A prognostic aging-related IncRNA risk model correlates with the immune microenvironment in HCC

Kun Mei a,1, Zilu Chen b,1, Qin Wang c, Akbar Ali d, Yan Huang e, *, Luo Yi f, g, *

a Department of Cardiothoracic Surgery, The Third Affiliated Hospital of Soochow University, Changzhou 213003, China
b Center for Molecular Imaging and Nuclear Medicine, Soochow University, School of Radiological & Interdisciplinary Sciences, Soochow University (RAD-X), Suzhou, China
c Nanjing University of Chinese Medicine, Nanjing 210023, China
d Nishtar Medial College, Multan, Pakistan
e Department of Ultrasound, Nanjing Hospital of Chinese Medicine Affiliated to Nanjing University of Chinese Medicine, Nanjing 210001, China
f Department of Oncology, Jiangsu Province Hospital on Integration of Chinese and Western Medicine, Nanjing 210028, China
g Department of Oncology, Affiliated Hospital of Integrated Traditional Chinese and Western Medicine, Nanjing University of Chinese Medicine, Nanjing, China

ABSTRACT

Background: Hepatocellular carcinoma (HCC) stands out as one of the most lethal cancers globally, given its complexity, recurrence following surgical resection, metastatic potential, and inherent heterogeneity. In recent years, researchers have systematically elucidated the significance of long non-coding RNA (lncRNA) in the initiation and progression of HCC. The introduction of The Cancer Genome Atlas (TCGA) and Gene Expression Omnibus (GEO) databases has significantly enhanced the prognostic assessment of HCC. However, the association between HCC and cell senescence has been infrequently explored in the literature. Method: We downloaded liver hepatocellular carcinoma (LIHC)-related messenger RNA and lncRNA expression levels from TCGA. Correlation analysis, Cox regression, and least absolute shrinkage and selection operator (LASSO) regression analysis were employed to validate the lncRNA risk model associated with cellular aging. Comparing the infiltration of diverse immune cells enabled the identification of distinct differences in the immunological microenvironments of the two risk groups. Subsequently, we conducted a real-time polymerase chain reaction (qPCR) experiment to confirm the accuracy of the selected lncRNAs. Results: A predictive framework for HCC was constructed based on the expression levels of five lncRNAs. Multivariate and univariate Cox regression analyses revealed that lncRNA signatures associated with...
Senescence were independently correlated with an increased risk of HCC. Additionally, the nomogram also provides a more refined and sensitive model. Further investigation into the variations in immune cells and functions between the high-risk and low-risk groups was conducted. Subsequently, a qPCR experiment results revealed underexpression of AC068756.1, AC090578.1, AC145343.1, and LINC0022 in Huh7 and LM3 cells. In contrast, AP003392.4 did not exhibit a significant difference between Huh7 and control cells. **Conclusion:** The prognostic features and nomogram, consisting of five aging-related lncRNAs (AC068756.1, AC090578.1, AC145343.1, AP003392.4, and LINC00221), may be useful in predicting the overall survival of HCC.

**KEYWORDS**

Cell aging; hepatocellular carcinoma; immune; lncRNA; TCGA
1. Introduction

Primary liver cancer, ranked as the sixth most common malignancy, stands as the third leading cause of death globally. The incidence of intrahepatic cholangiocarcinoma and hepatocellular carcinoma (HCC) has increased in recent times [1]. Currently, the clinical treatment methods include surgery, interventional therapy, local ablation [2], and targeted drugs [3]. Despite recent breakthroughs in treatment strategies, the prognosis of patients remains notably poor, with a high likelihood of recurrence and significant accompanying complications causing considerable distress to patients [4, 5]. Therefore, it is imperative to conduct accurate prognostic screenings to identify risk factors tailored to the specific conditions of patients with liver cancer. This effort aims to enhance the effectiveness of current clinical treatment and predict and ameliorate the prognosis for individuals affected by this disease.

