Insights into the Prognostic Value of Small Nucleolar RNA U81 and SNORA7B in Breast Cancer

Jun Zhou¹, Xuan Zhu², Jingpei Long¹

¹Department of Breast Surgery, The Women's Hospital, School of Medicine, Zhejiang University, Hangzhou, Zhejiang, People's Republic of China; ²Department of Radiotherapy, The First Affiliated Hospital of Zhejiang University, Hangzhou, Zhejiang, People's Republic of China

Correspondence: Jingpei Long, Department of Breast Surgery, Women's Hospital, College of Medicine, Zhejiang University, 1 Xueshi Road, Hangzhou, Zhejiang, 310006, People's Republic of China, Email: longjp@zju.edu.cn

Purpose: Small nucleolar RNAs (snoRNAs) has been closely associated with tumor development. However, its prognostic role and underlying molecular mechanisms in breast cancer (BC) remain unclear. This study aimed to identify prognostic snoRNAs by analyzing sequencing data from BC and para-cancer tissues, building a risk score model to predict breast cancer prognosis, and investigating the mechanisms by which they affect BC prognosis.

Patients and Methods: We analyzed the expression levels of snoRNAs between breast cancer and para-cancer tissues in the SNORic and Cancer Genome Atlas (TCGA) databases. We constructed a risk score model based on the differentially expressed snoRNAs. Receiver operating characteristic (ROC) curves and nomograms were plotted to assess the prognostic value of the risk score model. In this model, we divided breast cancer patients into high- and low-risk groups and studied each group’s infiltration of immune cells. Pearson’s correlation was used to identify differentially expressed mRNAs that were significantly correlated with prognostic snoRNAs.

Results: Small nucleolar RNAs, C/D box 81 (U81) and H/ACA box 7 B (SNORA7B), were identified as breast cancer prognosis-associated snoRNAs, and low-U81- or high-SNORA7B-expression groups had worse survival. A risk score model based on U81 and SNORA7B expression can predict BC survival. The infiltration of multiple immune cell types was significantly different between the high- and low-risk score groups. The high-risk score group may respond better to anti-immune checkpoint therapy. Many genes are related to U81 and SNORA7B, some of which are related to cell proliferation, metastasis, and changes in the extracellular matrix.

Conclusion: U81 and SNORA7B may serve as potential prognostic biomarkers of BC. Risk scoring models based on the expression of two snoRNAs can screen patients with breast cancer who can benefit from immunotherapy.

Keywords: TNBC, ncRNA, biomarker, prognosis

Introduction

Breast cancer (BC) has now surpassed lung cancer as the most common cancer worldwide, with approximately 2.3 million new cases in 2020.¹ Traditional markers for predicting breast cancer prognosis include tumor size, lymph node metastasis, etc. However, with the introduction of molecular typing, breast cancer is divided into different subtypes according to the expression of estrogen receptor (ER), progesterone receptor (PR), and human epidermal growth factor receptor-2 (Her-2).² Different types of breast cancer predict different prognoses, such as Her2-positive breast cancer’s poor prognosis.³ Existing biomarkers (ER, PR, and HER2) are beneficial for predicting breast cancer prognosis; however, new biomarkers for prognosis prediction are still needed for further breast cancer research.

Small nucleolar RNAs (snoRNAs) are a type of protein non-coding RNA (ncRNA), 60–300 nucleotides in length, playing an important role in many biological processes, such as rRNA processing,⁴ mRNA splicing and editing,⁴ stress response, and metabolic homeostasis.⁵ Increasing evidence has shown that some snoRNAs are aberrantly expressed in tumors and act as tumor-oncogenic factors that participate in the tumorigenesis.⁶,⁷ As reviewed by Jiang et al,⁸ snoRNA can promote the growth, metastasis, etc. of tumor cells, thereby affecting the survival of tumor patients.
The role of SnoRNAs in breast cancer is multifaceted, such as being used as diagnostic markers for breast cancer owing to their abnormal expression in BC tissue. It has also been used for classification, prediction of recurrence, and prediction of prognosis. Gee et al found that levels of snoRNA-RUN44 were associated with overall survival in breast cancer. Studies have also found that inhibition of SNORA7B can significantly reduce the invasion and migration of breast cancer cells. However, current data are still minimal, and the relationship between snoRNA and BC prognosis requires further studies.

