

# SCIENTIFIC REPORTS



OPEN

## A Potential Prognostic Long Noncoding RNA Signature to Predict Recurrence among ER-positive Breast Cancer Patients Treated with Tamoxifen

Kang Wang<sup>1</sup>, Jie Li<sup>1</sup>, Yong-Fu Xiong<sup>2</sup>, Zhen Zeng<sup>3</sup>, Xiang Zhang<sup>1</sup> & Hong-Yuan Li<sup>1</sup>

Limited predictable long noncoding RNA (lncRNA) signature was reported in tamoxifen resistance among estrogen receptor (ER)-positive breast cancer (BC) patients. The aim of this study was to identify and assess prognostic lncRNA signature to predict recurrence among ER-positive BC patients treated with tamoxifen. Cohorts from Gene Expression Omnibus (GEO) (n = 298) and The Cancer Genome Atlas (TCGA) (n = 160) were defined as training and validation cohort, respectively. BC relapse associated lncRNAs was identified within training cohort, and the predictable value of recurrence was assessed in both cohorts. A total of 11 lncRNAs were recognized to be associated with relapse free survival (RFS) of ER-positive BC patients receiving tamoxifen, who were divided into low-risk and high-risk group on basis of relapse risk scores (RRS). Multivariate cox regression analyses revealed that the RRS is an independent prognostic biomarker in the prediction of ER-positive BC patients' survival. GSEA indicated that high-risk group was associated with several signaling pathways in processing of BC recurrence and metastasis such as PI3K-Akt and Wnt signaling. Our 11-lncRNA based classifier is a reliable prognostic and predictive tool for disease relapse in BC patients receiving tamoxifen.

As a formidable health problem for women around the world<sup>1</sup>, breast cancer (BC) is the most frequently diagnosed cancer and the arch-criminal of cancer death among women. In the United States, estimated 252,710 new cases will occur, and 40,610 of them will die in 2017, accounting for 30% of new cases and 14% of deaths of all sites cancers<sup>2</sup>. BC can be basically divided into four different subtypes including luminal A and luminal B accounting for approximately 70% of all BCs and marked with estrogen receptor (ER)-positive<sup>3,4</sup>, human epidermal growth factor receptor 2 (HER-2) overexpression and triple negative with poorer outcomes<sup>5,6</sup>. Disparity of therapy strategies according to pathological characteristics has been well established.

Tamoxifen is a selective ER $\alpha$  antagonist, the drug of first-line endocrine treatment, which is still widely prescribed in pre-menopausal women Luminal BC patients. Although tamoxifen can reduce the risk of BC recurrence by 41% and mortality by 34% after 5 years' treatment<sup>7</sup>, almost 30% of ongoing treated with tamoxifen may develop resistance to the drug, which leads to cancer recurrence or metastasis with poor disease-free or overall survival<sup>8</sup>. Some biomarkers were proposed to predict the drug resistance of these ER-positive patients primarily treated with tamoxifen, such as ER expression, progesterone receptor (PR) expression, HER2 and multigene signatures<sup>9</sup>. Additionally, age<sup>10,11</sup>, American Joint Committee on Cancer (AJCC) stage<sup>12</sup> and lymph nodes ratio (LNR)<sup>13,14</sup> that is the ratio of the numbers of metastatic lymph nodes to those of the dissected lymph nodes were also important prognostic factors on prediction of BC recurrence, which were also considered as potential confounding factors when developing new biomarkers.

<sup>1</sup>Department of the Endocrine and Breast Surgery, The First Affiliated Hospital of Chongqing Medical University, Chongqing Medical University, Chongqing, 400016, China. <sup>2</sup>Department of Gastrointestinal Surgery, The First Affiliated Hospital of Chongqing Medical University, Chongqing Medical University, Chongqing, 400016, China. <sup>3</sup>Department of the Breast and Thyroid Surgery, The Traditional Chinese Medicine Hospital, Chongqing, 400021, China. Kang Wang, Jie Li and Yong-Fu Xiong contributed equally to this work. Correspondence and requests for materials should be addressed to X.Z. (email: [cq\\_zhangxiang@163.com](mailto:cq_zhangxiang@163.com)) or H.-Y.L. (email: [hongyu\\_li@hotmail.com](mailto:hongyu_li@hotmail.com))

Variable	GSE17705 cohort	TCGA cohort
No. of patients	298	160
Age, year, mean (SD)	64(10)	53(13)
<b>T stage, No. (%)</b>		
I	133(45)	37(23)
II	151(51)	102(64)
III	23(8)	21(13)
NR	1(1)	0(0)
<b>Lymph node status, No. (%)</b>		
Positive	118(40)	75(47)
Negative	180(60)	58(36)
NR	0(0)	27(17)
<b>AJCC stage, No. (%)</b>		
I	88(30)	25(16)
II	183(61)	95(59)
III	22(7)	36(23)
NR	5(2)	4(2)
<b>PR status, No. (%)</b>		
Positive	77(26)	124(78)
Negative	22(7)	19(12)
NR	199(67)	17(10)

**Table 1.** Clinical Characteristics of the Patients. Abbreviation: TCGA, The Cancer Genome Atlas; SD, standard deviation; NR, not recorded; AJCC, American Joint Committee on Cancer; PR, progesterone receptor.

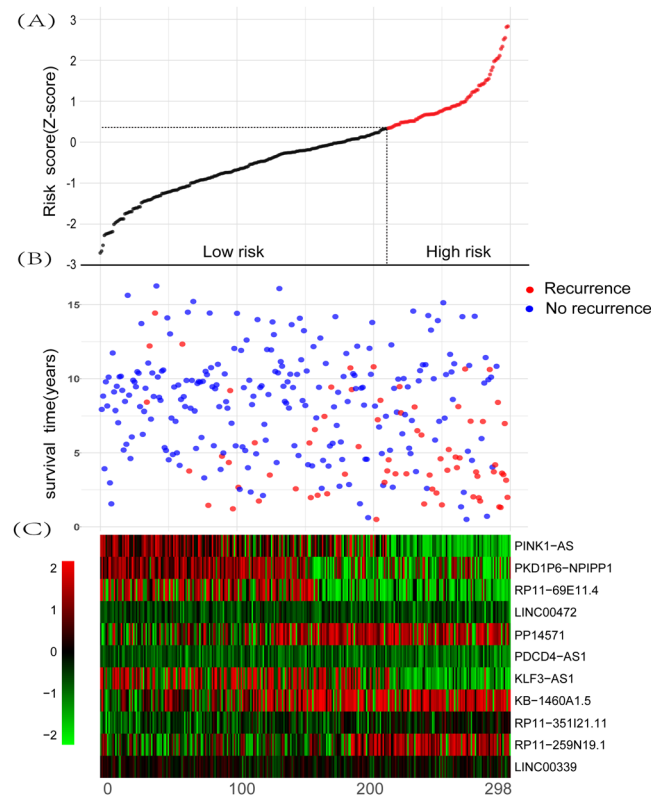
Along with the exploration going to a deeper and finer status, the role of long noncoding RNAs (lncRNAs) in the formation of tamoxifen resistance attracts much more attention<sup>15–17</sup>. lncRNAs are a class of RNA molecules more than 200 nucleotides in length and without protein-coding capacity<sup>18,19</sup>. They participate in all levels of nuclear architecture and gene expression, involving in telomere stability, long range DNA looping and trans-chromosomal interactions, X-chromosome inactivation, imprinting, formation of nuclear sub-compartments, fine regulation of epigenetic modification, regulation of RNA splicing and maturation, control of mRNA and protein stability<sup>20–22</sup>. It has been demonstrated that the dysregulation of lncRNA is associated with various kinds of human diseases, such as cancers, diabetes, chronic hepatitis C, ankylosing spondylitis, autism spectrum disorder<sup>23–27</sup>. Its predicted value for prognosis has been confirmed in several cancers including lung cancer, gallbladder cancer, ovarian cancer, gastric cancer and colorectal cancer<sup>28–33</sup>. Accordingly, previous studies had identified that overexpression of lncRNA BCAR4 in tamoxifen-sensitive ZR-75-1 cells blocked the anti-proliferative effects of tamoxifen, increasing the tamoxifen resistance<sup>34,35</sup>. Nevertheless, limited number of lncRNAs were proposed as clinically relevant biomarkers for tamoxifen resistance, and searching a lncRNA signature might be concrete predictive and prognostic value in the management of ER-positive BC receiving tamoxifen.

