Identification of a prognostic long non-coding RNA signature in breast cancer

Gaochen Lan

The First Affiliated Hospital of Zhejiang Chinese Medical University

Xiaoling Yu

The First Affiliated Hospital of Zhejiang Chinese Medical University

Yanna Zhao

The First Affiliated Hospital of Zhejiang Chinese Medical University

Jinjian Lan

The First Affiliated Hospital of Zhejiang Chinese Medical University

Wan Li

The First Affiliated Hospital of Zhejiang Chinese Medical University

Ruilan Gao (✉ gaoruilan@126.com)

<https://orcid.org/0000-0002-9346-5653>

Primary research

Keywords: Breast cancer, Long non-coding RNA, Prognostic signature

Posted Date: February 20th, 2020

DOI: <https://doi.org/10.21203/rs.2.24062/v1>

License: © ⓘ This work is licensed under a Creative Commons Attribution 4.0 International License.

[Read Full License](#)

Abstract

Background: Breast cancer is the most common malignant disease among women. At present, more and more attention has been paid to long non-coding RNAs (lncRNAs) in the field of breast cancer research. We aimed to investigate the expression profiles of lncRNAs and construct a prognostic lncRNA for predicting the overall survival (OS) of breast cancer.

Methods: The expression profiles of lncRNAs and clinical data with breast cancer were obtained from The Cancer Genome Atlas (TCGA). Differentially expressed lncRNAs were screened out by R package (limma). The survival probability was estimated by the Kaplan-Meier Test. The Cox Regression Model was performed for univariate and multivariate analysis. The risk score (RS) was established on the basis of the lncRNAs' expression level (exp) multiplied regression coefficient (β) from the multivariate cox regression analysis with the following formula: $RS = \exp a_1 * \beta a_1 + \exp a_2 * \beta a_2 + \dots + \exp a_n * \beta a_n$. Functional enrichment analysis was performed by Metascape.

Results: A total of 3404 differentially expressed lncRNAs were identified. Among them, CYTOR, MIR4458HG and MAPT-AS1 were significantly associated with the survival of breast cancer. Finally, The RS could predict OS of breast cancer ($RS = \exp \text{CYTOR} * \beta \text{CYTOR} + \exp \text{MIR4458HG} * \beta \text{MIR4458HG} + \exp \text{MAPT-AS1} * \beta \text{MAPT-AS1}$). Moreover, it was confirmed that the three-lncRNA signature could be an independent prognostic biomarker for breast cancer (HR=3.040, P=0.000).

Conclusions: This study established a three-lncRNA signature, which might be a novel prognostic biomarker for breast cancer.

Background

Breast cancer is the most common malignant disease and the second largest cause of mortality among women all over the world [1, 2]. According to the SEER database, it estimated 268,600 new cases and 41,760 deaths, and accounted for 15.2% and 6.9% out of all cancers, respectively in the United States in 2019 [3]. As is known to all, some conventional clinicopathological parameters, including age, histological grade, tumor size, lymph node status, vascular tumor thrombus, hormone receptor expression, human epidermal growth factor receptor 2 (HER2) expression have been established to predict the prognosis in breast cancer [4, 5]. However, breast cancer is a heterogeneous disease which makes just a part of patients could match predictive value of these factors. A study showed that their predictive ability was only effective in approximately 30% of patients [6]. In addition, above clinicopathological parameters are too general to cater to the precise management of breast cancer. In recent years, multigene detection such as 21-gene detection, 12-gene detection, 50-gene detection, and so on, with the molecular and genetic levels have displayed greater accuracy than traditional prognostic biomarkers in a cohort of patients with HR-positive and HER2-negative [7–10].

Nowadays, more and more attention has been paid to long non-coding RNAs (lncRNAs) which are involved in biological signal pathways and pathogenesis in malignant diseases, such as breast cancer.

