Prognostic and predictive values of long non-coding RNA \textit{LINC00472} in breast cancer

Yi Shen\textsuperscript{1}, Dionyssios Katsaros\textsuperscript{2}, Lenora W. M. Loo\textsuperscript{1}, Brenda Y. Hernandez\textsuperscript{1}, Clayton Chong\textsuperscript{1}, Emilie Marion Canuto\textsuperscript{2}, Nicoletta Biglia\textsuperscript{3}, Lingeng Lu\textsuperscript{4}, Harvey Risch\textsuperscript{4}, Wen-Ming Chu\textsuperscript{5}, Herbert Yu\textsuperscript{1}

\textsuperscript{1}Cancer Epidemiology Program, University of Hawaii Cancer Center, Honolulu, Hawaii, USA
\textsuperscript{2}Department of Surgical Sciences, Azienda Ospedaliero-Universitaria, Turin, Italy
\textsuperscript{3}Department of Surgical Sciences, University of Turin, Torino, Italy
\textsuperscript{4}Department of Chronic Disease Epidemiology, Yale School of Public Health, New Haven, Connecticut, USA
\textsuperscript{5}Cancer Biology Program, University of Hawaii Cancer Center, Honolulu, Hawaii, USA

Correspondence to: Herbert Yu, e-mail: hyu@cc.hawaii.edu

Keywords: lincRNA, breast, LINC00472

Received: January 05, 2015  Accepted: February 07, 2015  Published: March 25, 2015

ABSTRACT

\textit{LINC00472} is a novel long intergenic non-coding RNA. We evaluated \textit{LINC00472} expression in breast tumor samples using RT-qPCR, performed a meta-analysis of over 20 microarray datasets from the Gene Expression Omnibus (GEO) database, and investigated the effect of \textit{LINC00472} expression on cell proliferation and migration in breast cancer cells transfected with a \textit{LINC00472}-expressing vector. Our qPCR results showed that high \textit{LINC00472} expression was associated with less aggressive breast tumors and more favorable disease outcomes. Patients with high expression of \textit{LINC00472} had significantly reduced risk of relapse and death compared to those with low expression. Patients with high \textit{LINC00472} expression also had better responses to adjuvant chemo- or hormonal therapy than did patients with low expression. Results of meta-analysis on multiple studies from the GEO database were in agreement with the findings of our study. High \textit{LINC00472} was also associated with favorable molecular subtypes, Luminal A or normal-like tumors. Cell culture experiments showed that up-regulation of \textit{LINC00472} expression could suppress breast cancer cell proliferation and migration. Collectively, our clinical and \textit{in vitro} studies suggest that \textit{LINC00472} is a tumor suppressor in breast cancer. Evaluating this long non-coding RNA in breast tumors may have prognostic and predictive value in the clinical management of breast cancer.

INTRODUCTION

Breast cancer is the most common female malignancy, annually accounting for more than a million new diagnoses worldwide [1]. Although breast cancer has been studied extensively for decades, mechanisms of tumor progression have remained largely elusive. Tumor heterogeneity is a major challenge to understanding the disease. Breast cancer varies not only from patient to patient, but also within the same tumor. The disease evolves over time and changes in response to chemo- and hormonal therapies [2–5]. The unstable genome of tumor cells is partially responsible for these changes [3, 6, 7]. Many studies have been conducted to search for disease features that can help to predict disease outcome and treatment response, but no robust tumor markers have yet been identified.

Proteins are considered to be the major molecules carrying out essential biologic actions [8], yet only 2% of the human genome contains the codes for proteins. Most genomic sequences are now understood to be transcribed, though without translation capability [9]. These non-protein coding transcripts are involved in many biologic processes and cellular activities. Recently, long
non-coding RNAs (lncRNAs) were recognized as a new class of non-coding RNAs with important biologic functions [10–16]. LncRNAs exert their actions through interactions with chromatin in the regulation of gene expression [17–20], modulation of epigenetic regulation pre- and post-transcriptionally [21–23], and influences on activities and locations of other functional molecules such as proteins and other RNA species [12, 14, 24–29]. Studies have shown that disruption of lncRNA action occurs in certain diseases including cancer [30–34].