In this study, we used data acquired from The Cancer Genome Atlas (TCGA, <http://cancergenome.nih.gov/>) and Gene Expression Omnibus (GEO, <https://www.ncbi.nlm.nih.gov/gds>) whose gene expression microarray data can be mined to develop lncRNAs profiling<sup>36</sup> to identify the predicted effects of lncRNAs on the relapse of ER-positive BC patients treated with tamoxifen.

## Materials and Methods

**ER-Positive BC treated with tamoxifen Datasets.** ER-positive BC datasets corresponding clinical data were downloaded from the GEO and TCGA as training and external validation series, respectively. Only the ER-positive BC patients who were primarily treated with tamoxifen as the unique endocrine therapy other than mixed endocrine therapy such as alternative aromatase inhibitors (AIs) were included. The endpoint is defined relapse free survival (RFS), which is an interval from time of BC diagnosed to disease relapse or metastasis. GSE17705 dataset was screened, and 298 eligible patients treated with tamoxifen for 5 years with clinicopathological and survival data were included in our study as training cohort. After removing corresponding BC patients from TCGA database without clinicopathological data such as age, AJCC stage, LNR, chemotherapy and relapse or progress status, 160 ER-positive BC patients were enrolled finally, which were applied for validation cohort and nomogram construction. Eligible samples were identified from GSE17705 (n = 298) on basis of Affymetrix HG-U133A platform and TCGA dataset (n = 160), whose clinicopathological characteristics in two datasets are presented in Table 1.

**lncRNA Expression Data Processing and Profile Mining.** Raw microarray data files of GSE17705 was retrieved from GEO database, and we adjusted the background using the Robust Multichip Average (RMA) algorithm actualized by R package ‘affy’<sup>37</sup>. lncRNA profile mining was mainly referenced to method proposed by Du *et al.*<sup>38</sup>. Briefly, we mapped the probe sets IDs from Affymetrix HG-U133A to human genome, and SeqMap<sup>39</sup> was used to avoid mismatching. Then, the chromosomal positions according these probe IDs were used to



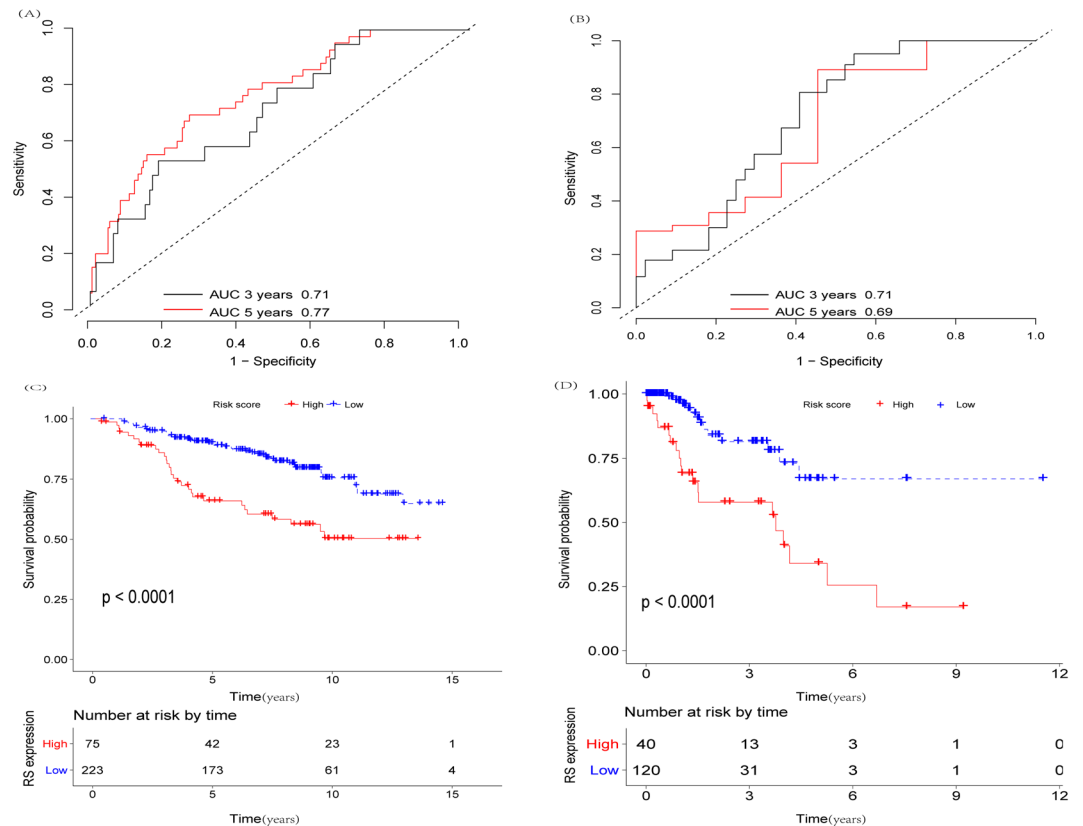
**Figure 1.** 11-LncRNA signature biomarker characteristics in the GSE17705 cohort (N = 298). (A) 11-LncRNA expression and risk score distribution by z-score. The risk scores for all patients in GSE17705 cohort are plotted in ascending order and marked as low risk (black) or high risk (red), as divided by the threshold (vertical black line). (B) Patients' recurrence status and time; (C) heatmap of the lncRNA expression profiles. Rows represent lncRNAs, and columns represent patients. Red indicated higher expression and black indicated lower expression.

identify corresponding positions of lncRNA. Finally, 569 annotated lncRNA transcripts with 742 corresponding Affymetrix probe IDs were generated. The expression profiling of multiple probe IDs mapped to unique lncRNA was calculated as arithmetic mean of the values. Similarly, we downloaded expression information from TCGA BC RNA-sequencing database of lncRNA using The Atlas of Noncoding RNAs Cancer (TANRIC)<sup>40</sup>, and then annotated each included sample according to barcode ID.

**Statistical.** To identify lncRNAs predictive of a RFS, a univariate Cox proportional hazards regression analysis was performed to assess relationship between expression of lncRNAs level and RFS among training series, which was considered statistically significant if their P values were less than or equal to 0.01. Then we conducted stepwise multivariate Cox hazard regression analysis using above lncRNAs to yield a set of genes ultimately that had independent effects on RFS, and a relapse risk score (RRS) for predicting RFS was calculated by taking into account the expression of gene and correlation coefficient as following:

$$RRS = \sum_{i=1}^n (E(i) * C(i)) \quad (1)$$

where n is the number of selected lncRNAs, E is the expression level of lncRNAs and C is the coefficient generated from multivariate Cox regression. Considering the practice of this prognostic model, we assigned each patient a RRS combined linearly by expression of significant lncRNAs and their respective coefficients. Then, we divided training and validation cohort into low-risk group and high-risk group according to RRS, respectively, and the third quantile RRS is defined as the cut-off value depended on receiver operating characteristic (ROC). RFS curve was estimated between high-risk group and low-risk group using Kaplan Meier method, and log-rank tests were used to evaluate the differences between them. ROC curve was conducted to assess sensitivity and specificity of this RRS on basis of selected lncRNAs. To further analysis the independent association between RRS and RFS, univariate and multivariate Cox proportional hazards regressions were applied using clinicopathological factors and RRS. This action was executed only within TCGA cohort, which had relatively comprehensive clinical data. Additionally, a nomogram including factors mentioned before was constructed, and validated internally through 200 bootstrap resamples. Concordance index (C-index) was calculated for the evaluation of the performance of predicting and discrimination ability by test concordance between predicted probability and actual outcome. Similarly, to achieve visualization, calibration of this nomogram was conducted by comparing



**Figure 2.** ROC curves and Kaplan-Meier estimates of 3-year and 5-year RFS. ROC curves showed sensitivity and specificity of 3-year and 5-year RFS prediction by the 11-lncRNA risk score within (A) GSE17705 cohort (n = 298) and (B) TCGA cohort (n = 160). Kaplan-Meier plots were used to visualize the survival probabilities for the low-risk versus high-risk group of patients determined on the basis of the third quartile risk score from the training-set patients within (C) GSE17705 cohort (n = 298) and (D) TCGA cohort (n = 160) (all  $P < 0.0001$ ).

