Long non-coding RNA-based signatures to improve prognostic prediction of breast cancer

Yi Zhang, MMa, Yuzhi Wang, MMb, Gang Tian, MDb, Tianhua Jiang, MDb

Abstract
Breast cancer (BC) is a disease of high mortality rate because of high malignant, while early diagnosis and personal management may make a better prognosis possible. This study aimed to establish and validate IncRNAs signatures to improve the prognostic prediction for BC.

RNA sequencing data along with the corresponding clinical information of patients with BC were gained from The Cancer Genome Atlas (TCGA). Prognostic differentially expressed IncRNAs were obtained using differentially expressed IncRNAs analysis (P value < .01 and fold change > 2) and univariate cox regression (P value < .05). By applying least absolute shrinkage and selection operation (LASSO) Cox regression analysis along with 10-fold cross-validation, 2 IncRNA-based signatures were constructed in the training, test and whole set.

A 14-IncRNAs signature and a 10-IncRNAs signature were built for overall survival (OS) and relapse-free survival (RFS) respectively in the 3 sets. BC patients were divided into high-risk groups and low-risk groups depended on median risk score value. Significant differences were found for OS and RFS between 2 groups in the 3 sets. The time-dependent receiver operating characteristic (ROC) curves analysis demonstrated that our IncRNAs signatures had better predictive capacities of survival and recurrence for BC patients as well as enhancing the predictive ability of the tumor node metastasis (TNM) stage system.

These results indicate that the 2 IncRNAs signatures with the potential to be biomarkers to predict the prognosis of BC for OS and RFS.

Abbreviations: AUCs = area under curves, BC = breast cancer, DElncRNAs = differentially expressed long non-coding RNAs, ER = estrogen receptor, GO = Gene Ontology, Her2 = Human epidermal growth factor receptor 2, KEGG = Kyoto Encyclopedia of Genes and Genomes, IncRNA = long non-coding RNA, LASSO = least absolute shrinkage and selection operation, OS = overall survival, PR = progesterone receptor, RFS = relapse free survival, ROC = receiver operating characteristic, TCGA = The Cancer Genome Atlas, TNM = tumor node metastasis.

Keywords: long non-coding RNA, breast cancer, overall survival, relapse free survival

1. Introduction
Breast cancer (BC) is one of the most common cancers and the leading cause of cancer death in females around the world, estimating that 24% of diagnosed cases and 15% of death cases among them.[11] There are various treatment strategies for BC patients, including surgery, chemotherapy, radiation therapy and hormone-blocking therapy,[12] but the prognosis of the patients remains dismal.[13] Therefore, it is particularly important to explore the potential molecular mechanisms in the development of BC and find new biomarkers and therapeutic targets to improve the prognosis of patients with BC. The outcome of BC varies depending on quite a few factors, such as age, stage and grade of cancer.[4] As is known to all, tumor node metastasis (TNM) stage system is a tool extensively used in predicting prognosis of patients and making treatment plans in clinical practice.[1] A single biomarker or a clinical characteristic often is lack of adequate accuracy for predicting outcome of patients, while integrating multiple biomarkers and clinical characteristics can significantly improve prognostic performance.[16,17] It is highly necessary that making a comprehensive approach of combining biomarkers and traditional clinicopathological factors to achieve more reliable prognostic assessment.

Long non-coding RNA (lncRNA), a type of RNA molecules with a length of more than 200 nucleotides, is a major class of ncRNA which can regulate gene expression at the levels of transcription, epigenetics and translation.[8–10] Increasing studies
have found that lncRNAs exist abnormal expression and related to prognosis in BC.[11–13] However, there are few studies on the prognostic value of uniting biomarkers and other clinical indicators.

In the present study, by using least absolute shrinkage and selection operation (LASSO) Cox regression analysis and 10-fold cross-validation, a 14-lncRNAs signature for overall survival (OS) and a 10-lncRNAs signature for relapse-free survival (RFS) were built to appraise predictive value of lncRNAs for survival and recurrence of BC. Besides, Correlation analysis and Cox regression analysis between lncRNAs signatures and clinicopathological characteristics were performed. Moreover, a comparison of predictive ability between lncRNA signatures and TNM system was evaluated. Through this research, we hope to develop reliable prognostic predictors from lncRNAs signatures for BC patients.

2. Material and methods

2.1. Data collection

A flowchart showing the major steps in the study process (Fig. S1, http://links.lww.com/MD/E902). The RNA sequencing datasets of BC patients and related clinical profiles were obtained from The Cancer Genome Atlas (TCGA) website (https://cancer genome.nih.gov/), which contained 1109 BC tissue samples and 113 adjacent non-cancerous tissue samples. No ethical approval is required since all raw data came from public databases for this study.

