Identification of Novel Pyroptosis-Related IncRNAs Associated with the Prognosis of Breast Cancer Through Interactive Analysis

Lili Gao*
Qing Li*

Department of Pathology, Pudong New Area People’s Hospital, Shanghai, People’s Republic of China

*These authors contributed equally to this work

Background: The role of pyroptosis and IncRNAs in breast cancer remains controversial. This study aimed to explore the pyroptosis-related IncRNAs in breast cancer.

Methods: All the data used for bioinformatics analysis were downloaded from The Cancer Genome Atlas database. Limma package was used to perform difference analysis, and distinguish mRNA and IncRNA. Survival package was used to conduct prognosis analysis. LASSO algorithm, univariate cox analysis and multivariate cox analysis were used to construct the prognosis model. P value <0.05 was regarded as statistically significant.

Results: Based on the seven pyroptosis-related IncRNAs tightly associated with patients’ prognosis, a prognostic prediction model was finally developed, which showed powerful effectiveness (Training cohort, one-year AUC = 0.82, 95% CI = 0.69–0.95, three-year AUC = 0.77, 95% CI = 0.68–0.85, five-year AUC = 0.74, 95% CI = 0.66–0.82; Validation cohort, one-year AUC = 0.68, 95% CI = 0.53–0.84, three-year AUC = 0.72, 95% CI = 0.64–0.81, five-year AUC = 0.67, 95% CI = 0.57–0.77). GSEA analysis demonstrated that the protein secretion, angiogenesis, TGF-β signaling and MTORC1 signaling might be involved in the high-risk patients. Moreover, immune infiltration analysis showed that the risk score was positively correlated with Tgd and Th2 cells, yet negatively correlated with CD8+ T cells, cytotoxic cells and T helper cells, which might partly explain the poor prognosis of high-risk patients. Finally, the expression level of seven model IncRNAs in the real world was validated by qRT-PCR using four cancer cell lines (MCF-7, T47D, MDA-MB-231, MDA-MB-469).

Conclusion: In conclusion, our study identified IncRNAs that are remarkably correlated with patients’ survival and might participate in the pyroptosis process, which might be underlying tumor biomarker and therapeutic targets. This study may provide direction for future research.

Keywords: breast cancer, pyroptosis, IncRNAs

Introduction
Breast cancer is the most frequent malignant tumor with about 2.08 million newly diagnosed cases in 2018 and has the most cancer-related deaths in females, accounting for 6.6% of all sites.¹ Nowadays, the primary treatment methods of BC include surgery, radiotherapy, and chemotherapy.² Despite advancements in early diagnosis and treatment modalities, breast cancer still suffers from high relapse rates and progress rates.³ Meanwhile, many factors could influence the prognosis of breast cancer patients, such as cancer subtypes (in particular triple-negative breast cancer),
no optimal prevention, terrible lifestyle, sub-health, and so on. Consequently, it is critical to characterize promising and novel biomarkers to predict prognosis and further targeted therapy.

As previous studies showed the expression changes of non-coding genes may impact the initiation and progression of cancers. Recent researches have confirmed that long-noncoding RNAs (lncRNAs), containing over 200 nucleotides, regulate early-stage breast cancer development through multiple mechanisms. For example, lncRNA MAFG-AS1 contributes to the tumorigenesis through the miR-574-5p/SOD2 axis in breast cancer, while lncRNA TMPO-AS1 acts as the oncogene in breast cancer by regulating the miR-1179/TRIM37 axis. Besides, Wang et al revealed that apoptosis-associated transcript of lncRNA AATBC could facilitate breast cancer invasion and metastasis through activating the YAP1/Hippo signaling route.

Damage-associated molecular pattern molecules (DAMPs) play a possible role in the pathogenesis of aging and cancer, but remain largely unknown. DAMPs, sometimes referred to as alarmins or danger signals, are endogenous molecules and endogenous stressors released by cells in response to exogenous and endogenous stresses, particularly after injury or cell death. They can act as inducers, sensors and mediators of stress and immune responses, either through individual plasma membrane receptors, cell membrane/intracellular recognition receptors, or after endocytic uptake. Pyroptosis, programmed cell necrosis activated by inflammasomes and caspases-1, is correlated with tumor cell evolution. A study showed that lncRNA HOTTIP holds up ovarian cancer cell pyroptosis by modulating the miR-148a-3p/AKT2 signaling pathway. In addition, another work showed that lncRNA ADAMTS9-AS2 induces pyroptotic cell death and strengthens gastric cancer epithelial pyroptosis by regulating the miR-18a-3p/GSDMD pathway in LPS-induced AKI. Therefore, pyroptosis-related lncRNAs might play a crucial role in tumor cell hyperplasia and transference in various cancers. However, the precise principle of pyroptosis-related lncRNAs in breast cancer still not been elucidated clearly.