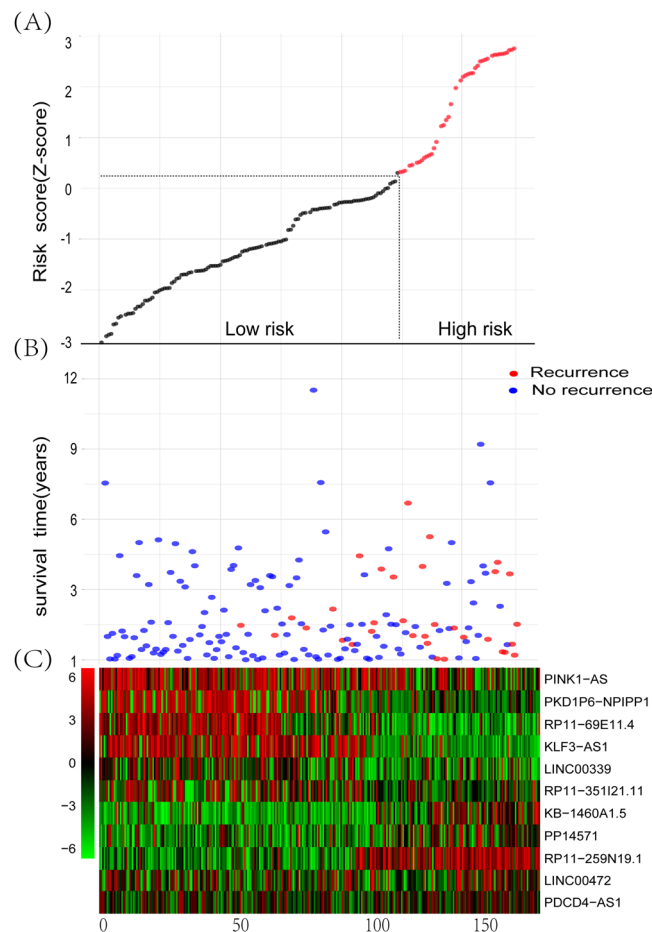
the predicted survival with the observed survival in both, and a 45-degree diagonal line was deemed to a perfectly calibrated model.

**Gene Set Enrichment Analysis.** To reversely validate the functional prediction of identified lncRNAs set, we conducted gene set enrichment analysis (GSEA) based on GSE17705 cohort. GSEA was performed by the JAVA program (<http://www.broadinstitute.org/gsea>) through MSigDB C2 CP: Canonical pathways gene set collection (1320 gene sets available). Gene sets with a false discovery rate (FDR) value  $< 0.05$  after performing 1,000 permutations were considered to be significantly enriched<sup>41</sup>, and selected recurrence or metastasis related enrichment results were visualized by Cytoscape (version 2.8.2) and the Enrichment Map software<sup>42</sup>.

All P or adjusted P values reported were two-sided, which less than 0.05 were considered statistically significant. All of statistical tests were done with SPSS 23 (SPSS, Chicago, IL, USA), R software (version 3.2.5) with Bioconductor<sup>43</sup> and rms packages.

## Results

**Identification of lncRNA genes associated with relapse in the training cohort.** GSE17705 (n = 298) was used to explore prognostic lncRNAs, which was larger database in this study. A univariate Cox proportional hazards regression analysis was performed to identify whether each lncRNA gene expression level was associated with RFS, and we identified a set of 23 genes (PINK1.AS, RP11.93209.9, RP11.259N19.1, RP3.339A18.6, KLF3.AS1, LINC00339, LINC00472, RP11.351I21.11, RP11.72I8.1, KB.1460A1.5, RP1.223E5.4, PKD1P6.NPIPP1, RP11.499P20.2, PDCD4.AS1, RP6.74O6.2, KB.1592A4.15, PP14571, RP11.69E11.4, MGC16275, RP11.105C19.2, AC093110.3, LLNLR.284B4.1, RP11.488C13.7) were significantly correlated with RFS ( $p < 0.05$ , Supplementary Table S1). Stepwise multivariate Cox proportional hazards regression analysis was conducted to select final 11 included lncRNAs, including PINK1.AS, RP11.259N19.1, KLF3.AS1, LINC00339, LINC00472, RP11.351I21.11, KB.1460A1.5, PKD1P6.NPIPP1, PDCD4.AS1, KLF3.AS1 PP14571, RP11.69E11.4. Among those genes, the negative coefficients for the eight genes (PINK1.AS, KLF3.AS1, LINC00339, LINC00472, RP11.351I21.11, PKD1P6.NPIPP1, PDCD4.AS1, RP11.69E11.4) indicated high level of expression were associated with longer RFS. The positive coefficients for remaining 3 genes (RP11.259N19.1, KB.1460A1.5, PP14571) revealed that high expression of them was associated with shorter RFS. Accordingly, a RRS was



**Figure 3.** 11-LncRNA signature biomarker characteristics in the TCGA cohort (N = 160). **(A)** 11-LncRNA expression and risk score distribution by z-score. The risk scores for all patients in TCGA cohort are plotted in ascending order and marked as low risk (black) or high risk (red), as divided by the threshold (vertical black line). **(B)** Patients' recurrence status and time; **(C)** heatmap of the lncRNA expression profiles. Rows represent lncRNAs, and columns represent patients. Red indicated higher expression and black indicated lower expression.

established to predict the relapse risks of BC patients treated with tamoxifen, as following:  $RRS = (-0.67223 * \text{expression value of PINK1.AS}) + (0.38401 * \text{expression value of RP11.259N19.1}) + (-0.18630 * \text{expression value of KLF3.AS1}) + (-0.19157 * \text{expression value of LINC00339}) + (-0.21613 * \text{expression value of LINC00472}) + (-0.20857 * \text{expression value of RP11.351I21.11}) + (0.22666 * \text{expression value of KB.1460A1.5}) + (-0.50148 * \text{expression value of PKD1P6.NPIPP1}) + (-0.23588 * \text{expression value of PDCD4.AS1}) + (0.16569 * \text{expression value of PP14571}) + (-0.23448 * \text{expression value of RP11.69E11.4})$ . Distribution of the eleven-lncRNA risk score, relapse status, and lncRNAs expression of training patients were shown in Fig. 1. After conducting ROC analysis, the prognostic power of 3-year or 5-year RFS predicted by eleven-lncRNAs risk score was evaluated within GSE17705. As shown in Fig. 2A, the sensitivity, specificity of RRS were 0.67 and 0.74, respectively, and the area under curves (AUCs) of RRS for prediction 3-year and 5-year RFS were 0.71 and 0.77, respectively, indicating that the performance of eleven-lncRNAs based RRS was mediocre. The training cohort was divided into high-risk (n = 74) and low-risk group (n = 224) according to the third quartile risk score as the cut-off value. Kaplan-Meier analysis revealed that patients in high-risk group had significantly shorter median DFS than those in low-risk group (log-rank test  $P < 0.0001$ ) (Fig. 2C).

**The 11-lncRNA signature and the patients' survival in the external validation cohort.** To validate the prognostic power of the eleven-lncRNA signature for the prediction of relapse risk, the TCGA cohort (n = 160) use the same formula to classify patients into high-risk group (n = 40) and low-risk group (n = 120) using the third quartile RRS of validation series as the cut-off point. ROC suggested that AUC of 0.71 and 0.69 revealed general predictive performances of 3-year and 5-year RFS, respectively (Fig. 2B). In the consistency with the findings described above, patients in the high-risk group had significantly shorter median RFS than those in the low-risk group (log-rank test  $P < 0.0001$ ) (Fig. 2D). The distribution of the eleven-lncRNAs risk score, the relapse status of the BC patients and the lncRNA expression signature were also obtained (Fig. 3).