2.2. Construction of survival-predicting models

The raw data were converted into an expression matrix, and lncRNAs names from the dataset were annotated by Ensembl (http://asia.ensembl.org/index.html). We utilized “edgeR” package in R studio to screen the differentially expressed lncRNAs (DElncRNAs) with fold change > 2 between BC tissues and adjacent non-cancerous tissue samples. Besides, we carried out univariate Cox regression analysis for lncRNAs in BC tissues and the lncRNAs with P value <.05 were identified to be prognostic lncRNAs. Prognostic DElncRNAs were obtained from the intersection between and DElncRNAs and prognostic lncRNAs for the next analysis. Tumor samples (the whole set) were randomly divided into the training set and the test set at a 2:1 ratio. Then, survival-predicting models were built by combining LASSO Cox regression analysis and 10-fold cross-validation in the 3 sets.[14,15]

2.3. Functional enrichment analyses of lncRNAs in the signatures

Expression correlation analysis between the lncRNAs and genes was used to find the putative genes of lncRNAs in the signatures. The absolute value of the Pearson correlation coefficient >0.3 and P value <.001 were set as the cutoff values. Gene Ontology (GO) and Kyoto Encyclopedia of Genes and Genomes (KEGG) enrichment analysis were performed by the R package “clusterProfiler” classifiers.

2.4. Statistical analysis

The lncRNAs prognosis risk score value was calculated by using the formula: risk score value = EXP1 * β1lncRNA1 + EXP2 * β2lncRNA2 + EXPn * βnlncRNA1, where EXP presents the expression level of each lncRNA and β presents the regression coefficient from LASSO Cox selection method at 10-fold cross-validation. Among them, the median risk score value is the criterion for splitting BC patients into high-risk groups and low-risk groups. Pearson’s Chi-Squared test was applied to investigate the relationship between the models and clinicopathologic characteristics. Prognostic factors were selected through univariate and multivariate Cox regression analysis in the above parameters. To further evaluate the prognostic value and clinical application value of the risk scores in each model and stage, we performed Kaplan–Meier survival analysis and time-dependent receiver operating characteristic (ROC) curves.[16,17]

3. Results

3.1. Derivation of the lncRNAs signatures for OS and RFS from BC patients

Firstly, there were 1042 DElncRNAs in BC tissue samples compare with normal tissue samples based on “edgeR” package at the threshold of P value <.01 and [fold change] > 2 (Fig. 1A). Moreover, lncRNAs with log-rank P <.05 were screened from all lncRNAs by univariate Cox regression analysis. Totally, 93 prognostic DElncRNAs (OS, Fig. 1B) and 87 prognostic DElncRNAs (RFS, Fig. 1C) were retained for subsequent analysis. For OS, BC samples (n = 1069) were randomly divided into training set (n = 712) and test set (n = 357) at a 2:1 ratio. Likewise, BC samples (n = 488) were randomly divided into training set (n = 325) and test set (n = 163) at the same ratio for RFS. Survival-predicting models were built by combining LASSO Cox regression analysis and 10-fold cross-validation in the training sets (OS: Fig. 1D, 1E; RFS: Fig. 1F, 1G). As a result, a 14-lncRNAs signature for OS and a 10-lncRNAs signature for RFS were established. To confirm our results, the same signatures for OS and RFS were also constructed in the test set and the whole set. Detailed information about the lncRNAs of signatures as shown in Table 1S (http://links.lww.com/MD/E903). BC patients were ranked on the basis of risk score value and split into high-risk groups or low-risk groups using the median risk score value of the sets as the cut-off point. Kaplan–Meier survival plots revealed that the low-risk group had better OS and RFS in the training sets (Fig. 2A and B), the test sets (Fig. 2C and D) and the whole sets (Fig. 2E and F). GO and KEGG analyses showed that the lncRNAs might involve in immune cells activation, formation of membrane and vesicle, cytokine activity and immune-related pathways for OS (Fig. 3A and B). According to the same analysis, the lncRNAs were enriched in immune cells activation, membrane formation, antigen binding and immune-related pathways for RFS (Fig. 3C and D).

3.2. lncRNAs signatures associated with clinical and pathologic features

To assess the association between lncRNAs signatures and clinical and pathologic features, we performed Pearson’s Chi-Squared test for the 3 sets independently. As shown in Table 1, 2 groups existed significant differences in 4 characteristics (subtype, Her2, ER, and PR) in the 3 sets for OS. Besides, 2 groups existed significant differences on only 1 characteristic (ER) in 3 sets for RFS (Table 2). We also analyzed associations between risk scores and clinical characteristics. Patients with
Her2 positive, distant metastases, ER negative, PR negative tended more towards high-risk score (Fig. 4).

