Prognostic and predictive values of long non-coding RNA *LINC00472* in breast cancer

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

[1] Cancer Epidemiology Program, University of Hawaii Cancer Center, Honolulu, Hawaii, USA

[2] Department of Surgical Sciences, Azienda Ospedaliero-Universitaria, Turin, Italy

[3] Department of Surgical Sciences, University of Turin, Torino, Italy

[4] Department of Chronic Disease Epidemiology, Yale School of Public Health, New Haven, Connecticut, USA

[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

***LINC00472* is a novel long intergenic non-coding RNA. We evaluated *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 *LINC00472* expression on cell proliferation and migration in breast cancer cells transfected with a *LINC00472*-expressing vector. Our qPCR results showed that high *LINC00472* expression was associated with less aggressive breast tumors and more favorable disease outcomes. Patients with high expression of *LINC00472* had significantly reduced risk of relapse and death compared to those with low expression. Patients with high *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 *LINC00472* was also associated with favorable molecular subtypes, Luminal A or normal-like tumors. Cell culture experiments showed that up-regulation of *LINC00472* expression could suppress breast cancer cell proliferation and migration. Collectively, our clinical and *in vitro* studies suggest that *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 1: Associations of *LINC00472* with Clinicopathological Factors in Turin Study

Patient Features	Total No. (%)	Low LincRNA No. (%)	Mid LincRNA No. (%)	High LincRNA No. (%)	*P* value
Disease Stage	346				0.007
Stage 1	128 (36.99)	18 (14.06)	71 (55.47)	39 (30.47)	
Stage 2	183 (52.89)	58 (16.76)	83 (23.99)	42 (22.95)	
Stage 3 & 4	35 (10.12)	10 (28.57)	19 (54.29)	6 (17.14)	
Tumor Grade	343				< 0.0001
Grade 1	57 (16.62)	3 (5.26)	27 (47.37)	27 (47.37)	
Grade 2	141 (41.11)	22 (15.60)	80 (56.74)	39 (27.66)	
Grade 3	145 (42.27)	62 (42.76)	62 (42.76)	21 (14.48)	
Histology Type	347				0.043
Ductal	219 (63.11)	65 (29.68)	103 (47.03)	51 (23.29)	
Lobular	56 (16.14)	6 (10.71)	29 (51.79)	21 (37.50)	
Mix	35 (10.09)	6 (17.14)	21 (60.00)	8 (22.86)	
Others	37 (10.66)	10 (27.03)	20 (54.05)	7 (18.92)	
ER Status	342				< 0.0001
Positive	222 (64.91)	32 (14.41)	118 (53.15)	72 (32.43)	
Negative	120 (35.09)	54 (45.00)	51 (42.50)	15 (12.50)	
PR Status	341				< 0.0001
Positive	178 (52.20)	26 (14.61)	98 (55.06)	54 (30.34)	
Negative	163 (47.80)	59 (36.20)	71 (43.56)	33 (20.25)	

(*Continued*)

Patient Features	Total No. (%)	Low LincRNA No. (%)	Mid LincRNA No. (%)	High LincRNA No. (%)	*P* value
Nodal Status	347				0.475
Positive	160 (46.11)	45 (28.13)	77 (48.13)	38 (23.75)	
Negative	187 (53.89)	42 (22.46)	96 (51.34)	49 (26.20)	
Tumor Size	346				< 0.0001
T1	201 (58.09)	31 (15.42)	106 (52.74)	64 (31.84)	
T2	120 (34.68)	46 (38.33)	54 (45.00)	20 (16.67)	
T3/T4	25 (7.23)	9 (36.00)	13 (52.00)	3 (12.00)	
Adjuvant Endo-therapy	173				0.003
Complete Response	135 (78.03)	17 (12.59)	71 (52.59)	47 (34.81)	
No complete response	38 (21.97)	13 (34.21)	19 (50)	6 (15.79)	
Adjuvant Chemo-therapy	219				0.021
Complete Response	158 (72.15)	43 (27.22)	75 (47.47)	40 (25.32)	
No complete response	61 (27.85)	25 (40.98)	30 (49.18)	6 (9.84)	

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

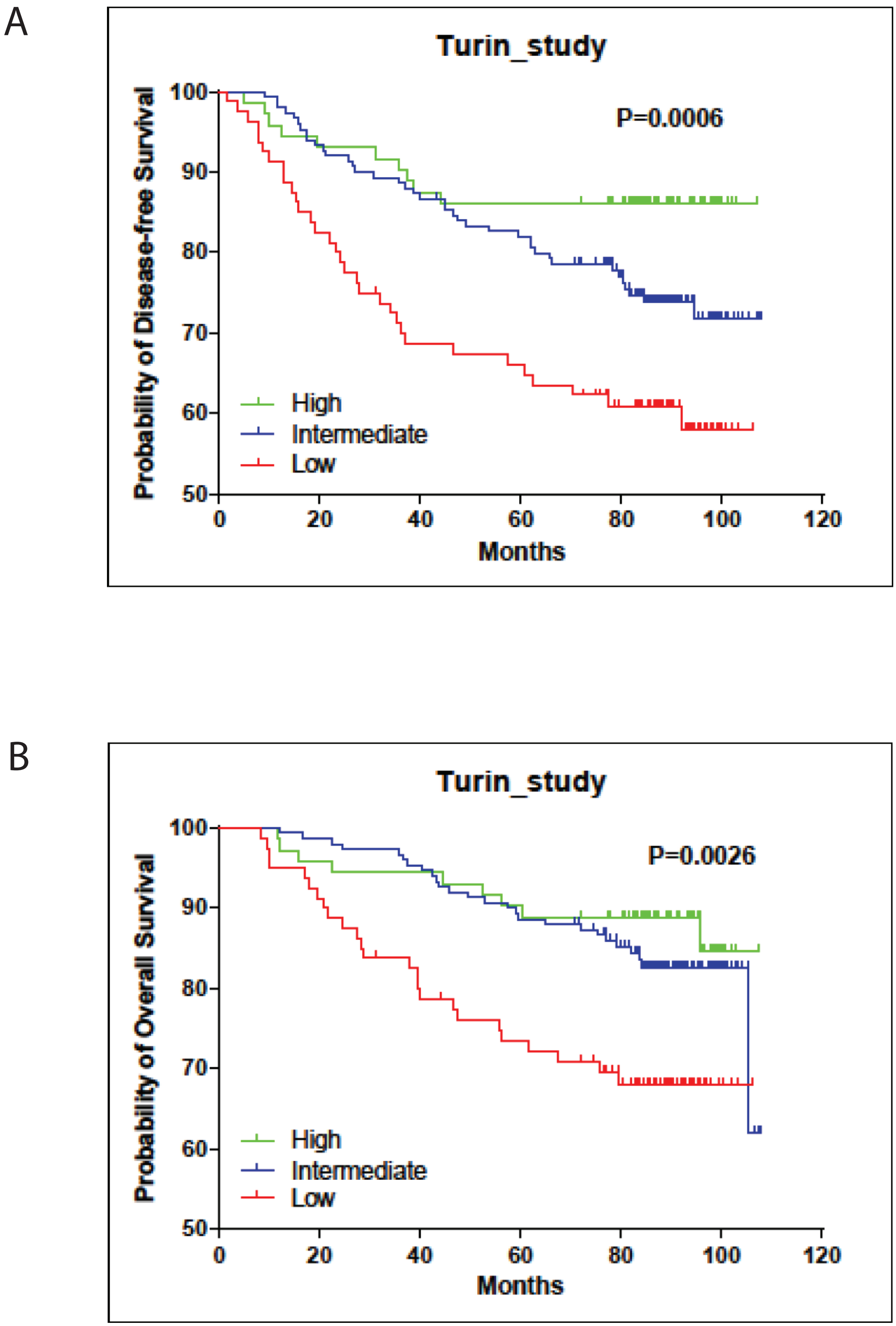


Figure 1: Associations of *LINC00472* expression with patient survival in turin study. A. Kaplan-Meier estimates for disease-free survival by *LINC00472* expression. **B.** Kaplan-Meier estimates for overall survival by *LINC00472* expression.

