Long non-coding RNA, LINC01614 as a potential biomarker for prognostic prediction in breast cancer

Yaozong Wang , Baorong Song , Zhu Leilei , Xia Zhang

1 Department of General Surgery, HwaMei Hospital (Ningbo No.2 Hospital), University of Chinese Academy of Sciences, Ningbo, China
2 Department of Radiotherapy, Shanghai East Hospital, Tongji University, Shanghai, China
3 Breast Cancer Center, Shanghai East Hospital, Tongji University, Shanghai, China

Corresponding Author: Xia Zhang
Email address: zhangxia010203@126.com

Background Dysregulated lncRNAs may serve as potential biomarkers of cancers including breast cancer. This study aimed to identify lncRNAs with strong prognostic value for breast cancer.

Methods LncRNA expression profiles of 929 tissue samples were downloaded from TANRIC database. We performed differential expression analysis between paired breast cancer and adjacent normal tissues. Survival analysis was used to identify lncRNAs with prognostic value. Univariate and multivariate Cox regression analyses were performed to confirm the independent prognostic value of potential lncRNAs. Dysregulated signaling pathways associated with lncRNA expression were evaluated using GSEA.

Results We found that a total of 398 lncRNAs were significantly differentially expressed between breast cancer and adjacent normal tissues (adjusted P value <= 0.0001 and |logFC| >=1). Additionally, 381 potential lncRNAs were correlated Overall Survival (P value < 0.05). 48 lncRNAs remained when differentially expressed lncRNAs overlapped with lncRNAs that had prognostic value. Among the 48 lncRNAs, one lncRNA (LINC01614) had stronger prognostic value and was highly expressed in breast cancer tissues. LINC01614 expression was validated as an independent prognostic factor using univariate and multivariate analyses. Higher LINC01614 expression was observed in several molecular subgroups including ER+, PR+ and HER2+ subgroup, respectively. Also, breast cancer carrying one of four gene mutations had higher expression of LINC01614 including AOAH, CIT, HER2 and ODZ1. Higher expression of LINC01614 was positively correlated with several gene sets including TGF-beta1 response, CDH1 signals and cell adhesion pathways.

Conclusions A novel lncRNA LINC01614 was identified as a potential biomarker for prognosis prediction of breast cancer. This study emphasized the importance of LINC01614 and further research should be focused on it.
Long non-coding RNA, LINC01614 as a potential biomarker for prognostic prediction in breast cancer

Yaozong Wang¹, Baorong Song¹, Leilei Zhu², Xia Zhang³

¹ Department of General Surgery, HwaMei Hospital (Ningbo No.2 Hospital), University of Chinese Academy of Sciences, Ningbo, China.

² Department of Radiotherapy, Shanghai East Hospital, Tongji University, Shanghai, China.

³ Breast Cancer Center, Shanghai East Hospital, Tongji University, Shanghai, China

Corresponding author:
Xia Zhang
Breast Cancer Center, Shanghai East Hospital, Tongji University, Shanghai, China

Email address: zhangxia010203@126.com
Abstract

Background Dysregulated lncRNAs may serve as potential biomarkers of cancers including breast cancer. This study aimed to identify lncRNAs with strong prognostic value for breast cancer.

Methods LncRNA expression profiles of 929 tissue samples were downloaded from TANRIC database. We performed differential expression analysis between paired breast cancer and adjacent normal tissues. Survival analysis was used to identify lncRNAs with prognostic value. Univariate and multivariate Cox regression analyses were performed to confirm the independent prognostic value of potential lncRNAs. Dysregulated signaling pathways associated with lncRNA expression were evaluated using GSEA.

Results We found that a total of 398 lncRNAs were significantly differentially expressed between breast cancer and adjacent normal tissues (adjusted P value <= 0.0001 and |logFC| >=1). Additionally, 381 potential lncRNAs were correlated Overall Survival (P value < 0.05). 48 lncRNAs remained when differentially expressed lncRNAs overlapped with lncRNAs that had prognostic value. Among the 48 lncRNAs, one lncRNA (LINC01614) had stronger prognostic value and was highly expressed in breast cancer tissues. LINC01614 expression was validated as an independent prognostic factor using univariate and multivariate analyses. Higher LINC01614 expression was observed in several molecular subgroups including ER+, PR+ and HER2+ subgroup, respectively. Also, breast cancer carrying one of four gene mutations had higher expression of LINC01614 including AOAH, CIT, HER2 and ODZ1. Higher expression of LINC01614 was positively correlated with several gene sets including TGF-beta1 response, CDH1 signals and cell adhesion pathways.