Among the most fundamental biological processes, aging is a ubiquitous phenomenon observed in all organisms [6]. In contemporary understanding, cellular senescence has emerged as a pivotal risk factor in the advanced stages of numerous malignancies [7-9]. This is attributed to its role in increasing the susceptibility of cancer cells to external stress and their sensitivity to treatment. During the process of malignant transformation, tumor cells undergo a continuous acquisition of new functions [10], contrary to the aging process characterized by functional degeneration and loss of tissue function [11, 12]. Therefore, understanding the senescence process of HCC cells is important [13].

Transcripts exceeding 200 nucleotides in length are classified as long non-coding RNA (lncRNA). The occurrence and progression of various cancer types have been associated with lncRNA [14-16]. Abnormal sequences and expression levels of lncRNA are often linked to processes such as epithelial-mesenchymal transition [17], chemoradiotherapy sensitivity [18], and cell proliferation and migration [19]. Despite the recent surge in research on lncRNAs, the functions of numerous long intergenic non-coding RNAs (lincRNAs) remain unknown [20], especially their role in regulating the prognosis of liver cancer cells, which has been rarely researched.

2. Materials and Methods

2.1. Data acquisition

The Cancer Genome Atlas (TCGA) (http://portal.gdc.cancer.gov/) was used to download HCC expression patterns that matched clinical data. A total of 425 patients’ messenger RNA (mRNA) and lncRNA expression patterns were considered. To validate the outcomes of the TCGA data analysis, two datasets (Test1 and Test2) were randomly grouped using the "Caret" package in R (version 4.2.0).

2.2. To screen aging-related genes and construct a lncRNA co-expression network

A total of 279 aging genes were downloaded from the Molecular Signatures Database (MSigDB) v7.1 (https://www.gsea-msigdb.org/gsea/index.js). Subsequently, the Veen diagram package (version 1.7.1) was used to screen 272 genes associated with aging. The "limma" package in R (version 4.2.0) was utilized to examine the correlation of aging-related genes (corFilter=0.4, pvalueFilter=0.001) and exclude aging-related lncRNAs. The visualization of the genes was accomplished using the "igraph" package in R (version 4.2.0).

2.3. Differential gene analysis

The "limma" package was employed to investigate the differences in lncRNAs in the R (version 4.2.0). The filter criteria were set at $P < 0.05$ and $|\log2FC| > 0.6$. Ultimately, 18 differentially expressed aging-related lncRNAs
(ARGLnExp) were identified.

2.4. Construction and assessment of prognostic risk score model

Prognostic screening for ARGLnExp was performed through univariate Cox regression analysis. The least absolute shrinkage and selection operator (LASSO) algorithm was then executed to determine the value of the minimum cross-validation error and identify the optimal prognostic gene for the model. The final prognostic model was constructed using stable ARGLnExp. Kaplan–Meier and receiver operating characteristic (ROC) curves were employed to evaluate differences in prognosis among groups and to ascertain the survival rates at 1, 3, and 5 years. Finally, the correlation between patients in low- and high-risk groups and clinical information was calculated.

2.5. Establishment of nomogram prognosis prediction model

To plot the lipopograph model in conjunction with patient age, sex, grade, stage, and risk score, the "RMS" package in R (version 4.2.0) was utilized. Calibration curves were generated to illustrate the alignment between the actual survival probability at 1, 3, and 5 years and the predictions of the nomogram. Finally, the model was validated using the ROC curve, multivariate Cox regression, and univariate Cox regression.

2.6. Comparison of high-risk and low-risk groups for gene enrichment

A group of R-packets Different expressions of ARGLnExp were analyzed using Profiler’s gene body Gene Ontology (GO) enrichment and Kyoto Encyclopedia of Genes and Genomes (KEGG) pathway analyses to see which biological pathways might be involved. To investigate the connection between biology and genetics, a gene set enrichment analysis (GSEA) was performed on the C2 (C2. Cp. Kegg. 7.5.1. Entrez. GMT) MSigDB. P <0.05 was considered significant, and 1000 permutations of the sample size were executed. The correlation between high- and low-risk groups for immune cell infiltration was assessed. To quantify the presence of tumor-infiltrating immune cells in HCC tumor samples, the Cell-type Identification by Estimating Relative Subsets of RNA Transcripts (CIBERSORT) R package was utilized. Finally, we examined the morphological and functional differences of tumor immune cells using the R software packages "reshape2," "GSVA," and "GSEABase" (version 4.2.0).