LncRNAs are a kind of RNA transcripts with a length of more than 200 bp and no function of encoding protein. A number of biological processes such as cell proliferation, differentiation, chromosome remodeling, epigenetic modulation, transcriptional and posttranscriptional modifications are related to lncRNAs [11]. Various studies have indicated that disorderly expression of lncRNAs in the occurrence and progression of multiple cancers including breast cancer [11, 12]. Moreover, more and more dysregulated lncRNAs have been found and demonstrated their potential as possible biomarkers for predicting clinical outcome in breast cancer. For instance, the over-expression lncRNAs of HOTAIR and BCAR4 would promote the proliferation and metastasis of tumor cells in breast cancer [13, 14]. Wang et al. [15] found that higher expression of lncRNA LINC01614 was associated with shorter overall survival (OS). Meanwhile, some lncRNAs, such as NKILA, MEG3, XIST, and PTENP1, etc. possessed the activity to suppress development and progression of tumor [16–21]. If these lncRNAs were absent or low-expression, breast cancer would be more aggressive. More and more lncRNAs have been found and identified with the development of gene sequencing technology. However, prognostic signature with lncRNAs has been rarely reported. In addition, the biological functions and signal pathways of lncRNA remain unclear in breast cancer. Therefore, the relationship between lncRNAs and breast cancer deserves further exploration.

In this study, we performed the lncRNA expression profiles by analyzing a cohort of previously published gene expression profiles from The Cancer Genome Atlas (TCGA, <http://cancergenome.nih.gov/>) to establish a prognostic lncRNA signature in breast cancer. Finally, we identify three lncRNAs as the prognostic signature related to OS in breast cancer. These findings may provide a novel prognostic model and research concept at the genetic levels in the field of breast cancer study.

Methods

Differentially expressed lncRNAs and breast cancer cases from TCGA

The RNA sequencing and clinical data were downloaded from TCGA data portal (<https://cancergenome.nih.gov/>). The RNA-Seq data contained the expression profiles of 60483 mRNAs and clinical data covered 1222 cases, including 1109 patients of breast cancer and 113 cases of normal breast tissue. Next, we extracted 14847 lncRNAs from the expression data of 60483 mRNAs. The data were obtained from TCGA, which is a community resource project providing available data for community research. Approval by a local ethics committee was not required because the current study adhered to the TCGA publication guidelines and data access policies. Then, we selected 3404 significant differentially expressed lncRNAs by R package (limma) with the standard of $P \leq 0.01$ and the absolute $\log_2 FC \geq 1$. The clinical data of breast cancer were also obtained from TCGA as mentioned above. The cases with incomplete clinical data would be excluded, such as OS, age, tumor size (T), lymph nodes (N), distant metastasis (M), TNM stage, HER2 expression, and estrogen/progesterone receptor (ER/PR) expression.

Finally, the prognostic analysis included a total of expression data of 3404 lncRNAs derived from 565 cases with breast cancer.

Statistical analysis

SPSS (version 23.0; SPSS, Inc., Chicago, IL, USA) was performed to analyze data. The survival probability was estimated by the KaplanMeier Test, and the statistical significance was performed using the LogRank Test. The univariate and multivariate analysis for the expression of CYTOR, MIR4458HG and MAPT-AS1 were analyzed by the Cox Regression Model. In the meanwhile, the Cox Regression Model was also performed the multivariate analysis for RS with the three-lncRNA signature and other clinicopathological parameters. The risk score (RS) for predicting OS was established on the basis of the expression level (exp) multiplied regression coefficient (β) from the multivariate cox regression analysis with the following formula: $RS = \exp_{CYTOR} * \beta_{CYTOR} + \exp_{MIR4458HG} * \beta_{MIR4458HG} + \exp_{MAPT-AS1} * \beta_{MAPT-AS1}$. 565 cases of breast cancer were divided into two groups of low-risk and high-risk according to the median RS as cut-off. The relationships between RS groups and the expression of CYTOR, MIR4458HG and MAPT-AS1 were analyzed by the T Test. Receiver operating characteristic (ROC) curves predicted the sensitivity and specificity of the RS with the three-lncRNA signature for predicting OS. The area under characteristic (AUC) values was calculated from the ROC curves. All tests were two sided. $P \leq 0.01$ was considered to indicate a statistically significant difference. Multivariate analysis for the expression of CYTOR, MIR4458HG and MAPT-AS1 was also statistically significant if $P \leq 0.05$.

Co-expression Analysis Of lncRNAs With mRNAs

Co-expression network analysis (<http://bio-bigdata.hrbmu.edu.cn/Co-LncRNA/>) was performed to evaluate the prognostic lncRNAs and mRNAs of breast cancer. Linear Regression was performed and $P < 0.01$ was considered to define significant correlations.