Table 2: Associations of *LINC00472* with Breast Cancer Survival in Turin Study

Unadjusted Cox Regression Model

LincRNA	HR[1] for relapse	95% CI[2]	*p* value	HR for death	95% CI	*p* value
Low	1			1		
Mid	0.55	0.34–0.88	0.012	0.47	0.27–0.81	0.006
High	0.28	0.14–0.58	< 0.001	0.34	0.16–0.72	0.005
Continuous	0.54	0.39–0.74	< 0.001	0.55	0.38–0.80	0.002

Adjusted Cox Regression Model*

LincRNA	HR for relapse	95% CI	*p* value	HR for death	95% CI	*p* value
Low	1			1		
Mid	0.82	0.48–1.39	0.45	0.62	0.33–1.16	0.134
High	0.44	0.20–0.97	0.043	0.51	0.21–1.22	0.129
Continuous	0.70	0.48–1.00	0.050	0.69	0.45–1.07	0.098

*Adjusted for age, stage, grade, histology, ER, PR, nodal status and adjuvant treatment.

[1]HR: Hazards Ratio

[2]CI: Confidence Interval

suppress tumor cell proliferation and migration, which is consistent with the findings of our clinical studies.

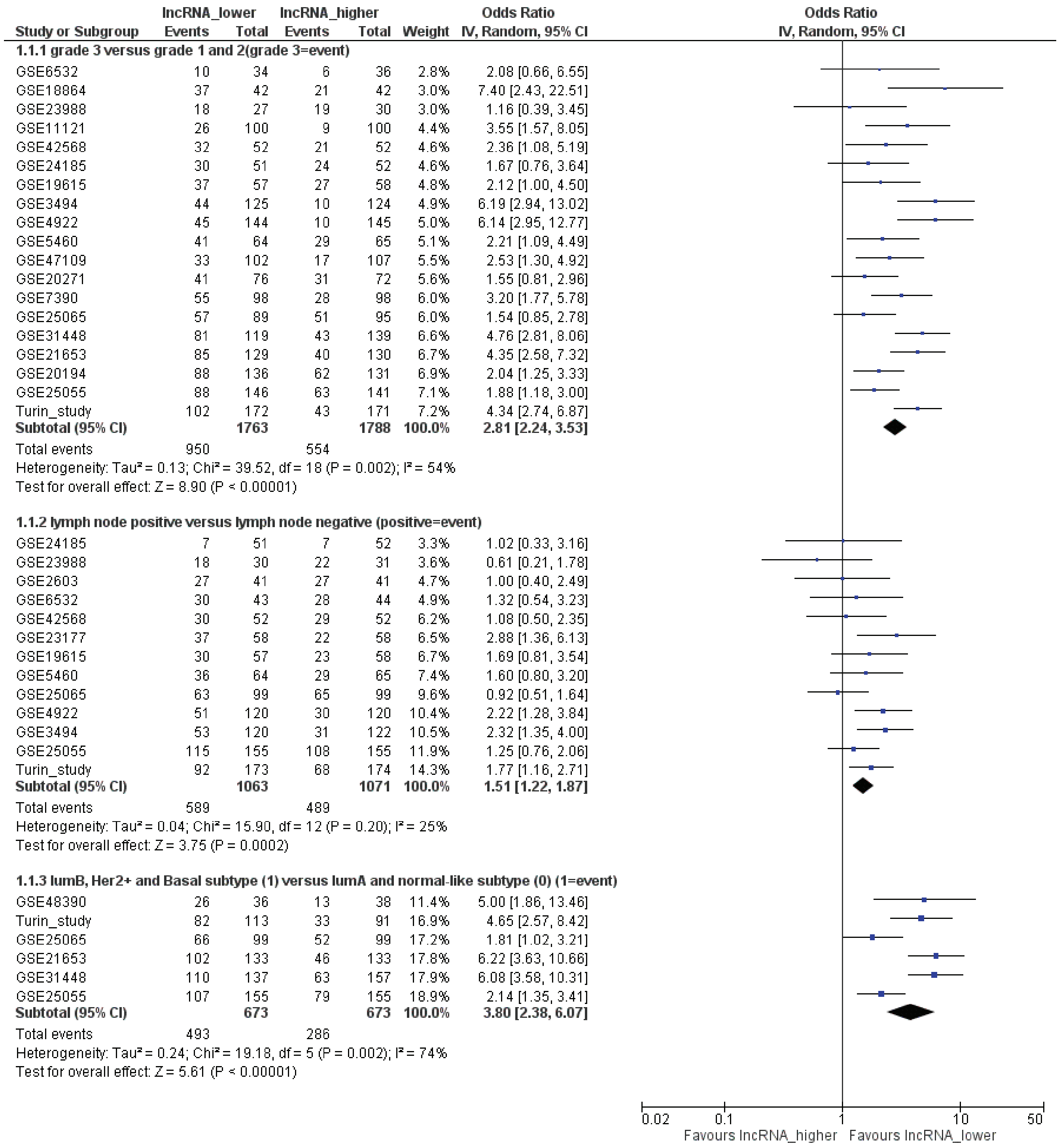
DISCUSSION

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*.