Variable	Relapse Free Survival			
	Univariate Analysis		Multivariate Analysis	
	HR (95% CI)	P value	HR (95% CI)	P value
<b>11-LncRNA risk score (n = 160)</b>				
Low	Reference		Reference	
High	3.30(1.66–6.57)	0.001	3.56(1.46–8.66)	0.005
<b>Age (year)</b>				
<40	Reference		Reference	
40–50	0.64(0.24–1.69)	0.37	0.44(0.15–1.34)	0.15
50–60	0.42(0.12–1.44)	0.17	0.35(0.09–1.31)	0.12
>60	1.19(0.46–3.04)	0.72	0.36(0.10–1.36)	0.13
<b>AJCC stage</b>				
I	Reference		Reference	
II	1.97(0.57–6.82)	0.29	4.06(0.82–29.09)	0.09
III	4.89(1.39–17.15)	0.01	6.60(1.17–37.34)	0.03
LNR*	3.60(0.97–13.43)	0.06	2.28(0.28–18.76)	0.45
<b>Chemotherapy</b>				
Yes	Reference		Reference	
No	0.64(0.31–1.32)	0.23	0.41(0.14–1.22)	0.11

**Table 2.** Univariate and Multivariable Cox Regression Analysis in TCGA Cohort. \*In Cox regression analysis, LNR was considered as continuous variable. Abbreviations: HR, hazard ratio; CI, confidence interval; LNR, positive lymph node ratio.

**Independence of RFS prediction by the 11-lncRNA signature from other clinicopathological variables.** To further test whether the 11-lncRNA signature was independent of other clinicopathological variables, we conducted multivariate Cox regression analysis using RRS, age, AJCC stage, chemotherapy status and LNR as covariables in the TCGA cohort (n = 160). The results suggested that the 11-lncRNA signature risk score (RRS) maintained a significant correlation with RFS after adjustment for other clinical variables (HR = 3.56, 95%CI = 1.46–8.66) (Table 2), indicating that RRS is an independent relapse predictive factor. Additionally, a nomogram was constructed using above variables, which was shown as Fig. 4A. Corresponding ROC analyses were conducted according to 3-year and 5-year RFS (Fig. 4B), and yielded AUC of 0.71 and 0.73 indicated that RRS combined clinicopathological variables can generally predict 3-year and 5-year RFS, respectively. In internal validation, C-indexes for the nomograms to predict RFS were 0.78 (95% CI, 95%CI = 0.67 to 0.89), whose calibration plots for prediction of 3-year and 5-year RFS were presented in Fig. 4C.

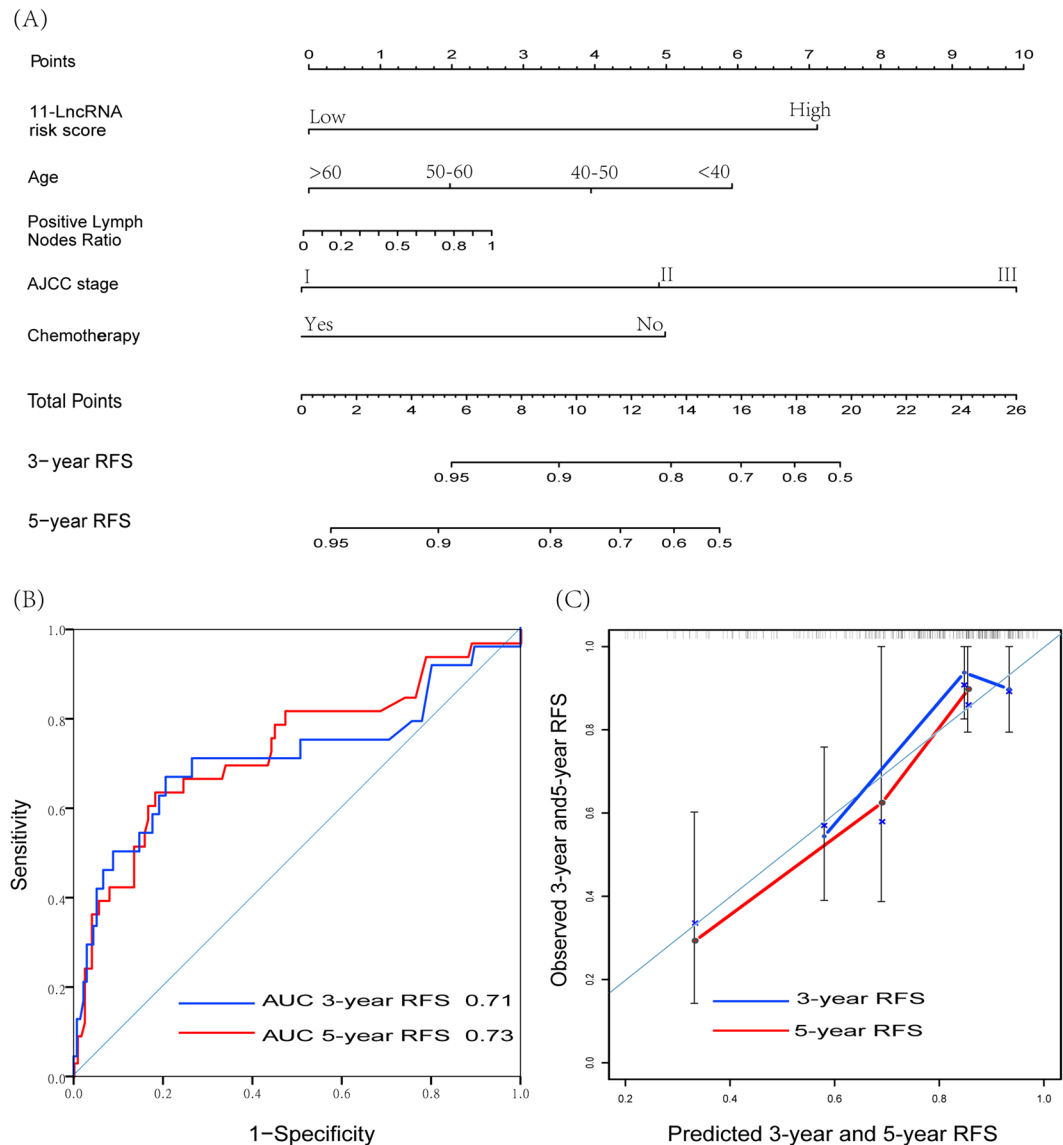
**Functional prediction of the 11-lncRNA signature.** To infer the potential function of the 11-lncRNAs signature in prediction of BC relapse, we performed GSEA to identify associated biological processes and signaling pathways based on GSE17705 datasets. The gene sets from high-risk group compared with low-risk defined before was employed to select significant individuals (adjusted  $P < 0.05$ ) (Fig. 5A) (Supplementary Table S2), and then relapse or metastasis related sets of them were visualized as interaction networks with Cytoscape and Enrichment Map (Fig. 5B). Several relapse or metastasis-related pathways like PI3K-Akt signaling, Wnt signaling and focal adhesion were found from up-regulation gene sets (Fig. 5B).

## Discussion

In this study, we developed and validated a novel prognostic tool based on 11-lncRNAs to improve the prediction of disease recurrence for ER-positive BC patients who were primarily treated with tamoxifen as unique endocrine therapy. Our results showed that this tool can successfully classify patients into high-risk and low-risk groups, and differences in 3-year/5-year RFS were identified in training and validation cohorts. In addition, the 11-lncRNAs signature was proved to be independent from other clinicopathological variables, and GSEA was conducted on basis of RRS to validate the functional differences. A comprehensive and practical nomogram was construct to predict the individualized recurrence, which provided opportunities for clinicians to classify the patients according to these valid tools.