Conclusions A novel lncRNA LINC01614 was identified as a potential biomarker for prognosis.
prediction of breast cancer. This study emphasized the importance of LINC01614 and further
research should be focused on it.

Introduction

Although the management of breast cancer (BRCA) has been significantly improved over the
decades of years, it is still one of the most common cancers worldwide, especially for women. The
morbidity and mortality of BRCA is continuously increasing in recent years especially in Europe
countries such as the United States (Anderson et al. 2011). Based on the statistics from American
Cancer Society, an estimated more than 260,000 BRCA cases occurred and more than 41,000
patients died because of BRCA in 2018 (Siegel et al. 2018). There are several common treatments
for BRCA including surgical resection, chemotherapy, radiotherapy (De Vita 1989). Recently,
immunotherapy is becoming a new strategy for some cancer types including BRCA (Hadden 1999).
Since BRCA is a disease with high heterogeneity, the underlying mechanism of BRCA has not been
totally elucidated (Garcia-Closas et al. 2008). Some BRCA cases were resistant to chemotherapy
and these patients generally had worse survival. Recently, several molecules have been identified
as biomarkers that can be used to determine response to therapy and prognostic prediction such as
estrogen receptors (ERs), progesterone receptors (PRs) and (human epidermal growth factor
receptor 2) HER2 (Colleoni & Montagna 2012; Davies et al. 2011; Network 2012).

With the development of high throughput technique such as next generation sequencing, the
biological function of long non-coding RNAs (IncRNAs) has been increasingly identified (Core et al.
2008). Indeed, more than 75% of the genome are transcribed into noncoding RNAs including
IncRNAs and small RNAs such as tRNAs, rRNAs, miRNAs. Among these, IncRNAs are noncoding transcripts more than 200 nucleotide (nt) in length and are lack of protein-coding abilities (Huarte 2015). According to the Encyclopedia of DNA Elements (ENCODE) Project Consortium, more than 120,000 IncRNAs have been annotated (Jalali et al. 2016). Recently, dysregulated expression of IncRNAs have been found to be involved with cancer development and progression. Some critical IncRNAs have been identified as biomarkers and prognostic factors such as HOTAIR (Gupta et al. 2010), SAMMSON (Leucci et al. 2016), MALAT1 (Sun & Ma 2019) and H19 (Zhang et al. 2016a).

Specifically, HOTAIR could increase cancer invasiveness and metastasis by reprogramming chromatin state in a manner dependent on Polycomb repressive complex 2 (PRC2) (Gupta et al. 2010). LncRNA SAMMSON was found to interact with p32 and regulate homeostasis and metabolism of mitochondria. This effect could increase mitochondrial targeting and pro-oncogenic function of p32 (Leucci et al. 2016). Some other IncRNAs are identified to be involved with cell cycle control, tumor growth, drug response. Since IncRNAs may play important roles in the cancer carcinogenesis, it is necessary to identify potential IncRNAs that may serve as useful biomarkers and prognostic predictors by bioinformatic analysis.

This study was aimed to identify oncogenic IncRNAs with prognostic value in BRCA. LncRNAs were analyzed by correlating with tumorigenesis and Overall Survival (OS). 48 IncRNAs were differentially expressed between BRCA and adjacent normal tissues and also associated with OS. LncRNA LINC01614 had higher expression in BRCA tissues and had stronger prognostic value than other candidate IncRNAs, thus, was selected for further analyses including correlation with clinical parameters, univariate and multivariate analyses using Cox regression model and identifying
dysregulating signaling pathways using Gene Set Enrichment Analysis (GSEA).

Materials and Methods

Clinical data collection

LncRNA expression profile of BRCA was downloaded from TANRIC database (https://ibl.mdanderson.org/tanric/_design/basic/index.html). All of these samples in TANRIC database were from the Cancer Genomic Atlas (TCGA, https://cancergenome.nih.gov/). A total of 929 tissue samples were collected and 105 of them were derived from paired normal tissues. The clinical information of these samples were downloaded from TCGA. 824 of these samples had survival information (OS). A total of 12727 lncRNAs were annotated in the expression profile. The clinical characteristics of these patients including age, gender, TNM stage, ER status, PR status, HER2 status, treatment history, PAM50 subtypes.