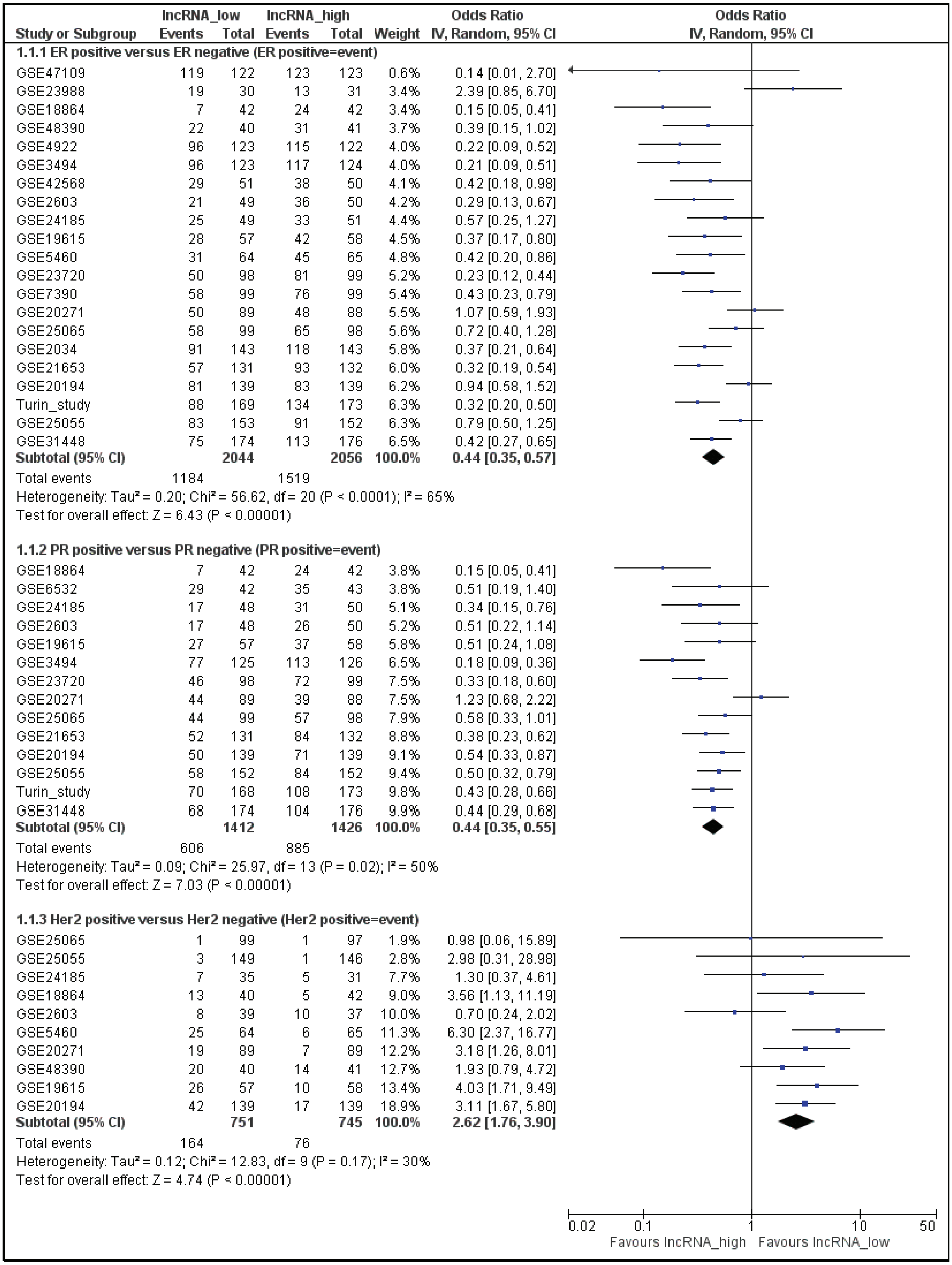
Long non-coding RNAs are suspected to exert their functions through interacting with ribonucleoprotein



Study or Subgroup	lncRNA_lower Events	lncRNA_lower Total	lncRNA_higher Events	lncRNA_higher Total	Weight	Odds Ratio IV, Random, 95% CI
1.1.1 grade 3 versus grade 1 and 2(grade 3=event)						
GSE6532	10	34	6	36	2.8%	2.08 [0.66, 6.55]
GSE18864	37	42	21	42	3.0%	7.40 [2.43, 22.51]
GSE23988	18	27	19	30	3.0%	1.16 [0.39, 3.45]
GSE11121	26	100	9	100	4.4%	3.55 [1.57, 8.05]
GSE42568	32	52	21	52	4.6%	2.36 [1.08, 5.19]
GSE24185	30	51	24	52	4.6%	1.67 [0.76, 3.64]
GSE19615	37	57	27	58	4.8%	2.12 [1.00, 4.50]
GSE3494	44	125	10	124	4.9%	6.19 [2.94, 13.02]
GSE4922	45	144	10	145	5.0%	6.14 [2.95, 12.77]
GSE5460	41	64	29	65	5.1%	2.21 [1.09, 4.49]
GSE47109	33	102	17	107	5.5%	2.53 [1.30, 4.92]
GSE20271	41	76	31	72	5.6%	1.55 [0.81, 2.96]
GSE7390	55	98	28	98	6.0%	3.20 [1.77, 5.78]
GSE25065	57	89	51	95	6.0%	1.54 [0.85, 2.78]
GSE31448	81	119	43	139	6.6%	4.76 [2.81, 8.06]
GSE21653	85	129	40	130	6.7%	4.35 [2.58, 7.32]
GSE20194	88	136	62	131	6.9%	2.04 [1.25, 3.33]
GSE25055	88	146	63	141	7.1%	1.88 [1.18, 3.00]
Turin_study	102	172	43	171	7.2%	4.34 [2.74, 6.87]
Subtotal (95% CI)		**1763**		**1788**	**100.0%**	**2.81 [2.24, 3.53]**
Total events	950		554			
Heterogeneity: Tau² = 0.13; Chi² = 39.52, df = 18 (P = 0.002); I² = 54%						
Test for overall effect: Z = 8.90 (P < 0.00001)						
1.1.2 lymph node positive versus lymph node negative (positive=event)						
GSE24185	7	51	7	52	3.3%	1.02 [0.33, 3.16]
GSE23988	18	30	22	31	3.6%	0.61 [0.21, 1.78]
GSE2603	27	41	27	41	4.7%	1.00 [0.40, 2.49]
GSE6532	30	43	28	44	4.9%	1.32 [0.54, 3.23]
GSE42568	30	52	29	52	6.2%	1.08 [0.50, 2.35]
GSE23177	37	58	22	58	6.5%	2.88 [1.36, 6.13]
GSE19615	30	57	23	58	6.7%	1.69 [0.81, 3.54]
GSE5460	36	64	29	65	7.4%	1.60 [0.80, 3.20]
GSE25065	63	99	65	99	9.6%	0.92 [0.51, 1.64]
GSE4922	51	120	30	120	10.4%	2.22 [1.28, 3.84]
GSE3494	53	120	31	122	10.5%	2.32 [1.35, 4.00]
GSE25055	115	155	108	155	11.9%	1.25 [0.76, 2.06]
Turin_study	92	173	68	174	14.3%	1.77 [1.16, 2.71]
Subtotal (95% CI)		**1063**		**1071**	**100.0%**	**1.51 [1.22, 1.87]**
Total events	589		489			
Heterogeneity: Tau² = 0.04; Chi² = 15.90, df = 12 (P = 0.20); I² = 25%						
Test for overall effect: Z = 3.75 (P = 0.0002)						
1.1.3 lumB, Her2+ and Basal subtype (1) versus lumA and normal-like subtype (0) (1=event)						
GSE48390	26	36	13	38	11.4%	5.00 [1.86, 13.46]
Turin_study	82	113	33	91	16.9%	4.65 [2.57, 8.42]
GSE25065	66	99	52	99	17.2%	1.81 [1.02, 3.21]
GSE21653	102	133	46	133	17.8%	6.22 [3.63, 10.66]
GSE31448	110	137	63	157	17.9%	6.08 [3.58, 10.31]
GSE25055	107	155	79	155	18.9%	2.14 [1.35, 3.41]
Subtotal (95% CI)		**673**		**673**	**100.0%**	**3.80 [2.38, 6.07]**
Total events	493		286			
Heterogeneity: Tau² = 0.24; Chi² = 19.18, df = 5 (P = 0.002); I² = 74%						
Test for overall effect: Z = 5.61 (P < 0.00001)						