Therefore, this study aimed to identify prognostic snoRNAs by analyzing sequencing data from BC and para-cancer tissues and building a risk score model to predict breast cancer prognosis. We also aimed to investigate the mechanisms by which snoRNAs affect BC prognosis.

Materials and Methods

Data Source

Expression data of 1524 snoRNAs in 1077 tumors and 104 para-cancer samples were downloaded from the SNORic database. The mRNA expression levels of 1072 tumors and 99 para-cancer samples were obtained from The Cancer Genome Atlas (TCGA) database. Immune cell-related gene sets were downloaded from a study published in 2013 by Bindea et al.

Identification of Prognostic snoRNA Signature

The Limma package was used to screen DEsnoRNAs between tumor and control samples with \(| \log_{2}(\text{fold change}) | > 1 \) and \(P\)-value < 0.05. To investigate the prognostic value of these DEsnoRNAs, we randomly assigned 874 BC cases with clinical information into training (N = 612) and validation (N = 262) sets at a ratio of 7:3. For each DEsnoRNA, patients with BC in the training set were divided into high- and low-snoRNA expression groups. Kaplan–Meier analysis (K-M) and Log rank test were performed to analyze and compare the overall survival of the high- and low-snoRNA expression groups. Statistical significance was set at \(P < 0.05\). DEsnoRNAs were further filtered according to the K-M results. Univariate Cox regression was used to screen DEsnoRNAs significantly associated with survival (p <0.05). Multivariate Cox regression was applied to obtain two robust prognostic snoRNA signatures (p <0.05).

Construction of the Risk Score Model and Nomogram

The risk score was calculated using the following formula:

\[
\text{ExpGene1} \times \text{Coef1} + \text{ExpGene2} \times \text{Coef2}
\]

where Coef is the regression coefficient of genes, and Exp is the normalized expression value of each prognostic snoRNA signature. Patients with BC were divided into high- and low-risk score groups according to the median value of the risk score. The overall survival of the high- and low-risk groups was analyzed using K-M and compared using the Log rank test. Statistical significance was set at \(P < 0.05\). The accuracy of the risk score model was evaluated by ROC curves using the “survival ROC” function in the R package. Univariate and multivariate Cox regression analyses were performed to determine BC patients’ independent prognostic factors (p <0.05). A nomogram was established to predict BC patients’ 1-, 2-, 3-, 4-, and 5-year survival. The performance of the nomogram was evaluated using calibration curves.

Immune Characterization of High- and Low-Risk Groups

In high- and low-risk groups, immune infiltration was assessed by single-sample gene set enrichment analysis (ssGSEA) using 24 immune-associated gene sets. This included aDCs, B cells, CD8+ T cells, cytotoxic cells, DCs, eosinophils, iDCs, macrophages, mast cells, neutrophils, NK CD56bright cells, NK CD56dim cells, NK cells, pDCs, T cells, T helper cells, Tcm, Tem, Tfh, Tgd, Th1 cells, Th2 cells, Th17 cells, and Tregs. International Prognostic Score (IPS) and Tumor Immune Dysfunction and Exclusion (TIDE) prediction scores were determined in the high- and low-risk groups to assess
the response to immune checkpoint inhibitors in BC patients. Higher IPS and lower TIDE scores indicate a better response to anti-immune checkpoint therapy. 18,19

Related Gene Expression of the snoRNAs
Pearson’s correlations between each prognostic snoRNA signature and DE mRNAs were calculated. Gene ontology (GO) and Kyoto Encyclopedia of Genes and Genomes (KEGG) pathway enrichment analyses were performed to investigate the biological functions of the top 50 mRNAs significantly correlated with each prognostic snoRNA signature.