Gene set enrichment analysis (GSEA)

We used GSEA to identify dysregulated gene sets between high- and low-LINC01614 expression groups. GSEA was conducted by a JAVA program (http://software.broadinstitute.org/gsea/index.jsp) based on molecular signature database C2 CP (v6.2) which consisted of 4762 gene sets. The number of random sample permutations was set as 1000. Generally, if the majority of members of a given gene set had higher expression accompanied with higher risk score, then gene set would have positive enrichment score and considered as enriched based on the nominal P value (< 0.05), FDR (< 25%) and normalized enrichment score (NES > 1) (Subramanian et al. 2005).

Statistical analysis
All of the statistical analyses were conducted by R (version 3.5.1). Differential expression analysis between BRCA and adjacent normal tissues was performed using “Limma” R package (Ritchie et al. 2015). Two-sample t-test was used to test the statistical differences of LINC01614 expression between two groups and ANOVA test for more than two groups (e.g., PAM50 subtypes). Chi-squared test or Fisher’s exact test was used to compare the clinical differences between high- and low-LINC01614 expression groups. Kaplan–Meier estimate was used to measure the proportion of patients living for a certain time period after surgery. Kaplan–Meier estimate was efficient to deal with censored data. We visualized Kaplan–Meier estimate using “survival” R package. Survival difference was compared using the log-rank test. Univariate and Multivariate Cox analysis were used to test whether LINC01614 expression was an independent prognostic factor. Significance difference was confirmed as P value was less than 0.05.

Results

Identification of candidate IncRNAs with prognostic value in BRCA

The work flow of this study was shown in Fig. S1. According to the TANRIC database, a total of 12727 lncRNAs were annotated. Firstly, we excluded lncRNAs that could not be detected in at least 80% of these 824 samples. By doing so, 4537 lncRNAs remained for further analysis. Since paired adjacent normal tissues in 105 of these samples were available, we performed differential expression analysis between these 105 paired samples. By doing so, 398 IncRNAs were differentially expressed between BRCA and adjacent normal tissues (|lgFC| >= 1 and adjusted P value <= 0.0001, Fig. 1A). To further screen out IncRNAs with prognostic value, we correlated these 4537 IncRNAs with OS. A total of 381 IncRNAs were found associated with OS (P value < 0.05 and
Subsequently, we identified 48 IncRNAs that were differentially expressed and had prognostic value as well (Fig. 1B). Among six IncRNAs that were highly expressed in BRCA tissues, one IncRNA named LINC01614 had stronger prognostic value than others (HR > 1 and P value < 0.01, Fig. 1C, Table. S1). Thus, we selected LINC01614 for further analysis.

**Correlation between LINC01614 expression and clinical characteristics in BRCA**

The clinical characteristics of these 824 BRCA patients were shown in Table. S2. The age of these patients ranged from 26 to 90 and the median was 58. There were only 9 males in this cohort. Stage I patients were found in 140 patients, stage II in 471, stage III in 178 and stage IV in 14. Among these patients, 573 were ER positive, 500 were PR positive and 110 were HER2 positive. Only 65 patients had treatment information and 41 of them were treated. In terms of PAM50 molecular subtypes according to one paper in 2012 (Network 2012), 22 patients were classified as Normal-like subtype, 138 patients were classified as Basal-like subtype, 64 patients were classified as HER2-enriched subtype, 413 patients were classified as Luminal-A subtype and 187 patients were classified as Luminal-B subtype.

These patients were classified into two groups based on the expression of LINC01614 and the median expression was set as the cutoff value. The proportions of ER status (p < 0.0001), PR status (p < 0.0001), HER2 status (p = 0.0006) and PAM50 subtypes (p < 0.0001) were significantly different between high- and low-expression group as shown in Table. S2. Although there was no significant difference between these two groups (P = 0.076), there were more stage III and IV cases in high-expression group (n=102) than that in low-expression group (n=90). We further compared the expression differences among different molecular subgroups. Higher expression of LINC01614 was...
observed in ER positive group (P < 0.0001, Fig. 2A), PR positive group (P < 0.0001, Fig. 2B), and
HER2 positive group (P = 0.0057, Fig. 2C). Significant difference was also observed among PAM50
subtypes (P < 0.0001, Fig. 2D). Specifically, Luminal A and B had higher expression of LINC01614
followed by HER2, Normal and Basal subtypes.