3.3. Prognostic value of IncRNAs signatures for BC

Cox regression analysis was applied to investigate the independent predictive power of IncRNAs signatures and other clinicopathological data. Univariate Cox regression analysis found that age, stage, pM, and the 14-IncRNAs signature were obviously related to OS in the 3 sets (Table 3). But age, stage and the 14-IncRNAs signature were remarkably related to OS after multivariate Cox regression analysis of BC in the 3 sets (Table 3). However, in both univariate and multivariate Cox regression analyses, only the 10-IncRNAs signature remained an independent predictor for RFS in the 3 sets (Table 4). In addition, time-dependent ROC curves demonstrated prediction power of the 14-IncRNAs signature for OS (Fig. 5A) in the 1 year, 3 years and 5 years achieved area under curves (AUCs) were 0.711, 0.674, and 0.691 separately. As for RFS, the AUCs based on the 10-IncRNAs signature in the 1 year, 3 years and 5 years were 0.741, 0.752, and 0.781 separately (Fig. 5B). The curves indicate that 2 signatures had a considerable predictive performance for BC survival and recurrence. TNM stage system is the most frequently used tool to distinguish poor prognosis and good prognosis for cancer patients in clinical practice. Therefore, we compared predicting prognostic ability of the IncRNAs signatures, TNM stage and their combination. The multi-index time-dependent ROC curves analysis revealed that the 14-IncRNAs signature and the 10-
lncRNAs signature had better predictive power than TNM stage for OS and RFS. What is more, combining the lncRNAs signatures and TNM stage may increase predictive accuracy for survival and recurrence (Fig. 5C and D). Kaplan–Meier curves and log-rank tests demonstrated that differences between groups were significant after separating BC patients into 4 groups according to lncRNAs risk scores and TNM stage (Fig. 5E and F).

Figure 2. Functional enrichment analyses for lncRNAs of the 2 signatures. (A and B) GO and KEGG analysis of lncRNAs in 14-lncRNA-based classifier. (C and D) GO and KEGG analysis of lncRNAs in 10-lncRNA-based classifier. lncRNA = long non-coding RNA, GO = Gene Ontology, KEGG = Kyoto Encyclopedia of Genes and Genomes.
Figure 3. Kaplan–Meier survival analysis between 2 groups in the 3 sets. (A, C, and E): Survival curves of BC patients from high-risk groups and low-risk groups in training, test and whole sets based on 14-lncRNA-based classifier risk score level for OS. (B, D, and F): Survival curves of BC patients from high-risk groups and low-risk groups in training, test and whole sets based on 10-lncRNA-based classifier risk score level for RFS. lncRNA = long non-coding RNA, OS = overall survival, RFS = relapse free survival.
4. Discussion

There is a large population with cancer in China, and the incidence and mortality have been increasing at a rapid rate.\textsuperscript{[18]} BC is the first cause of new cancer diagnoses at the same time the most commonly diagnosed cancer at young and middle age in females.\textsuperscript{[19]} Improvement in the clinical management of breast cancer has alleviated mortality rate in recent years, but conventional diagnostic criteria and treatment means are far from mainly satisfactory because of tumor heterogeneity in molecular level.\textsuperscript{[20,21]} Fortunately, the rapid development of molecular biology makes it possible to discover highly specific markers of diagnosis and treatment for patients with BC. Cumulative studies have indicated that lncRNAs are aberrantly expressed in numerous cancers and play significant roles in tumorigenesis, cancer recurrence and metastasis.\textsuperscript{[22,23]} For BC, emerging investigations have emphasized the vital function of lncRNAs in tumorigenesis and development.\textsuperscript{[24–26]} For example, UASR1 has a high expression in BC tissues and promotes cancer cell growth, proliferation, wound healing and migration through the AKT/mTOR signaling pathway.\textsuperscript{[27]} MALAT1 is considered an oncogenic regulator of BC, which accelerates angiogenesis through miR-145 in MCF-7 cells to enhance aggressiveness of BC.\textsuperscript{[28]} Its worth noting that some lncRNAs have been regarded as potential prognosis markers for BC, according to the previous researches.\textsuperscript{[29,30]} But they have limited prognostic value and insufficient credibility because of small sample sizes and