2.7. Culture and therapy of cell lines

Huh7, LM3, and HL-702 HCC cells were cultured in a humidified atmosphere at 37 °C and 5% carbon dioxide using Dulbecco’s Modified Eagle Medium supplemented with 10% fetal bovine serum. The cell cultures utilized a complete set of Gibco, USA, cell culture supplies. To minimize the impact of cell line passage on experimental results, cell lines within the last ten generations were selected.

2.8. Separation of RNA and real-time reverse transcription-polymerase chain reaction (qRT-PCR)

The total RNA from the cells was isolated using an RNA Isolation Kit (Vazyme, Nanjing, China). Reverse transcription was performed using the HiScript II Q RT SuperMix for qPCR (Vazyme, Nanjing, China). The mRNA concentration was determined by adhering to the manufacturer's instructions for the ChamQ Universal SYBR qPCR Master Mix (Vazyme, Nanjing, China). Primer fragments are provided in Supplemental Material 1. The analysis of mRNA expression levels was performed using the $2^{\Delta\Delta Ct}$ method, with normalization to beta-actin.
2.9. Statistical tests

All statistical tests were performed in GraphPad Prism 8.0. (GraphPad Software, San Diego, CA, USA). To determine whether there was a statistically significant difference among the means of multiple groups, a one-way analysis of variance was conducted. The mean and standard deviation were displayed for all statistical data. Statistical significance was considered when $P < 0.05$.

3. Result

3.1. Screening and co-expression network construction of aging-related lncRNA

We screened 271 aging-related gene expressions from TCGA using the Veen diagram package. Through co-expression analysis, we identified 64 lncRNAs with significant associations and constructed an mRNA-lncRNA co-expression network (FIG. 1). Subsequently, 18 differentially expressed lncRNAs were identified through differential analysis.

![Figure 1. Aging-related co-expression regulatory network](image)

3.2. Prognostic risk signature construction of ARGLnExps

Based on univariate Cox regression analysis, seven lncRNAs were identified as prognostic factors (FIG. 2A). Finally, we screened five lncRNA analysis features using LASSO analysis (FIG. 2B, C). AC068756.1, AC090578.1, AC145343.1, AP003392.4, and LINC00221 were identified as risk factors for the prognosis model. The following equation was used for the risk score of each patient with HCC: risk score = \((0.19 \times \text{AC068756.1 expression}) + (0.22 \times \text{AC090578.1 expression}) + ...\)
\(\times AC090578.1 \text{ expression}) + (0.10 \times AC145343.1 \text{ expression}) + (0.75 \times AP003392.4 \text{ expression}) + (0.02 \times LINC00221 \text{ expression})\). Ultimately, the median risk score was used to classify patients with HCC from TCGA into high-risk and low-risk groups.

**Figure 2.** Construction of the prognostic model in TCGA-LIHC. (A) According to the results of univariate Cox regression analysis, a total of 7 LncRNA were identified as prognostic genes; (B) LASSO coefficient profiles of the prognostic genes; (C) Turning optimal parameter (lambda) screening in the LASSO model.