Functional Prediction Of The Three-lncRNA Signature

Finally, the lncRNAs and the co-expressed mRNAs were inputed in the online tool: Metascape (<http://metascape.org/gp/index.html>) for functional enrichment analysis. Gene Ontology (GO) terms for the biological process (BP), Kyoto Encyclopedia of Genes and Genomes (KEGG) pathways, and so on, were enriched. $P < 0.01$ was considered as a statistically significant term. All the resultant terms were then grouped into clusters based on their similarities.

Results

Differentially expressed lncRNAs in breast cancer

We extracted 14847 lncRNAs from the expression data of 60483 mRNAs. 3404 differentially expressed lncRNAs were obtained by R package (limma) ($P \leq 0.01$), and these lncRNAs were presented with Volcano

Plot (Fig. 1).

Differentially Expressed IncRNAs Association With Os In Breast Cancer

The clinical data of 1222 cases was downloaded from TCGA, and 565 cases were selected for prognostic analysis according to completed information of clinicopathological parameters including OS, age, T, N, M, TNM stage, HER2 expression, and ER/PR expression. The Kaplan-Meier was performed to evaluate the relationship between differentially expressed IncRNAs and OS, then, we obtained 30 IncRNAs related to prognostic value in breast cancer ($P\leq 0.01$).

Prognostic IncRNA Signature In Breast Cancer

Among these 30 IncRNAs with prognostic value, co-expression network analysis (<http://bio-bigdata.hrbmu.edu.cn/Co-LncRNA/>) revealed that three IncRNAs: ENSG00000222041 (CYTOR), ENSG00000247516 (MIR4458HG) and ENSG00000264589 (MAPT-AS1) were co expressed with mRNAs. Further univariate and multivariate cox regression analysis indicated that CYTOR, MIR4458HG and MAPT-AS1 were independent prognostic factors in breast cancer patients ($P\leq 0.05$; Table 1). Next, we established the RS for predicting OS basing on the three IncRNAs above: $RS = \exp_{CYTOR} * \beta_{CYTOR} + \exp_{MIR4458HG} * \beta_{MIR4458HG} + \exp_{MAPT-AS1} * \beta_{MAPT-AS1}$. The “ β ” means the coefficient from the multivariate cox regression analysis and the “exp” means the expression level of IncRNA. According to above RS formula, we calculated the RS for each patient in selected 565 cases of breast cancer group, and divided the patients into two groups of low-risk and high-risk basing on the median RS as cut-off (Fig. 2).

Table 1
The detailed information of three prognostic IncRNAs in breast cancer

LncRNA	Coef	HR	P-value	Low 95% CI	High 95% CI
MIR4458HG	-0.43741	0.64571	0.02657	0.64571	1.54869
CYTOR	0.08687	1.09071	0.01288	1.09075	0.91680
MAPT-AS1	-1.01247	0.36332	0.00951	0.36332	2.75240
Coef; Regression coefficient; HR; Hazard ratio; CI; Confidence interval					

The Kaplan-Meier was performed to evaluate the correlation between RS and OS, and we found that patients in the high-risk group had significantly shorter OS than those in the low-risk group ($P = 4e-04$; Fig. 3). The RS with the three-IncRNA signature could strongly predict OS in breast cancer. Furthermore, multivariate cox regression analysis showed that the RS with the three-IncRNA signature was an independent predictor of OS in breast cancer, which was as same as TNM stage when multiple factors,

including age, T, N, M, TNM stage, HER-2 expression, ER expression, and PR expression were controlled ($P \leq 0.01$; Table 2). We also evaluated the relationships between the three lncRNAs expression and RS groups (high-risk and low-risk group) by T Test. The results showed that MIR4458HG and MAPT-AS1 was lower expression in the high-risk group compared with the low-risk group, respectively, while CYTOR was higher expression in the high-risk group (Fig. 4).

Table 2. Multivariate Cox regression analysis of clinicopathological parameters and RS in breast cancer

Clinicopathological parameters	HR	P-value	Low 95% CI	High 95% CI
RS	3.040	0.000	1.634	5.658
TNM stage	0.404	0.002	0.229	0.713
Age	0.611	0.628	0.084	4.470
T	1.097	0.781	0.571	2.107
N	1.098	0.797	0.539	2.236
M	0.171	0.002	0.056	0.518
HER-2	1.207	0.936	0.541	1.948
ER	1.802	0.214	0.712	4.559
PR	0.837	0.693	0.346	2.026

HR; Hazard ratio; CI; Confidence interval

Evaluation RS Performance By ROC Curve Analysis

We analyzed ROC to show sensitivity and specificity of survival prediction in our model. ROC curves indicated that AUC of the three-lncRNA signature was 0.802 (Fig. 5).