| Parameter                      | Low risk | High risk | Pearson $\chi^2$ | $P$  |
|--------------------------------|----------|-----------|------------------|------|
| Training set                   |          |           |                  |      |
| Age                            |          |           |                  |      |
| <60                            | 204      | 192       | 0.819            | .365 |
| >60                            | 152      | 164       |                  |      |
| Stage                          |          |           |                  |      |
| S1-2                           | 262      | 262       | 0.071            | .79  |
| S3-4                           | 83       | 87        |                  |      |
| Subtype                        |          |           |                  |      |
| Infiltrating ductal carcinoma  | 225      | 283       | 33.950           | <.001|
| Infiltrating lobular carcinoma | 96       | 36        |                  |      |
| others                         | 35       | 37        |                  |      |
| pT                             |          |           |                  |      |
| T0-2                           | 289      | 303       | 1.774            | .183 |
| T3-4                           | 65       | 52        |                  |      |
| pN                             | 182      | 162       | 2.286            | .131 |
| N1-3                           | 168      | 188       |                  |      |
| pM                             | 291      | 301       | 5.767            | .016 |
| M0                             | 3 13     |            |                  |      |
| M1                             | 186      | 173       | 8.938            | .003 |
| Her2                           | 47       | 103       | 26.025           | <.001|
| ER                             | 289      | 237       |                  |      |
| Negative                       | 72       | 144       | 33.502           | <.001|
| Positive                       | 261      | 195       |                  |      |
| Test set                       |          |           |                  |      |
| Age                            |          |           |                  |      |
| <60                            | 96       | 106       | 1.107            | .203 |
| >60                            | 65       | 90        |                  |      |
| Stage                          |          |           |                  |      |
| S1-2                           | 129      | 134       | 6.691            | .01  |
| S3-4                           | 30       | 60        |                  |      |
| Subtype                        |          |           |                  |      |
| Infiltrating ductal carcinoma  | 102      | 156       | 14.684           | .001 |
| Infiltrating lobular carcinoma | 44       | 23        |                  |      |
| others                         | 14       | 17        |                  |      |
| pT                             |          |           |                  |      |
| T0-2                           | 144      | 160       | 4.263            | .039 |
| T3-4                           | 17       | 36        |                  |      |
| pN                             | 75       | 83        | 0.611            | .434 |
| N1-3                           | 84       | 110       |                  |      |
| pM                             | 136      | 162       | 1.994            | .158 |
| M0                             | 5 1       |            |                  |      |
| Her2                           | 89       | 101       | 7.541            | .006 |
| ER                             | 14       | 40        |                  |      |
| Negative                       | 28       | 57        | 6.540            | .01  |
| Positive                       | 127      | 133       |                  |      |
| PR                             | 35       | 84        | 17.363           | <.001|
| Negative                       | 120      | 107       |                  |      |

| Parameter                      | Low risk | High risk | Pearson $\chi^2$ | $P$  |
|--------------------------------|----------|-----------|------------------|------|
| Stage                          |          |           |                  |      |
| S1-2                           | 389      | 398       | 3.159            | .075 |
| S3-4                           | 112      | 148       |                  |      |
| Subtype                        |          |           |                  |      |
| Infiltrating ductal carcinoma  | 326      | 440       | 47.225           | <.001|
| Infiltrating lobular carcinoma | 139      | 60        |                  |      |
| others                         | 48       | 55        |                  |      |
| pT                             |          |           |                  |      |
| T0-2                           | 430      | 466       | 0.003            | .953 |
| T3-4                           | 82       | 88        |                  |      |
| pN                             | 255      | 247       | 2.8              | .094 |
| N1-3                           | 251      | 299       |                  |      |
| pM                             | 424      | 466       | 7.481            | .006 |
| M0                             | 4 18     |            |                  |      |
| Her2                           | 273      | 276       | 15.477           | <.001|
| ER                             | 51       | 108       |                  |      |
| Negative                       |          |           |                  |      |
| PR                             | 107      | 228       | 49.964           | <.001|
| Positive                       | 379      | 304       |                  |      |

lncRNA = long non-coding RNA, OS = overall survival, RFS = relapse free survival.

Table 1 (continued).