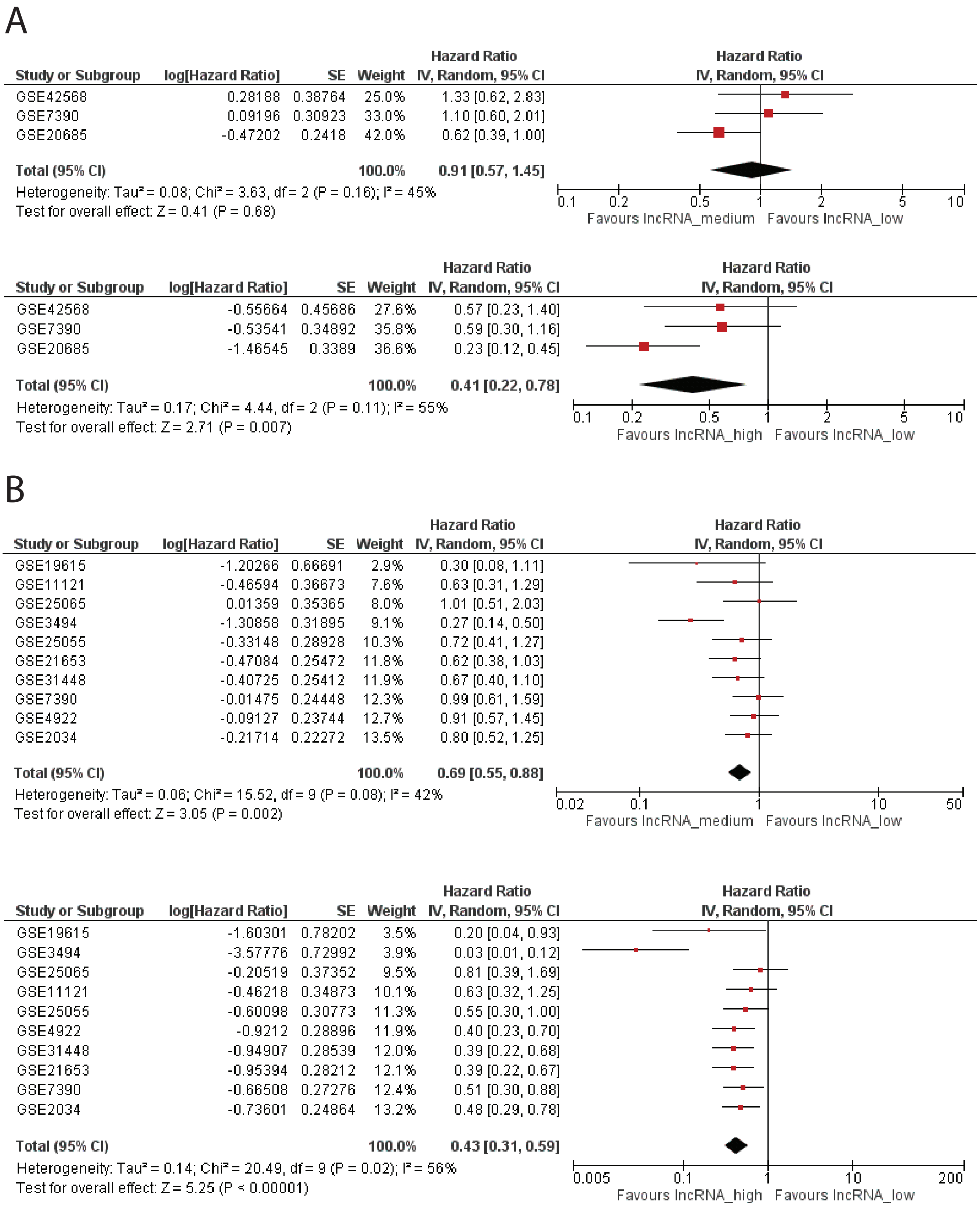
Figure 2: Meta-analysis of associations between *LINC00472* expression and clinicopathological features of breast cancer. Summarized odds ratios were estimated using the random-effect model, and the odds ratio in each study was weighted with the variance of probe values (inverse-variance weighted method). **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*)



Study or Subgroup	lncRNA_low Events	lncRNA_low Total	lncRNA_high Events	lncRNA_high Total	Weight	Odds Ratio IV, Random, 95% CI
1.1.1 ER positive versus ER negative (ER positive=event)						
GSE47109	119	122	123	123	0.6%	0.14 [0.01, 2.70]
GSE23988	19	30	13	31	3.4%	2.39 [0.85, 6.70]
GSE18864	7	42	24	42	3.4%	0.15 [0.05, 0.41]
GSE48390	22	40	31	41	3.7%	0.39 [0.15, 1.02]
GSE4922	96	123	115	122	4.0%	0.22 [0.09, 0.52]
GSE3494	96	123	117	124	4.0%	0.21 [0.09, 0.51]
GSE42568	29	51	38	50	4.1%	0.42 [0.18, 0.98]
GSE2603	21	49	36	50	4.2%	0.29 [0.13, 0.67]
GSE24185	25	49	33	51	4.4%	0.57 [0.25, 1.27]
GSE19615	28	57	42	58	4.5%	0.37 [0.17, 0.80]
GSE5460	31	64	45	65	4.8%	0.42 [0.20, 0.86]
GSE23720	50	98	81	99	5.2%	0.23 [0.12, 0.44]
GSE7390	58	99	76	99	5.4%	0.43 [0.23, 0.79]
GSE20271	50	89	48	88	5.5%	1.07 [0.59, 1.93]
GSE25065	58	99	65	98	5.6%	0.72 [0.40, 1.28]
GSE2034	91	143	118	143	5.8%	0.37 [0.21, 0.64]
GSE21653	57	131	93	132	6.0%	0.32 [0.19, 0.54]
GSE20194	81	139	83	139	6.2%	0.94 [0.58, 1.52]
Turin_study	88	169	134	173	6.3%	0.32 [0.20, 0.50]
GSE25055	83	153	91	152	6.3%	0.79 [0.50, 1.25]
GSE31448	75	174	113	176	6.5%	0.42 [0.27, 0.65]
Subtotal (95% CI)		**2044**		**2056**	**100.0%**	**0.44 [0.35, 0.57]**
Total events	1184		1519			
Heterogeneity: Tau² = 0.20; Chi² = 56.62, df = 20 (P < 0.0001); I² = 65%						
Test for overall effect: Z = 6.43 (P < 0.00001)						
1.1.2 PR positive versus PR negative (PR positive=event)						
GSE18864	7	42	24	42	3.8%	0.15 [0.05, 0.41]
GSE6532	29	42	35	43	3.8%	0.51 [0.19, 1.40]
GSE24185	17	48	31	50	5.1%	0.34 [0.15, 0.76]
GSE2603	17	48	26	50	5.2%	0.51 [0.22, 1.14]
GSE19615	27	57	37	58	5.8%	0.51 [0.24, 1.08]
GSE3494	77	125	113	126	6.5%	0.18 [0.09, 0.36]
GSE23720	46	98	72	99	7.5%	0.33 [0.18, 0.60]
GSE20271	44	89	39	88	7.5%	1.23 [0.68, 2.22]
GSE25065	44	99	57	98	7.9%	0.58 [0.33, 1.01]
GSE21653	52	131	84	132	8.8%	0.38 [0.23, 0.62]
GSE20194	50	139	71	139	9.1%	0.54 [0.33, 0.87]
GSE25055	58	152	84	152	9.4%	0.50 [0.32, 0.79]
Turin_study	70	168	108	173	9.8%	0.43 [0.28, 0.66]
GSE31448	68	174	104	176	9.9%	0.44 [0.29, 0.68]
Subtotal (95% CI)		**1412**		**1426**	**100.0%**	**0.44 [0.35, 0.55]**
Total events	606		885			
Heterogeneity: Tau² = 0.09; Chi² = 25.97, df = 13 (P = 0.02); I² = 50%						
Test for overall effect: Z = 7.03 (P < 0.00001)						
1.1.3 Her2 positive versus Her2 negative (Her2 positive=event)						
GSE25065	1	99	1	97	1.9%	0.98 [0.06, 15.89]
GSE25055	3	149	1	146	2.8%	2.98 [0.31, 28.98]
GSE24185	7	35	5	31	7.7%	1.30 [0.37, 4.61]
GSE18864	13	40	5	42	9.0%	3.56 [1.13, 11.19]
GSE2603	8	39	10	37	10.0%	0.70 [0.24, 2.02]
GSE5460	25	64	6	65	11.3%	6.30 [2.37, 16.77]
GSE20271	19	89	7	89	12.2%	3.18 [1.26, 8.01]
GSE48390	20	40	14	41	12.7%	1.93 [0.79, 4.72]
GSE19615	26	57	10	58	13.4%	4.03 [1.71, 9.49]
GSE20194	42	139	17	139	18.9%	3.11 [1.67, 5.80]
Subtotal (95% CI)		**751**		**745**	**100.0%**	**2.62 [1.76, 3.90]**
Total events	164		76			
Heterogeneity: Tau² = 0.12; Chi² = 12.83, df = 9 (P = 0.17); I² = 30%						
Test for overall effect: Z = 4.74 (P < 0.00001)						