LncRNA LINC00472 is located on chromosome 6q13, and the verified transcript has 2,933 bp. Previous *in vitro* study with meta-analysis suggested that LINC00472 is a tumor suppressor in breast cancer, whose potential prognostic and predictive value should be evaluated<sup>44</sup>. Similarly, Shen *et al.* indicated that LINC00472 expression in breast tumors predicted the recurrence of grade II breast cancer<sup>45</sup>. Interestingly, high expression of LINC00472 was often more frequently in low grade and early stage epithelial ovarian cancer compared to high grade tumors and late stage<sup>46</sup>. Moreover, integrated analysis of LncRNA-associated competing endogenous RNAs (ceRNAs) network revealed that LINC00472 was positively correlated with OS of lung adenocarcinoma, and the correlation between LINC00472 and AFAP1-AS1, one of the most common biomarkers to predict the outcomes of cancer patients, was also confirmed<sup>47,48</sup>. LncRNA KB.1460A1.5 was only proposed in Alzheimer's disease, suggesting this lncRNA with high topological features in the intracellular neurofibrillary tangles may play an important role in Alzheimer's disease. Unexpectedly, related reports were not found in field of oncology. Among the 11 lncRNAs,



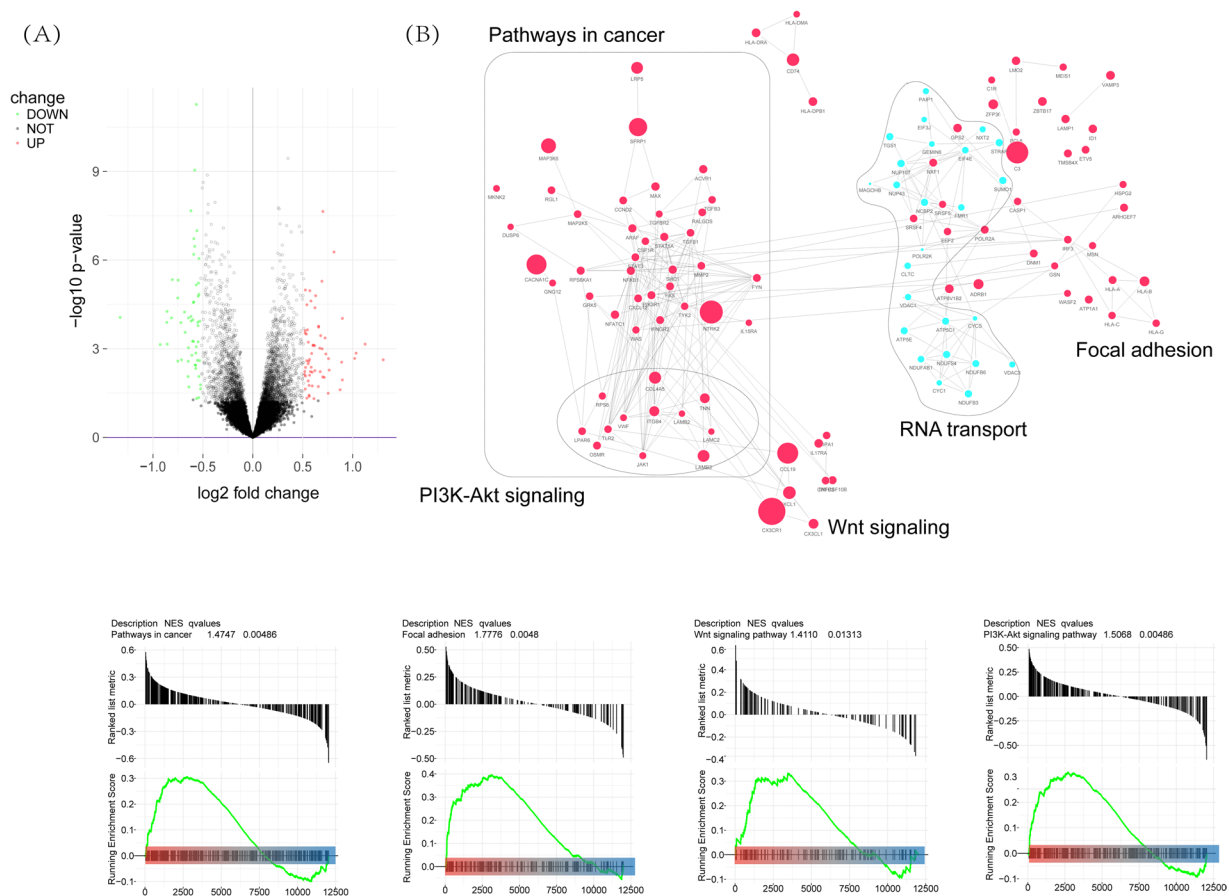


**Figure 4.** Nomogram, ROC curves and the internal calibration for predicting 3-year and 5-year RFS within TCGA cohort ( $n = 160$ ). The nomogram was built using 11-lncRNA risk score, age, positive lymph nodes ratio, AJCC stage, and chemotherapy. **(A)** Nomogram for predicting proportion of patients with 3-year and 5-year RFS. **(B)** ROC curve showed sensitivity and specificity of the nomogram. **(C)** Plots depict the calibration of model in terms of agreements between predicted and observed 3-year and 5-year RFS. Model performance is shown by the plot, relative to the 45-degree line, which represents perfect prediction.

except for those mentioned above, other lncRNAs are either poorly investigated or have not been reported, which means our findings suggested further research for them is imperative.

With the development of next generation sequencing technologies and the rise of transcriptomics, a class of noncoding RNA attracts people's attention such as MicroRNAs (miRNAs) and lncRNAs. Several miRNAs are thought to involve in tamoxifen resistance by directly inhibiting the expression of ER- $\alpha$  like miR-221/222<sup>49</sup> and mi-342<sup>50</sup>, or mediately regulating genes related to cell survival and metastasis, such as miR-451<sup>51</sup>, miR-375<sup>52</sup>, miR-15a/16<sup>53</sup>, miR-200s<sup>54</sup>. Zhang H *et al.* reported that downregulation of lncRNA-ROR could enhance the sensibility of breast cancer cells to tamoxifen by increasing miR-205 expression<sup>15</sup>. lncRNA UCA1 could enhance tamoxifen resistance via inhibiting mTOR signaling pathway, activating wnt/ $\beta$ -catenin signaling pathway, or through a miR-18a-HIF1 $\alpha$  regulatory feedback loop<sup>16,55,56</sup>. lncRNA HOTAIR could promote the ligand-independent ER activities and thereby contribute to tamoxifen resistance<sup>57</sup>.

Accumulating evidences indicate that the altered expression of ER $\alpha$ , crosstalk with other signaling like HER-2, PI3K and MAPK signaling, mutation of ESR1 and decreased metabolism of the drug may contribute to the acquiring resistance of tamoxifen<sup>8,58-60</sup>, of which the phosphatidylinositol 3-kinase (PI3K)/Akt/mammalian target of rapamycin (mTOR) signaling pathway was identified within results of GSEA in this study. PI3K/Akt and mTOR are two important signaling pathways controlling series of biological processes like cell growth and



**Figure 5.** (A) Gene Set Enrichment Analysis Delineates Biological Pathways and Processes associated with risk score within GSE17705 cohort. (A) Volcano Plot with t-statistics. Circles show potentially differential genes between low-risk and high-risk group based on 11-lncRNA risk scores. The green and red circle were up and down genes, respectively. (B) Cytoscape and Enrichment Map were used for visualization of the GSEA results. Nodes represent enriched gene sets, which are grouped and annotated by their similarity according to related gene sets. Enrichment results were mapped as a network of gene sets (nodes). Node size is proportional to the total number of genes within each gene set. Proportion of shared genes between gene sets is represented as the thickness of the green line between nodes.

proliferation, apoptosis, cell-cycle, DNA repair, angiogenesis and glucose metabolism<sup>61</sup>. Ample preclinical studies have indicated that PI3K/Akt/mTOR signaling pathway is co-related with endocrine resistance. The crosstalk between ER and PI3K/Akt/mTOR signaling pathway is a key mechanism by which tamoxifen resistance happens. Yamnik, R. L. *et al.* found that S6 kinase 1, a substrate of mTOR complex 1 (mTORC1) could phosphorylate the function domain 1 of ER, which led to an estrogen-independent activation of the ER- $\alpha$ , thereby resulting in tamoxifen resistance<sup>62</sup>. Otherwise, clinical data showed that the activation of Akt or reduction of PTEN, a tumor suppressor mediating dephosphorylation of PIP3 to PIP2, were associated with tamoxifen resistance in both metastatic and recurrent breast cancer patients<sup>63,64</sup>. Everolimus, a mTOR inhibitors, has been approved by FDA for the treatment of hormone receptors positive and HER2 negative advanced postmenopausal breast cancer patients, hoping to overcome or slowdown the development of endocrine resistance.