**Figure 1** U81 and SNORA7B as a prognostic signature in breast cancer. (A) Differentially expressed snoRNAs between breast cancer and control samples shown in the volcano plot. (B) Differentially expressed snoRNAs between breast cancer and control samples shown in the heatmap. (C) Univariate cox regression analysis identified six snoRNA associated with prognosis of BC. (D) Multivariate cox regression analysis identified U81 and SNORA7B as a robust prognostic signature in BC. (E) Kaplan–Meier survival curves for U81 (E) and SNORA7B (F).
Statistical Analysis
All data were analyzed using the R software (version 4.0.0). The clinical characteristics and IPS were compared between the two groups using the Wilcoxon test. Comparison of TIDE scores between the high- and low-risk groups was performed using the Wilcoxon test and chi-square test. Statistical significance was set at $P < 0.05$ unless otherwise specified.

Ethics Approval
This study was approved by the Human Ethics Committee of Women’s Hospital of Zhejiang University (Hangzhou, China).

Results
Identification of U81 and SNORA7B as a Prognostic Signature in BC
A total of 129 DEsnoRNAs were identified between breast cancer and para-cancer samples, including 97 upregulated and 32 downregulated genes in the tumor samples (Figure 1A). The expression of DEsnoRNAs in each sample is displayed as a heatmap (Figure 1B). K-M showed that there were significant differences in survival between the high- and low-expression groups based on the expression of U81, SNORA7B, small nucleolar RNA, C/D box 49 B (U49B), small nucleolar RNA, C/D box 1A (snR38A), small nucleolar RNA, H/ACA box 7A (SNORA7A), and small nucleolar RNA, C/D box 114-14 (SNORD114-14; Table S1). We then investigated the prognostic value of U81, SNORA7B, U49B, snR38A, SNORA7A, and SNORD114-14 in BC. Univariate Cox regression analysis showed that U81, SNORA7A, and

![Diagram showing relationship between U81 and different clinical characteristics. The expression of U81 was significantly different between age ≥ 50 and age < 50, while there was no difference in other clinical features including gender, T, N, M and pathologic stage. *$P < 0.05$.](https://doi.org/10.2147/IJGM.S345945)
SNORA7B were significantly associated with prognosis \((P < 0.05, \text{Figure 1C})\). To obtain a more robust signature, U81, SNORA7A, and SNORA7B were further selected to conduct a multivariate Cox regression algorithm analysis, and U81 and SNORA7B were identified as key prognostic signatures in BC \((P < 0.05, \text{Figure 1D})\). We observed that SNORA7B acted as a risk factor \((HR > 1)\), whereas U81 played a protective role \((HR < 1)\) in BC \((\text{Figure 1D})\). In accordance with their roles in BC, we found that the high-U81-expression and low-SNORA7B-expression groups had better survival \((\text{Figure 1E and F})\).

Furthermore, we investigated the relationship between each prognostic signature (U81 and SNORA7B) and clinical characteristics, including sex (female and male), age \((\text{age} \geq 50 \text{ and age} < 50 \text{ years})\), M stage \((\text{M0 and M1})\), N stage \((\text{N0, N1, N2, and N3})\), T stage \((\text{T1, T2, T3, and T4})\), and pathologic stage \((\text{stages I, II, III, and IV})\). We found that the expression of U81 was significantly different between patients \(\geq 50\) and \(< 50\) years \((\text{Figure 2})\). The expression of SNORA7B was remarkably different between the groups divided by age, N stage, and T stage \((\text{Figure 3})\).

**Construction of the snoRNA-Related Risk Score Model and Nomogram in BC**

We calculated the risk scores for each patient according to the expression and coefficients of U81 and SNORA7B. Based on the median risk scores, the patients in the training set were divided into high- and low-risk groups \((\text{Figure 4A})\). The expression of SNORA7B was significantly higher, whereas the expression of U81 was much lower in the high-risk group than in the low-risk group, as shown in the heatmap \((\text{Figure 4B})\). The high-risk group had worse survival rates than the low-risk group \((P < 0.05, \text{Figure 4C})\). The ROC curves showed that the risk score model was accurate in predicting the 1-, 2-, 3-, 4-, and 5-year survival of BC patients with areas under the curve \((\text{AUC}) > 0.6\) \((\text{Figure 4D})\). Consensus results were
also obtained in the validation set (Figure 4E–H). Thereafter, we investigated the relationship between the risk score and clinical characteristics. We observed that the risk score was remarkably different between the groups divided by age, N stage, and T stage (Table 1).