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).



Study or Subgroup	log[Hazard Ratio]	SE	Weight	Hazard Ratio IV, Random, 95% CI
GSE42568	0.28188	0.38764	25.0%	1.33 [0.62, 2.83]
GSE7390	0.09196	0.30923	33.0%	1.10 [0.60, 2.01]
GSE20685	-0.47202	0.2418	42.0%	0.62 [0.39, 1.00]
Total (95% CI)			**100.0%**	**0.91 [0.57, 1.45]**

Heterogeneity: Tau² = 0.08; Chi² = 3.63, df = 2 (P = 0.16); I² = 45%
Test for overall effect: Z = 0.41 (P = 0.68)

Study or Subgroup	log[Hazard Ratio]	SE	Weight	Hazard Ratio IV, Random, 95% CI
GSE42568	-0.55664	0.45686	27.6%	0.57 [0.23, 1.40]
GSE7390	-0.53541	0.34892	35.8%	0.59 [0.30, 1.16]
GSE20685	-1.46545	0.3389	36.6%	0.23 [0.12, 0.45]
Total (95% CI)			**100.0%**	**0.41 [0.22, 0.78]**

Heterogeneity: Tau² = 0.17; Chi² = 4.44, df = 2 (P = 0.11); I² = 55%
Test for overall effect: Z = 2.71 (P = 0.007)

Study or Subgroup	log[Hazard Ratio]	SE	Weight	Hazard Ratio IV, Random, 95% CI
GSE19615	-1.20266	0.66691	2.9%	0.30 [0.08, 1.11]
GSE11121	-0.46594	0.36673	7.6%	0.63 [0.31, 1.29]
GSE25065	0.01359	0.35365	8.0%	1.01 [0.51, 2.03]
GSE3494	-1.30858	0.31895	9.1%	0.27 [0.14, 0.50]
GSE25055	-0.33148	0.28928	10.3%	0.72 [0.41, 1.27]
GSE21653	-0.47084	0.25472	11.8%	0.62 [0.38, 1.03]
GSE31448	-0.40725	0.25412	11.9%	0.67 [0.40, 1.10]
GSE7390	-0.01475	0.24448	12.3%	0.99 [0.61, 1.59]
GSE4922	-0.09127	0.23744	12.7%	0.91 [0.57, 1.45]
GSE2034	-0.21714	0.22272	13.5%	0.80 [0.52, 1.25]
Total (95% CI)			**100.0%**	**0.69 [0.55, 0.88]**

Heterogeneity: Tau² = 0.06; Chi² = 15.52, df = 9 (P = 0.08); I² = 42%
Test for overall effect: Z = 3.05 (P = 0.002)

Study or Subgroup	log[Hazard Ratio]	SE	Weight	Hazard Ratio IV, Random, 95% CI
GSE19615	-1.60301	0.78202	3.5%	0.20 [0.04, 0.93]
GSE3494	-3.57776	0.72992	3.9%	0.03 [0.01, 0.12]
GSE25065	-0.20519	0.37352	9.5%	0.81 [0.39, 1.69]
GSE11121	-0.46218	0.34873	10.1%	0.63 [0.32, 1.25]
GSE25055	-0.60098	0.30773	11.3%	0.55 [0.30, 1.00]
GSE4922	-0.9212	0.28896	11.9%	0.40 [0.23, 0.70]
GSE31448	-0.94907	0.28539	12.0%	0.39 [0.22, 0.68]
GSE21653	-0.95394	0.28212	12.1%	0.39 [0.22, 0.67]
GSE7390	-0.66508	0.27276	12.4%	0.51 [0.30, 0.88]
GSE2034	-0.73601	0.24864	13.2%	0.48 [0.29, 0.78]
Total (95% CI)			**100.0%**	**0.43 [0.31, 0.59]**

Heterogeneity: Tau² = 0.14; Chi² = 20.49, df = 9 (P = 0.02); I² = 56%
Test for overall effect: Z = 5.25 (P < 0.00001)

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).