Recently, the rapid growth of the next-generation sequencing unmasks the prelude of genome era and the massive data generated by this has brought great convenience in investigating disease essence. In this study, we identified pyroptosis-related lncRNAs by analyzing the public data downloaded from The Cancer Genome Atlas database. Then, an effective and stable prognosis model was established based on seven pyroptosis-related lncRNAs (AL513477.2, LINC01871, U73166.1, AL451085.2, AC005034.5, AC027307.2, AC121761.2), which might potentially predict patients survival and guide clinical therapy. The underlying clinical correlation and biological role of these lncRNAs were also explored. Finally, qRT-PCR was performed to detect the expression level of seven model lncRNAs in the real world using four breast cancer cell lines.

Methods and Materials
Publica Data Acquisition and Preprocessing
The transcriptional profiles and clinical data of breast cancer were downloaded from the TCGA database, a public database providing comprehensive cancer information. The transcriptional profiles were in FPKM form and preprocessed before analysis, including normalization, detection of low abundance probes, probes annotation and missing values completion. Clinical information was collated using Perl code. The mRNA and lncRNA were distinguished according to the biotype in the reference file “GRCh37.0.gtf”. The survival days lower than 30 days were deleted for their potential bias. The signature of pyroptosis-related genes was obtained from previous reviews.

Protein–Protein Interaction (PPI) and Enrichment Analysis
The STRING database was used to construct a PPI network to explore the interaction between these pyroptosis-related genes. Cytoscape v3.7.2 was used to visualize the PPI network and the hub nodes were identified with the cytohubba plug-in. Gene oncology (GO) and Kyoto Encyclopedia of Genes and Genomes (KEGG) enrichment analysis were performed using the clusterprofiler package in R software with the threshold of P < 0.05.

Identification of Pyroptosis-Related lncRNAs and Prognosis Model Construction
Based on the expression profile of pyroptosis-related genes, the limma package was used to identify pyroptosis-related lncRNAs with the following criteria: 1. For each pyroptosis-related gene, the lncRNAs meeting the
correlation filter >0.4 or <-0.4 were screened; 2. The statistical P value of correlation was <0.05. These pyroptosis-related lncRNAs were conducted univariate cox analysis to identify the lncRNAs associated with patients’ survival. Patients were randomly divided into training and validation cohorts (1:1). Subsequently, the least absolute shrinkage and selection operator (Lasso) algorithm was used to reduce dimensionality. Finally, multivariate cox analysis was performed based on the lncRNAs after dimensionality reduction to construct a prognostic prediction model. Each patient was assigned a risk score with the formula of “Riskscore = lncRNA1 * coef1 + lncRNA2 * coef2 + … + lncRNAX * coefx”. The patients with the risk score above the median value were defined as high-risk group.

Clinical Correlation Analysis
The clinical information of patients was combined with expression profile and risk score to explore the corresponding clinical correlation. The M classification and pathological grade were excluded for the reasons that most data were marked as unknown. Ultimately, age, clinical stage, T and N classification were included in the clinical correlation analysis.

Gene Set Enrichment Analysis (GSEA)
GSEA analysis was used to explore the biological pathway difference between high- and low-risk patients. In detail, the gene set was “Hallmark.v7.4.gmt”; the collapse was “No_Collapse”; the metric for ranking genes was “Signal2Noise”; the enrichment of statistic was “weighted”.

Immune Infiltration Analysis
Single sample gene enrichment analysis (ssGSEA) algorithm was used to quantify the immune microenvironment by calculating the abundance of 24 immune cells based on the specific gene mark, an extension of the GSEA algorithm.