Functional Prediction Of Three LncRNAs

Functional enrichment analysis tools, including GO terms and KEGG pathways were used to predict potential biological processes for three lncRNAs. The results revealed that 1110 terms were enriched, including 954 BP terms, 24 KEGG pathways, and so on. The three lncRNAs may partake following biological processes, such as protein phosphorylation of post-translational, positive regulation of cell migration, regulation of apoptotic signaling pathway, positive regulation of cell adhesion, etc., which may influence the onset and progress of breast cancer by means of the breast cancer-associated lncRNAs. The top 20 clusters of significantly enriched terms are shown in Fig. 6.

Discussion

LncRNAs are involved in carcinogenesis by performing their multiple biological functions. Scientists have revealed a number of mechanisms of lncRNAs in multiple malignant tumors including breast cancer. For example, lncRNAs act as a signal or decoy to promote or suppress gene expression [22]. The lncRNAs regulate the translation of mRNAs and control their stability via forming double-stranded RNA with mRNAs or regulate protein stability by binding [23]. The expression of more and more lncRNAs are abnormal in cancers, and their potential as possible prognostic biomarkers can be used as therapeutic targets [15, 24].

In this study, we firstly screened out the differentially expressed lncRNAs, and next identified three prognostic lncRNAs: CYTOR, MIR4458HG and MAPT-AS1. CYTOR is also called LINC00152, previous studies showed that the expression of lncRNA CYTOR was up-regulated in multiple malignant diseases, such as gastrointestinal cancer, liver cancer, lung adenocarcinoma and esophageal squamous cell carcinoma [25–28]. Yue et al. [29] showed that CYTOR promoted metastasis of colon cancer through Wnt/ β -catenin signal pathway. Reon et al. [30] reported that CYTOR promoted invasion via a 3'-hairpin structure, and as a effective biomarker for predicting survival of glioblastoma. As far as we knew, there were just a few of studies on CYTOR in breast cancer. It might be involved in the EGFR/mammalian target of rapamycin pathway, which promoted triple negative breast cancer progression by affecting the stability of PTEN protein [31]. Previous study indicated the high expression of MAPT-AS1 was correlated with better survival in breast cancer [32]. However, we have not found any report about MIR4458HG related to breast cancer. Our data showed above three lncRNAs were associated with OS in breast cancer. Furthermore, MAPT-AS1 and MIR4458HG were lower expression in the high-risk group compared with the low-risk group, respectively, while CYTOR was higher expression in the high-risk group, indicated that our results were consistent with previous study. However, the studies about the three lncRNAs are still rare reports in breast cancer, so they are worth to further study as the good targets in the future.

We also structured a prognostic signature with three lncRNAs according to the coefficient from the multivariate cox regression analysis and the expression profiles. We calculated the RS for each patient in selected 565 cases of breast cancer and divided them into two groups of low-risk and high-risk. We found that patients in the high-risk group showed significantly shorter OS than those in the low-risk group, it might be used as an independent predictor for OS. Moreover, the ROC curve analysis suggested that three lncRNAs revealed high sensitivity and specificity of survival prediction in our model. As far as we knew, this was the first investigation to establish a prognostic signature with RNA-sequence, for the survival evaluation of 565 cases with breast cancer basing on TCGA data. Previous studies explored a few of prognostic signatures with lncRNAs from another database-Gene Expression Omnibus (GEO). For instance, Sun et al. [4] analyzed lncRNA expression profiles in breast cancer via repurposing microarray probes from three GEO datasets (GSE25066, GSE4922, and GSE1456). Their results showed that an expression pattern of nine lncRNAs was associated with metastasis-free survival (MFS), including RP11-482H16.1, AC010729.1, RP11-983P16.4, FOXD3-AS1, LINC01249, AC096574.4, AC015971.2, AC012487.2 and RP11-15A1.2. Univariate analysis demonstrated that the nine-lncRNA signature was a prognostic biomarker to predict MFS of breast cancer. The ROC curve analysis showed the AUC of prognostic signature was 0.693. Similarly, Zhou et al. [33] identified a twelve –lncRNA prognostic signature in