So far, a series of studies have demonstrated that the wnt/ $\beta$ -catenin signaling pathway is co-related with embryogenesis, self-renewal of adult tissues and tumorigenesis via controlling cell cell morphology, differentiation, proliferation, migration, apoptosis, and even stem cell self-renewal<sup>65–71</sup>. And the accumulation of  $\beta$ -catenin in the nucleus or cytoplasm has been found in approximately 50% of breast cancers, which predicted a poor prognosis<sup>72</sup>. Meanwhile, wnt/ $\beta$ -catenin has the potential to promote cancer malignant development through motivating the EMT process to facilitate cancer metastasis or inducing therapeutic resistance<sup>73,74</sup>. Interestingly, lncRNA HNF1A-AS1, lncRNA CCND2-AS1, lncRNA CRNDE, lncRNA CASC2 were revealed that promotional effects on cell proliferation via activation of Wnt/ $\beta$ -catenin cascade in osteosarcoma, glioma, colorectal cancer and bladder cancer were documented, respectively<sup>75–78</sup>. Additionally, lncRNA UCA1 promotes epithelial-mesenchymal transition of BC cells via enhancing Wnt/ $\beta$ -catenin signaling pathway to facilitate its metastasis<sup>79</sup>, and better reactions to tamoxifen through inhibiting Wnt/ $\beta$ -catenin pathway were found when knock-downing of lncRNA UCA1 could<sup>86</sup>.



Some limitations of our study should be acknowledged. Firstly, GSE17705 data sets included in this study was profiled through Affymetrix Human Genome U133A Array, and only part of possible lncRNAs were mined. The incomplete candidate lncRNAs maybe cannot represent the whole biological function of this noncoding RNA, and more large-scale studies were urgently needed to validate reliability of our model. Secondly, eligible participants from two cohorts all received tamoxifen as primarily endocrine therapy, but status and duration of this prescription was poorly reported. The GSE17705 cohort involved 298 BC patients treated with 5-year tamoxifen, whereas TCGA cohort was identified by therapy data, which is one of possible reasons of imperfect validation. Lastly, no experimental validation was acquired to reveal the underlying mechanisms of 11-lncRNAs based prognostic model, and further experimental studies were needed to supply the better understanding of this 11-lncRNAs signature.

In conclusion, our findings showed that the 11-lncRNA based classifier is a reliable prognostic and predictive tool for disease relapse in BC patients receiving tamoxifen, which is independent from traditional clinicopathological factors. Developed nomograms is also available to predict whether patients can benefit from tamoxifen therapy, which are valid tools to illustrate the visualized results.

## References

1. Torre, L. A. *et al.* Global cancer statistics, 2012. *CA: a cancer journal for clinicians* **65**, 87–108, <https://doi.org/10.3322/caac.21262> (2015).
2. Siegel, R. L., Miller, K. D. & Jemal, A. Cancer Statistics, 2017. *CA: a cancer journal for clinicians* **67**, 7–30, <https://doi.org/10.3322/caac.21387> (2017).
3. Rossi, L. & Pagani, O. Adjuvant Endocrine Therapy in Breast Cancer: Evolving Paradigms in Premenopausal Women. *Curr Treat Options Oncol* **18**, 1534–6277 (2017).
4. Zwart, W., Terra, H., Linn, S. C. & Schagen, S. B. Cognitive effects of endocrine therapy for breast cancer: keep calm and carry on? *Nat Rev Clin Oncol* **12**, 597–606, <https://doi.org/10.1038/nrclinonc.2015.124> (2015).
5. Tao, Z. *et al.* Breast Cancer: Epidemiology and Etiology. *Cell biochemistry and biophysics*. <https://doi.org/10.1007/s12013-014-0459-6> (2014).
6. Sims, A. H., Howell, A., Howell, S. J. & Clarke, R. B. Origins of breast cancer subtypes and therapeutic implications. *Nature clinical practice. Oncology* **4**, 516–525, <https://doi.org/10.1038/ncponc908> (2007).
7. Hernandez-Aya, L. F. & Gonzalez-Angulo, A. M. Adjuvant systemic therapies in breast cancer. *The Surgical clinics of North America* **93**, 473–491, <https://doi.org/10.1016/j.suc.2012.12.002> (2013).
8. Fan, W., Chang, J. & Fu, P. Endocrine therapy resistance in breast cancer: current status, possible mechanisms and overcoming strategies. *Future Med. Chem.* **7**, 1511–1519, <https://doi.org/10.4155/fmc> (2015).
9. Selli, C., Dixon, J. M. & Sims, A. H. Accurate prediction of response to endocrine therapy in breast cancer patients: current and future biomarkers. *Breast cancer research: BCR* **18**, 118, <https://doi.org/10.1186/s13058-016-0779-0> (2016).
10. Liedtke, C. *et al.* The prognostic impact of age in different molecular subtypes of breast cancer. *Breast cancer research and treatment* **152**, 667–673, <https://doi.org/10.1007/s10549-015-3491-3> (2015).
11. van den Broek, A. J. *et al.* Impact of Age at Primary Breast Cancer on Contralateral Breast Cancer Risk in BRCA1/2 Mutation Carriers. *Journal of clinical oncology: official journal of the American Society of Clinical Oncology* **34**, 409–418, <https://doi.org/10.1200/jco.2015.62.3942> (2016).
12. Yi, M. *et al.* Novel staging system for predicting disease-specific survival in patients with breast cancer treated with surgery as the first intervention: time to modify the current American Joint Committee on Cancer staging system. *Journal of clinical oncology: official journal of the American Society of Clinical Oncology* **29**, 4654–4661, <https://doi.org/10.1200/jco.2011.38.3174> (2011).
13. Liao, G. S. *et al.* Prognostic value of the lymph node ratio in breast cancer subtypes. *American journal of surgery* **210**, 749–754, <https://doi.org/10.1016/j.amjsurg.2014.12.054> (2015).
14. Solak, M. *et al.* The lymph node ratio as an independent prognostic factor for non-metastatic node-positive breast cancer recurrence and mortality. *Journal of B.U.O.N.: official journal of the Balkan Union of Oncology* **20**, 737–745 (2015).
15. Zhang, H. Y. *et al.* Effects of long noncoding RNA-ROR on tamoxifen resistance of breast cancer cells by regulating microRNA-205. *Cancer chemotherapy and pharmacology* **79**, 327–337, <https://doi.org/10.1007/s00280-016-3208-2> (2017).
16. Li, X., Wu, Y., Liu, A. & Tang, X. Long non-coding RNA UCA1 enhances tamoxifen resistance in breast cancer cells through a miR-18a-HIF1alpha feedback regulatory loop. *Tumour biology: the journal of the International Society for Oncodevelopmental Biology and Medicine* **37**, 14733–14743, <https://doi.org/10.1007/s13277-016-5348-8> (2016).
17. Cai, Y., He, J. & Zhang, D. [Suppression of long non-coding RNA CCAT2 improves tamoxifen-resistant breast cancer cells' response to tamoxifen]. *Molekuliarnaia biologii* **50**, 821–827, <https://doi.org/10.7868/s0026898416030046> (2016).
18. Mercer, T. R., Dinger, M. E. & Mattick, J. S. Long non-coding RNAs: insights into functions. *Nature reviews. Genetics* **10**, 155–159, <https://doi.org/10.1038/nrg2521> (2009).
19. Lipovich, L., Johnson, R. & Lin, C. Y. MacroRNA underdogs in a microRNA world: evolutionary, regulatory, and biomedical significance of mammalian long non-protein-codingRNA. *Biochimica et biophysica acta* **1799**, 597–615, <https://doi.org/10.1016/j.bbgrm.2010.10.001> (2010).
20. Khorkova, O., Hsiao, J. & Wahlestedt, C. Basic biology and therapeutic implications of lncRNA. *Advanced drug delivery reviews* **87**, 15–24, <https://doi.org/10.1016/j.addr.2015.05.012> (2015).
21. Wang, Z. *et al.* Downregulation of the long non-coding RNA TUSC7 promotes NSCLC cell proliferation and correlates with poor prognosis. *American journal of translational research* **8**, 680–687 (2016).
22. Zhang, X. *et al.* Maternally expressed gene 3 (MEG3) noncoding ribonucleic acid: isoform structure, expression, and functions. *Endocrinology* **151**, 939–947, <https://doi.org/10.1210/en.2009-0657> (2010).
23. Zhang, Q. *et al.* The characteristic landscape of lncRNAs classified by RBP-lncRNA interactions across 10 cancers. *Molecular bioSystems*. <https://doi.org/10.1039/c7mb00144d> (2017).
24. Li, X. *et al.* The Diagnostic Value of Whole Blood lncRNA ENST00000550337.1 for Pre-Diabetes and Type 2 Diabetes Mellitus. *Molecular bioSystems*, <https://doi.org/10.1039/c7mb00144d>, 10.1055/s-0043-100018 (2017).
25. Sabry, D., Elamir, A., Mahmoud, R. H., Abdelaziz, A. A. & Fathy, W. Role of lncRNA-AF085935, IL-10 and IL-17 in Rheumatoid Arthritis Patients With Chronic Hepatitis C. *Journal of clinical medicine research* **9**, 416–425, <https://doi.org/10.14740/jocmr2896w> (2017).
26. Li, X. *et al.* Down-Regulation of lncRNA-AK001085 and its Influences on the Diagnosis of Ankylosing Spondylitis. *Medical science monitor: international medical journal of experimental and clinical research* **23**, 11–16 (2017).
27. Liu, L. *et al.* Progression-free survival as a surrogate endpoint for overall survival in patients with third-line or later-line chemotherapy for advanced gastric cancer. *Onco Targets Ther* **8**, 921–928 (2015).
28. Chen, S. *et al.* lncRNA CCAT2 predicts poor prognosis and regulates growth and metastasis in small cell lung cancer. *Biomed Pharmacother* **82**, 583–588, <https://doi.org/10.1016/j.biopha.2016.05.017> (2016).