(continued)
deficiency of verification from multiple data sets. Therefore, we explored and validated lncRNAs signatures to improve predictive power of prognosis in BC based on a large sample size from TCGA cohort. LASSO Cox regression model not only has advantages to solve the common multicollinearity problem but also holds some traits such as numerical stability and interpretability.\[31\] Besides, LASSO Cox regression model combined with 10-fold cross-validation can contribute to deal with the “curse-of-dimensionalit” in high-throughput biological data. Hence, we established a 14-lncRNAs signature for OS and a 10-lncRNAs signature for RFS by way of LASSO Cox regression model along with 10-fold cross-validation, which may have better predictive power for BC. BC patients could be split into high-risk groups and low-risk groups with remarkable differences through 2 signatures for OS and RFS in the training sets. What is more, 2 signatures could distinguish 2 groups in the test sets and the whole sets, further confirming the robustness and reliability of the 2 lncRNAs signatures in predicting BC prognosis. Interestingly, GO and KEGG analyses show that the lncRNAs in 2 signatures were both involved in some similar biological processes and pathways. Next, Pearsons Chi-Squared test was performed to evaluate correlations between risk scores and clinicopathological characteristics. The results had shown that risk scores of the 14-lncRNAs signature correlated with subtype, Her2, ER, and PR for OS, while risk scores of the 10-lncRNAs signature only correlated with ER. Afterward, multivariate Cox regression analyses revealed age, stage and the 14-lncRNAs signature were creditable independent predictors for OS and only the 10-lncRNAs signature was a creditable

| Table 2 Correlations of risk score of the 10-lncRNA-based classifier for RFS with clinicopathological characteristics in training set, test set, and whole set. |
|-----------------|-----------------|-----------------|-----------------|-----------------|
| Parameter       | Low risk        | High risk       | Pearson $\chi^2$ | $P$             |
| Training set    |                 |                 |                 |                 |
| Age             |                 |                 |                 |                 |
| $<60$           | 90              | 110             | 4.885           | .027            |
| $>60$           | 72              | 53              |                 |                 |
| Stage           |                 |                 |                 |                 |
| S1-2            | 114             | 124             | 2.531           | .112            |
| S3-4            | 46              | 33              |                 |                 |
| Subtype         |                 |                 |                 |                 |
| Infiltrating ductal carcinoma | 121 | 137             | 15.113          | .001            |
| Infiltrating lobular carcinoma | 27   | 6               |                 |                 |
| others          | 14              | 19              |                 |                 |
| pT              |                 |                 |                 |                 |
| T0-2            | 136             | 145             | 1.74            | .187            |
| T3-4            | 26              | 18              |                 |                 |
| pN              |                 |                 |                 |                 |
| N0              | 78              | 72              | 0.381           | .537            |
| N1-3            | 84              | 89              |                 |                 |
| pM              |                 |                 |                 |                 |
| M0              | 143             | 148             | 3.815           | .051            |
| M1              | 0               | 4               |                 |                 |
| Her2            |                 |                 |                 |                 |
| Negative        | 93              | 77              | 6.292           | .012            |
| Positive        | 11              | 24              |                 |                 |
| ER              |                 |                 |                 |                 |
| Negative        | 18              | 58              | 26.374          | <.001           |
| Positive        | 131             | 96              |                 |                 |
| PR              |                 |                 |                 |                 |
| Negative        | 34              | 66              | 13.788          | <.001           |
| Positive        | 114             | 87              |                 |                 |
| Test set        |                 |                 |                 |                 |
| Age             |                 |                 |                 |                 |
| $<60$           | 40              | 57              | 2.199           | .138            |
| $>60$           | 35              | 31              |                 |                 |
| Stage           |                 |                 |                 |                 |
| S1-2            | 61              | 64              | 1.625           | .202            |
| S3-4            | 12              | 21              |                 |                 |
| Subtype         |                 |                 |                 |                 |
| Infiltrating ductal carcinoma | 60 | 65             | 2.378           | .304            |
| Infiltrating lobular carcinoma | 9   | 9               |                 |                 |
| others          | 6               | 14              |                 |                 |
| pT              |                 |                 |                 |                 |
| T0-2            | 68              | 74              | 0.822           | .365            |
| T3-4            | 7               | 12              |                 |                 |
| pN              |                 |                 |                 |                 |
| N0              | 40              | 43              | 0.458           | .498            |
| N1-3            | 33              | 44              |                 |                 |
| pM              |                 |                 |                 |                 |
| M0              | 69              | 78              |                 |                 |
| M1              | 0               | 0               |                 |                 |
| Her2            |                 |                 |                 |                 |
| Negative        | 30              | 44              | 0.202           | .653            |
| Positive        | 7               | 13              |                 |                 |
| ER              |                 |                 |                 |                 |
| Negative        | 11              | 29              | 6.543           | .011            |
| Positive        | 58              | 56              |                 |                 |
| PR              |                 |                 |                 |                 |
| Negative        | 18              | 35              | 3.608           | .057            |
| Positive        | 50              | 50              |                 |                 |
| Whole set       |                 |                 |                 |                 |

(continued)
Figure 4. Correlation between risk score and clinical and pathologic features in the whole set. Boxplot of risk scores in BC patients with Her2 (A, OS), pM (B, OS), ER (C, OS; E, RFS) and PR (D, OS; F, RFS). "P < .05, ""P < .01, """"P < .001 and """"""P < .0001. OS = overall survival, RFS = relapse-free survival.