Next, we performed univariate and multivariate analyses to determine whether risk score was an independent prognostic factor. Univariate analysis showed that age, M stage, N stage, T stage, pathological stage, and risk score were significantly correlated with prognosis (Figure 5A). Multivariate analysis based on the above factors indicated that the risk score remained remarkably associated with the prognosis (Figure 5B). Furthermore, we constructed a nomogram to predict the 1-, 2-, 3-, 4-, and 5-year survival of BC patients with a C-index of 0.76 using factors significantly associated with prognosis in the multivariate analysis (age, stage, pathologic stage, and risk score) (Figure 5C). The calibration curves showed that the predicted overall survival was very close to the actual observed overall survival (Figure 5D), indicating the clinical use of the nomogram.

Immune Characterization of High- and Low-Risk Groups

The ssGSEA results revealed that the infiltration of CD8+ T cells, cytotoxic cells, DCs, eosinophils, iDCs, mast cells, neutrophils, NK CD56bright cells, NK cells, pDCs, T cells, T helper cells, Tcm, Tem, Thf, Th2 cells, and Th17 cells was significantly different between the high- and low-risk groups (Figure 6A). Thus, we investigated the immune response of the high- and low-risk groups using IPS and TIDE scores. We observed that the high-risk group had higher IPS-PD-1, IPS-PD-L1, and IPS-CTLA-4 scores (Figure 6B), lower TIDE scores (Figure 6C), and more responders (Figure 6D), indicating that the high-risk group may have a better response to immune checkpoint therapy.

Figure 4 The prognostic value of the risk score signature in training and testing cohorts. (A–D) Risk curve, gene expression profile, KM survival and time-dependent ROC curves of OS based on U81 and SNORA7B in training cohort. (E–H) Risk curve, gene expression profile, K-M survival and time-dependent ROC curves of OS based on U81 and SNORA7B in the testing cohort.
Related Gene Expression of the snoRNAs
To explore the potential mechanisms of U81 and SNORA7B in regulating BC, we first identified DEmRNAs between tumor and control samples and then calculated the correlations between DEmRNAs and U81 or SNORA7B. A total of 1603 DEmRNAs were identified, including 605 upregulated and 998 downregulated genes in the tumor samples (Figure 7A). The top 50 genes correlated with U81 or SNORA7B were obtained (Figure 7B). Thereafter, the biological functions of these genes were analyzed using GO and KEGG. We found that the top 50 U81-correlated genes were mainly enriched in ion channel-related biological functions (Figure 7C), such as cellular response to cadmium and metal ions, and immune-related signaling pathways (Figure 7D), such as the TNF, mitogen-activated protein kinase (MAPK), and IL-17 signaling pathways. The top 50 SNORA7B-correlated genes were mainly enriched in muscle development-related biological processes (Figure 7E), such as skeletal muscle tissue development, response to transforming growth factor beta, muscle organ development, and extracellular matrix (ECM)-related pathways (Figure 7F), such as focal adhesion, extracellular matrix organization, and signaling by receptor tyrosine kinases.

Discussion
Finding more biomarkers for breast cancer and developing drugs is critical for achieving precise cancer treatment. The range of biomarkers used to predict prognosis is very wide, which can be proteins, mutated genes or non-coding RNAs. Some studies have found that snoRNA plays a particular role in breast cancer and can be used as a new target for breast cancer therapy. Given the value of snoRNAs as biomarkers and therapeutic targets, we aimed to investigate potential snoRNAs involved in BC.

This study identified two differentially expressed snoRNAs (U81, SNORA7B) between BC and para-cancer tissues using the publicly available SNORic and TCGA databases. Our results showed that two snoRNAs could be used as
prognostic signatures in BC. Specifically, SNORA7B acts as a risk factor, whereas U81 is protective in BC. Our results further expand the role of snoRNAs as novel molecular markers of breast cancer.

