

# Bioinformatics analysis of prognosis-related long non-coding RNAs in invasive breast carcinoma

YUANYUAN HU, XIDONG GU, YIN DUAN, YONG SHEN and XIAOHONG XIE

Department of Breast Surgery, The First Affiliated Hospital of Zhejiang Chinese Medical University  
(Zhejiang Provincial Hospital of Traditional Chinese Medicine), Hangzhou, Zhejiang 310006, P.R. China

Received September 25, 2019; Accepted February 7, 2020

DOI: 10.3892/ol.2020.11558

**Abstract.** Breast cancer is one of the most common types of cancer among women worldwide and needs more sensitive prognostic biomarkers to improve its treatment. In the present study, differentially expressed long non-coding RNAs (lncRNAs) in invasive breast carcinoma from The Cancer Genome Atlas and cBioPortal database were investigated, identifying 292 differentially expressed lncRNAs in 1,100 cases. By analyzing the overall survival rate, 10 lncRNAs were significantly correlated with poor prognosis. To explore the underlying molecular mechanisms of the 10 prognosis-related lncRNAs, bioinformatic methods were used to predict the potential target miRNAs, mRNAs and proteins, and to construct a lncRNA-miRNA-mRNA regulatory network and lncRNA-protein interaction network. Finally, the functions of the target genes and proteins were investigated using Gene Ontology enrichment and Kyoto Encyclopedia of Genes and Genomes pathway analyses. The results showed that these 10 lncRNAs could be novel prognostic markers for invasive breast carcinoma and the present study aimed to provide novel insight into the diagnosis and treatment of breast cancer.

## Introduction

Breast cancer is one of the most common types of cancer worldwide among women (1). Although progress has been made in the diagnosis and treatment of breast cancer, it is the second leading cause of cancer-associated death among

women in the United States (1,2). To reduce the mortality rate of breast cancer, early detection and improved therapy is needed. It is important to identify more novel sensitive prognostic biomarkers for breast cancer, especially for invasive breast carcinoma.

Long non-coding RNAs (lncRNAs) are non-coding transcripts, usually >200 nucleotides in length (3). Increasing evidence suggests that lncRNAs are important regulators in multiple biological processes and that lncRNAs with aberrant expression levels are involved in cancer development and may be potential diagnostic biomarkers for cancer (3-5). MicroRNAs (miRNAs) are a class of small non-coding RNAs which serve an important role in the gene regulation network (6). miRNAs target the coding DNA sequence region or 3'untranslated region (3'UTR) of mRNAs, inducing mRNA degradation or translational repression (7,8).

lncRNAs have been identified as miRNA sponges that compete for miRNA binding with mRNAs (9). lncRNA H19 was found to promote the epithelial to mesenchymal transition in human colon cancer cells by sponging miR-138 and miR-200a and eliminating the inhibition of vimentin, ZEB1 and ZEB2 induced by the two miRNAs (9). Moreover, lncRNAs can exert their biological effects by interacting with proteins. For example, it was reported that lncRNA GATA6-AS negatively regulates nuclear LOXL2 function and regulates endothelial gene expression via interaction with the epigenetic regulator LOXL2 in endothelial cells (10).

In the present study, differentially expressed lncRNAs in invasive breast carcinoma from The Cancer Genome Atlas (TCGA) database and cBioPortal (11,12) were investigated, identifying 292 differentially expressed lncRNAs in 1,100 invasive breast carcinoma cases, including 10 prognosis-related lncRNAs. Subsequently, to explore the molecular mechanisms of these 10 prognosis-related lncRNAs, bioinformatics methods were used to predict the potential target miRNAs, mRNAs and proteins, and to construct a lncRNA-miRNA-mRNA regulatory network and lncRNA-protein interaction network. Finally, the target mRNAs and proteins were annotated using Gene Ontology (GO) enrichment and Kyoto Encyclopedia of Genes and Genomes (KEGG) pathway analyses using The Database for Annotation, Visualization and Integrated Discovery tools (DAVID) (13,14).

---

*Correspondence to:* Dr Yong Shen or Professor Xiaohong Xie, Department of Breast Surgery, The First Affiliated Hospital of Zhejiang Chinese Medical University (Zhejiang Provincial Hospital of Traditional Chinese Medicine), 54 Youdian Road, Hangzhou, Zhejiang 310006, P.R. China  
E-mail: shenyong@zcmu.edu.cn  
E-mail: xxh666857@163.com

**Key words:** invasive breast carcinoma, prognosis, overall survival, competitive endogenous RNA network, long non-coding RNA-protein interaction

## Materials and methods

**Identification of differentially expressed lncRNAs and prognosis-related lncRNAs.** The HUGO Gene Nomenclature Committee (HGNC) database ([genenames.org/](http://genenames.org/)) is a resource for approved human gene nomenclature and gene symbols of lncRNAs were obtained from this site. The cBioPortal database ([cbioportal.org/](http://cbioportal.org/)) was used to visualize, analyze and download several large-scale cancer genomics datasets from databases such as TCGA and GEO (11,12). In this study, the TCGA Breast Invasive Carcinoma dataset (TCGA, Firehose Legacy) was selected, which contains 1,100 invasive breast carcinoma samples with RNA-seq data (11,12). The subtypes of breast carcinoma were not distinguished in the present study. The gene symbols of lncRNAs were submitted to the cBioPortal database and the differentially expressed lncRNAs were obtained. To investigate whether these differentially expressed lncRNAs were correlated with invasive breast carcinoma prognosis, these lncRNAs were analyzed using the Kaplan-Meier estimate of overall survival and disease free survival rate in cBioPortal (11,12), and the log-rank test were statistically significant at a  $P < 0.05$  level. The median overall survival time was also obtained from the database.

**lncRNA-miRNA-mRNA regulatory network construction.** lncRNA sequences were obtained from National Center for Biotechnology Information database ([ncbi.nlm.nih.gov/](http://ncbi.nlm.nih.gov/)). To predict the miRNAs which could be sponged by these lncRNAs, the lncRNA sequences were submitted to the RegRNA database (<http://regrna2.mbc.nctu.edu.tw/>) (15), which selected the predicted miRNA target sites according to the follow criteria: System score  $\geq 150$  and minimum folding free energy  $\leq -25$ . To reduce the number of false positive results, the top 10 ranked miRNAs were selected as potential targets.