recurrence-free survival (RFS) of breast cancer from GEO datasets. These lncRNAs included RP1-34M23.5, RP11-202K23.1, RP11-560G2.1, RP4-613B23.1, RP11-360F5.1, CTD-2031P19.5, RP4-591L5.2, RP13-104F24.2, RP11-506D12.5, ERVH48-1, RP11-247A12.8 and SNHG7. However, they selected MFS and RFS as research endpoint, respectively, while we used OS as endpoint in this study, because OS directly reflected the survival of cancer patients. In addition, the sensitivity and specificity of our model were superior to the study conducted by Sun et al. Moreover, they only conducted univariate analysis without multivariate analysis, which might affect the final prognostic value. Zhou et al. only mined data from microarray without making in-depth analysis, such as co-expression network analysis with mRNA, enrichment analysis and prediction of biological function, and so on. We have made further analysis and screened out mRNAs, which were associated with CYTOR, MIR4458HG and MAPT-AS1 by co-expression network analysis. We also applied GO terms and KEGG pathways to predict potential biological processes of the three lncRNAs with mRNAs, and obtained 1110 terms enrichment, including 954 BP terms, 24 KEGG pathways, and so on. There were a number of functions, including post-translational protein phosphorylation, positive regulation of cell migration, regulation of apoptotic signaling pathway, positive regulation of cell adhesion, etc. in relationship with important biological processes and signaling pathways in cancer. Their functions above were associated with the three lncRNAs, therefore, three lncRNAs which we screened out might be as the targets to influence the onset and progress of breast cancer. Sun et al. and Zhou et al. applied additional independent GEO datasets to validate their lncRNA signatures. We also attempted to validate the prognostic value of the three-lncRNA signature from GEO database. Regrettably, insufficient cases of these three lncRNAs could be provided from these GEO datasets, therefore, the prognostic value of the three-lncRNA signature needs to be further validated with other methods, and confirmed with a larger sample size in the future.

Conclusions

In a word, we evaluated the lncRNA data of 565 cases with breast cancer from TCGA database and established the three-lncRNA signature, which might be a novel prognostic biomarker for breast cancer patients. However, it needs more clinical practice to validate prognostic value, and more experiments to explore biological function of these lncRNAs in breast cancer.

Abbreviations

HER2: human epidermal growth factor receptor 2; lncRNAs: long non-coding RNAs; OS: overall survival; TCGA: The Cancer Genome Atlas; T: tumor size; N: lymph nodes; M: distant metastasis; ER/PR: estrogen/progesterone receptor; RS: risk score; exp: expression level; β : regression coefficient; ROC: receiver operating characteristic; AUC: area under characteristic; GO: Gene Ontology; BP: biological process; KEGG: Kyoto Encyclopedia of Genes and Genomes; GEO: Gene Expression Omnibus; MFS: metastasis-free survival; RFS: recurrence-free survival.

Declarations

Authors' contributions

GCL designed the study. GCL, XLY, and RLG prepared the manuscript. YNZ, JYL, and WL collected and processed data. GCL, XLY, and RLG analyzed data. RLG acquired funding. All authors read and approved the final manuscript.

Acknowledgments

We are grateful to all patients and healthy donors for donating their samples. We thank TCGA for providing huge amounts of clinical and genetic data. G.L. would like to thank his wife, Lyra, for her support. It's because of her contribution that he can concentrate on his research.

Competing interests

The authors declare that they have no competing interests

Funding

This work was supported by the National Natural Science Foundation of China (No. 81774068), Natural Science Foundation of Zhejiang Province (No. LY20H290004).

Availability of data and materials

All data generated and analyzed during this study are included in this published article.

Ethics approval and consent to participate

Not applicable.

Consent for publication

Not applicable.

References

1. Lan, G.; Li, J.; Wen, Q.; Lin, L.; Chen, L.; Chen, L.; Chen, X. Cytotoxic T lymphocyte associated antigen 4 expression predicts poor prognosis in luminal B HER2-negative breast cancer. *Oncol Lett.* 2018, 15, 5093-5097, doi: 10.3892/ol.2018.7991.
2. Bray, F.; Ferlay, J.; Soerjomataram, I.; Siegel, R.L.; Torre, L.A.; Jemal, A. Global cancer statistics 2018: GLOBOCAN estimates of incidence and mortality worldwide for 36 cancers in 185 countries. *CA Cancer J Clin.* 2018, 68, 394-424, doi: 10.3322/caac.21492.
3. NATIONAL CANCER INSTITUTE: Surveillance, Epidemiology, and End Results (SEER) Program. Available online: <https://seer.cancer.gov/statfacts/html/breast.html>.