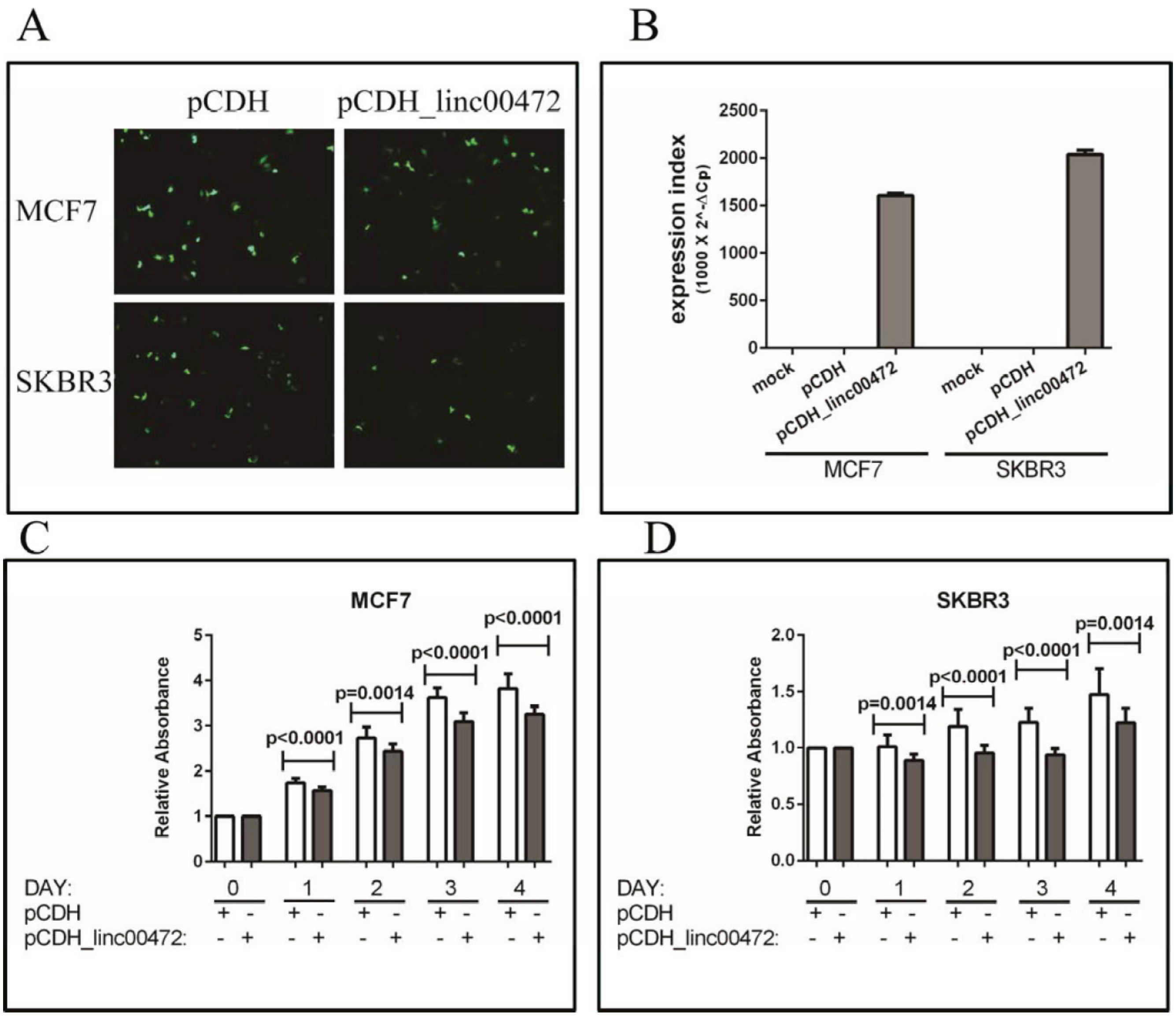


Figure 4: Effect of *LINC00472* expression on breast cancer cell proliferation. A. GFP fluorescence images in MCF7 and SKBR3 cells transfected with pCDH or pCDH_*LINC00472* vectors. **B.** RT-qPCR results of *LINC00472* expression in MCF7 and SKBR3 cells transfected with pCDH or pCDH_*LINC00472* or mock 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, *cat*RAPID *omics*, to search for the binding proteins of *LINC00472*. Since *cat*RAPID *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 *cat*RAPID *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

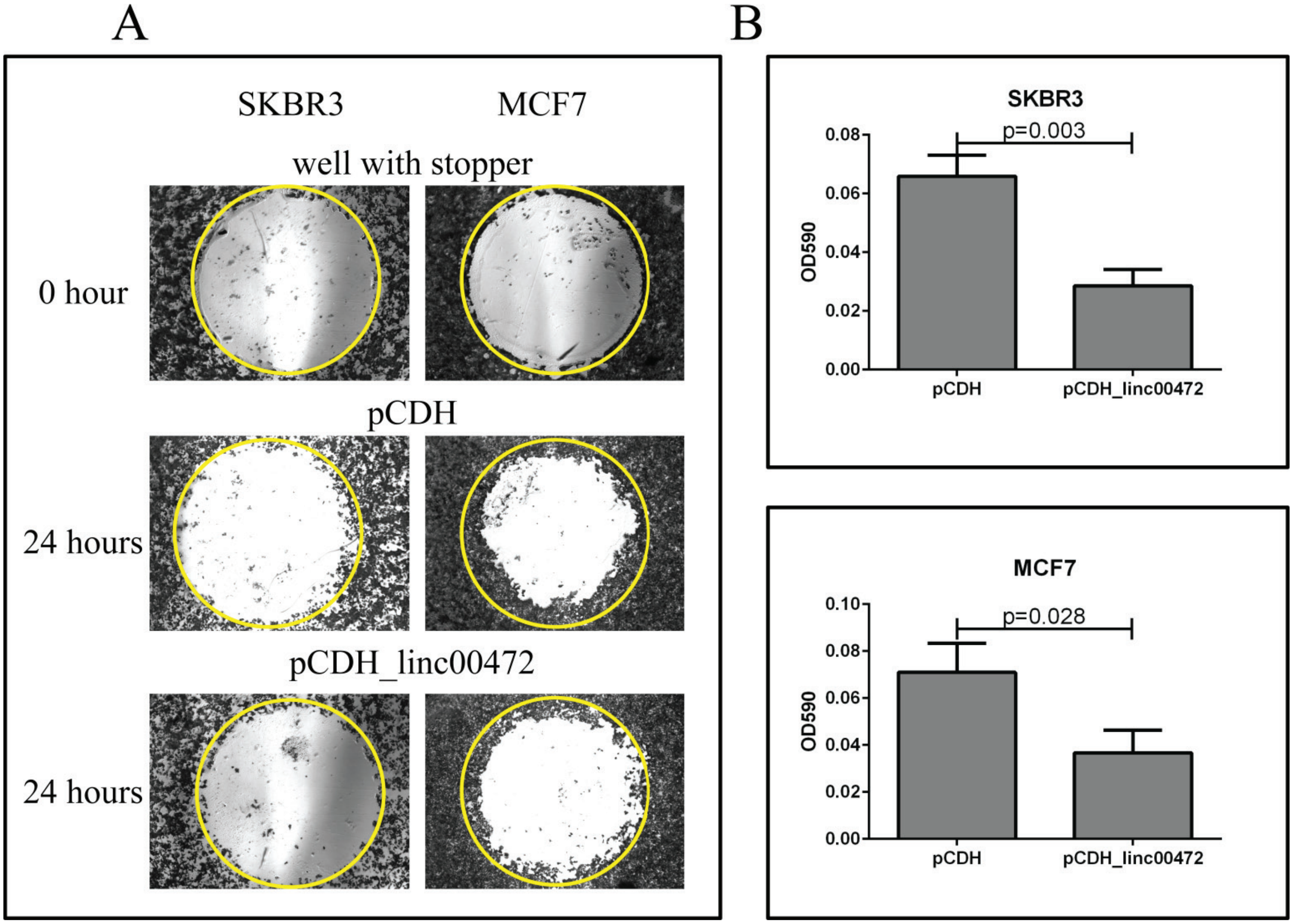


Figure 5: Effect of *LINC00472* expression on breast cancer cell migration. A. Microscopic views of cell migration before (0 hour) and after (24 hour) removal of stopper in SKBR3 and MCF7 cells transfected with pCDH or pCDH_*LINC00472*. Cells transfected with pCDH or pCDH_*LINC00472* vectors formed monolayers in Oris 96-well plate and started to migrate to the exposed area after removing the stoppers in the well. Twenty-four hours later, the cells were fixed and stained with 0.1% crystal violet stain. The photomicrograph of the entire 'wound' area was taken under the IX71 inverted microscope with 4X objective lens. The representative wells were presented and the yellow circles indicated the areas previously occupied by stoppers. **B.** Measurements of absorbance after removal of stopper in SKBR3 and MCF7 cells transfected with pCDH or pCDH_*LINC00472*. With the detection mask, the absorbance at 590 nm wavelength of each well, which was directly proportional to the number of cells that migrated into the 'wound' area, was measured. The bar charts showed the average absorbance (y axis) from the wells with different cells after subtracting the background, the absorbance of the reference wells. Error bars represent SEM, *n* = 8. *P* values were determined by the Mann-Whitney *U* test.