Subsequently, the target mRNAs of miRNAs were predicted using the TargetScan database ([http://www.targetscan.org/vert\\_72/](http://www.targetscan.org/vert_72/)) (16), which is widely used in the prediction of miRNA targets and has been updated with a modified set of representative transcripts and different miRNA annotations in March 2018. Using the TargetScan database, predictions were ranked according to the predicted efficacy of targeting and the probability of conserved targeting. Then, the top 10 ranked mRNAs were selected and submitted to the cBioPortal database to screen for aberrantly expressed mRNAs (define as alternation frequency  $\geq 5\%$ ) as potential targets of miRNAs. The lncRNA-miRNA-mRNA regulatory network was constructed using Cytoscape version 3.6.1 (17).

**lncRNA-protein interactions prediction.** In order to predict the lncRNA-binding proteins, the lncRNA symbols were submitted to the RNAc database (<https://rnact.crg.eu/>) (18) and the top 10 ranked proteins were selected based on the prediction score to minimize the likelihood of false positive results. The lncRNA-protein interaction network was constructed using Cytoscape version 3.6.1.

**GO enrichment and KEGG pathway analysis.** GO enrichment and KEGG pathway analyses were performed to further investigate the functions of the target genes and proteins. GO and KEGG analyses were performed using DAVID

Table I. The median overall survival of patients with invasive breast carcinoma with and without alterations in different genes.

Gene	Median survival, months	Log-rank test P-value
NORAD		$5.69 \times 10^{-3}$
With alteration	91.92	
Without alteration	130.06	
SNHG1		0.0453
With alteration	114.06	
Without alteration	129.6	
LINC00654		0.0242
With alteration	94.15	
Without alteration	129.6	
FAM157A		0.0232
With alteration	114.72	
Without alteration	129.6	
DLEU2		$2.17 \times 10^{-03}$
With alteration	93.76	
Without alteration	129.47	
LINC01559		0.0218
With alteration	77.56	
Without alteration	129.47	
SPATA8		0.0444
With alteration	74.67	
Without alteration	129.47	
TCL6		$2.72 \times 10^{-04}$
With alteration	81.57	
Without alteration	129.47	
LINC00632		0.0104
With alteration	33.9	
Without alteration	129.47	
PWRN1		0.0164
With alteration	74.67	
Without alteration	129.6	

version 6.8 (<https://david.ncifcrf.gov/>) (13,14) and GO analysis was limited to biological process and molecular function.  $P < 0.05$  was considered to indicate a statistically significant difference.

## Results

**Identification of differentially expressed lncRNAs associated with prognosis in invasive breast carcinoma.** A total of 4,525 lncRNA gene symbols were obtained from HGNC, of which 292 differentially expressed lncRNAs were identified in 1,100 patients with invasive breast carcinoma. Among these, 8 lncRNAs (PVT1, LINC00467, SNHG6, SNHG28, NORAD, SNHG20, CASC15 and LINC00847) exhibited aberrant expression levels in  $\geq 10\%$  of cases, accounting for 65% of patients with invasive breast carcinoma (Fig. 1A).