In search of the Gene Expression Omnibus (GEO) database from the NCBI website, we found that a novel long intergenic non-coding RNA, *LINC00472* [35] (Supplementary Figure S1), frequently appeared on top of the gene transcript lists that are associated with tumor grade or disease death in several breast cancer microarray datasets. We searched the literature and found no information on this lincRNA with regard to its biological functions and associations with cancer or other diseases. To determine the clinical relevance of *LINC00472* in breast cancer and to assess its biologic effects on breast cancer cells, we measured *LINC00472* expression in more than 300 breast tumor samples to analyze its association with clinical and pathological features of breast cancer, conducted a meta-analysis on more than 2 dozens independent clinical studies to confirm the findings of our clinical study, and transfected a *LINC00472* expression vector into breast cancer cells to assess the lincRNA’s effects on cell growth and migration. In this report, we describe the findings of *LINC00472* in our clinical study, meta-analysis and *in vitro* experiments.

**RESULTS**

**Clinical study**

In the clinical study of 348 tumor samples, high *LINC00472* expression occurred more often in patients with smaller tumors (*p* < 0.0001), lower tumor grades (*p* < 0.0001), and earlier stage disease (*p* = 0.007) (Table 1). Also, patients with positive hormone receptors had higher *LINC00472* expression compared to those

<table>
<thead>
<tr>
<th>Patient Features</th>
<th>Total No. (%)</th>
<th>Low LincRNA (%)</th>
<th>Mid LincRNA (%)</th>
<th>High LincRNA (%)</th>
<th>P value</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Disease Stage</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Stage 1</td>
<td>346</td>
<td>128 (36.99)</td>
<td>18 (14.06)</td>
<td>39 (30.47)</td>
<td>0.007</td>
</tr>
<tr>
<td>Stage 2</td>
<td>183 (52.89)</td>
<td>58 (16.76)</td>
<td>83 (23.99)</td>
<td>42 (22.95)</td>
<td></td>
</tr>
<tr>
<td>Stage 3 &amp; 4</td>
<td>35 (10.12)</td>
<td>10 (28.57)</td>
<td>19 (54.29)</td>
<td>6 (17.14)</td>
<td></td>
</tr>
<tr>
<td><strong>Tumor Grade</strong></td>
<td>343</td>
<td></td>
<td></td>
<td></td>
<td>&lt; 0.0001</td>
</tr>
<tr>
<td>Grade 1</td>
<td>57 (16.62)</td>
<td>3 (5.26)</td>
<td>27 (47.37)</td>
<td>27 (47.37)</td>
<td></td>
</tr>
<tr>
<td>Grade 2</td>
<td>141 (41.11)</td>
<td>22 (15.60)</td>
<td>80 (56.74)</td>
<td>39 (27.66)</td>
<td></td>
</tr>
<tr>
<td>Grade 3</td>
<td>145 (42.27)</td>
<td>62 (42.76)</td>
<td>62 (42.76)</td>
<td>21 (14.48)</td>
<td></td>
</tr>
<tr>
<td><strong>Histology Type</strong></td>
<td>347</td>
<td></td>
<td></td>
<td></td>
<td>0.043</td>
</tr>
<tr>
<td>Ductal</td>
<td>219 (63.11)</td>
<td>65 (29.68)</td>
<td>103 (47.03)</td>
<td>51 (23.29)</td>
<td></td>
</tr>
<tr>
<td>Lobular</td>
<td>56 (16.14)</td>
<td>6 (10.71)</td>
<td>29 (51.79)</td>
<td>21 (37.50)</td>
<td></td>
</tr>
<tr>
<td>Mix</td>
<td>35 (10.09)</td>
<td>6 (17.14)</td>
<td>21 (60.00)</td>
<td>8 (22.86)</td>
<td></td>
</tr>
<tr>
<td>Others</td>
<td>37 (10.66)</td>
<td>10 (27.03)</td>
<td>20 (54.05)</td>
<td>7 (18.92)</td>
<td></td>
</tr>
<tr>
<td><strong>ER Status</strong></td>
<td>342</td>
<td></td>
<td></td>
<td></td>
<td>&lt; 0.0001</td>
</tr>
<tr>
<td>Positive</td>
<td>222 (64.91)</td>
<td>32 (14.41)</td>
<td>118 (53.15)</td>
<td>72 (32.43)</td>
<td></td>
</tr>
<tr>
<td>Negative</td>
<td>120 (35.09)</td>
<td>54 (45.00)</td>
<td>51 (42.50)</td>
<td>15 (12.50)</td>
<td></td>
</tr>
<tr>
<td><strong>PR Status</strong></td>
<td>341</td>
<td></td>
<td></td>
<td></td>
<td>&lt; 0.0001</td>
</tr>
<tr>
<td>Positive</td>
<td>178 (52.20)</td>
<td>26 (14.61)</td>
<td>98 (55.06)</td>
<td>54 (30.34)</td>
<td></td>
</tr>
<tr>
<td>Negative</td>
<td>163 (47.80)</td>
<td>59 (36.20)</td>
<td>71 (43.56)</td>
<td>33 (20.25)</td>
<td></td>
</tr>
</tbody>
</table>