Quantitative Real-Time PCR (qRT-PCR)
Total RNA was extracted using a RNA simple total RNA extraction kit (TIANGEN Beijing, China) and then reversely transcribed to cDNA. The SyBr Green PCR system was used to perform qRT-PCR. The primer used were as following: AC121761.2, forward primer: 5’- TAGATGCT TGAAAATGTAGCC-3’, reverse primer: 5’- ACTCTG TCCATCCCCAGTCTCA-3’; AC027307.2, forward primer: 5’- CAAAAGTTGGGACTTCCTTTCC-3’, reverse primer: 5’- AACTACCTCGCTCCCTTTCC-3’, LINC01871, forward primer: 5’- GCAACTACAGTCATTGTCTTC-3’, reverse primer: 5’- AACTGGAAGCTCATATATGAAG-3’; U73166.1, forward primer: 5’- GGATGGCCTCCA GTCCAGCT-3’, reverse primer: 5’- CTTTTAGACCCT TCCCTTGTA-3’; AL513477.2, forward primer: 5’- TATTTGAGGAAGGGTTG-3’, reverse primer: 5’- CAGCTTTGGAGTAGTTAAGA-3’; AC005034.5, forward primer: 5’- TCTTTAGACCCT TCCCTTGTA-3’, reverse primer: 5’- CTTTTAGACCCT TCCCTTGTA-3’; AL451085.2, forward primer: 5’- GGCCTCCAGAAGG GTAAG-3’, reverse primer: 5’- GATGTTCACG GGTCTC-3’.

Statistical Analysis
All the statistical analyses were conducted in the R software and GraphPad Prism 8. The two-sided P value <0.05 was regarded as statistically significant. Student T test was to compare the difference between two groups.20

Results
Acquisition of Pyroptosis-Related Genes and Their Biological Role
According to the prior reviews, we finally identified 33 pyroptosis-related genes shown in Table S1. Based on the PPI network established from the STRING database, we found that the gene PYCARD, IL1B, CASP1, IL18 and NLRC4 were the top five key nodes in these pyroptosis-related genes, which may act through the interaction with other proteins (Figure 1A). GO and KEGG analysis were then applied to evaluate the biological function of these pyroptosis-related genes. GO analysis showed that in molecular function (MF), these genes were primarily enriched in terms of “cytokine receptor binding”, “cysteine-type peptidase activity” and “cysteine-type endopeptidase activity” (Figure 1B); in biological process (BP), the top three terms were “positive regulation of cytokine production”, “regulation of inflammatory response” and “interleukin-1 production” (Figure 1C); in cellular component (CC), these genes were mainly enriched in “inflammasome complex”, “membrane region” and “membrane microdomain” (Figure 1D). KEGG results indicated that these genes were involved in the “NOD–like receptor signaling pathway”, “Salmonella infection”, “Lipid and atherosclerosis” and “Pathogenic Escherichia coli infection” (Figure 1E).
Identification of Pyroptosis-Related lncRNAs and Prognosis Model Construction

The pyroptosis-related lncRNAs were identified according to the criteria mentioned in the methods part and 356 lncRNA met our requirements, which were defined as pyroptosis-related lncRNAs for further analysis (Table S2). Univariate cox analysis, LASSO algorithm and multivariate cox analysis were performed on these pyroptosis-related lncRNAs, respectively (Figure 2A). Finally, seven lncRNA AC121761.2, AC027307.2, LINC01871, U73166.1, AL513477.2, AC005034.5 and AL451085.2 were selected to construct a prognostic prediction model for their tight association with patients overall survival (OS). The formula was “Riskscore = AC121761.2 * −0.39 + AC027307.2 * 0.11 + LINC01871 * −0.21 + U73166.1 * −0.34 + AL513477.2 * −0.31 + AC005034.5 * 0.17 + AL451085.2 * −0.31” (Figure 2B). In training cohort, the model showed high stability and efficacy (ROC curve: 1-year AUC = 0.82, 95% CI = 0.69–0.95; 3-year-AUC = 0.77, 95% CI = 0.65–0.85, 5-year-AUC = 0.74, 95% CI = 0.66–0.82), which was also validated in validation cohorts (ROC curve: 1-year AUC = 0.68, 95% CI = 0.53–0.84; 3-year-AUC = 0.72, 95% CI = 0.64–0.81, 5-year-AUC = 0.67, 95% CI = 0.57–0.77) (Figure 2C and D). Meanwhile, Kaplan-Meier survival curves demonstrated that the patients with high-risk score tend to had a worse prognosis (Figure 2C and D). Further regression analysis indicated that our model is an independent factor to predict the OS of breast cancer patients (Univariate analysis: Riskscore, HR = 1.81, 95% CI = 1.52–2.15, P < 0.001; Multivariate analysis: Riskscore, HR = 1.66, 95% CI = 1.39–1.97, P < 0.001) (Figure 3A and B).