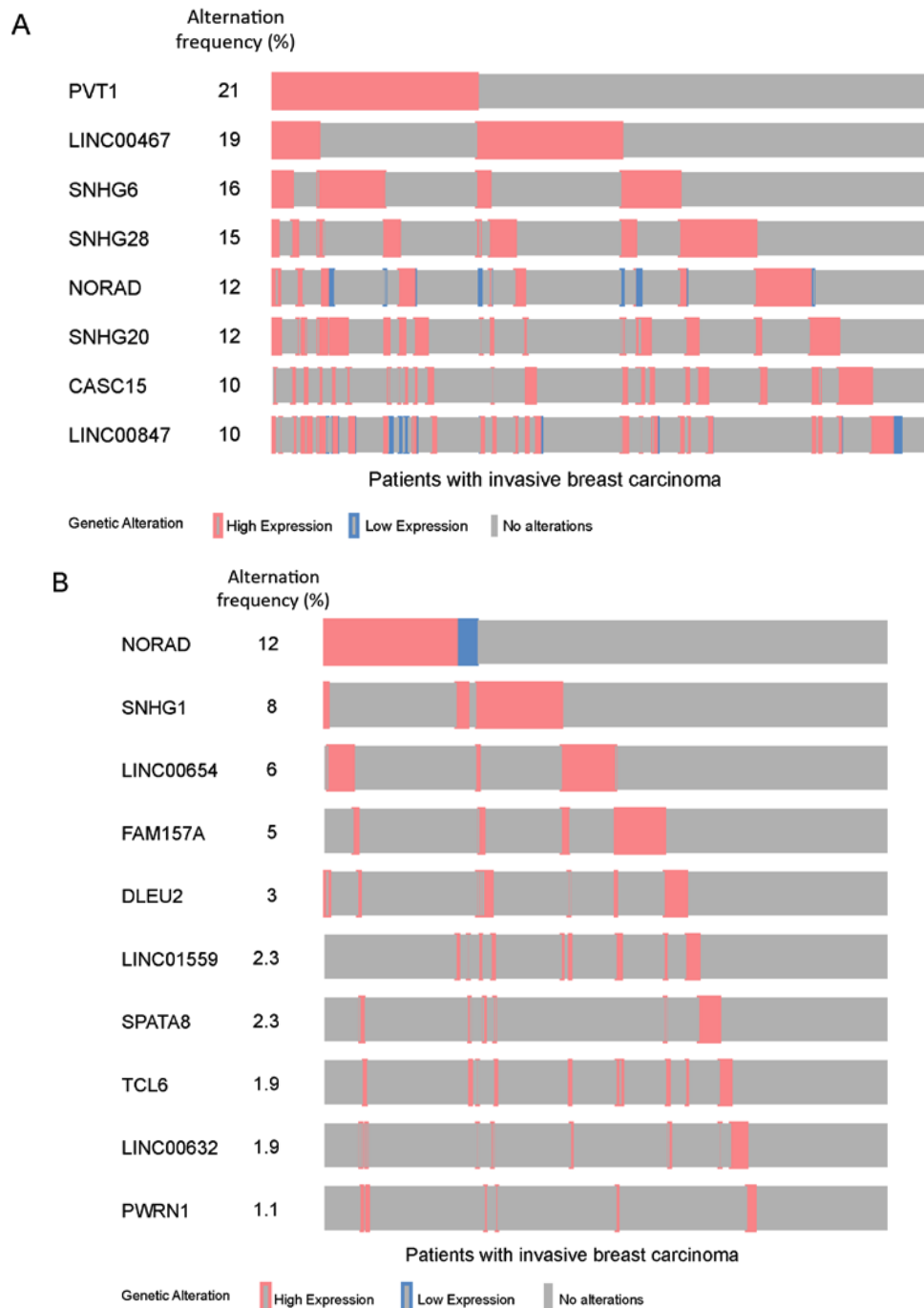


Figure 1. Identification of differentially expressed lncRNAs associated with prognosis in 1,100 patients with invasive breast carcinoma. (A) Differentially expressed lncRNAs with alteration frequency  $\geq 10\%$ . (B) Differentially expressed lncRNAs associated with prognosis.

To investigate whether differentially expressed lncRNAs were associated with the prognosis of invasive breast carcinoma, the 292 differentially expressed lncRNAs were submitted to the cBioPortal database and overall survival of patients with invasive breast carcinoma was analyzed. As shown in Fig. 1B, the 10 lncRNAs (NORAD, SNHG1, LINC00654, FAM157A, DLEU2, LINC01559, SPATA8, TCL6, LINC00632 and PWRN1) with expression alterations were significantly associated with poor prognosis in invasive breast carcinoma. The overall survival rate in patients with expression alterations of the 10 lncRNAs were lower than in patients without expression alterations (log-rank test  $P < 0.05$ ; Fig. 1C), and the

median overall survival time in cases with expression alterations of these 10 lncRNAs were shorter compared with cases without expression alterations (Table I). Moreover, among the 10 lncRNAs, DLEU2, TCL6 and PWRN1 expression alterations were significantly correlated with disease free survival rate, the disease free survival rate in cases with expression alterations of the 3 lncRNAs were lower than cases without expression alterations (log-rank test  $P < 0.05$ ; Fig. 1D).

**lncRNA-miRNA-mRNA regulatory network analysis.** Firstly, to explore the underlying molecular mechanisms of the 10 prognosis-related lncRNAs, 88 miRNAs which could be sponged

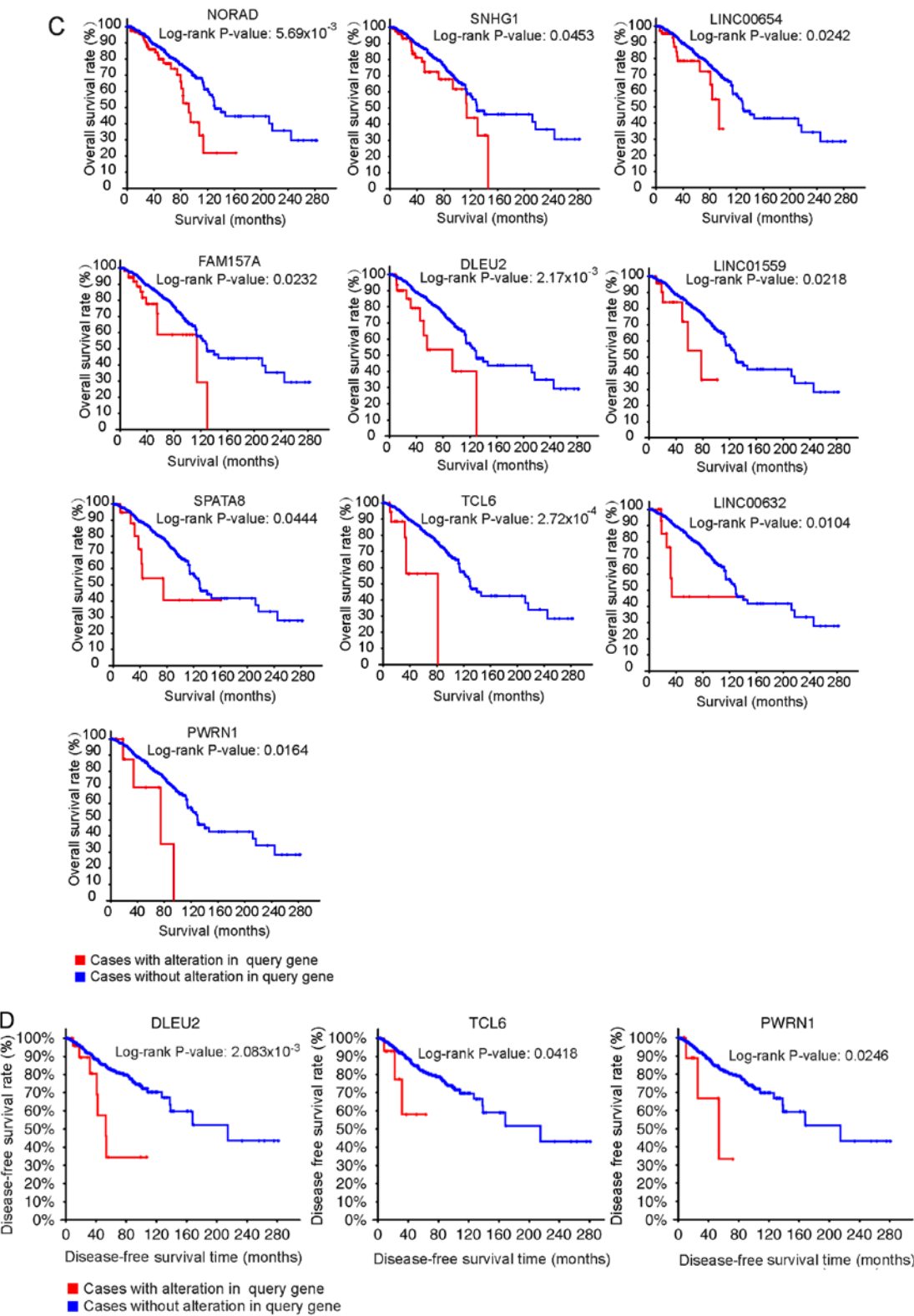


Figure 1. Continued. (C) Kaplan-Meier overall survival analysis results of 10 prognosis-related lncRNAs. Expression alterations of the 10 lncRNAs were associated with lower overall survival rate. (D) Kaplan-Meier disease free survival analysis results of 3 prognosis-related lncRNAs. Expression alterations of the three lncRNAs were associated with lower disease free survival rate. lncRNA, long non-coding RNA.