Table 3

| Parameters | Univariate COX | Multivariate COX |
|------------|---------------|-----------------|
|            | HR (95% CI)   | P               | HR (95% CI)   | P               |
| Training set |               |                 |               |                 |
| Age (<60 vs >60) | 1.488 (1.000, 2.213) | .049 | 2.331 (1.044, 4.515) | .012 |
| Subtype (Infiltrating Ductal Carcinoma vs Infiltrating lobular carcinoma vs others) | 0.946 (0.705, 1.270) | .714 | 0.992 (0.617, 1.595) | .974 |
| Stage (T0-2 vs T3-4) | 3.204 (2.115, 4.855) | <.001 | 2.864 (1.044, 7.858) | .041 |
| pN (N0 vs N1-3) | 2.318 (1.494, 3.598) | <.001 | 0.873 (0.374, 2.038) | .754 |
| pM (M0 vs M1) | 3.167 (1.662, 6.033) | <.001 | 1.158 (0.329, 4.076) | .819 |
| Her2 (Negative vs Positive) | 2.545 (1.407, 4.605) | .002 | 1.509 (0.758, 3.004) | .242 |
| ER (Negative vs Positive) | 0.592 (0.328, 0.916) | .019 | 0.797 (0.261, 2.430) | .690 |
| PR (Negative vs Positive) | 0.683 (0.451, 1.035) | .072 | 1.299 (0.474, 3.560) | .612 |
| 14-marker-based classifier (high risk vs low risk) | 14.932 (7.559, 29.500) | <.001 | 24.598 (7.282, 83.093) | <.001 |
| Test set |               |                 |               |                 |
| Age (<60 vs >60) | 3.387 (1.823, 6.292) | .001 | 7.751 (2.659, 22.599) | .001 |
| Subtype (Infiltrating Ductal Carcinoma vs Infiltrating lobular carcinoma vs others) | 1.401 (0.931, 2.106) | .105 | 1.401 (0.931, 2.106) | .105 |
| Stage (T0-2 vs T3-4) | 2.023 (1.083, 3.779) | .027 | 2.022 (1.082, 3.779) | .027 |
| pN (N0 vs N1-3) | 1.575 (0.774, 3.204) | .210 | 1.575 (0.774, 3.204) | .210 |
| pM (M0 vs M1) | 1.766 (0.926, 3.370) | .084 | 1.766 (0.925, 3.369) | .084 |
| Her2 (Negative vs Positive) | 2.549 (1.407, 4.605) | .002 | 1.509 (0.758, 3.004) | .242 |
| ER (Negative vs Positive) | 0.692 (0.328, 0.916) | .019 | 0.797 (0.261, 2.430) | .690 |
| PR (Negative vs Positive) | 0.683 (0.451, 1.035) | .072 | 1.299 (0.474, 3.560) | .612 |
| 14-marker-based classifier (high risk vs low risk) | 14.932 (7.559, 29.500) | <.001 | 24.598 (7.282, 83.093) | <.001 |
| Whole set |               |                 |               |                 |
| Age (<60 vs >60) | 1.917 (1.377, 2.669) | <.001 | 2.726 (1.649, 4.506) | <.001 |
| Subtype (Infiltrating Ductal Carcinoma vs Infiltrating lobular carcinoma vs others) | 1.045 (0.824, 1.325) | .713 | 1.115 (0.780, 1.503) | .549 |
| Stage (T0-2 vs T3-4) | 1.973 (1.934, 3.856) | <.001 | 3.365 (1.501, 7.544) | .003 |
| pN (N0 vs N1-3) | 1.073 (1.167, 2.484) | .005 | 1.647 (0.797, 3.301) | .177 |
| pM (M0 vs M1) | 2.161 (1.503, 3.107) | <.001 | 0.775 (0.395, 1.519) | .458 |
| Her2 (Negative vs Positive) | 1.677 (1.016, 2.768) | .043 | 0.901 (0.499, 1.625) | .729 |
| ER (Negative vs Positive) | 0.726 (0.501, 1.054) | .032 | 0.699 (0.298, 1.638) | .410 |
| PR (Negative vs Positive) | 0.754 (0.533, 1.066) | .111 | 0.971 (0.436, 2.162) | .943 |
| 14-marker-based classifier (high risk vs low risk) | 8.037 (4.672, 13.824) | <.001 | 7.494 (3.310, 16.906) | <.001 |

lncRNA = long non-coding RNA, OS = overall survival, RFS = relapse-free survival.
Table 4
Cox regression analysis of the 10-lncRNA-based classifier with RFS in training set, test set and whole set.