(Continued)
with negative receptor status ($p < 0.0001$ for ER or PR) (Table 1). Furthermore, patients with high expression had better responses to adjuvant chemotherapy ($p = 0.021$) and hormonal therapy ($p = 0.003$) than those with low expression (Table 1). Finally, survival analysis suggested that patients with high expression had better disease-free ($p < 0.001$) and overall survival ($p = 0.005$) compared to those with low expression (Figures 1A, 1B, and Table 2). Risk reduction in relapse was also observed when disease stage, tumor grade, receptor status and other clinical features of the patients were adjusted in analysis ($p = 0.043$) (Table 2).

### Meta-analysis of GEO data

Twenty-seven datasets containing clinicopathologic information were identified in the GEO database (Supplementary Table S1). In a meta-analysis, we found higher LINC00472 expression to be associated with well-differentiated tumors (low grades) and less aggressive disease (positive ER or PR, negative lymph nodes, luminal A and normal-like molecular subtypes) (Figures 2A, 2B). In GEO, 4 datasets compared gene expression between normal breast tissues and tumors. The comparison showed higher LINC00472 in normal than in tumor tissues (Supplementary Figure S2). Furthermore, 15 datasets in GEO contained information on various survival outcomes, of which 12 included more than 100 patients individually. We performed survival analysis on these 12 datasets after LINC00472 expression was grouped into 3 categories similar to those of our own study. Of the 12 studies, 9 showed significant individual associations between high LINC00472 and favorable survival outcomes (Supplementary Figure S3). Meta-analysis showed both overall survival (Figure 3A) and disease-free survival (Figure 3B) were significantly improved in patients with high LINC00472 compared to those with low expression. Overall, the results of these analyses were in agreement with our own study.

### In vitro experiments

To examine the effect of LINC00472 on breast cancer cells, we analyzed LINC00472 expression in MCF7 and SKBR3, and found that the expression was low in these cell lines. We created an expression vector, pCDH_LINC00472 (Supplementary Figure S4), and transfected the vector into breast cancer cells to increase LINC00472 expression. The transfection was successful based on the GFP signal and RT-qPCR analysis (Figure 4). Cell proliferation assays showed that LINC00472 expression significantly inhibited tumor cell growth (Figure 4). Furthermore, in the cell migration experiments, we found that both MCF7 and SKBR3 cell lines exhibited reduced migration after being transfected with LINC00472 expressing vectors (Figure 5). These in vitro experiments suggest that increased LINC00472 expression may...
Figure 1: Associations of LINC00472 expression with patient survival in Turin study. A. Kaplan-Meier estimates for disease-free survival by LINC00472 expression. B. Kaplan-Meier estimates for overall survival by LINC00472 expression.
suppress tumor cell proliferation and migration, which is consistent with the findings of our clinical studies.