by these 10 lncRNAs were predicted using the RegRNA database. Secondly, 765 target mRNAs of these 88 miRNAs were predicted using the TargetScan database. Thirdly, to improve the prediction accuracy, 322 aberrantly expressed target mRNAs were screened as potential targets using the

cBioPortal database. Overall, the lncRNA-miRNA-mRNA regulatory network was composed of 10 lncRNAs, 88 miRNAs and 322 mRNAs, the red triangles represent lncRNAs, the blue squares represent miRNAs, and green circles represent mRNAs (Fig. 2).



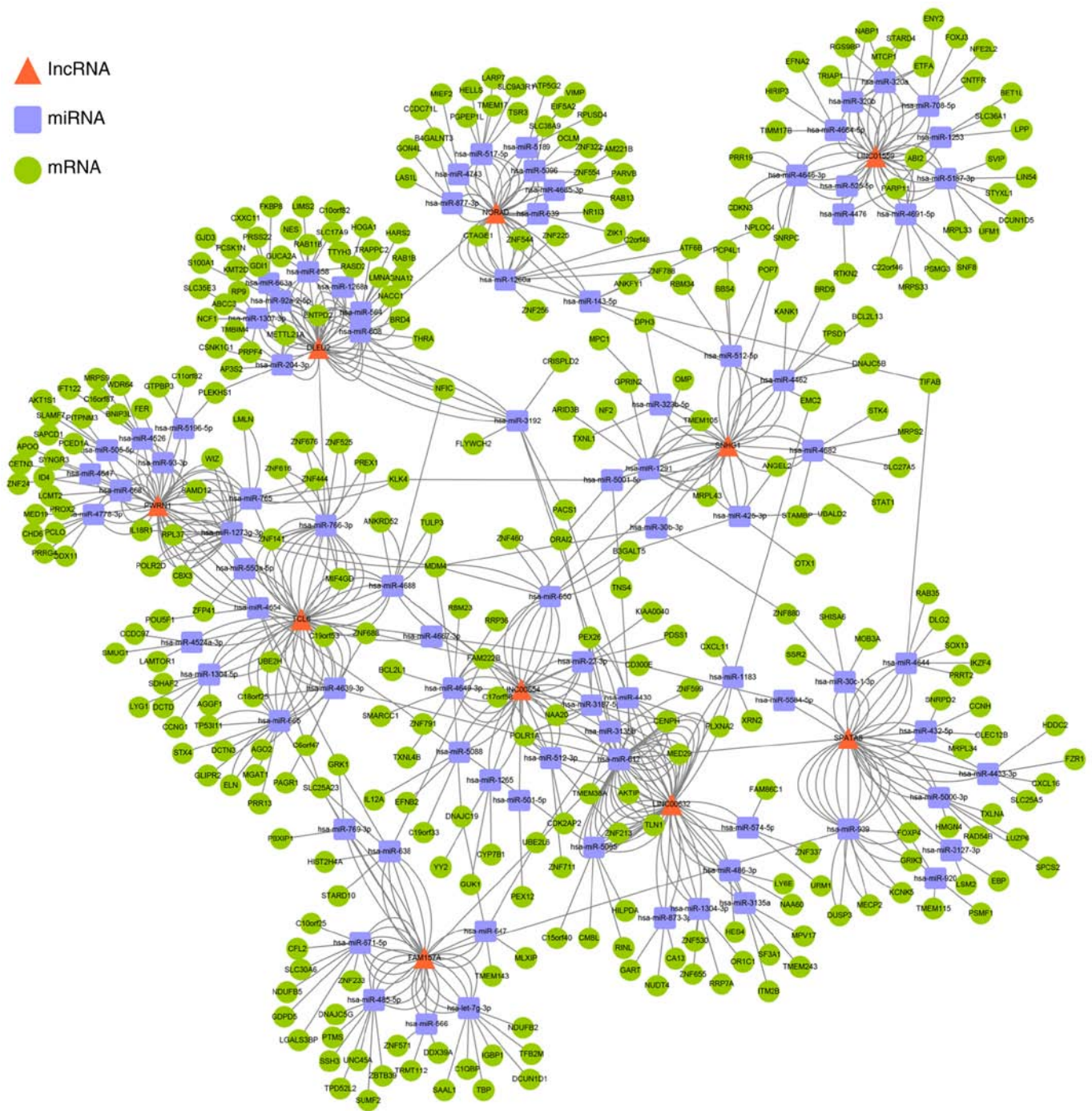


Figure 2. Interaction network of lncRNA-miRNA-mRNA in patients with invasive breast carcinoma. lncRNA, long non-coding RNA; miRNA, microRNA.

**lncRNA-protein interaction analysis.** RNA-binding proteins are important to a number of cellular processes, including mRNA splicing, export, stability and translation (19), and lncRNAs could bind to proteins forming RNA-binding proteins to exert their biological functions. The RNAc database was a useful tool for RNA-binding proteins prediction and the results showed that there were 25 potential target proteins of the 10 lncRNAs (Fig. 3) and different lncRNAs might target the same protein, such as, PWRN1, TCL6, LINC00632, SPATA8 and SNHG1 had the same target protein RBICC1.

Combining the above two interaction networks, lncRNAs could indirectly regulate mRNAs through sponging miRNAs, abrogating the inhibition of mRNA expression levels or

translation induced by miRNAs. In addition, the lncRNAs could regulate the target proteins by forming RNA-protein complexes (Fig. 4).