Prognostic value of LINC01614 in BRCA

The median follow-up time was 32.5 months and only 15% of these patients were dead (n=125).
Initially, we found that high-expression group had shorter survival than low-expression group (P =
0.0076, HR = 1.63, 95%CI = 1.13-2.33, Fig. 3A). We also found that the prognostic value of
LINC01614 was also significant only in TNM stage II and III patients (P = 0.01, HR = 1.71, 95%CI =
1.14-2.66, Fig. 3B). We included age, TNM stage, PR status, ER status, HER2 status and PAM50
subtypes for Univariate and Multivariate analysis using Cox regression model. Age was included as
a continuous variable in the model while the others were included as categorical variables. We did
not include gender in Cox regression model as there were only nine male patients and only one was
in low-expression group.

As for univariate Cox regression analysis, we found that high LINC01614 expression was
significantly associated with worse survival (P = 0.008, HR = 1.63, 95%CI = 1.13-2.33, Fig. 3C) and
also for age (P < 0.0001, HR = 1.03, 95%CI = 1.02-1.05), stage III (P = 0.001, HR = 2.72, 95%CI =
1.48-5.00), stage IV (P < 0.0001, HR = 10.03, 95%CI = 4.58-21.90), PAM50 HER2 subtype (P =
0.007, HR = 2.66, 95%CI = 1.31-5.39), PAM50 Luminal B (P = 0.017, HR = 2.04, 95%CI = 1.14-
3.67). We then performed multivariate Cox regression analysis and found that LINC01614
expression was an independent prognostic factor for BRCA (P = 0.007, HR = 1.88, 95%CI = 1.19-
2.97, Fig. 3D) as well as for age (P < 0.0001, HR = 1.04, 95%CI = 1.02-1.06), stage III (P = 0.001, HR = 3.72, 95%CI = 1.77-7.79), stage IV (P < 0.0001, HR = 7.24, 95%CI = 2.94-17.86), PAM50 HER2 subtypes (P = 0.021, HR = 3.31, 95%CI = 1.20-9.13) and HER2 status (P = 0.026, HR = 0.39, 95%CI = 0.17-0.89).

We further performed subgroup analyses based on PR status, ER status and HER2 status (Fig. S2). Our results indicated that prognostic value of LINC01614 could be observed in ER positive group (P = 0.032, HR = 1.62, 95%CI = 1.04-2.53) and HER2 negative group (P = 0.037, HR = 1.56, 95%CI = 1.02-2.37). Although there were no significant differences in other subgroups, we observed the trend that high-expression group had shorter survival than low-expression group among these subgroups. More samples were needed to validate these results.

**Gene mutations correlated with LINC01614 expression**

We also explored the correlations between LINC01614 expression and gene mutations status using the inbuilt functions of the TANRIC database. We found that the expression of LINC01614 was significant associated with four gene mutations including AOHA, CIT, HER2 and ODZ1. Consistently, patients carrying any one of these gene mutations had higher expression of LINC01614 (P = 0.0064 for AOHA, P = 0.0004 for CIT, P = 0.0057 for HER2 and P = 0.0011 for ODZ1, Fig. S3).

**Dysregulated pathways associated with LINC01614 expression**

We performed GSEA to identify dysregulated gene sets between high- and low-expression group. 799 and 42 gene sets were significantly (FDR < 25%, normal P value < 0.05) enriched in high- and low-expression group, respectively (Table. S3). Among these gene sets, some of them were critical
in the development and progression of cancer such as TGF-β1 response, CDH1 signaling via CTNNB1, integrin3 and integrin cell surface interaction pathways. High-expression group was also enriched with one gene signature associated with ductal carcinoma. Finally, we also found that genes associated with Estradiol treatment were enriched in high-expression group (Fig. 4A-F).

**Discussion**

Taking the advantage of high-throughput sequencing technologies, increasingly comprehensive transcriptomes have been recognized and identified. The roles of lncRNAs have been investigated and emphasized (Huarte 2015). It has been found that lncRNAs play important roles in multiple biological processes including chromatin modification, transcription activation, transcription regulation (Yang et al. 2014). Accumulated studies have suggested that lncRNAs might be novel biomarkers for cancer diagnosis and prognosis prediction. The prognostic value of lncRNAs have been identified in several cancer types including colorectal cancer (Hu et al. 2014), non-small cell lung cancer (Zhou et al. 2015), gastric cancer (Zhu et al. 2016).