**DISCUSSION**

Our clinical study showed that LINC00472 expression was significantly different by tumor grade, tumor size, disease stage, receptor status and molecular subtype. The expression was also associated with treatment response and survival outcomes, suggesting potential prognostic and predictive implications in clinical management of breast cancer. More importantly, our findings are very consistent across multiple datasets. This consistency is established on the basis of more than 20 independent studies which collectively included thousands of patients. The consistent finding was also based on two different laboratory methods, qPCR and microarray chips. Furthermore, our in vitro experiments on breast cancer cell lines supported the findings of our clinical studies. The overall consistency among multiple clinical studies and between in vivo and in vitro systems underscores the potential significance of LINC00472’s involvement in breast cancer, especially in tumor progression.

The gene encoding LINC00472 is located on chromosome 6q13, and the verified transcript has 2,933 bp (NR_026807.1). The earlier annotation also contained a predicted transcript (XR_241853.1) that is longer than 6,000 bp. We had designed two pairs of RT-qPCR primers to analyze both transcripts, and found that they were highly correlated \((r = 0.76, p < 0.0001)\). Therefore, we chose the one which covers both transcripts for our measurement of expression in the study. Downstream from the LINC00472 gene, there are two microRNA genes, miR30c and miR30a (Supplementary Figure S1), which have been reported to have possible effects on breast cancer invasion, metastasis and patient response to chemotherapy [36–42]. Although LINC00472 has not yet been functionally characterized for cellular activities or molecular processes, the results from multiple independent datasets and different platforms suggest that this lincRNA may behave like a tumor suppressor. Our in vitro experiments also indicates that LINC00472 can suppress cell proliferation and migration.

We notice that the deletion of this gene is rare in the TCGA data [43, 44], and therefore the low expression of LINC00472 in breast cancer is likely due to the down-regulation of expression. The biological mechanisms that control the expression of LINC00472 are still unknown.

Very recently a new clone (RP1–288M22) on chromosome 6q12–13 was submitted to NCBI. This clone expands the LINC00472 transcript to more than 9,000 bp (NR_026807.2) with three additional transcript variants of similar size. Our RT-qPCR primers were able to detect all four transcripts according to the NCBI Blast results (Supplementary Figure S5). Concerning the probes in the microarray chip from which the GEO data were generated, one probe (235771_at) represents the new clone. We compared the expression of this probe with the one used in our study (220324_at) in 15 datasets where data were available, and found a strong correlation of expression between these probes (Supplementary Figure S6), suggesting that our findings of LINC00472 expression in association with breast cancer based on 220324_at be applicable to the recent submission or renewal of LINC00472.

Long non-coding RNAs are suspected to exert their functions through interacting with ribonucleoprotein
Figure 2: Meta-analysis of associations between LINC00472 expression and clinicopathological features of breast cancer. Summarized odds ratios were estimated using the random-effect model, and the odds ratio in each study was weighted with the variance of probe values (inverse-variance weighted method). A. Lower LINC00472 expression associated with high tumor grade (OR = 2.81; 95% CI: 2.24–3.53), positive lymph node (OR = 1.51; 95% CI: 1.22–1.87), or molecular subtypes of luminal B, Her2 positive and basal-like (OR = 3.80; 95% CI: 2.38–6.07).

(Continued)
Figure 2 (Continued): B. Higher LINC00472 expression associated with ER positive tumors (OR = 0.44; 95% CI: 0.35–0.57), PR positive tumors (OR = 0.44; 95% CI: 0.35–0.55), or Her2 negative tumors (OR = 2.62; 95% CI: 1.76–3.90).
Figure 3: Meta-analysis of associations between LINC00472 expression and patient survival. A. Overall survival: HR = 0.91 (95% CI: 0.57–1.45) between mid and low LINC00472 expression (Upper Figure), and HR = 0.41 (95% CI: 0.22–0.78) between high and low LINC00472 expression (Lower Figure). B. Disease-free Survival: HR = 0.69 (95% CI: 0.55–0.88) between mid and low LINC00472 expression (Upper Figure), and HR = 0.43 (95% CI: 0.31–0.59) between high and low LINC00472 expression (Lower Figure).
Figure 4: Effect of LINC00472 expression on breast cancer cell proliferation. A. GFP fluorescence images in MCF7 and SKBR3 cells transfected with pCDH or pCDH_LINC00472 vectors. B. RT-qPCR results of LINC00472 expression in MCF7 and SKBR3 cells transfected with pCDH or pCDH_LINC00472 or mock transfected. C. Cell growth inhibition by LINC00472 in MCF7 cells. Twenty hours after seeding in 96-well plates, cells were transiently transfected with pCDH or pCDH_LINC00472 vectors, and kept in culture for up to 96 hours. Absorbance at 450nm of each well, which was directly proportional to the number of living cells in the well, was measured by SpectraMax M3 Multimode Plate Reader. The y axis showed the relative absorbance of corresponding wells from different days compared to that from day 0. Error bars represent SEM, n = 15. P values were determined by the Mann-Whitney U test. D. Inhibition of cell growth by LINC00472 in SKBR3 cells. The experiment and analysis are same as to those in C.