To date, relatively few studies have been conducted on SNORA7B. In a previous study, SNORA7B was found to act as an oncogene in BC, promoting the proliferation and metastasis of breast cancer cells. A recent study found that snoRNA 7B is involved in the growth regulation of lung cancer, and the knockout of SNORA7B inhibited the proliferation of non-small cell lung cancer cell lines. SNORA7B also promotes the self-renewal of human umbilical cord mesenchymal stem cells. Based on existing findings, SNORA7B seems involved in many biological processes. Our study also confirmed that high expression of SNORA7B indicates poor prognosis, and its specific molecular mechanism requires further investigation.

This study is the first to show that U81 is significantly associated with BC prognosis and plays a protective role in BC to the best of our knowledge. Few studies have found that U81 snoRNA has been shown to mediate nucleotide modifications in eukaryotic 28S rRNA, however, the specific mechanism has not been clarified. At present, there is no research on U81 in breast cancer, and we will further study its function.

Previous studies have used univariate analysis to find that SNORA7B is associated with BC prognosis. Unlike previous studies, we calculated a risk score based on the expression of two snoRNAs (SNORA7B and U81), which was an independent prognostic factor. The nomogram also confirmed that this risk index model could predict the survival rate of patients from 1 to 5 years. This finding complements the breast cancer prognostic prediction tools, which are currently commonly used.

Immunotherapy represents a tremendous advancement in human cancer treatment in recent decades. Due to the lack of specific molecular markers, triple-negative breast cancer, which is defined as negative for ER, PR, and Her-2, cannot
be effectively targeted. The most significant treatment progress for triple-negative breast cancer has been immunotherapy in the past decade.\textsuperscript{23,24} Although some clinical classifications, such as the Fudan\textsuperscript{25} and Lehmann classifications\textsuperscript{26–30} can screen out patients who can benefit from immunotherapy. Primary and acquired resistance to immunotherapy has been observed in several patients.\textsuperscript{27} More molecular markers are needed to screen these populations that benefit from immunotherapy.

Studies have found that snoRNAs are involved in tumorigenic immune processes and have special significance in anti-tumor immune responses.\textsuperscript{28} According to many research reports, immune infiltration is closely associated with cancer immunotherapy.\textsuperscript{29,30} Based on the above information, we hypothesized that U81 and SNORA7B are associated with tumor immunity and can guide therapeutic immunotherapy. According to our risk score model, the high-risk group showed high PD-1 and PD-L1 expression, indicating that this model may be used to screen breast cancer patients for whom immunotherapy is effective. However, more breast cancer samples are needed to validate our model so that it can be used to predict the effects of immunotherapy in the future.

snoRNAs play an essential role in regulating gene expression.\textsuperscript{31,32} This study found that the genes associated with U81 and SNORA7B are comprehensive, ranging from cell proliferation to cell adhesion. These findings indicate that the potential role of snoRNAs deserves further investigation.

**Conclusion**

In conclusion, we identified two snoRNAs (U81 and SNORA7B) using bioinformatics associated with breast cancer prognosis. Here, for the first time, we demonstrated that snoRNA-U81 functions as a protective gene and re-verified SNORA7B functions as an oncogene in BC. However, this study also has certain limitations, such as lacking larger-scale sample verification, and we will further verify the biological effects of these two snoRNAs in the following studies.
Figure 7. Construction of snoRNA-related gene-drug interaction network. (A) Differentially expressed mRNAs between control and BC samples. (B) The top 50 genes correlated with U81 or SNORA7B. (C and D) GO (C) and KEGG pathway (D) enrichment analyses of the top 50 U81-correlated genes. These genes were mainly enriched in ion channel-related biological functions and immune-related signaling pathways. (E and F) GO (E) and KEGG pathway (F) enrichment analyses of the top 50 SNORA7B-correlated genes. These genes were primarily enriched in muscle development-related biological processes and ECM-related pathways.
Acknowledgments
We would like to thank Editage (www.editage.cn) for English language editing.

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

References
22. Mattijssen S, Omnekin C, Puijin G. The role of RNase MRP and RNase P in human ribosomal RNA synthesis. From a small RNA to a small man. 2010:79.