In this study, we made full use of comprehensive lncRNA expression profiles derived from TANRIC database. Among these lncRNAs, one lncRNA named LINC01614 (ENSG00000230838.1) had stronger prognostic value than others. High-expression group had shorter survival than that in low-expression group. Further univariate and multivariate Cox regression analyses indicated that LINC01614 was an independent prognostic factor of BRCA. We further correlated the expression of LINC01614 with BRCA subtypes. Higher expression of LINC01614 was consistently observed in ER, PR, and HER2 positive group, respectively. Additionally, Luminal A, Luminal B and HER2 subgroups had relative higher expression of LINC01614 than that in Basal-like and Normal-like subgroups. The
prognostic value of LINC01614 could be observed in ER positive and HER2 negative groups, marginally in PR positive and negative subgroups, respectively. Some of the statistics were not significant probably because of the small sample size of these subgroups. When correlated the expression of LINC01614 with gene mutation status, we found the mutation status of four genes was correlated with LINC01614 expression including AOAH, CIT, HER2 and ODZ1. Higher expression of LINC01614 was observed in mutated samples. HER2 has been recognized as biomarker not only just in BRCA (Chmielecki et al. 2015; Connell & Doherty 2017). Most of HER2 mutation patients responded to HER2 inhibitors (Ben-Baruch et al. 2015). AOHA was reported to have higher conditional frequencies than marginal frequencies in early time intervals. This might indicate that patients with AOAH mutation was more progressive (Sakoparnig et al. 2015). Citron kinase (CIT-K) was also highly expressed in BRCA and could promote cell growth and colony formation, indicating its oncogenic function in BRCA (Meng et al. 2018). Finally, ODZ1 was involved with increased cell proliferation and invasion in glioblastoma (Talamillo et al. 2017) and was identified as a biomarker in papillary thyroid carcinoma (Cheng et al. 2017).

Several critical pathways associated with TGF-β1 response, CDHI signaling via CTNNB1, integrin3 and integrin cell surface interaction pathways were enriched in subgroup with high expression of LINC01614. These dysregulated signaling pathways played important roles in the development and progression of cancer (Liu & Chu 2014; Seguin et al. 2015; Syed 2016). These results might imply the potential function of LINC01614 in BRCA. In terms of the prognostic value, previous studies have emphasized the prognostic value of LINC01614 in cancer. For instance, Huang et al identified three IncRNAs with prognostic value in esophageal squamous cell carcinoma including RP11-366H4.1.1,
LINC00460 and LINC01614 (Huang et al. 2018). These three IncRNAs-based signature could facilitate the prognosis prediction of esophageal squamous cell carcinoma. Moreover, Zhang et al also found that LINC01614 was highly expression in ER positive BRCA which was in line with our result (Zhang et al. 2016b). These findings highlighted the potential role of LINC01614 in BRCA and it might be an useful biomarker for BRCA prognostic prediction.

There are several limitations for this study. Firstly, the prognostic value of LINC01614 should be further validated in cohorts with larger sample size in the future. Secondly, the differential expression of LINC01614 between BRCA and normal tissue should be validated through experiment. Thirdly, the potential role of LINC01614 in BRCA should be experimentally explored and dysregulated pathways identified in this study should also validated. Finally, as for the clinical practice, after the potential role of LINC01614 is further confirmed, the clinicians can use it to facilitate prognosis assessment of BRCA patients after surgery.

Conclusions: This research provided an example to identify candidate IncRNAs with prognostic value in cancers. A novel IncRNA LINC01614 was identified as a potential biomarker for prognosis prediction in BRCA. This study emphasized the importance of LINC01614 and further more research should be focused on it.

Acknowledgments: The authors are grateful for all the subjects who participated in the study. Y. W. and X. Z designed this research. Y. W., B. S. and L. Z. collected data and performed the analyses. Y. W. drafted the manuscript. Y. W., B. S., L. Z. and X. Z read and revised the manuscript.

Funding: This work was supported by grants from Shanghai Pudong New Area Health and Family
Planning Commission Health Science and Technology Project, NO# PW2016B-5.

**Competing Interests:** The authors declare that there are no conflicts of interest.

**References:**

Anderson WF, Katki HA, and Rosenberg PS. 2011. Incidence of breast cancer in the United States: current and future trends. *J Natl Cancer Inst* 103:1397-1402. 10.1093/jnci/djr257

Ben-Baruch NE, Bose R, Kavuri SM, Ma CX, and Ellis MJ. 2015. HER2-Mutated Breast Cancer Responds to Treatment With Single-Agent Neratinib, a Second-Generation HER2/EGFR Tyrosine Kinase Inhibitor. *J Natl Compr Canc Netw* 13:1061-1064.