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 breast cancer 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 (TMN), 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 (cyclophosfamide, methotrexate, 5 fluorouracil; 600/60/600 mg/mq every three weeks for six cycles), CEF (cyclophosfamide, epirubicin, 5 fluorouracil; 600/90/600 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 (doxetaxel-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^{o}C for 2 min to activate UDG, 95^{o}C for 2 min to activate Taq polymerase, and 40 cycles of 95^{o}C for 15 s and 60^{o}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 χ^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 evaluation was repeated in 48 h. After that, total RNA was extracted from the cells, and reverse-transcribed for qPCR analysis of *LINC00472* expression which was normalized to GAPDH and compared between cells with different transfection conditions.

Cell proliferation assay

Twenty hours after seeding in 96-well plates, cells were transiently transfected with plasmids or mock-transfected, and kept growing for up to 96 h. At the end of each incubation period (24, 48, 72 and 96 h), numbers of viable cells were assessed by Cell Counting Kit-8 (Sigma-Aldrich).

Cell migration assay

Cell migration was assessed with the modified Oris Cell Migration Assay (AMSBIO). After 24 hours of transfection with plasmids pCDH or pCDH_*LINC00472*, equal amount of cells were plated into the Oris 96-well plate which was incubated overnight with stoppers to permit cell attachment. Stoppers were removed from wells except the reference wells in which the stoppers were remained in place until results were read as pre-migration controls. Media was removed and wells were gently washed with sterile PBS. After that, fresh culture media was added to each well, and the plate was incubated for another 24 hours to allow for cell migration. The cultured cells were fixed and stained with 0.1% crystal violet stain. The photomicrograph of the entire wound area was taken using the IX71 inverted microscope with 4X objective lens. Using a detection mask, the absorbance at 590 nm wavelength in each well, which was directly proportional to the number of cells that migrated into the 'wound' area, 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 × $2^{(-\Delta Ct)}$, where $\Delta Ct = Ct$ (*LINC00472*) – Ct (GAPDH). For data analysis, EI was analyzed as a categorical variable with 3 ordinal levels, low (< 25 percentile), Intermediate (25–75 percentile, labeled as "Mid" in Table 1), and high expression (> 75 percentile). Associations of *LINC00472* with clinical, pathologic and treatment variables and survival outcomes were determined using the Chi-square test or Cox proportional hazards regression model, as appropriate. Kaplan-Meier survival curves were constructed to show survival differences according to *LINC00472* expression. The survival time for either overall 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 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 SI, 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 for Research on Cancer. 2013.
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 evolution. Nature. 2013; 501:338–345.

4. Roylance R, Endesfelder D, Gorman P, Burrell RA, Sander J, Tomlinson I, Hanby AM, Speirs V, Richardson AL, Birkbak NJ, Eklund AC, Downward J, Kschischo M, Szallasi Z, Swanton C. Relationship of extreme chromosomal instability with long-term survival in a retrospective analysis of primary breast cancer. Cancer epidemiology, biomarkers & prevention: a publication of the American Association for Cancer Research, cosponsored by the American Society of Preventive Oncology. 2011; 20:2183–2194.

5. Stephens PJ, Tarpey PS, Davies H, Van Loo P, Greenman C, Wedge DC, Nik-Zainal S, Martin S, Varela I, Bignell GR, Yates LR, Papaemmanuil E, Beare D, Butler A, Cheverton A, Gamble J, et al. The landscape of cancer genes and mutational processes in breast cancer. Nature. 2012; 486:400–404.

6. Ding L, Ellis MJ, Li S, Larson DE, Chen K, Wallis JW, Harris CC, McLellan MD, Fulton RS, Fulton LL, Abbott RM, Hoog J, Dooling DJ, Koboldt DC, Schmidt H, Kalicki J, et al. Genome remodelling in a basal-like breast cancer metastasis and xenograft. Nature. 2010; 464:999–1005.

7. Nik-Zainal S, Van Loo P, Wedge DC, Alexandrov LB, Greenman CD, Lau KW, Raine K, Jones D, Marshall J, Ramakrishna M, Shlien A, Cooke SL, Hinton J, Menzies A, Stebbings LA, Leroy C, et al. The life history of 21 breast cancers. Cell. 2012; 149:994–1007.

8. International Human Genome Sequencing C. Finishing the euchromatic sequence of the human genome. Nature. 2004; 431:931–945.

9. Kapranov P, Willingham AT, Gingeras TR. Genome-wide transcription and the implications for genomic organization. Nature reviews Genetics. 2007; 8:413–423.

10. Huarte M, Rinn JL. Large non-coding RNAs: missing links in cancer? Human molecular genetics. 2010; 19:R152–161.

11. Ji P, Diederichs S, Wang W, Boing S, Metzger R, Schneider PM, Tidow N, Brandt B, Buerger H, Bulk E, Thomas M, Berdel WE, Serve H, Muller-Tidow C. MALAT-1, a novel noncoding RNA, and thymosin beta4 predict metastasis and survival in early-stage non-small cell lung cancer. Oncogene. 2003; 22:8031–8041.

12. Huarte M, Guttman M, Feldser D, Garber M, Koziol MJ, Kenzelmann-Broz D, Khalil AM, Zuk O, Amit I, Rabani M, Attardi LD, Regev A, Lander ES, Jacks T, Rinn JL. A large intergenic noncoding RNA induced by p53 mediates global gene repression in the p53 response. Cell. 2010; 142:409–419.

13. Mourtada-Maarabouni M, Pickard MR, Hedge VL, Farzaneh F, Williams GT. GAS5, a non-protein-coding RNA, controls apoptosis and is downregulated in breast cancer. Oncogene. 2009; 28:195–208.

14. Yu W, Gius D, Onyango P, Muldoon-Jacobs K, Karp J, Feinberg AP, Cui H. Epigenetic silencing of tumour suppressor gene p15 by its antisense RNA. Nature. 2008; 451:202–206.

15. Gupta RA, Shah N, Wang KC, Kim J, Horlings HM, Wong DJ, Tsai MC, Hung T, Argani P, Rinn JL, Wang Y, Brzoska P, Kong B, Li R, West RB, van de Vijver MJ, et al. Long non-coding RNA HOTAIR reprograms chromatin state to promote cancer metastasis. Nature. 2010; 464:1071–1076.

16. Prensner JR, Iyer MK, Balbin OA, Dhanasekaran SM, Cao Q, Brenner JC, Laxman B, Asangani IA, Grasso CS, Kominsky HD, Cao X, Jing X, Wang X, Siddiqui J, Wei JT, Robinson D, et al. Transcriptome sequencing across a prostate cancer cohort identifies PCAT-1, an unannotated lincRNA implicated in disease progression. Nature biotechnology. 2011; 29:742–749.

17. Wang X, Arai S, Song X, Reichart D, Du K, Pascual G, Tempst P, Rosenfeld MG, Glass CK, Kurokawa R. Induced ncRNAs allosterically modify RNA-binding proteins in cis to inhibit transcription. Nature. 2008; 454:126–130.

18. Feng J, Bi C, Clark BS, Mady R, Shah P, Kohtz JD. The Evf-2 noncoding RNA is transcribed from the Dlx-5/6 ultraconserved region and functions as a Dlx-2 transcriptional coactivator. Genes & development. 2006; 20:1470–1484.

19. Martianov I, Ramadass A, Serra Barros A, Chow N, Akoulitchev A. Repression of the human dihydrofolate reductase gene by a non-coding interfering transcript. Nature. 2007; 445:666–670.

