>qRT-PCR</td>
<td>quantitative real-time PCR</td>
</tr>
<tr>
<td>BC</td>
<td>breast cancer</td>
</tr>
<tr>
<td>NSCLC</td>
<td>non-small cell lung cancer</td>
</tr>
</tbody>
</table>

**Declarations**

**CONSENT FOR PUBLICATION**

All participants signed the patient consent-to-disclose forms.

**AVAILABILITY OF DATAS AND MATERIALS**

The data supporting the findings of the article are available in the TCGA at https://cancergenome.nih.gov/ and the GSEA website at https://www.gsea-msigdb.org/gsea/msigdb/index.jsp.

**CONFLICT OF INTEREST**

This work was supported by the Chongqing Science and Health Joint Medical Research Project for Young High-end Medical Talents (2018GDRC007) and the Fundamental Research Funds for Central Universities (2019CDYGYB005). The authors declare no conflict of interest.

**ACKNOWLEDGEMENTS**

The authors are grateful to Dr. Ahmad Ud Din for language proofreading (Southwest Medical University).

**ETHICS APPROVAL**

The ethics has been approved by the Ethics Committee of Chongqing University Cancer Hospital (protocol code: 2019.120, review date: 19 July 2019).

**SUPPORTIVE/SUPPLEMENTARY MATERIAL**

Table S1. The primer sequences.

Table S2. CT values for qRT-PCR experiments.
References


**Figures**

![The RNA-Seq data of THAC from TCGA](#)

- **Immune-related gene set from GSEA**
- **Protein RNA**
- **IncRNA**
  - **Immune-related genes**
  - **Correlation analysis**
  - **Immune-related IncRNAs**
  - **The clinical data of THAC from TCGA**
  - **COX regression analysis**
  - **The prognostic model**
  - **qRT-PCR Validation in THAC tissues**
  - **Clinical relevance**
  - **Principal component analysis**

**Figure 1**

The process of building the prognostic model.
Figure 2

The prognostic model composed of 7 IncRNAs related to immunity in THCA.

A. Distribution of risk scores and survival status in the high-risk and low-risk groups of THCA patients.

B. The heat map of the expression profile of the 7 immune-related IncRNAs.
Figure 3

Overall survival analysis of THCA patients in high-risk group and low-risk group.
Figure 4

Different immune status between high-risk and low-risk groups. Principal component analysis between high-risk and low-risk groups in all genes (A), immune-related genes, (B), immune-related lncRNAs, (C), and immune-related lncRNAs in the prognosis model (D).
Figure 5

Immune-related IncRNA prognostic model as an independent prognostic factor. (A) Univariate Cox regression analysis to test THCA prediction tool. (B) Multiple Cox regression analysis to test THCA prediction tool.
Figure 6

Relationship between seven immune-related lncRNAs and clinical characteristics of THCA patients. (A) Clinical stage. (B) T stage.
Figure 7

gRT-PCR validation of immune-related lncRNAs expression in paracancerous and cancer tissues of THCA patients. The experiment was repeated three times independently.

Supplementary Files

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

- TableS1Theprimersequences.docx
- TableS2CTvaluesforRTqPCRvalidationexperiments.xlsx