Cheng SP, Chen MJ, Chien MN, Lin CH, Lee JJ, and Liu CL. 2017. Overexpression of teneurin transmembrane protein 1 is a potential marker of disease progression in papillary thyroid carcinoma. *Clin Exp Med* 17:555-564. 10.1007/s10238-016-0445-y

Chmielecki J, Ross JS, Wang K, Frampton GM, Palmer GA, Ali SM, Palma N, Morosini D, Miller VA, Yelensky R, Lipson D, and Stephens PJ. 2015. Oncogenic alterations in ERBB2/HER2 represent potential therapeutic targets across tumors from diverse anatomic sites of origin. *Oncologist* 20:7-12. 10.1634/theoncologist.2014-0234

Colleoni M, and Montagna E. 2012. Neoadjuvant therapy for ER-positive breast cancers. *Ann Oncol* 23 Suppl 10:x243-248. 10.1093/annonc/mds305

Connell CM, and Doherty GJ. 2017. Activating HER2 mutations as emerging targets in multiple solid cancers. *ESMO Open* 2:e000279. 10.1136/esmoopen-2017-000279

Core LJ, Waterfall JJ, and Lis JT. 2008. Nascent RNA sequencing reveals widespread pausing and divergent initiation at human promoters. *Science* 322:1845-1848. 10.1126/science.1162228

Davies C, Godwin J, Gray R, Clarke M, Cutter D, Darby S, McGale P, Pan HC, Taylor C, Wang YC, Dowsett M, Ingle J, and Peto R. 2011. Relevance of breast cancer hormone receptors and other factors to the efficacy of adjuvant tamoxifen: patient-level meta-analysis of randomised trials. *Lancet* 378:771-784. 10.1016/s0140-6736(11)60993-8

De Vita VT, Jr. 1989. Breast cancer therapy: exercising all our options. *N Engl J Med* 320:527-529. 10.1056/nejm198902233200812

Garcia-Closas M, Hall P, Nevanlinna H, Pooley K, Morrison J, Richesson DA, Bojesen SE, Nordestgaard BG, Axelsson CK, Arias JI, Milne RL, Ribas G, Gonzalez-Neira A, Benitez J, Zamora P, Brauch H, Justenhoven C, Hamann U, Ko YD, Bruening T, Haas S, Dork T, Schurmann P, hillemanns P, Bogdanova N, Bremer M, Karstens JH, Fagerholm R, Aaltonen K, Aittomaki K, von Smitten K, Blomqvist C, Mannermaa A, Uusitupa M, Eskelinen M, Tengstrom M, Kosma VM, Kataja V, Chenevix-Trench G, Spurdle AB, Beesley J, Chen X, Devilee P, van Asperen CJ, Jacob C, Poller AA, Ylitalo KA, Klijn JG, Chang-Claude J, Kropp S, Slanger T, Flesch-Janys D, Mutschelknauss E, Salazar R, Wang-Gohrke S, Couch F, Goode EL, Olson JE, Vachon C, Fredericksen ZS, Giles GG, Baglietto L, Severi G, Hopper JL, English DR, Southey MC, Haiman CA, Henderson BE, Kolonel LN, Le Marchand L, Stram DO, Hunter DJ, Hankinson SE, Cox DG, Tamimi R,
Kraft P, Sherman ME, Chanock SJ, Lissowska J, Brinton LA, Peplonska B, Klijn JG, Hoo ning MJ, Meijers-Heijboer H, Collee JM, van den Ouweland A, Uitterlinden AG, Liu J, Lin LY, Yuqing L, Humphreys K, Czene K, Cox A, Balasubramanian SP, Cross SS, Reed MW, Blows F, Driver K, Dunning A, Tyrer J, Ponder BA, Sangrajrang S, Brennan P, McKay J, Odefrey F, Gabrieau V, Sigurdson A, Doody M, Struwen ing JP, Alexander B, Easton DF, and Pharoah PD. 2008. Heterogeneity of breast cancer associations with five susceptibility loci by clinical and pathological characteristics. *PLoS Genet* 4:e1000054. 10.1371/journal.pgen.1000054.

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, Sukumar S, and Chang HY. 2010. Long non-coding RNA HOTAIR reprograms chromatin state to promote cancer metastasis. *Nature* 464:1071-1076. 10.1038/nature08975.