4. Sun, J.; Chen, X.; Wang, Z.; Guo, M.; Shi, H.; Wang, X.; Cheng, L.; Zhou, M. A potential prognostic long non-coding RNA signature to predict metastasis-free survival of breast cancer patients. *Sci Rep*. 2016, 5, 16553, doi: 10.1038/srep16553.
5. Rodenhiser, D.I.; Andrews, J.D.; Vandenberg, T.A.; Chambers, A.F. Gene signatures of breast cancer progression and metastasis. *Breast Cancer Res*. 2011, 13, 201, doi: 10.1186/bcr2791.
6. Network, T.C.G.A. Comprehensive molecular portraits of human breast tumors. *Nature*. 2012, 490, 61-70, doi: 10.1038/nature11412.
7. Sparano, J.A.; Gray, R.J.; Makower, D.F.; Pritchard, K.I. Adjuvant Chemotherapy Guided by a 21-Gene Expression Assay in Breast Cancer. *N Engl J Med*. 2018, 379, 111-121, doi: 10.1056/NEJMoa1804710.
8. Paik, S.; Tang, G.; Shak, S.; Kim, C.; Baker, J.; Kim, W.; Cronin, M.; Baehner, F. L.; Watson, D.; Bryant, J. Gene Expression and Benefit of Chemotherapy in Women With Node-Negative, Estrogen Receptor-Positive Breast Cancer. *J Clin Oncol*. 2006, 24, 3726-3734, doi: 10.1200/JCO.2005.04.7985.
9. Lænkholm, A.V.; Jensen, M.B.; Eriksen, J.O.; Rasmussen, B.B.; Knoop, A.S.; Buckingham, W.; Ferree, S.; Schaper, C.; Nielsen, T.O.; Haffner, T. PAM50 Risk of Recurrence Score Predicts 10-Year Distant Recurrence in a Comprehensive Danish Cohort of Postmenopausal Women Allocated to 5 Years of Endocrine Therapy for Hormone Receptor–Positive Early Breast Cancer. *J Clin Oncol*. 2018, 36, 735-740, doi: 10.1200/JCO.2017.74.6586.
10. Sestak, I.; Richard, B.; Cuzick, J.; Dubsky, P.; Kronenwett, R.; Denkert, C.; Ferree, S.; Sgroi, D.; Schnabel, C.; Baehner, F.L.; Mallon, E.; Dowsett, M. Comparison of the Performance of 6 Prognostic Signatures for Estrogen Receptor–Positive Breast Cancer. *Jama Oncol*. 2018, 4, 545-553, doi: 10.1001/jamaoncol.2017.5524.
11. Quinn, J.J.; Howard, Y. Unique features of long non-coding RNA biogenesis and function. *Nat Rev Genet*. 2016, 17, 47-62, doi: 10.1038/nrg.2015.10.
12. Zhang, R.; Xia, L.; Lu, Q.; Wen, W.; Zhang, J.; Zhu, J. LncRNAs and cancer. *Oncol Lett*. 2016, 12, 1233-1239, doi: 10.3892/ol.2016.4770.
13. Zhang, X.; Sun, S.; Lam, K.; Kiang, K.; Pu, J.; Ho, A.; Lui, W.; Fung, C.; Wong, T.; Leung, G. A long non-coding RNA signature in glioblastoma multiforme predicts survival. *Neurobio Dis*. 2013, 58, 123-131, doi: 10.1016/j.nbd.2013.05.011.
14. Xing, Z.; Lin, A.; Li, C.; Liang, K.; Wang, S.; Liu, Y.; Park, P.K.; Qin, L.; Wei, Y.; Hawke, D.H.; Hung, M.C.; Lin, C.; Yang, L. lncRNA Directs Cooperative Epigenetic Regulation Downstream of Chemokine Signals. *Cell*. 2014, 159, 1110-1125, doi: 10.1016/j.cell.2014.10.013.
15. Wang, Y.; Song, B.; Zhu, L.; Zhang, X. Long non-coding RNA, LINC01614 as a potential biomarker for prognostic prediction in breast cancer. *Peer J*. 2019, 7, e7976, doi: 10.7717/peerj.7976.
16. Liu, B.; Sun, L.; Liu, Q.; Gong, C.; Yao, Y.; Lv, X.; Lin, L.; Yao, H.; Su, F.; Li, D.; Zeng, M.; Song, E. A Cytoplasmic NF-κB Interacting Long Noncoding RNA Blocks IκB Phosphorylation and Suppresses Breast Cancer Metastasis. *Cancer Cell*. 2015, 27, 370-381, doi: 10.1016/j.ccell.2015.02.004.