(RNP), but experimentally elucidating the actions of RNP complexes is difficult and time-consuming. To date, only a few RNP complexes have been characterized experimentally. Thus, computational prediction may be a valuable alternative to help predict the possible actions of LINC00472. We used a sequence-based method, catRAPID omics, to search for the binding proteins of LINC00472. Since catRAPID omics uses an algorithm that combines multiple features in addition to primary structure, its prediction is considered to be relatively reliable. Importantly this method has been used previously for testing several long non-coding RNAs, such as Xist, the long non-coding RNA X-inactive-specific transcript, and the results of prediction have been consistent with experimental findings. The LINC00472-binding proteome returned from the catRAPID omics analysis contained a total of 579 proteins, from which 127 showed high interaction propensities (Interaction Strength > 50% and Discriminative Power > 75%, Z-score > 1.0) (Supplementary Figure S7). We uploaded these proteins
onto the Ingenuity Pathway Analysis (IPA) System for a core analysis using the Ingenuity Knowledge Base as a reference set. The IPA analysis suggested that proteins interacting with LINC00472 may involve several diseases and disorders, of which cancer is on top of the list (Supplementary Figure S7). IPA analysis also indicated that these proteins may be functionally involved in RNA post-transcriptional modification and protein synthesis and associated with 3 signaling networks where multiple biomarkers are known for breast cancer diagnosis, prognosis and treatment prediction (Supplementary Figures S8A–S8C). Knowing the potential RNA binding-proteins may allow us not only to predict the possible targets of the lincRNA, but also to design additional experiments to assess its biologic functions.

Microarray technologies have substantially enhanced the search of biomarkers for cancer prognosis, but long non-coding RNAs in data generated by microarray analysis have not been well interrogated. Using GEO2R, an interactive web tool in GEO, we analyzed the relationships of LINC00472 expression and breast cancer characteristics in multiple GEO datasets generated from the most recent microarray chips, the Affymetrix Human Genome U133 plus 2.0 array and the U133A array. Our finding of a consistent association between LINC00472 expression and breast cancer survival among various datasets supports the validity of this association and indicates such public databases invaluable to biomedical research. The consistency in the associations with clinicopathologic variables further strengthens the finding of our study.
Overall, our investigation was based on multiple studies from diverse patient populations, and involved a large number of patients. We also demonstrated the results with different methods of lncRNA analysis (microarray and qPCR). These differences in patient populations and lab techniques lend strong support to our findings of LINC00472 in breast cancer. However, at present, very little is known about the exact functions of this long non-coding RNA. Therefore, we have not discussed the possible biologic mechanisms that may explain why LINC00472 behaves like a tumor suppressor. Whether our findings reflect direct effects of LINC00472 on breast cancer or indirect actions via other molecules is unclear. More work is needed to elucidate the function of LINC00472 and its role in breast tumorigenesis.

MATERIALS AND METHODS

The role of LINC00472 in clinical studies was investigated through a) a clinical study in which we analyzed LINC00472 expression in 348 tumors using RT-qPCR, b) meta-analysis of microarray data deposited in the GEO database, and c) manipulation of LINC00472 expression in breast cancer cell lines.