Hadden JW. 1999. The immunology and immunotherapy of breast cancer: an update. *Int J Immunopharmacol* 21:79-101.

Hu Y, Chen HY, Yu CY, Xu J, Wang JL, Qian J, Zhang X, and Fang JY. 2014. A long non-coding RNA signature to improve prognosis prediction of colorectal cancer. *Oncotarget* 5:2230-2242. 10.18632/oncotarget.1895.

Huang GW, Xue YJ, Wu ZY, Xu XE, Wu JY, Cao HH, Zhu Y, He JZ, Li CQ, Li EM, and Xu LY. 2018. A three-lncRNA signature predicts overall survival and disease-free survival in patients with esophageal squamous cell carcinoma. *BMC Cancer* 18:147. 10.1186/s12885-018-4058-6.

Huarte M. 2015. The emerging role of lncRNAs in cancer. *Nat Med* 21:1253-2061. 10.1038/nm.3981.

Jalali S, Gandhi S, and Scaria V. 2016. Navigating the dynamic landscape of long noncoding RNA and protein-coding gene annotations in GENCODE. *Hum Genomics* 10:35. 10.1186/s40246-016-0090-2.

Leucci E, Vendramin R, Spinazzi M, Laurette P, Fiers M, Wouters J, Radaelli E, Eyckerman S, Leonelli C, Vanderheyden K, Rogiers A, Hermans E, Baatsen P, Aerts S, Amant F, Van Aelst S, van den Oord J, de Strooper B, Davidson I, Lafontaine DL, Gevaert K, Vandesompele J, Mestdagh P, and Marine JC. 2016. Melanoma addiction to the long non-coding RNA SAMMSON. *Nature* 531:518-522. 10.1038/nature17161.

Liu H, and Chu KM. 2014. E-cadherin and gastric cancer: cause, consequence, and applications. *Biomed Res Int* 2014:637308. 10.1155/2014/637308.

Meng D, Yu Q, Feng L, Luo M, Shao S, Huang S, Wang G, Jing X, Tong Z, Zhao X, and Liu R. 2018. Citron kinase (CIT-K) promotes aggressiveness and tumorigenesis of breast cancer cells in vitro and in vivo: preliminary study of the underlying mechanism. *Clin Transl Oncol*. 10.1007/s12094-018-02003-9.

Network CGA. 2012. Comprehensive molecular portraits of human breast tumours. *Nature* 490:61-70. 10.1038/nature11412.

Ritchie ME, Phipson B, Wu D, Hu Y, Law CW, Shi W, and Smyth GK. 2015. limma powers differential expression analyses for RNA-sequencing and microarray studies. *Nucleic Acids Res* 43:e47. 10.1093/nar/gkv007.

Sakoparnig T, Fried P, and Beerenwinkel N. 2015. Identification of constrained cancer driver genes based on mutation timing. *PLoS Comput Biol* 11:e1004027. 10.1371/journal.pcbi.1004027.

Seguin L, Desgrosellier JS, Weis SM, and Cheresh DA. 2015. Integrins and cancer: regulators of cancer stemness, metastasis, and drug resistance. *Trends Cell Biol* 25:234-240. 10.1016/j.tcb.2014.12.006.

Siegel RL, Miller KD, and Jemal A. 2018. Cancer statistics, 2018. *CA: A Cancer Journal for Clinicians* 68:7-30. 10.3322/caac.21442.

Subramanian A, Tamayo P, Mootha VK, Mukherjee S, Ebert BL, Gillette MA, Paulovich A, Pomeroy SL, Golub TR,
Lander ES, and Mesirov JP. 2005. Gene set enrichment analysis: a knowledge-based approach for interpreting genome-wide expression profiles. *Proc Natl Acad Sci U S A* 102:15545-15550. 10.1073/pnas.0506580102

Sun Y, and Ma L. 2019. New Insights into Long Non-Coding RNA MALAT1 in Cancer and Metastasis. *Cancers (Basel)* 11. 10.3390/cancers11020216

Syed V. 2016. TGF-beta Signaling in Cancer. *J Cell Biochem* 117:1279-1287. 10.1002/jcb.25496