29. Zhao, Q. S. *et al.* Over-expression of lncRNA SBF2-AS1 is associated with advanced tumor progression and poor prognosis in patients with non-small cell lung cancer. *European review for medical and pharmacological sciences* **20**, 3031–3034 (2016).
30. Ma, F. *et al.* Overexpression of lncRNA AFAP1-AS1 predicts poor prognosis and promotes cells proliferation and invasion in gallbladder cancer. *Biomed Pharmacother* **84**, 1249–1255, <https://doi.org/10.1016/j.biopha.2016.10.064> (2016).
31. Chen, Z. J., Zhang, Z., Xie, B. B. & Zhang, H. Y. Clinical significance of up-regulated lncRNA NEAT1 in prognosis of ovarian cancer. *European review for medical and pharmacological sciences* **20**, 3373–3377 (2016).
32. Fan, Z. Y. *et al.* Identification of a five-lncRNA signature for the diagnosis and prognosis of gastric cancer. *Tumour Biol* **37**, 13265–13277, <https://doi.org/10.1007/s13277-016-5185-9> (2016).
33. Liu, Y., Zhang, M., Liang, L., Li, J. & Chen, Y. X. Over-expression of lncRNA DANCR is associated with advanced tumor progression and poor prognosis in patients with colorectal cancer. *International journal of clinical and experimental pathology* **8**, 11480–11484 (2015).
34. Godinho, M. F. *et al.* Relevance of BCAR4 in tamoxifen resistance and tumour aggressiveness of human breast cancer. *British journal of cancer* **103**, 1284–1291, <https://doi.org/10.1038/sj.bjc.6605884> (2010).
35. Meijer, D., van Agthoven, T., Bosma, P. T., Nooter, K. & Dorssers, L. C. Functional screen for genes responsible for tamoxifen resistance in human breast cancer cells. *Molecular cancer research: MCR* **4**, 379–386, <https://doi.org/10.1158/1541-7786.mcr-05-0156> (2006).
36. Li, R., Qian, J., Wang, Y. Y., Zhang, J. X. & You, Y. P. Long noncoding RNA profiles reveal three molecular subtypes in glioma. *CNS neuroscience & therapeutics* **20**, 339–343, <https://doi.org/10.1111/cns.12220> (2014).
37. Irizarry, R. A. *et al.* Exploration, normalization, and summaries of high density oligonucleotide array probe level data. *Biostatistics (Oxford, England)* **4**, 249–264, <https://doi.org/10.1093/biostatistics/4.2.249> (2003).
38. Du, Z. *et al.* Integrative genomic analyses reveal clinically relevant long noncoding RNAs in human cancer. *Nature structural & molecular biology* **20**, 908–913, <https://doi.org/10.1038/nsmb.2591> (2013).
39. Jiang, H. & Wong, W. H. SeqMap: mapping massive amount of oligonucleotides to the genome. *Bioinformatics (Oxford, England)* **24**, 2395–2396, <https://doi.org/10.1093/bioinformatics/btn429> (2008).
40. Li, J. *et al.* TANRIC: An Interactive Open Platform to Explore the Function of lncRNAs in Cancer. *Cancer research* **75**, 3728–3737, <https://doi.org/10.1158/0008-5472.can-15-0273> (2015).
41. Subramanian, A. *et al.* Gene set enrichment analysis: a knowledge-based approach for interpreting genome-wide expression profiles. *Proceedings of the National Academy of Sciences of the United States of America* **102**, 15545–15550, <https://doi.org/10.1073/pnas.0506580102> (2005).
42. Merico, D., Isserlin, R., Stueker, O., Emili, A. & Bader, G. D. Enrichment map: a network-based method for gene-set enrichment visualization and interpretation. *PLoS one* **5**, e13984, <https://doi.org/10.1371/journal.pone.0013984> (2010).
43. Gentleman, R. C. *et al.* Bioconductor: open software development for computational biology and bioinformatics. *Genome biology* **5**, R80, <https://doi.org/10.1186/gb-2004-5-10-r80> (2004).
44. Shen, Y. *et al.* Prognostic and predictive values of long non-coding RNA LINC00472 in breast cancer. *Oncotarget* **6**, 8579–8592, <https://doi.org/10.18632/oncotarget.3287> (2015).
45. Shen, Y. *et al.* LINC00472 expression is regulated by promoter methylation and associated with disease-free survival in patients with grade 2 breast cancer. *Breast cancer research and treatment* **154**, 473–482, <https://doi.org/10.1007/s10549-015-3632-8> (2015).
46. Fu, Y. *et al.* Long non-coding RNAs, ASAP1-IT1, FAM215A, and LINC00472, in epithelial ovarian cancer. *Gynecologic oncology* **143**, 642–649, <https://doi.org/10.1016/j.ygyno.2016.09.021> (2016).
47. Sui, J. *et al.* Integrated analysis of long non-coding RNA-associated ceRNA network reveals potential lncRNA biomarkers in human lung adenocarcinoma. *International journal of oncology* **49**, 2023–2036, <https://doi.org/10.3892/ijo.2016.3716> (2016).
48. Liu, F. T. *et al.* Long noncoding RNA AFAP1-AS1, a potential novel biomarker to predict the clinical outcome of cancer patients: a meta-analysis. *OncoTargets and therapy* **9**, 4247–4254, <https://doi.org/10.2147/ott.s107188> (2016).
49. Miller, T. E. *et al.* MicroRNA-221/222 confers tamoxifen resistance in breast cancer by targeting p27Kip1. *The Journal of biological chemistry* **283**, 29897–29903, <https://doi.org/10.1074/jbc.M804612200> (2008).
50. He, Y. J. *et al.* miR-342 is associated with estrogen receptor-alpha expression and response to tamoxifen in breast cancer. *Experimental and therapeutic medicine* **5**, 813–818, <https://doi.org/10.3892/etm.2013.915> (2013).
51. Bergamaschi, A. & Katzenellenbogen, B. S. Tamoxifen downregulation of miR-451 increases 14-3-3zeta and promotes breast cancer cell survival and endocrine resistance. *Oncogene* **31**, 39–47, <https://doi.org/10.1038/nc.2011.223> (2012).
52. Ward, A. *et al.* Re-expression of microRNA-375 reverses both tamoxifen resistance and accompanying EMT-like properties in breast cancer. *Oncogene* **32**, 1173–1182, <https://doi.org/10.1038/nc.2012.128> (2013).
53. Chu, J. *et al.* E2F7 overexpression leads to tamoxifen resistance in breast cancer cells by competing with E2F1 at miR-15a/16 promoter. *Oncotarget* **6**, 31944–31957, <https://doi.org/10.18632/oncotarget.5128> (2015).
54. Manavalan, T. T. *et al.* Reduced expression of miR-200 family members contributes to antiestrogen resistance in LY2 human breast cancer cells. *PLoS one* **8**, e62334, <https://doi.org/10.1371/journal.pone.0062334> (2013).
55. Wu, C. & Luo, J. Long Non-Coding RNA (lncRNA) Urothelial Carcinoma-Associated 1 (UCA1) Enhances Tamoxifen Resistance in Breast Cancer Cells via Inhibiting mTOR Signaling Pathway. *Medical Science Monitor* **22**, 3860–3867, <https://doi.org/10.12659/msm.900689> (2016).
56. Liu, H. *et al.* Knockdown of Long Non-Coding RNA UCA1 Increases the Tamoxifen Sensitivity of Breast Cancer Cells through Inhibition of Wnt/beta-Catenin Pathway. *PLoS one* **11**, e0168406, <https://doi.org/10.1371/journal.pone.0168406> (2016).
57. Xue, X. *et al.* lncRNA HOTAIR enhances ER signaling and confers tamoxifen resistance in breast cancer. *Oncogene* **35**, 2746–2755, <https://doi.org/10.1038/nc.2015.340> (2016).
58. Ma, G. *et al.* Mitogen-activated protein kinase phosphatase 1 is involved in tamoxifen resistance in MCF7 cells. *Oncol Rep* **34**, 2423–2430, <https://doi.org/10.3892/or.2015.4244> (2015).
59. Jeselsohn, R., Buchwalter, G., De Angelis, C., Brown, M. & Schiff, R. ESR1 mutations—a mechanism for acquired endocrine resistance in breast cancer. *Nat Rev Clin Oncol* **12**, 573–583, <https://doi.org/10.1038/nrclinonc.2015.117> (2015).
60. Motamedi, S. *et al.* Tamoxifen resistance and CYP2D6 copy numbers in breast cancer patients. *Asian Pacific journal of cancer prevention: APJCP* **13**, 6101–6104 (2012).
61. Nicolini, A., Ferrari, P., Kotlarova, L., Rossi, G. & Biava, P. M. The PI3K-AKT-mTOR Pathway and New Tools to Prevent Acquired Hormone Resistance in Breast Cancer. *Curr Pharm Biotechnol* **16**, 804–815 (2015).
62. Yammik, R. L. & Holz, M. K. mTOR/S6K1 and MAPK/RSK signaling pathways coordinately regulate estrogen receptor alpha serine 167 phosphorylation. *FEBS Lett* **584**, 124–128, <https://doi.org/10.1016/j.febslet.2009.11.041> (2010).
63. Kirkegaard, T. *et al.* AKT activation predicts outcome in breast cancer patients treated with tamoxifen. *The Journal of pathology* **207**, 139–146, <https://doi.org/10.1002/path.1829> (2005).
64. Shoman, N. *et al.* Reduced PTEN expression predicts relapse in patients with breast carcinoma treated by tamoxifen. *Modern pathology: an official journal of the United States and Canadian Academy of Pathology, Inc* **18**, 250–259, <https://doi.org/10.1038/modpathol.3800296> (2005).
65. Fodde, R. & Brabletz, T. Wnt/beta-catenin signaling in cancer stemness and malignant behavior. *Curr Opin Cell Biol* **19**, 150–158, <https://doi.org/10.1016/j.ceb.2007.02.007> (2007).
66. Clevers, H. Wnt/beta-catenin signaling in development and disease. *Cell* **127**, 469–480, <https://doi.org/10.1016/j.cell.2006.10.018> (2006).