Clinical and Prognosis Correlation of Model lncRNAs

Kaplan-Meier survival curves indicated that high expression level of lncRNA AL513477.2, LINC01871,
U73166.1, AL451085.2 and AC121761.2 were linked to a better prognosis of patients, but the lncRNA AC005034.5 and AC027307.2 were opposite (Figure 3C–I). The clinical correlation analysis showed that the risk score was remarkably correlated with the poor clinical stage and N classification, but not significant in age and T classification (Figure 4A–D). Moreover, we explored the clinical correlation in these seven model lncRNAs. The results indicated that the patients older than 60 years tend to have a high expression of lncRNA LINC01871, contrary to U7366.1 (Figure 4E); the high values of AL451085.2 might lead to a worse T classification (T3-4, Figure 4F); the high expression level of AL513477.2 was related to a poor N classification (N2-3, Figure 4G); the high value of AC027307.2 was correlated with a poor clinical stage (Stage IV/III, Figure 4H).

GSEA and Immune Infiltration Analysis
GSEA analysis was performed to explore the biological difference between high- and low-risk patients. The result showed that the top six pathways the risk score involved in were “protein secretion”, “androgen response”, “angiogenesis”, “early estrogen response”, “TGF-β signaling pathway” and “mTORC1 signaling pathway” (Figure 5). ssGSEA algorithm was used to calculate the abundance of the immune cell in breast cancer (Figure 6A). Multiple immune cells were found differentially expressed in high- and low-risk groups (Figure 6B). The results demonstrated that the risk
score was positively correlated with Tgd and Th2 cells, yet negatively correlated with CD8+ T cells, cytotoxic cells and T helper cells (Figure 6C–G).

The Expression Level of Model lncRNAs in Breast Cancer Cell Lines
We further explore the expression level of seven model lncRNAs in real world using cell lines (normal cell line MCF-10A and breast cancer cell lines MCF-7, T47D, MDA-MB-231 and MDA-MB-469). The result showed that the lncRNA AC005034.5 was downregulated in all breast cancer cell lines, especially in T47D and MDA-MB-231 (Figure 7A). The lncRNA AC027307.2 had a higher expression level in MDA-MB-231 and MDA-MB-469 cell lines than MCF-10A cell lines (Figure 7B). Meanwhile, MCF-7, T47D and MDA-MB-469 cell lines had a lower expression level of AC121761.2 than MCF-10A cell lines (Figure 7C). However, no consistent trends were observed on the expression level of AL451085.2, AL513477.2, LINC01871 and U73166.1 in breast cancer cell lines (Figure 7D–G).

Discussion
Despite the death rate of breast cancer among females dropped by nearly 40% from 1989 to 2016, the survival rate for patients with distant metastasis is only 27%.21 Furthermore, the high metastatic heterogeneity of breast cancer brings an extreme challenge to the treatment of patients.22 Therefore, it is increasingly urgent to find promising therapeutic targets aiming to improve patients’ prognosis effectively.

To the best of our knowledge, our study is the first systematic and comprehensive analysis of pyroptosis-related lncRNA in breast cancer and these results may assist in the diagnosis and treatment of breast cancer.
Pyroptosis, triggered by different pathological stimuli, is an inherent inflammation and crucial for the proliferation and migration of various cancer cells. In addition, numerous studies have focused on the molecules promoting pyroptosis, which includes non-coding RNAs and brings hope for more effective treatment of different cancers.

Besides, updates confirm that assorted lncRNA may fulfill distinct functional roles in different cancers. For instance, lncRNA HDAC2 is highly expressed in hepatocellular carcinoma and could serve as a biomarker for the diagnosis. By way of activating the Wnt/β-catenin/EMT signaling cascade, LncRNA ROR1-AS1 predicts a poor prognosis and exerts its oncogenic function in cervical cancer. A recent study demonstrated that MACC1-AS1 was highly expressed in pancreatic carcinoma tissues and could promote pancreatic carcinoma progression through the PAX8/NOTCH1 signaling pathway. In addition, the novel lncRNA OXCT1-AS1, considered as a new potential prognostic factor, significantly increased glioblastoma cell proliferation ability and inhibited cell invasion via regulating miR-195/CDC25A axis. Notably, an
accumulating amount of researches indicate a link between lncRNA and pyroptosis in cancer.\textsuperscript{28} In our work, analysis was performed to identify a signature of 7 pyroptosis-related lncRNAs, including AC121761.2, AC027307.2, LINC01871, U73166.1, AL513477.2, AC005034.5 and AL451085.2. Among these pyroptosis-related lncRNAs, LINC01871 has been previously discovered as one of the prognostic genes in breast cancer and AC027307.2 was considered an immune-related lncRNA to predict the survival of patients in colon adenocarcinoma.\textsuperscript{29} However, these 7 pyroptosis-related lncRNAs have not been thoroughly studied in tumors.