### 3.3. Assessment of the predictive performance of the ARGLnExps signature

Using the median risk score as a basis, both the training and validation sets were separated into low-risk and high-risk subgroups (FIG. 3A, B, C). Patients in the low-risk group exhibited a more favorable prognosis, surviving longer than those in the high-risk group, as illustrated by the Kaplan–Meier survival curve (FIG. 3D). Furthermore, we investigated the accuracy of the prognostic risk profile using external validation datasets Test1 and Test2, revealing consistent overall survival (OS) differences between low- and high-risk groups (FIG. 3E, F). The scatter plot in FIG. 3G, along with the validation set, corroborates the conclusion that high-risk patients with HCC experienced a higher mortality rate than low-risk patients with HCC (FIG. 3H, I). In a ROC analysis, our prognostic prediction demonstrated excellent sensitivity and specificity, as indicated by the area under the curves (AUCs) of 0.643 for 1-year survival rates, 0.580 for 3-year survival rates, and 0.570 for 5-year survival rates. A Cox regression study revealed that the ARGLnExps prognostic risk model is an independent predictor of HCC prognosis (FIG. 4A, B). Using the ROC analysis, we sought to assess the relationship between clinical features and prognostic value. A
comparison of the AUC for the model with those of age, sex, and stage—yielding values 0.532, 0.502, and 0.624, respectively—reveals that the model's sensitivity and specificity were inadequate. Further clinical correlation analysis (FIG. 4D-I) revealed differences in the stage between the high- and low-risk groups, particularly notable in Stage I and Stage III with the most significant discrepancies ($P = 0.0026$); however, no significant differences were observed in terms of sex, grade, or stage (except T1 and T3).

Figure 3. Evaluation of the prognostic signature in TCGA-LIHC. (A-I) Distribution of risk scores, Kaplan-Meier (KM) curves, and survival status of the aging-related risk lncRNAs; (J) ROC curves of 1-year, 3-year and 5-year survival rates
3.4. Construction and assessment of nomograms

Nomograms were constructed to predict the 1-, 3-, and 5-year survival rates for patients with colorectal cancer based on age, sex, stage, and the risk score of the five ARGLnExps. This approach was adopted owing to the poor sensitivity and specificity in the aforementioned model (FIG. 5A). The fifth-year survival prediction closely aligns with the actual value, as indicated by the correction curve (FIG. 5B). The AUC value in the ROC curve (FIG. 5C) was 0.74, signifying the accuracy of the model. Subsequently, single-variable and multiple-variable Cox regression analyses demonstrated that the nomogram model accurately predicted the prognosis for patients with HCC (FIG. 5D, E).
Figure 5. Nomogram prediction model and evaluation. (A) Nomogram of age, gender, stage, TNM, and risk score for predicting 1-, 3-, and 5-year survival. (B) 1-year, 3-year and 5-year calibration curves for the TCGA. (C) ROC curve of Nomogram; (D) Cox regression analysis of risk score and other clinical characteristics of Nomogram (age, gender, stage).

3.5. Gene enrichment analysis

Using GSEA enrichment analysis, the enrichment difference between the low- and high-risk groups was observed. Notably, the low-risk group exhibited enrichment in pathways such as cell cycle, complement and coagulation cascade, fatty acid metabolism, N-glycan biosynthesis, and oocyte meiosis. In contrast, the high-risk...
The group displayed enrichment in pathways including primary bile acid biosynthesis, proteasome, spliceosome, tryptophan metabolism, and valine leucine and isoleucine degradation (FIG. 6A). Numerous biological processes were found to involve ribonucleoprotein complexes, including mRNA processing, RNA splicing, ncRNA processing, ribosome biogenesis, ribosomal RNA processing, and cytoplasmic translation. The cells exhibit abundance in components related to various cellular structures and processes, including the U2-type spliceosome complex, the spliceosome catalytic step 2, the ribosomal subunit, the cytosolic ribosome, the nuclear envelope, the nuclear speck, the cell-substrate junction, the focal adhesion, the chromosomesal region, and the spliceosome complex. Several molecular activities were significantly enhanced, including transcription-enabling factor activity, DNA-binding transcription factor binding, RNA catalytic activity, cadherin binding, ubiquitin-like protein ligase binding, ubiquitin protein ligase binding, histone binding, transferase activity, transmitting one-carbon groups, ribosomal functional constituent, and transfer RNA catalytic activity (FIG. 6B). Additionally, KEGG enrichment analysis revealed significant enhancements in pathways such as spliceosome, cell cycle, ribosome, coronavirus disease-COVID-19, proteasome, nucleocytoplasmic transport, endocytosis, complement, and coagulation cascades (FIG. 6C).

**Figure 6.** Enrichment analysis. (A) GSEA enrichment analysis. (B) GO enrichment analysis; (C) KEGG enrichment analysis.