67. Clevers, H. & Nusse, R. Wnt/beta-catenin signaling and disease. *Cell* **149**, 1192–1205, <https://doi.org/10.1016/j.cell.2012.05.012> (2012).
68. Chirillo, M. A. Role of the Wnt/b-catenin pathway in gastric cancer: An in-depth literature review. *World J Exp Med* **5**, 84–102, <https://doi.org/10.5493/wjem> (2015).
69. Arend, R. C., Londono-Joshi, A. I., Straughn, J. M. Jr. & Buchsbaum, D. J. The Wnt/beta-catenin pathway in ovarian cancer: a review. *Gynecologic oncology* **131**, 772–779, <https://doi.org/10.1016/j.ygyno.2013.09.034> (2013).
70. Kypka, R. M. & Waxman, J. Wnt/beta-catenin signalling in prostate cancer. *Nature reviews. Urology* **9**, 418–428, <https://doi.org/10.1038/nrurol.2012.116> (2012).
71. Peifer, M. & Polakis, P. Wnt signaling in oncogenesis and embryogenesis—a look outside the nucleus. *Science* **287**, 1606–1609 (2000).
72. Lin, S. Y. *et al.* Beta-catenin, a novel prognostic marker for breast cancer: its roles in cyclin D1 expression and cancer progression. *Proceedings of the National Academy of Sciences of the United States of America* **97**, 4262–4266, <https://doi.org/10.1073/pnas.060025397> (2000).
73. Shan, S. *et al.* Wnt/ $\beta$ -catenin pathway is required for epithelial to mesenchymal transition in CXCL12 over expressed breast cancer cells. *International journal of clinical and experimental pathology* **8**, 12357–12367 (2015).
74. Cui, J., Jiang, W., Wang, S., Wang, L. & Xie, K. Role of Wnt/beta-catenin signaling in drug resistance of pancreatic cancer. *Current pharmaceutical design* **18**, 2464–2471 (2012).
75. Zhao, H. *et al.* Upregulation of lncRNA HNF1A-AS1 promotes cell proliferation and metastasis in osteosarcoma through activation of the Wnt/ $\beta$ -catenin signaling pathway. *Am J Transl Res* **8**, 3503–3512 (2016).
76. Zhang, H., Wei, D. L., Wan, L., Yan, S. F. & Sun, Y. H. Highly expressed lncRNA CCND2-AS1 promotes glioma cell proliferation through Wnt/beta-catenin signaling. *Biochemical and biophysical research communications* **482**, 1219–1225, <https://doi.org/10.1016/j.bbrc.2016.12.016> (2017).
77. Han, P. *et al.* The lncRNA CRNDE promotes colorectal cancer cell proliferation and chemoresistance via miR-181a-5p-mediated regulation of Wnt/beta-catenin signaling. *Molecular cancer* **16**, 9, <https://doi.org/10.1186/s12943-017-0583-1> (2017).
78. Pei, Z. *et al.* Down-regulation of lncRNA CASC2 promotes cell proliferation and metastasis of bladder cancer by activation of the Wnt/ $\beta$ -catenin signaling pathway. *Oncotarget* **8**, 18145–18153 (2017).
79. Xiao, C., Wu, C. H. & Hu, H. Z. LncRNA UCA1 promotes epithelial-mesenchymal transition (EMT) of breast cancer cells via enhancing Wnt/beta-catenin signaling pathway. *European review for medical and pharmacological sciences* **20**, 2819–2824 (2016).

## Acknowledgements

This study was supported by grants from National Key Clinical Specialty Construction Program of China, Chinese Academy of Medical Sciences & Peking Union Medical College (2014BAI08B03), Zhejiang Province Key Project of Science and Technology (2014BAI08B00). The funder of this study had no role in the decisions about the design and conduct of the study; collection, management, analysis, or interpretation of the data; or the preparation, review, or approval of the manuscript. The views expressed in this review are the opinions of the authors.

## Author Contributions

H.Y.L. and K.W. conceived the study idea. J.L. and Y.F.X. performed data mining, and statistical analyses. X.Z. and Z.Z. interpreted results of statistical analyses. K.W. drafted the initial manuscript. H.Y.L. made critical comment and revision for the initial manuscript. H.Y.L. had primary responsibility for the final content. All authors reviewed and approved the final manuscript.

## Additional Information

**Supplementary information** accompanies this paper at <https://doi.org/10.1038/s41598-018-21581-w>.

**Competing Interests:** The authors declare no competing interests.

**Publisher's note:** Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.



**Open Access** This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons license, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons license and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this license, visit <http://creativecommons.org/licenses/by/4.0/>.

© The Author(s) 2018