17. Wu, W.; Chen, F.; Cui, X.; Yang, L.; Chen, J.; Zhao, J.; Huang, D.; Liu, J.; Yang, L.; Zeng, J. LncRNA NKILA suppresses TGF- β -induced epithelial-mesenchymal transition by blocking NF- κ B signaling in breast cancer. *Int J Cancer*. 2018, 143, 2213-2224, doi: 10.1002/ijc.31605.
18. Shi, X.; Tang, X.; Su, L. Overexpression of Long Noncoding RNA PTENP1 Inhibits Cell Proliferation and Migration via Suppression of miR-19b in Breast Cancer Cells. *Oncol Res*. 2018, 26, 869-878, doi: 10.3727/096504017X15123838050075.
19. Xing, F.; Liu, Y.; Wu, S.; Wu, K.; Sharma, S.; Mo, Y.; Feng, J.; Sanders, S.; Jin, G.; Singh, R.; Vidi, P.A.; Tyagi, A.; Chan, M.D.; Ruiz, J.; Debinski, W.; Pasche, B.C.; Lo, H.W.; Metheny-Barlow, L.J.; D'Agostino, R.B.J.; Watabe, K. Loss of XIST in Breast Cancer Activates MSN-c-Met and Reprograms Microglia via Exosomal miRNA to Promote Brain Metastasis. *Cancer Res*. 2018, 78, 4316-4330, doi: 10.1158/0008-5472.CAN-18-1102.
20. Zheng, R.; Lin, S.; Guan, L.; Yuan, H.; Liu, K.; Liu, C.; Ye, W.; Liao, Y.; Jia, J.; Zhang, R. Long non-coding RNA XIST inhibited breast cancer cell growth, migration, and invasion via miR-155/CDX1 axis. *Biochem Biophys Res Commun*. 2018, 498, 1002-1008, doi: 10.1016/j.bbrc.2018.03.104.
21. Zhang, W.; Shi, S.; Jiang, J.; Li, X.; Lu, H.; Ren, F. LncRNA MEG3 inhibits cell epithelial-mesenchymal transition by sponging miR-421 targeting E-cadherin in breast cancer. *Biomed Pharmacother*. 2017, 91, 312-319, doi: 10.1016/j.biopha.2017.04.085.
22. Hung, T.; Wang, Y.; Lin, M.; Koegel, A.K.; Kotake, Y.; Grant, G.D.; Horlings, H.M.; Shah, N.; Umbricht, C.; Wang, P.; Wang, Y.; Kong, B.; Langerød, A.; Børresen-Dale, A.L.; Kim, S.K.; van de, V.M.; Sukumar, S.; Whitfield, M.L.; Kellis, M.; Xiong, Y.; Wong, D.J.; Chang, H.Y. Extensive and coordinated transcription of noncoding RNAs within cell-cycle promoters. *Nat Genet*. 2011, 43, 621-629, doi: 10.1038/ng.848.
23. Yuan, J.; Yang, F.; Wang, F.; Ma, J.; Guo, Y.; Tao, Q.; Liu, F.; Pan, W.; Wang, T.; Zhou, C.; Wang, S.; Wang, Y.; Yang, Y.; Yang, N.; Zhou, W.; Yang, G.; Sun, S. A long noncoding RNA activated by TGF- β promotes the invasion-metastasis cascade in hepatocellular carcinoma. *Cancer Cell*. 2014, 25, 666-681, doi: 10.1016/j.ccr.2014.03.010.
24. Meng, Y.; Liu, Y.; Li, K.; Fu, T. Prognostic value of long non-coding RNA breast cancer anti-estrogen resistance 4 in human cancers: A meta-analysis. *Medicine*. 2019, 98, e15793, doi: 10.1097/MD.00000000000015793.
25. Neumann, O.; Kesselmeier, M.; Geffers, R.; Pellegrino, R.; Radlwimmer, B.; Hoffmann, K.; Ehemann, V.; Schemmer, P.; Schirmacher, P.; Lorenzo Bermejo, J.; Longerich, T. Methyloome analysis and integrative profiling of human HCCs identify novel protumorigenic factors. *Hepatology*. 2012, 56, 1817-27, doi: 10.1002/hep.25870.
26. Cao, W.; Wu, H.; He, B.; Zhang, Y.; Zhang, Z. Analysis of long non-coding RNA expression profiles in gastric cancer. *World J Gastroenterol*. 2013, 19, 3658-3664, doi: 10.3748/wjg.v19.i23.3658.
27. Yang, S.; Ning, Q.; Zhang, G.; Sun, H.; Wang, Z.; Li, Y. Construction of differential mRNA-lncRNA crosstalk networks based on ceRNA hypothesis uncover key roles of lncRNAs implicated in esophageal squamous cell carcinoma. *Oncotarget*. 2016, 7, 85728-85740, doi: 10.18632/oncotarget.13828.