### 3.6. Relationship between tumor immune cell invasion and threat score

The CIBERSORT algorithm was used to determine the score of immune cells in the tumor group, and the
A difference in immune cell composition between the high- and low-risk groups was computed (FIG. 7A). As depicted in the figure, T cell CD4 memory activation is more evident in the high-risk groups (P 0.05), whereas natural killer (NK) cell activation and monocyte infiltration are dramatically increased in the low-risk group. In the analysis of immune cell function (FIG. 7B), it was observed that major histocompatibility complex class I was more significant in the high-risk group, whereas Type I and Type II interferon responses were more significant in the low-risk group.

Figure 7. Immunoinfiltration analysis. (A) Correlation between risk score model and tumor infiltrating immune cells; (B) Correlation between risk score model and tumor infiltrating immune cell function. * P < 0.05; ** P < 0.01; *** P < 0.001.
3.7. To determine whether aging-related lncRNA is expressed in HCC cells

To verify the validity of the TCGA dataset, we measured the expression levels of five senescence-related lncRNAs in cells using qRT-PCR. Among them, AC068756.1, AC090578.1, AC145343.1, and LINC00221 exhibited low expression in Huh7 and LM3 cells; however, AP003392.4 did not exhibit a significant difference between Huh7 and control cells (FIG. 8).

Figure 8. Validation of prognostic LncRNA.

4. Discussion

In recent years, given the rising incidence of HCC, the prevailing approach for diagnosing and prognostic evaluation primarily hinges on pathological evaluation and tumor-node-metastasis staging [21]. However, owing to the considerable heterogeneity of HCC, these assessment methods lack the desired sensitivity [22]. The sensitivity in diagnosing early HCC was reported to be only 47% [23]. Therefore, there is an urgent need for a more effective diagnostic and prognostic marker to improve clinical outcomes.

In this study, we developed a unique predictive model by combining five lncRNA signatures associated with aging. In this model, AC068756.1, AC090578.1, AC145343.1, AP003392.4, and LINC00221 were identified as prognostic risk factors. The developed model demonstrates a commendable ability to distinguish the OS rate between the high- and low-risk groups. Notably, the survival rate among low-risk groups is significantly better than that among high-risk groups. The ROC curve analysis further highlights that the 5- and 3-year survival rates of the patients are both lower than the 1-year survival rate. However, the clinical correlation analysis of patients revealed that the ROC of the model was 0.570, indicating suboptimal sensitivity and specificity. Therefore, we further analyzed the clinical correlation and observed no statistically significant correlation between age, sex, and stage.

Consequently, we constructed a nomogram, establishing a more accurate model. The ROC curve of the nomogram indicates reasonable specificity and sensitivity. This suggests that the model has great potential in predicting the prognosis of patients with HCC.

Indeed, DNA damage, telomere malfunction, oncogene activation, and organelle stress represent just a subset of external and internal variables capable of inducing cellular senescence. Cellular senescence is intricately linked to functions such as tumor suppression, tissue repair, embryogenesis, and organismal aging [24, 25]. Cellular senescence is an inherently unpredictable and difficult-to-manage dynamic process [26]. Aging has been implicated
in numerous studies on chronic diseases and common malignancies, such as colorectal cancer [27]. In recent years, there has been a growing body of research dedicated to understanding the processes associated with aging. HE [12] explored the cellular regulation of the body through single-cell omics. Meanwhile, Johnson [8] demonstrated that age-dependent variations in lipid metabolism are associated with aging. Despite these insights, our understanding of cancer cell aging remains incomplete, and the development of an accurate prediction model would significantly contribute to unraveling the relationship between aging and cancer. GO analysis reveals that senescence may exert control over cells through various cellular components, including the nuclear envelope, nuclear speck, cell-substrate junction, focal adhesion, chromosomal region, spliceosome complex, ribosomal subunit, cytosolic ribosome, U2-type spliceosome complex, and catalytic step 2 spliceosome. Moreover, senescence may influence cells through processes such as transcription coregulator activity and DNA binding. Additionally, in terms of the KEGG pathway, aging may predominantly regulate the aging process by influencing pathways such as the cell cycle and nucleocytoplasmic transport. Notably, there is growing evidence that cellular senescence plays a pivotal role in the immune surveillance of cancer [26, 28].