**Functional analysis of target mRNAs and proteins.** To further investigate the functions of the target genes and proteins, the official gene symbols of the targets were submitted to the DAVID database, in which GO enrichment and KEGG pathways analyses were performed. The result of GO analysis showed that a total of 284 targets were enriched in 18 GO terms, including 'transcription', 'regulation of transcription', 'translation', 'RNA processing', 'mRNA splicing', 'negative regulation of release of cytochrome c from mitochondria',

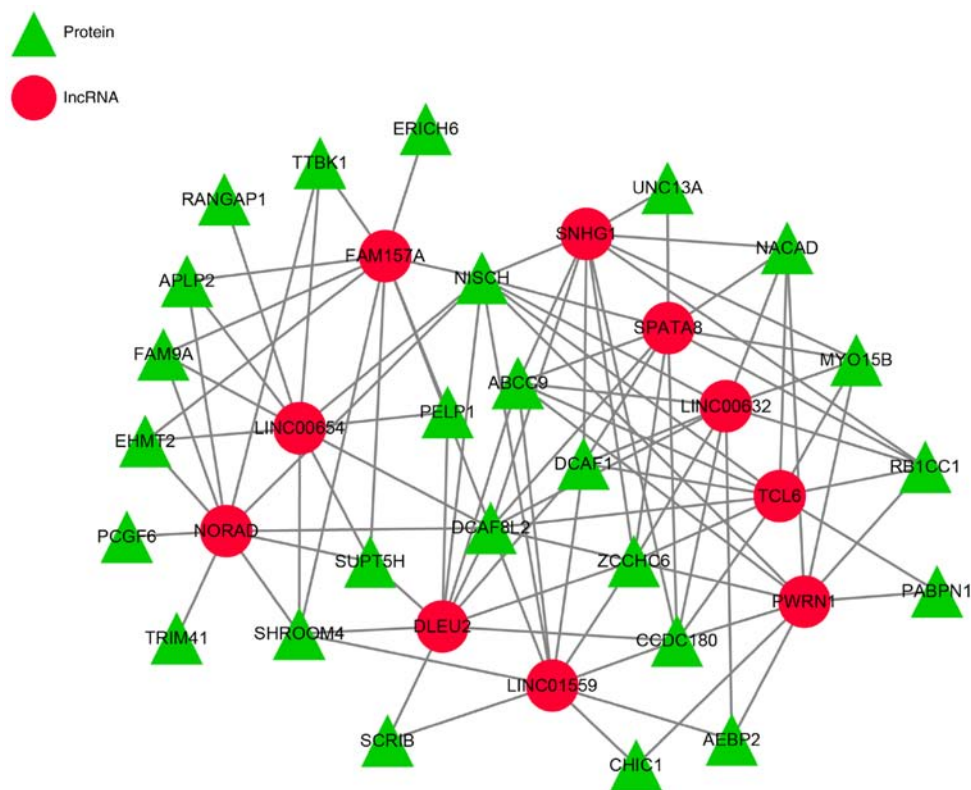


Figure 3. Interaction network of lncRNA-protein in patients with invasive breast carcinoma. lncRNA, long non-coding RNA. Interaction network of lncRNA-protein in patients with invasive breast carcinoma. lncRNA, long non-coding RNA.

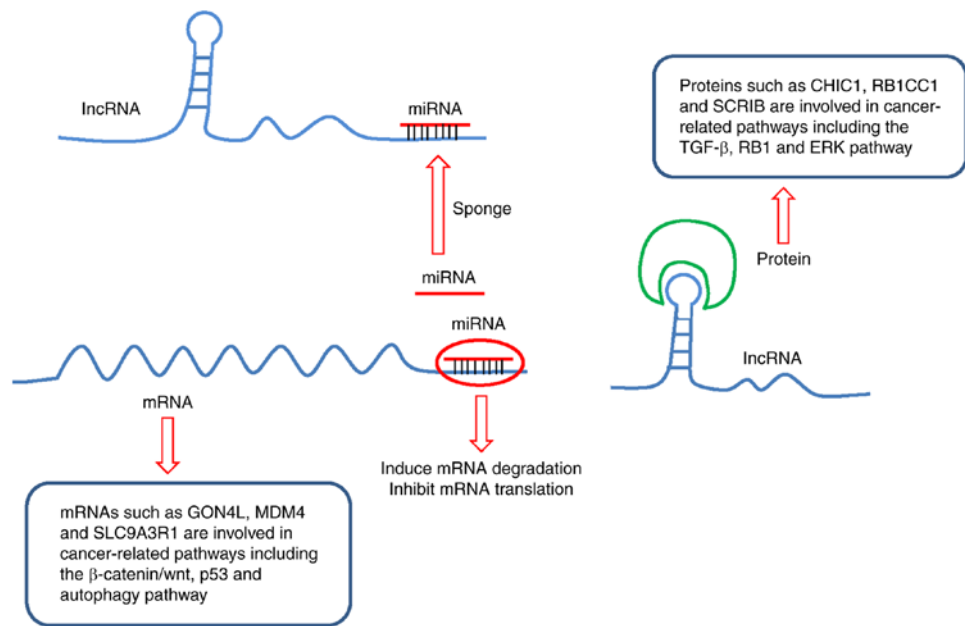


Figure 4. Underlying molecular mechanisms of the prognosis-related lncRNAs in patients with invasive breast carcinoma. lncRNA, long non-coding RNA; miRNA, microRNA; CHIC1, cysteine rich hydrophobic domain I; RB1CC1, retinoblastoma transcriptional corepressor 1 inducible coiled-coil I; SCRIB, scribble planar cell polarity protein; TGF-β, transforming growth factor β; RB1, retinoblastoma transcriptional corepressor 1; ERK, extracellular regulated MAP kinase.

‘covalent chromatin modification’, ‘endosome localization’, ‘bile acid secretion’, ‘negative regulation of smoothened signaling pathway involved in ventral spinal cord patterning’, ‘nucleic acid binding’, ‘DNA binding’, ‘protein binding’, ‘metal ion binding’, ‘transcription factor activity, sequence-specific

DNA binding’, ‘structural constituent of ribosome’, ‘protein tyrosine/serine/threonine phosphatase’ and ‘single-stranded RNA binding’ (Table II). The results of KEGG pathways analysis showed the targets were significantly enriched in ‘protein processing in endoplasmic reticulum’.

Table II. GO enrichment analysis of target mRNAs and proteins of patients with invasive breast carcinoma.