Talamillo A, Grande L, Ruiz-Ontanon P, Velasquez C, Mollinedo P, Torices S, Sanchez-Gomez P, Aznar A, Esparis-Ogado A, Lopez-Lopez C, Lafita C, Berciano MT, Montero JA, Vazquez-Barquero A, Segura V, Villagra NT, Pandiella A, Lafarga M, Leon J, Martinez-Climent JA, Sanz-Moreno V, and Fernandez-Luna JL. 2017. ODZ1 allows glioblastoma to sustain invasiveness through a Myc-dependent transcriptional upregulation of RhoA. *Oncogene* 36:1733-1744. 10.1038/onc.2016.341

Yang L, Froberg JE, and Lee JT. 2014. Long noncoding RNAs: fresh perspectives into the RNA world. *Trends Biochem Sci* 39:35-43. 10.1016/j.tibs.2013.10.002

Zhang K, Luo Z, Zhang Y, Zhang L, Wu L, Liu L, Yang J, Song X, and Liu J. 2016a. Circulating IncRNA H19 in plasma as a novel biomarker for breast cancer. *Cancer Biomark* 17:187-194. 10.3233/cbm-160630

Zhang Y, Wagner EK, Guo X, May I, Cai Q, Zheng W, He C, and Long J. 2016b. Long intergenic non-coding RNA expression signature in human breast cancer. *Sci Rep* 6:37821. 10.1038/srep37821

Zhou M, Guo M, He D, Wang X, Cui Y, Yang H, Hao D, and Sun J. 2015. A potential signature of eight long non-coding RNAs predicts survival in patients with non-small cell lung cancer. *J Transl Med* 13:231. 10.1186/s12967-015-0556-3

Zhu X, Tian X, Yu C, Shen C, Yan T, Hong J, Wang Z, Fang JY, and Chen H. 2016. A long non-coding RNA signature to improve prognosis prediction of gastric cancer. *Mol Cancer* 15:60. 10.1186/s12943-016-0544-0
Figure 1

Fig. 1. Expression profiles of candidate lncRNAs.

**Fig. 1.** Expression profiles of candidate lncRNAs. (A) Heatmap of 398 lncRNA expression profiles that were differentially expressed between BRCA and adjacent normal tissues (adjusted P value <= 0.0001, |logFC| >= 1). Rows represent lncRNAs, and columns represent patients. (B) Heatmap of 48 lncRNA expression profiles that were differentially expressed in (A) with prognostic value (P value < 0.05). (C) Boxplot of candidate lncRNA (LINC01614) between BRCA and paired normal tissues.
Figure 2

Fig. 2. Expression comparison of LINC01614 among molecular subgroups or subtypes of BRCA.

Fig. 2. Expression comparison of LINC01614 among molecular subgroups or subtypes of BRCA. (A) ER positive versus ER negative. (B) PR positive versus PR negative. (C) HER2 positive versus HER2 negative. (D) PAM50 molecular subtypes.
Figure 3

Fig. 3. Survival analysis of LINC01614 based on two groups classified using the median expression of LINC01614.

Fig. 3. Survival analysis of LINC01614 based on two groups classified using the median expression of LINC01614. (A) Kaplan-Meier estimates of Overall Survival in 824 BRCA samples. (B) Kaplan-Meier estimates of Overall Survival stage II and III BRCA samples (n = 647). The tick marks on the Kaplan-Meier curves indicated the censored cases. The differences between the two curves were determined by the two-side log-rank test. (C) Univariate Cox regression proportional hazards regression analysis between clinical features. (D) Multivariate Cox regression proportional hazards regression analysis of LINC0614 expression groups adjusted by other clinical features including age, TNM stage, PAM50, ER, PR and HER2 status. Age was regarded as continuous variable. Gender was not included because there were only nine male subjects and only one in low-expression group. Solid squares indicated the hazard ratio (HR) of death, and close-ended horizontal lines represent the 95% confidence intervals (CI). All P values were calculated using Cox regression hazards analysis.
Fig. 4. Enrichment plots of GSEA based on LINC01614 expression group.

GSEA indicated that TGF-beta1 response (A), CDHI signaling via CTNNB1 (B), integrin3 (C) and integrin cell surface interaction (D), ductal carcinoma signature (E) and Estradiol treatment response signature were significantly enriched in high-expression group. NES, normalized ES; FDR: false discovery rate.
Figure 5

Fig. 2. Expression comparison of LINC01614 among molecular subgroups or subtypes of BRCA.

Fig. 2. Expression comparison of LINC01614 among molecular subgroups or subtypes of BRCA. (A) ER positive versus ER negative. (B) PR positive versus PR negative. (C) HER2 positive versus HER2 negative. (D) PAM50 molecular subtypes.