Clinical study (Turin_Study)

Breast cancer patients

A clinical study of breast cancer was conducted in the University Hospital at University of Turin, in Italy, between January 1998 and July 1999. The study was approved by the university’s ethics review committee. During the study, 348 breast cancer patients were enrolled and provided informed consent. All study patients underwent surgical resection for breast cancer. Average age at surgery was 57 years (range: 23–84 years). Of the patients enrolled, 302 had follow-up information available through February 2007. Median follow-up was 86 months (range: 8–108 months). During follow-up, 81 patients developed recurrent or metastatic disease, and 55 of them died from the disease. A total of 60 died by the end of follow-up. Among the patients enrolled, 36.4% had Stage I disease (TNM), 53.4% had Stage II, and 10.3% had Stage III or IV. Patients diagnosed with Grade 1 (well differentiated) tumors comprised 16.6%, Grade 2 41.1%, and Grade 3 42.3%. Ductal carcinoma accounted for the majority of cases (63.1%), followed by lobular carcinoma (16.1%), other specific types (10.7%), and mixed histologic types (10.1%). Two hundred one patients (58.1%) had tumors smaller than 2 cm, 120 (34.7%) had tumors between 2 and 5 cm, and 25 (7.2%) had tumors greater than 5 cm. One hundred sixty patients (46.8%) had lymph node-positive tumors. Sixty-five percent of the patients had ER positive tumors, and 52.2% had PR positive tumors (at 10% cutoff for receptor positivity).

Of the 348 patients, 303 received adjuvant therapy after surgery. Of those treated, 119 (34.2%) had chemotherapy, 77 (22.1%) had hormonal therapy, and 107 (30.8%) received both. The chemotherapy protocols administered included CMF (cyclophosphamide, methotrexate, 5 fluorouracil; 600/60/600 mg/mq every three weeks for six cycles), CEF (cyclophosphamide, epirubicin, 5 fluorouracil; 600/60/600 mg/mq every three weeks for six cycles), EPI-TAX (epirubicin-paclitaxel; 90/175 mg/mq every three weeks for six cycles), EPI-VNB (epirubicin-vinorelbine; 90/30 mg/mq every three weeks for six cycles), DTX-EPI-VNB (docetaxel-epirubicin-vinorelbine; 75/90/30 mg/mq every three weeks for six cycles), and TXT-EPI-VNB (paclitaxel-epirubicin-vinorelbine; 175/90/30 mg/mq every three weeks for cycles). Tamoxifen was the only agent used for endocrine therapy at that time. The dose was one 20 mg tablet per day for five years or until disease progression or intolerable toxicity.

Analysis of LINC00472 expression

Fresh tumor samples collected from patients during surgery were snap-frozen in liquid nitrogen immediately after resection and stored at -80°C until analysis. All tissue samples were examined by pathologists to confirm at least 80% tumor content. The tissue samples (~30 mg each) were homogenized with ceramic beads in PowerLyzer (MO BIO), and processed to extract total RNA using Allprep DNA/RNA Kit (Qiagen). The RNA samples were treated with RNase-free DNase and quantified using a spectrophotometer.

PCR primers were designed using the sequence NR_026807.1 (Supplementary Figure S1) and the synthesis was done by IDT (San Diego, CA). Total RNA (1 μg) was reverse transcribed using the cDNA Reverse Transcription Kit (LifeTech); RT-qPCR was performed in triplicate using SYBR Select Master Mix (LifeTech). In PCR reaction (10 μl), cDNA template (0.5 μl) was mixed with 200 nM primers and 5 μl SYBR PCR master mix. PCR conditions were incubation at 50°C for 2 min to activate UDG, 95°C for 2 min to activate Taq polymerase, and 40 cycles of 95°C for 15 s and 60°C for 1 min. LINC00472 measurement was normalized to GAPDH using the formula described in Statistical Analysis.