A, biological process		
GO term	Gene counts	P-value
GO:0006351~transcription, DNA-templated	69	3.29x10 <sup>-09</sup>
GO:0006355~regulation of transcription, DNA-templated	41	0.00258
GO:0006412~translation	11	0.010969
GO:0006396~RNA processing	6	0.024103
GO:0090201~negative regulation of release of cytochrome c from mitochondria	3	0.032585
GO:0000398~mRNA splicing, via spliceosome	9	0.034864
GO:0016569~covalent chromatin modification	6	0.04242
GO:0032439~endosome localization	2	0.049715
GO:0032782~bile acid secretion	2	0.049715
GO:0021914~negative regulation of smoothened signaling pathway involved in ventral spinal cord patterning	2	0.049715
B, molecular function		
GO term	Gene counts	P-value
GO:0003676~nucleic acid binding	37	8.88x10 <sup>-06</sup>
GO:0003677~DNA binding	46	0.00099
GO:0005515~protein binding	172	0.002573
GO:0046872~metal ion binding	51	0.004791
GO:0003700~transcription factor activity, sequence-specific DNA binding	28	0.006082
GO:0003735~structural constituent of ribosome	10	0.012777
GO:0008138~protein tyrosine/serine/threonine phosphatase activity	4	0.023806
GO:0003727~single-stranded RNA binding	4	0.035156
GO, Gene Ontology.		

## Discussion

In recent years, lncRNAs have been shown to be regulators of onset and progression of different cancer types due to their functions at both transcriptional and post-transcriptional levels (19). At the transcriptional level, lncRNAs have been reported to act as inducers of the transcriptional activity of their targets (20). At post-transcriptional level, lncRNAs can serve as ceRNAs to compete for miRNA binding sites on mRNAs (9). Furthermore, lncRNAs can bind to their target proteins, forming a localized complex at specific targets or binding proteins together to form RNA-protein complexes (21).

Aberrant expression levels of lncRNAs may lead to abnormal miRNA and mRNA expression levels, which contributes to an increased risk of diseases and several types of cancer (22,23). For instance, lncRNA growth arrest-specific transcript 5 (GAS5) was found to be downregulated in human breast cancer tissues and was associated with advanced Tumor-Node-Metastasis stage and poor prognosis (24).

Moreover, GAS5 was associated with tamoxifen resistance in breast cancer and overexpression of GAS5 increased breast cancer cell sensitivity to tamoxifen by regulating the expression levels of miR-222 (24).

In the present study, the expression levels of 4,525 human lncRNAs in 1,100 invasive breast carcinoma cases were investigated and 292 differentially expressed lncRNAs were identified, including 8 lncRNAs with expression alteration frequency  $\geq 10\%$ . These results indicated that these 8 lncRNAs could be potential biomarkers for invasive breast carcinoma diagnosis.

Previously, studies have proven that lncRNAs could predict the prognosis of breast cancer. He *et al* (25) used Cox regression analysis and a robust likelihood-based survival model to screen prognosis-related lncRNAs in 1,052 patients with breast cancer from TCGA database. The group reported that 11 lncRNAs could classify the patients into high and low risk groups with different overall survival rate. Sun *et al* (26) demonstrated that 9 lncRNAs were associated with metastasis-free survival rate of patients with breast cancer by analyzing the expression

profiles of lncRNAs in 916 breast cancer cases from the Gene Expression Omnibus (GEO) database. Liu *et al* (27) investigated the prognostic value of 2,730 lncRNAs in ~1,000 invasive breast carcinoma cases from the cBioPortal database, finding 11 lncRNAs with copy number alterations, 4 lncRNAs with aberrant expression levels associated with poor overall survival and 9 lncRNAs which could predict recurrence of breast cancer. In addition, Liu *et al* demonstrated that lncRNAs could also predict better prognosis of breast cancer. Vishnubalaji *et al* (28) analyzed the lncRNA expression profiles of 837 patients with invasive breast cancer and 105 healthy patients from TCGA database, identifying 6 lncRNAs associated with more favorable disease-free survival rate and 4 lncRNAs associated with more favorable overall survival.

In the present study, altered expression levels of 10 lncRNAs (NORAD, SNHG1, LINC00654, FAM157A, DLEU2, LINC01559, SPATA8, TCL6, LINC00632 and PWRN1) were significantly correlated with shorter overall survival time in patients with invasive breast carcinoma, suggesting that these 10 lncRNAs may be independent prognostic markers for the disease. Zhou *et al* (29) demonstrated that lncRNA NORAD serves a carcinogenic role in breast cancer. By analyzing the overall survival time of 21 patients with breast cancer, Zhou *et al* (29) demonstrated that patients with high expression levels of NORAD had less favorable survival compared with the low expression group, which supports the findings of the present study. Studies have reported that increased SNHG1 expression levels could predict less favorable overall survival rate in various types of cancer, such as gastric cancer, hepatocellular carcinoma, colorectal cancer and non-small cell lung cancer (30). In the present study, among the 10 prognosis associated lncRNAs, DLEU2, TCL6 and PWRN1 were significantly correlated with shorter disease-free survival time of patients with invasive breast carcinoma. Recently, it was reported that DLEU2 was involved in several types of cancer and DLEU2 exon 9 was an independent marker of poor prognosis in patients with esophageal adenocarcinoma (31). Aberrant expression levels of TCL6 were involved in clear cell renal carcinoma by regulating the PI3K/AKT pathway and PWRN1 inhibits gastric cancer cell proliferation and metastasis via p53 signaling pathway (32,33).