Our results showed that the pathway of protein secretion, androgen response, angiogenesis, TGF-β signaling and mTORC1 were enriched in the high-risk patients. This predicts that 7 lncRNAs in breast cancer may be associated with pathways such as protein secretion, androgen response,
angiogenesis, TGF-β signaling, and mTORC1. Inhibition of mTORC1 or STAT3 triggers pyroptosis. mTORC1 regulates the expression of pyroptosis genes via management of nuclear localization of STAT3. mTORC1/STAT3 axis may play a moderating role in pyroptosis of macrophages. As a hormone-sensitive disease, the response of breast cancer to androgen was associated with cancer progression. Feng and their colleagues found that a high response rate of androgen in breast cancer could lead to cancer development and metastasis. Angiogenesis has been proved to play a crucial role in breast cancer progression. De Heer and their colleagues demonstrated that hypoxia-inducible factors (HIFs) marking angiogenesis and metabolic rewiring could significantly accelerate cell aggressiveness. Meanwhile, Tang and their colleagues found that the protein deacetylase SIRT7 deficiency could promote SMAD4 degradation and therefore result in the activation of TGF-β signaling, further promoting breast cancer metastasis. mTORC1 is involved in protein secretion and...
In breast cancer, co-inhibition of mTORC1, HDAC and ESR1α could inhibit the proliferation of breast cancer. Another aspect, immune infiltration analysis demonstrated that the risk score was positively correlated with Tgd and Th2 cells, yet negatively related with CD8+ T cells, cytotoxic cells and T helper cells. CD8+ T-cells intensify the activity of dendritic cells and T-helper type 1 cells, and enhance a lower density of immunosuppressive cells in the neoplasm, which further impacts tumor cells. The inflammasome triggers pyrogenesis upon activation by various inflammatory stimuli, including lipopolysaccharide (LPS) and inappropriate pH. This may lead to programmed death of affected cells. In the present study, pyroptosis-related lncRNAs were positively correlated with Tgd and Th2 cells, which predicts the possibility that this process is mediated through pyrogenesis. However, our study still has limitations with not taking other lncRNAs other than those above into consideration and discussion fully. Moreover, in our qPCR results, it is observed that the expression level trend of AL451085.2, AL513477.2, LINC01871 and U73166.1 in breast cancer cell lines was not consistent, which means the effect of these lncRNA molecules on breast cancer patients is still equivocal and needs to be further explored.

**Conclusion**

In all, our study identified seven pyroptosis-related lncRNAs, AC121761.2, AC027307.2, LINC01871, U73166.1, AL513477.2, AC005034.5 and AL451085.2, which were tightly correlated with breast cancer patients’ survival. Based on these seven lncRNA, an effective prognosis model was constructed to predict patients prognosis. GSEA analysis and immune infiltration analysis were performed to explore the underlying mechanisms of poor prognosis in high-risk patients. qPCR was conducted to assess the expression level of seven lncRNA in the real world.

**Abbreviations**

lncRNAs, long-noncoding RNAs; PPI, Protein-protein interaction; GO, Gene oncology; KEGG, Kyoto Encyclopedia of Genes and Genomes; GSEA, Gene set enrichment analysis; qRT-PCR, Quantitative Real-time PCR.

**Data Sharing Statement**

The transcriptional profiles and clinical data of breast cancer were downloaded from the TCGA database and are available from the website (https://www.cancer.gov/about-nci/organization/ccg/research/structural-genomics).
The datasets used and/or analyzed during the current study are available from the corresponding author on reasonable request.

Author Contributions
All authors made a significant contribution to the work reported, whether that is in the conception, study design, execution, acquisition of data, analysis and interpretation, or in all these areas; took part in drafting, revising or critically reviewing the article; gave final approval of the version to be published; have agreed on the journal to which the article has been submitted; and agree to be accountable for all aspects of the work.

Funding
This study was supported by Shanghai Pudong New Area Science and Technology Development Fund (PKJ2020-Y40) and Important Weak Subject Construction Project of Pudong Health and Family Planning Commission of Shanghai (PWZbr 2017-22).

Disclosure
The authors declare that they have no competing interests.

References