| Parameters | Univariate COX | Multivariate COX |
|------------|----------------|------------------|
|            | HR (95% CI) | P | HR (95% CI) | P |
| Training set |            |    |            |    |
| Age (<60 vs ≥60) | 1.417 (0.709, 2.831) | .323 | 2.331 (1.204, 4.515) | .012 |
| Subtype (Infiltrating Ductal Carcinoma vs infiltrating lobular carcinoma vs others) | 1.119 (0.715, 1.749) | .621 | 1.528 (0.609, 3.840) | .366 |
| Stage (S1-2 vs S3-4) | 2.091 (0.996, 4.387) | .051 | 11.685 (1.468, 52.993) | .020 |
| pT (T0-2 vs T3-4) | 0.946 (0.389, 2.300) | .903 | 2.798 (0.824, 7.586) | .898 |
| pH (N0 vs N1-3) | 2.073 (0.062, 4.467) | .062 | 0.511 (0.087, 3.001) | .457 |
| PM (M0 vs M1) | 2.311 (0.768, 6.945) | .135 | 1.158 (0.329, 4.076) | .819 |
| Her2 (Negative vs Positive) | 1.216 (0.252, 5.868) | .805 | 1.509 (0.756, 3.004) | .242 |
| ER (Negative vs Positive) | 0.587 (0.281, 1.231) | .159 | 1.292 (0.189, 8.807) | .793 |
| PR (Negative vs Positive) | 0.541 (0.268, 1.091) | .085 | 1.231 (0.186, 8.118) | .828 |
| 10-marker-based classifier (high risk vs low risk) | 3.187 (1.751, 5.799) | <.001 | 6.894 (1.672, 28.424) | .007 |
| Test set |            |    |            |    |
| Age (<60 vs ≥60) | 2.619 (0.358, 19.120) | .343 | 3.371 (0.142, 80.951) | .452 |
| Subtype (Infiltrating Ductal Carcinoma vs infiltrating lobular carcinoma vs others) | 2.082 (0.705, 6.144) | .105 | 1.543 (0.204, 11.693) | .671 |
| Stage (S1-2 vs S3-4) | 6.206 (0.743, 90.563) | .085 | 3.842 (1.082, 14.352) | .256 |
| pT (T0-2 vs T3-4) | 1.698 (0.171, 16.784) | .650 | 0.921 (0.062, 14.606) | .956 |
| pH (N0 vs N1-3) | 3.187 (0.280, 36.179) | .349 | 1.445 (0.468, 4.579) | .532 |
| PM (M0 vs M1) | 11.358 (7.259, 47.528) | .854 | 15.627 (9.457, 54.837) | .821 |
| Her2 (Negative vs Positive) | 0.695 (0.521, 5.624) | .957 | 1.302 (0.678, 2.534) | .446 |
| ER (Negative vs Positive) | 0.595 (0.079, 4.455) | .613 | 1.586 (0.556, 4.591) | .408 |
| PR (Negative vs Positive) | 0.885 (0.118, 6.709) | .912 | 1.618 (0.278, 9.656) | .607 |
| 10-marker-based classifier (high risk vs low risk) | 1.972 (0.296, 1.117) | .042 | 5.56 (3.124, 10.464) | .047 |
| Whole set |            |    |            |    |
| Age (<60 vs ≥60) | 1.465 (0.770, 2.789) | .244 | 1.406 (0.346, 6.471) | .589 |
| Subtype (Infiltrating Ductal Carcinoma vs infiltrating lobular carcinoma vs others) | 1.214 (0.808, 1.825) | .348 | 1.504 (0.649, 3.486) | .341 |
| Stage (S1-2 vs S3-4) | 2.552 (1.265, 5.145) | .008 | 10.278 (1.646, 64.180) | .012 |
| pT (T0-2 vs T3-4) | 0.992 (0.431, 2.270) | .985 | 1.266 (0.512, 3.168) | .628 |
| pH (N0 vs N1-3) | 2.597 (1.248, 5.405) | .010 | 0.643 (0.112, 3.699) | .621 |
| PM (M0 vs M1) | 3.178 (1.074, 9.406) | .036 | 2.421 (1.108, 1.976) | .832 |
| Her2 (Negative vs Positive) | 1.039 (0.221, 4.898) | .961 | 0.697 (0.130, 3.725) | .673 |
| ER (Negative vs Positive) | 0.707 (0.355, 1.406) | .324 | 0.913 (0.123, 6.870) | .929 |
| PR (Negative vs Positive) | 0.671 (0.347, 1.297) | .236 | 1.246 (0.167, 9.294) | .830 |
| 10-marker-based classifier (high risk vs low risk) | 2.619 (1.575, 5.045) | <.001 | 5.535 (1.986, 17.277) | .007 |