For a better understanding of the 10 lncRNAs in the present study, further investigation of the underlying molecular mechanisms are necessary. Competitive endogenous RNA (ceRNA) networks have been reported in several types of cancer, including breast cancer. For example, lncRNA CDC6 was upregulated in breast cancer tissues and its high expression levels were associated with more advanced clinical stages of patients with breast cancer (34). In addition, overexpression of CDC6 promoted proliferation and migration of breast cancer cells by directly sponging of miR-215 (34). The present study constructed a lncRNA-miRNA-mRNA regulatory network, which was composed of 10 lncRNAs, 88 miRNAs and 322 mRNAs. Overall, 11 target mRNAs (EMC2, TFB2M, ENY2, RAD54B, ANGEL2, RBM34, AGO2, GON4L, NPLOC4, ZFP41 and SSR2) were aberrantly expressed in  $\geq 20\%$  patients. Previously, several target mRNAs, such as GON4L, MDM4 and SLC9A3R1, were reported to be involved in cancer-related signaling pathways (35-37). For

instance, MDM4, a regulator of p53 tumor suppressor protein, suppressed the activity of p53 to promote tumor development and some molecular agents such as SJ-172550 and XI-006 have been developed to eliminate the inhibition of p53 by MDM4 (38). The findings of the present study indicated that the target mRNAs could be potential diagnostic markers or therapeutic targets for invasive breast carcinoma.

lncRNA-protein interactions and lncRNA-protein complexes serve important roles in a variety of biological processes and abnormal expression levels of RNA-binding proteins can contribute to tumor development (39). lncRNA HULC was reported to function as an oncogene and interact with the ATG7 protein in epithelial ovarian carcinoma, wherein overexpression of HULC promoted ovarian carcinoma cells proliferation, migration and invasion by suppressing mRNA and protein expression of ATG7 (40). In the present study, the proteins which could interact with the 10 lncRNAs were predicted and there were 25 potential target proteins, including CHIC1, RB1CC1 and SCRIB. Previous findings have shown that these target proteins are involved in several cancer-related pathways, including the TGF- $\beta$ , RB1 and ERK pathways (41-43). Therefore, the 10 lncRNAs identified in the present study may be potentially novel targets for invasive breast carcinoma therapy. To better understand the function of these lncRNAs and provide further insights into the ceRNA and lncRNA-protein regulatory networks, function enrichment analysis was performed on the 322 target mRNAs and 25 target proteins, reporting that the targets were part of a number of functional classes, including 'transcription', 'regulation of transcription', 'translation', 'RNA processing', 'mRNA splicing and protein processing'.

In conclusion, the present study analyzed the association between lncRNA expression level alterations and prognosis of 1,100 patients with invasive breast carcinoma, identifying 10 prognosis-related lncRNAs, whose expression level alterations predicted lower overall survival rate and shorter overall survival time. Therefore, therapeutically targeting these lncRNAs might be effective in improving the treatment of invasive breast carcinoma. To further understand the molecular basis of the lncRNAs, a regulatory network was constructed and the target mRNAs and proteins were predicted, of which some targets were associated with well-known signaling pathways. Overall, this regulatory network could identify novel targets to aid in the development of new drugs to improve the efficacy of breast cancer therapy. However, breast cancer is divided into four main subtypes: Luminal A, luminal B, ERBB2 positive and triple negative, and gene expression profiles are not the same for all these subtypes (44). As all these four subtypes were included in the present study, it would be important to explore the differentially expressed lncRNAs in each subtype of breast cancer, allowing for the identification of more diagnostic and prognostic biomarkers. In the present study, the relevance of lncRNA expression levels to invasive breast carcinoma were based on RNA sequencing (RNA-Seq) data, while in some other databases, the cancer-related lncRNAs are identified using microarrays (45). For example, the lncAR database extracts clinical information and gene expression microarray data from the GEO database (46). RNA-Seq and microarrays are two of the most common methods of transcript



expression detection. Microarrays are rapid, accurate, sensitive, cost-effective and specific to detect transcript expressions, and RNA-Seq is a high-throughput sequencing technology, which can be used to quantify, profile, and discover RNA transcripts, and the transcripts can be mapped to the reference genome to get comprehensive genetic information (47). Due to different analysis methods, heterogeneity of samples and different data processing algorithms, the results obtained from different databases may differ, therefore the results of the present study need to be further validated using *in vitro* and *in vivo* experiments. The primary limitation of the study present was the lack of validation of the findings *in vitro* and *in vivo*. To confirm the prognostic role of the lncRNAs, expression levels of the lncRNAs in a number of patients with invasive breast carcinoma with follow-up records should be analyzed. To validate the underlying molecular mechanisms of the lncRNAs, a luciferase reporter assay should be performed and lncRNA knockdown cell models and a tumor xenograft model should also be established to further identify the function of the lncRNAs (48,49).

### Acknowledgements

Not applicable.

### Funding

No funding was received.

### Availability of data and materials

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

### Authors' contributions

YH and YS searched the literature and designed the study; XG screened the lncRNAs; YD and YS constructed the interaction networks and performed the functional analysis; XX reviewed the data and drafted the manuscript. All authors have read and approved the final manuscript.

### Ethics approval and consent to participate

Not applicable.

### Patient consent for publication

Not applicable.

### Competing interests

The authors declare that they have no competing interests.