20. Mariner PD, Walters RD, Espinoza CA, Drullinger LF, Wagner SD, Kugel JF, Goodrich JA. Human Alu RNA is a modular transacting repressor of mRNA transcription during heat shock. Molecular cell. 2008; 29:499–509.

21. Rinn JL, Kertesz M, Wang JK, Squazzo SL, Xu X, Brugmann SA, Goodnough LH, Helms JA, Farnham PJ, Segal E, Chang HY. Functional demarcation of active and silent chromatin domains in human HOX loci by noncoding RNAs. Cell. 2007; 129:1311–1323.

22. Nagano T, Mitchell JA, Sanz LA, Pauler FM, Ferguson-Smith AC, Feil R, Fraser P. The Air noncoding RNA epigenetically silences transcription by targeting G9a to chromatin. Science. 2008; 322:1717–1720.

23. Pandey RR, Mondal T, Mohammad F, Enroth S, Redrup L, Komorowski J, Nagano T, Mancini-Dinardo D, Kanduri C. Kcnq1ot1 antisense noncoding RNA mediates lineage-specific transcriptional silencing through chromatin-level regulation. Molecular cell. 2008; 32:232–246.

24. Yap KL, Li S, Munoz-Cabello AM, Raguz S, Zeng L, Mujtaba S, Gil J, Walsh MJ, Zhou MM. Molecular interplay of the noncoding RNA ANRIL and methylated histone H3 lysine 27 by polycomb CBX7 in transcriptional silencing of INK4a. Molecular cell. 2010; 38:662–674.

25. Zhou Y, Zhong Y, Wang Y, Zhang X, Batista DL, Gejman R, Ansell PJ, Zhao J, Weng C, Klibanski A. Activation of p53 by MEG3 non-coding RNA. The Journal of biological chemistry. 2007; 282:24731–24742.

26. Prasanth KV, Prasanth SG, Xuan Z, Hearn S, Freier SM, Bennett CF, Zhang MQ, Spector DL. Regulating gene expression through RNA nuclear retention. Cell. 2005; 123:249–263.

27. Bernard D, Prasanth KV, Tripathi V, Colasse S, Nakamura T, Xuan Z, Zhang MQ, Sedel F, Jourdren L, Coulpier F, Triller A, Spector DL, Bessis A. A long nuclear-retained non-coding RNA regulates synaptogenesis by modulating gene expression. The EMBO journal. 2010; 29:3082–3093.

28. Tripathi V, Ellis JD, Shen Z, Song DY, Pan Q, Watt AT, Freier SM, Bennett CF, Sharma A, Bubulya PA, Blencowe BJ, Prasanth SG, Prasanth KV. The nuclear-retained noncoding RNA MALAT1 regulates alternative splicing by modulating SR splicing factor phosphorylation. Molecular cell. 2010; 39:925–938.

29. Wilusz JE, Freier SM, Spector DL. 3′ end processing of a long nuclear-retained noncoding RNA yields a tRNA-like cytoplasmic RNA. Cell. 2008; 135:919–932.

30. Penny GD, Kay GF, Sheardown SA, Rastan S, Brockdorff N. Requirement for Xist in X chromosome inactivation. Nature. 1996; 379:131–137.

31. Fitzpatrick GV, Soloway PD, Higgins MJ. Regional loss of imprinting and growth deficiency in mice with a targeted deletion of KvDMR1. Nature genetics. 2002; 32:426–431.

32. Sleutels F, Zwart R, Barlow DP. The non-coding Air RNA is required for silencing autosomal imprinted genes. Nature. 2002; 415:810–813.

33. Li L, Liu B, Wapinski OL, Tsai MC, Qu K, Zhang J, Carlson JC, Lin M, Fang F, Gupta RA, Helms JA, Chang HY. Targeted disruption of Hotair leads to homeotic transformation and gene derepression. Cell reports. 2013; 5:3–12.

34. Zhang B, Arun G, Mao YS, Lazar Z, Hung G, Bhattacharjee G, Xiao X, Booth CJ, Wu J, Zhang C, Spector DL. The lncRNA Malat1 is dispensable for mouse development but its transcription plays a cis-regulatory role in the adult. Cell reports. 2012; 2:111–123.

35. Mungall AJ, Palmer SA, Sims SK, Edwards CA, Ashurst JL, Wilming L, Jones MC, Horton R, Hunt SE, Scott CE, Gilbert JG, Clamp ME, Bethel G, Milne S, Ainscough R, Almeida JP, et al. The DNA sequence and analysis of human chromosome 6. Nature. 2003; 425:805–811.

36. Bockhorn J, Yee K, Chang YF, Prat A, Huo D, Nwachukwu C, Dalton R, Huang S, Swanson KE, Perou CM, Olopade OI, Clarke MF, Greene GL, Liu H. MicroRNA-30c targets cytoskeleton genes involved in breast cancer cell invasion. Breast cancer research and treatment. 2013; 137:373–382.

37. Zhou H, Xu X, Xun Q, Yu D, Ling J, Guo F, Yan Y, Shi J, Hu Y. microRNA-30c negatively regulates endometrial cancer cells by targeting metastasis-associated gene-1. Oncology reports. 2012; 27:807–812.

38. Bockhorn J, Dalton R, Nwachukwu C, Huang S, Prat A, Yee K, Chang YF, Huo D, Wen Y, Swanson KE, Qiu T, Lu J, Park SY, Dolan ME, Perou CM, Olopade OI, et al. MicroRNA-30c inhibits human breast tumour chemotherapy resistance by regulating TWF1 and IL-11. Nature communications. 2013; 4:1393.

39. Rodriguez-Gonzalez FG, Sieuwerts AM, Smid M, Look MP, Meijer-van Gelder ME, de Weerd V, Sleijfer S, Martens JW, Foekens JA. MicroRNA-30c expression level is an independent predictor of clinical benefit of endocrine therapy in advanced estrogen receptor positive breast cancer. Breast cancer research and treatment. 2011; 127:43–51.

40. Franzetti GA, Laud-Duval K, Bellanger D, Stern MH, Sastre-Garau X, Delattre O. MiR-30a-5p connects EWS-FLI1 and CD99, two major therapeutic targets in Ewing tumor. Oncogene. 2013; 32:3915–3921.

41. Baraniskin A, Birkenkamp-Demtroder K, Maghnouj A, Zollner H, Munding J, Klein-Scory S, Reinacher-Schick A, Schwarte-Waldhoff I, Schmiegel W, Hahn SA. MiR-30a-5p suppresses tumor growth in colon carcinoma by targeting DTL. Carcinogenesis. 2012; 33:732–739.

42. Zhang N, Wang X, Huo Q, Sun M, Cai C, Liu Z, Hu G, Yang Q. MicroRNA-30a suppresses breast tumor growth and metastasis by targeting metadherin. Oncogene. 2014; 33:3119–3128.

43. Gao J, Aksoy BA, Dogrusoz U, Dresdner G, Gross B, Sumer SO, Sun Y, Jacobsen A, Sinha R, Larsson E, Cerami E, Sander C, Schultz N. Integrative analysis of complex cancer genomics and clinical profiles using the cBioPortal. Science signaling. 2013; 6:pl1.

44. Cerami E, Gao J, Dogrusoz U, Gross BE, Sumer SO, Aksoy BA, Jacobsen A, Byrne CJ, Heuer ML, Larsson E, Antipin Y, Reva B, Goldberg AP, Sander C, Schultz N. The cBio cancer genomics portal: an open platform for exploring multidimensional cancer genomics data. Cancer discovery. 2012; 2:401–404.