lncRNA = long non-coding RNA; OS = overall survival; RFS = relapse free survival.

In the present study, the 14-lncRNAs signature and the 10-lncRNAs signature has been proven to be significantly connected with the OS and RFS of BC. The functions of a majority of the lncRNAs in the signatures have not been totally expounded up to now. But there are some reports about several lncRNAs in our signatures from the former researches. FEZF1-AS1 has been identified as a nuclear-restricted lncRNA. The expression levels of FEZF1-AS1 are downregulated in human prostate cancer tissues, and it paly tumor-promotive roles in prostate cancer via Notch signaling pathway.[122] LINC00536 is a newly identified lncRNA and named by Nakajima that located at chromosome 8q23.3 in 2014. It could induce malignant phenotypes to promote BC cell proliferation, migration, and invasion.[133] Gong et al found that LINC01224 regulated the expression levels of CHEK1 by competitively binding to miR-330-5p, thus inhibiting hepatocellular carcinoma progression.[134] LINC00668 is overexpressed in BC tissues and contributes to the progression of BC by accelerating cell cycle progression and inhibiting cell apoptosis.[135] In view of the outstanding performance of prognostic prediction, the roles of these lncRNAs in cancers should be investigated in the future, especially for BC.

However, some shortcomings in this study should not be ignored. On the 1 hand, this study was a retrospective nature because all the data came from the public database which lacked further test in a prospective clinical trial. Besides, TCGA cohort could not provide some helpful clinicopathologic characteristics, such as treatments. On the other hand, the identified functions and potential mechanisms of the lncRNAs in our signatures are still veiled. Therefore, further studies are required to complete the clinical applications and explore the roles of lncRNAs, as this may help us to understand the significance of them in BC occurrence and development. Although these deficiencies are inevitable, our results remain to provide reliable lncRNAs signatures to predict BC survival and recurrence.
In summary, we assessed the genome-wide lncRNA expression profiles from TCGA database and constructed a 14-lncRNAs signature and a 10-lncRNAs signature by the LASSO Cox regression analysis and 10-fold cross-validation. The 2 lncRNAs signatures with the potential to be biomarkers to predict the prognosis of BC for OS and RFS. In addition, our study could also complement clinical and clinicopathologic characteristics analysis in an attempt to facilitate the personalized management of patients with BC.

Acknowledgments

We thank all the researchers involved in the consolidation and submission of the data from the TCGA database, which may provide convenience and possibility of tumors studies in a large cohort.

Author contributions

All authors read and approved the final manuscript.

Conceptualization: Yi Zhang, Tianhua Jiang.
Data curation: Yuzhi Wang.
Formal analysis: Yuzhi Wang, Gang Tian, Yi Zhang.
Funding acquisition: Gang Tian.
Investigation: Yi Zhang, Yuzhi Wang, Gang Tian.
Methodology: Yi Zhang, Gang Tian, Tianhua Jiang.
Project administration: Yi Zhang, Tianhua Jiang.
Resources: Yi Zhang, Tianhua Jiang.
Software: Yuzhi Wang, Gang Tian.
Supervision: Gang Tian, Yi Zhang.
Validation: Yuzhi Wang, Tianhua Jiang.
Visualization: Yuzhi Wang, Gang Tian.
Writing – original draft: Yi Zhang.
Writing – review & editing: Gang Tian, Tianhua Jiang.

References

[1] Bray F, Ferlay J, Soerjomataram I, et al. Global cancer statistics 2018: GLOBOCAN estimates of incidence and mortality worldwide for 36 cancers in 185 countries. CA 2018;68:394–424.
[2] Radice D, Redaelli A. Breast cancer management: quality-of-life and cost considerations. Pharmacoeconomics 2003;21:383–96.
[3] Pelosi A, Careccia S, Saggese G, et al. Dual promoter usage as regulatory mechanism of let-7c expression in leukemic and solid tumors. Mol Cancer Res 2014;12:878–89.