Meta-analysis of GEO data

Gene expression data generated from the Affymetrix Human Genome U133 plus 2.0 array and U133A array were selected from the GEO database. These datasets include 4, 628 breast cancer samples (Supplementary Table S1) and 193 normal breast tissues (Supplementary Figure S2). Expression data from probe 220324_at (which targets LINC00472) were used for meta-analysis, in which normalized expression was dichotomized using study-specific median expression as cutoff to define “LINC00472_higher” at or above median versus “LINC00472_lower” below median. Clinical and pathologic variables were also dichotomized. For each variable, summary odds ratios
and their 95% confidence intervals were estimated using the inverse variance weighted method. Because the meta-analysis involved expression data assessed by different methods, we used the random-effects model. Forest plots were used to present the results. The meta-analysis was conducted with the use of Review Manager (Revman Version 5.3, Copenhagen, Denmark). Cochran $\chi^2$ test and $I^2$ statistic were used to assess the heterogeneity among the studies involved. For datasets with more than 100 patients, Kaplan-Meier survival analyses were performed on individual studies (Supplementary Figure S3).

**In vitro experiments**

**Preparation of LINC00472 construct**

A LINC00472 transcript (2933 bp, NR_026807.1) was assembled using the EST clones, EHS1001–
207275390, EHS1001–207498495, EHS1001–207533792,
EHS1001–207590772, EHS1001–210281579, and
EHS1001–211231922 (Thermo Scientific Open
Biosystems). Restriction digestion sites for NheI and
BamHI were incorporated into each site of the LINC00472
transcript for cloning and verification. The entire
sequence was inserted into a lentiviral expression
vector, pCDH-EF1-MCS-pA-PGK-copGFP-T2A-Puro
(System Biosciences). The construction was done by
Science Exchange (Supplementary Figure S4). The clone
authenticity was confirmed by restriction enzyme digestion
(Supplementary Figure S4) and sequencing analysis.

**Cell culture experiments**

Breast cancer cell lines, SKBR3 and MCF-7, were a kind gift from Jun Panee (JABSOM, University of
Hawaii) and were grown in DMEM medium containing
10% FBS and 100 units/ml of penicillin/streptomycin
(Pen/Strep). We transfected the cells with the LINC00472
vector (pCDH_LINC00472), along with mock transfection
and transfection of a control vector (pCDH). After 24 h
incubation, cells were examined for GFP fluorescence; the
incubation, cells were examined for GFP fluorescence; the
photomicrograph of the entire wound area was taken
photomicrograph of the entire wound area was taken
using the IX71 inverted microscope with 4X objective
using the IX71 inverted microscope with 4X objective
lens. Using a detection mask, the absorbance at 590 nm
lens. Using a detection mask, the absorbance at 590 nm
wavelength in each well, which was directly proportional
wavelength in each well, which was directly proportional
to the number of cells that migrated into the ‘wound’ area,
to the number of cells that migrated into the ‘wound’ area,
was measured by a spectrophotometer (SpectraMax M3
was measured by a spectrophotometer (SpectraMax M3
Multimode Plate Reader).

**Statistical analysis**

LINC00472 expression was calculated as an expression index (EI), using the formula $1,000 \times 2^{-\Delta Ct}$,
where $\Delta Ct = Ct (LINC00472) – Ct (GAPDH)$. For data
analysis, EI was analyzed as a categorical variable with 3
d Ordinal levels, low (< 25 percentile), Intermediate (25–75
percentile, labeled as “Mid” in Table 1), and high
eI was analyzed as a categorical variable with 3
percentile, labeled as “Mid” in Table 1), and high
expression. The survival time for either overall
expression. The survival time for either overall
disease or disease-free was calculated as the time from surgery
disease or disease-free was calculated as the time from surgery
until the occurrence of death and relapse, respectively. The
Mann-Whitney U test was used for comparing differences
Mann-Whitney U test was used for comparing differences
in cell counts and migration. Spearman correlation
coefficients were calculated for correlation analysis. All
statistical tests were two-sided, and a $p$ value less than 0.05
was considered as statistical significance.

**REFERENCES**

1. Ferlay J, Ervik M, Dikshit R, Eser S, Mathers C,
Rebelo M, Parkin DM, Forman D, Bray F. GLOBOCAN
2012 v1.0, Cancer Incidence and Mortality Worldwide:
IARC CancerBase No. 11 [Internet]. International Agency

2. Bertos NR, Park M. Breast cancer - one term, many
entities? The Journal of clinical investigation. 2011;
121:3789–3796.

3. Burrell RA, McGranahan N, Bartek J, Swanton C. The
causes and consequences of genetic heterogeneity in cancer