28. Zhang, J.; Li, W. Long noncoding RNA CYTOR sponges miR-195 to modulate proliferation, migration, invasion and radiosensitivity in nonsmall cell lung cancer cells. *Biosci Rep.* 2018, 38, pii: BSR20181599, doi: 10.1042/BSR20181599.
29. Yue, B.; Liu, C.; Sun, H.; Liu, M.; Song, C.; Cui, R.; Qiu, S.; Zhong, M. A Positive Feed-Forward Loop between LncRNA-CYTOR and Wnt/ β -Catenin Signaling Promotes Metastasis of Colon Cancer. *Mol Ther.* 2018, 26, 1287-1298, doi: 10.1016/j.ymthe.2018.02.024.
30. Reon, B.J.; Takao Real Karia, B.; Kiran, M.; Dutta, A. LINC00152 Promotes Invasion through a 3'-Hairpin Structure and Associates with Prognosis in Glioblastoma. *Mol Cancer Res.* 2018, 16, 1470-1482, doi: 10.1158/1541-7786.MCR-18-0322.
31. Shen, X.; Zhong, J.; Yu, P.; Zhao, Q.; Huang, T. YY1-regulated LINC00152 promotes triple negative breast cancer progression by affecting on stability of PTEN protein. *Biochem Biophys Res Commun.* 2019, 509, 448-454, doi: 10.1016/j.bbrc.2018.12.074.
32. Wang, D.; Li, J.; Cai, F.; Xu, Z.; Li, L.; Zhu, H.; Liu, W.; Xu, Q.; Cao, J.; Sun, J.; Tang, J. Overexpression of MAPT-AS1 is associated with better patient survival in breast cancer. *Biochem Cell Biol.* 2019, 97, 158-164, doi: 10.1139/bcb-2018-0039.
33. Zhou, M.; Zhong, L.; Xu, W.; Sun, Y.; Zhang, Z.; Zhao, H.; Yang, L.; Sun, J. Discovery of potential prognostic long non-coding RNA biomarkers for predicting the risk of tumor recurrence of breast cancer patients. *Sci Rep.* 2016, 6, 31038 doi: 10.1038/srep31038.

Figures

Figure 1

Volcano plot of the differentially expressed lncRNAs between breast cancer and normal tissues. Red and green dot indicates high expression and low expression, respectively. Black dot shows the lncRNA expression with both the $\log_2\text{FC} < 1$ and $\text{P-value} < 0.01$. The Y axis represents a P-value and the X axis represents a $\log_2\text{FC}$.

Figure 2

LncRNA risk score analysis of breast cancer. (a) The low-risk and high-risk score groups for the three-lncRNA signature in breast cancer patients; (b) The survival status and duration of breast cancer patients; (c) Heatmap of the three lncRNAs expression in breast cancer. The color from blue to red shows a trend from low expression to high expression.

Figure 3

Kaplan–Meier curve for the three-lncRNA signature in breast cancer. Kaplan-Meier survival curve shows OS outcomes according to relative high-risk and low-risk patients.

Figure 4

Different expression of the three lncRNAs between high-risk and low-risk group. (a) MIR4458HG; (b) MAPT-AS1; (c) CYTOR.

Figure 5

ROC curve for the three-lncRNA signature in breast cancer. Time-dependent ROC curve analysis for survival prediction by the three-lncRNA signature.

Figure 6

Functional enrichment results. (a) Network of top 20 clusters of enriched terms. Each node represents one enriched term colored by cluster ID, nodes share the same cluster are typically close to each other. The thicker edge displayed, the higher the similarity is. The term with the best P-value was selected to represent each cluster; (b) Heatmap of top 20 clusters, colored by P-values. The smaller the P-value is the deeper the color is.