Our study demonstrated the relationship between cell aging and immune infiltration. Using LASSO regression, we constructed a model for immune cell infiltration scores, which proved capable of predicting survival times with a high degree of accuracy. The intricate interplay between lncRNAs and the molecular pathways involved in NK cell activation underscores their potential as therapeutic targets for enhancing antitumor and antiviral immune responses. Moreover, the influence of lncRNAs extends to monocyte infiltration, as these molecules actively participate in modulating chemotactic signals and adhesion molecules, thereby influencing the recruitment and behavior of monocytes within tissues. In a study conducted by Wu et al., the effect of AC145343.1 on the progression of liver cancer was investigated using a genomic instability-derived lncRNA signature. The findings revealed that the high-risk group exhibited increased immune cell infiltration, including B-cell memory, macrophage M0, and neutrophils. The risk score model derived from this investigation represents a novel prognostic tool designed to improve survival prediction following HCC diagnosis.

LncRNA can be used as a potential biomarker for HCC diagnosis. Beyond its diagnostic value, extensively studied in previous research [29, 30], lncRNA also plays a role in predicting the prognosis of HCC. In our study, the model constructed with five lncRNAs demonstrates effective prognostic prediction for patients. The tests conducted on Huh7 and LM3 cells demonstrated significant under-expression of AC068756.1, AC090578.1, AC145343.1, and LINC00221.1. However, AP003392.4 did not exhibit a statistically significant difference. We will continue to further investigate the specific role of these five lncRNAs in HCC.

Although this study provides important evidence on the use of the risk score model in the prognosis of HCC, it has some limitations. The study relied on retrospective data and might have missed some important information for each patient. More data need to be collected prospectively to further validate these outcomes. Nevertheless, understanding cell aging using the risk score provides important insight that will improve the diagnosis and prognosis of patients with HCC.

In conclusion, the risk characteristic associated with aging can predict the severity and immune cell infiltration of patients with HCC.

5. Conclusion

In conclusion, a prognosis model for aging based on five lncRNAs that have a high predictive value was developed. This study offers a novel approach to the customized therapy of patients with HCC.

Data Availability Statement
The corresponding author can provide the data that were utilized to support the study’s conclusions upon request.

**Funding statement**

This study was supported by the National Natural Science Foundation of China (grant nos. 81773947), Foundation for The Top Talent Program of Jiangsu Commission of Health's "Six-One Project" for High-Level Personnel (grant no. LGY2020003).

**Conflict of interest**

The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

**Author contributions**

Each author has made a substantial contribution to the research described, whether in the conception, study design, execution, data acquisition, analysis, interpretation, or a combination of these aspects. They have all participated in drafting, revising, or critically reviewing the article. Furthermore, they have given their final approval for the version to be published and have reached an agreement on the journal to which the article has been submitted. Additionally, the authors accept full responsibility for all aspects of the work and are willing to be held accountable.

**Appendix**

<table>
<thead>
<tr>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>ACD68756.1</td>
</tr>
<tr>
<td>Forward primer</td>
</tr>
<tr>
<td>Reverse primer</td>
</tr>
<tr>
<td>AC090678.1</td>
</tr>
<tr>
<td>Forward primer</td>
</tr>
<tr>
<td>Reverse primer</td>
</tr>
<tr>
<td>AC145343.1</td>
</tr>
<tr>
<td>Forward primer</td>
</tr>
<tr>
<td>Reverse primer</td>
</tr>
<tr>
<td>LINC00221</td>
</tr>
<tr>
<td>Forward primer</td>
</tr>
<tr>
<td>Reverse primer</td>
</tr>
<tr>
<td>Ap003392.4</td>
</tr>
<tr>
<td>Forward primer</td>
</tr>
<tr>
<td>Reverse primer</td>
</tr>
</tbody>
</table>

**References**