### References

- Siegel RL, Miller KD and Jemal A: Cancer statistics, 2018. *CA Cancer J Clin* 68: 7-30, 2018.
- Colzani E, Liljegren A, Johansson AL, Adolfsson J, Hellborg H, Hall PF and Czene K: Prognosis of patients with breast cancer: Causes of death and effects of time since diagnosis, age, and tumor characteristics. *J Clin Oncol* 29: 4014-4021, 2011.
- Avgeris M, Tsilimantou A, Levis PK, Tokas T, Sideris DC, Stravodimos K, Ardavanis A and Scorilas A: Loss of GAS5 tumour suppressor lncRNA: An independent molecular cancer biomarker for short-term relapse and progression in bladder cancer patients. *Brit J Cancer* 119: 1477-1486, 2018.
- Chakraborty S, Andrieux G, Hasan AMM, Ahmed M, Hosen MI, Rahman T, Hossain MA and Boerries M: Harnessing the tissue and plasma lncRNA-peptidome to discover peptide-based cancer biomarkers. *Sci Rep* 9: 12322, 2019.
- Yarmishyn AA, Batagov AO, Tan JZ, Sundaram GM, Sampath P, Kuznetsov VA and Kurochkin IV: HOXD-AS1 is a novel lncRNA encoded in HOXD cluster and a marker of neuroblastoma progression revealed via integrative analysis of noncoding transcriptome. *BMC Genomics* (15 Suppl 9): S7, 2014.
- Sakaue S, Hirata J, Maeda Y, Kawakami E, Nii T, Kishikawa T, Ishigaki K, Terao C, Suzuki K, Akiyama M, *et al*: Integration of genetics and miRNA-target gene network identified disease biology implicated in tissue specificity. *Nucleic Acids Res* 46: 11898-11909, 2018.
- Fu X, Mao X, Wang Y, Ding X and Li Y: Let-7c-5p inhibits cell proliferation and induces cell apoptosis by targeting ERCC6 in breast cancer. *Oncol Rep* 38: 1851-1856, 2017.
- Hausser J, Syed AP, Bilen B and Zavolan M: Analysis of CDS-located miRNA target sites suggests that they can effectively inhibit translation. *Genome Res* 23: 604-615, 2013.
- Shuwen H, Qing Z, Yan Z and Xi Y: Competitive endogenous RNA in colorectal cancer: A systematic review. *Gene* 645: 157-162, 2018.
- Neumann P, Jaé N, Knau A, Glaser SF, Fouani Y, Rossbach O, Kruger M, John D, Bindereif A, Grote P, *et al*: The lncRNA GATA6-AS epigenetically regulates endothelial gene expression via interaction with LOXL2. *Nat Commun* 9: 237, 2018.
- Cerami E, Gao J, Dogrusoz U, Gross BE, Sumer SO, Aksoy BA, Jacobsen A, Byrne CJ, Heuer ML, Larsson E, *et al*: The cBio cancer genomics portal: An open platform for exploring multi-dimensional cancer genomics data. *Cancer Discov* 2: 401-404, 2012.
- Gao J, Aksoy BA, Dogrusoz U, Dresdner G, Gross B, Sumer SO, Sun Y, Jacobsen A, Sinha R, Larsson E, *et al*: Integrative analysis of complex cancer genomics and clinical profiles using the cBioPortal. *Sci Signal* 6: pii. 2013.
- Huang da W, Sherman BT and Lempicki RA: Systematic and integrative analysis of large gene lists using DAVID bioinformatics resources. *Nat Protoc* 4: 44-57, 2009.
- Huang DW, Sherman BT, Tan Q, Kir J, Liu D, Bryant D, Guo Y, Stephens R, Baseler MW, Lane HC, *et al*: DAVID Bioinformatics Resources: Expanded annotation database and novel algorithms to better extract biology from large gene lists. *Nucleic Acids Res* 35: W169-W175, 2007.
- Chang TH, Huang HY, Hsu JB, Weng SL, Horng JT and Huang HD: An enhanced computational platform for investigating the roles of regulatory RNA and for identifying functional RNA motifs. *BMC Bioinformatics* 14 (Suppl): S4, 2013.
- Agarwal V, Bell GW, Nam JW and Bartel DP: Predicting effective microRNA target sites in mammalian mRNAs. *ELife* 4, 2015.
- Shannon P, Markiel A, Ozier O, Baliga NS, Wang JT, Ramage D, Amin N, Schwikowski B and Ideker T: Cytoscape: A software environment for integrated models of biomolecular interaction networks. *Genome Res* 13: 2498-2504, 2003.
- Lang B, Armaos A and Tartaglia GG: RNAct: Protein-RNA interaction predictions for model organisms with supporting experimental data. *Nucleic Acids Res* 47: D601-D606, 2019.
- Guo J, Li P, Liu X and Li Y: NOTCH signaling pathway and non-coding RNAs in cancer. *Pathol Res Pract* 215: 152620, 2019.
- Youness RA and Gad MZ: Long non-coding RNAs: Functional regulatory players in breast cancer. *Non-coding RNA Res* 4: 36-44, 2019.
- Kazan H, Ray D, Chan ET, Hughes TR and Morris Q: RNAcontext: A new method for learning the sequence and structure binding preferences of RNA-binding proteins. *PLoS Comput Boil* 6: e1000832, 2010.
- Weir BA, Woo MS, Getz G, Perner S, Ding L, Beroukheim R, Lin WM, Province MA, Kraja A, Johnson LA, *et al*: Characterizing the cancer genome in lung adenocarcinoma. *Nature* 450: 893-898, 2007.
- Sanchez Calle A, Kawamura Y, Yamamoto Y, Takeshita F and Ochiya T: Emerging roles of long non-coding RNA in cancer. *Cancer Sci* 109: 2093-2100, 2018.

24. Gu J, Wang Y, Wang X, Zhou D, Shao C, Zhou M and He Z: Downregulation of lncRNA GAS5 confers tamoxifen resistance by activating miR-222 in breast cancer. *Cancer Lett* 434: 1-10, 2018.
25. He Y, Li X, Meng Y, Fu S, Cui Y, Shi Y and Du H: A prognostic 11 long noncoding RNA expression signature for invasive breast carcinoma. *J Cell Biochem* 120: 16692-16702, 2019.
26. Sun J, Chen X, Wang Z, Guo M, Shi H, Wang X, Cheng L and Zhou M: A potential prognostic long non-coding RNA signature to predict metastasis-free survival of breast cancer patients. *Sci Rep* 5: 16553, 2015.
27. Liu H, Li J, Koirala P, Ding X, Chen B, Wang Y, Wang Z, Wang C, Zhang X and Mo YY: Long non-coding RNAs as prognostic markers in human breast cancer. *Oncotarget* 7: 20584-20596, 2016.
28. Vishnubalaji R, Shaath H, Elkord E and Alajez NM: Long non-coding RNA (lncRNA) transcriptional landscape in breast cancer identifies LINC01614 as non-favorable prognostic biomarker regulated by TGF $\beta$  and focal adhesion kinase (FAK) signaling. *Cell Death Discov* 5: 109, 2019.
29. Zhou K, Ou Q, Wang G, Zhang W, Hao Y and Li W: High long non-coding RNA NORAD expression predicts poor prognosis and promotes breast cancer progression by regulating TGF- $\beta$  pathway. *Cancer Cell Int* 19: 63, 2019.
30. Yu J, Yan Y, Hua C and Ming L: Upregulation of lncRNA SNHG1 is associated with metastasis and poor prognosis in cancers: A meta-analysis. *Medicine* 98: e15196, 2019.
31. Ma W, Zhang CQ, Dang CX, Cai HY, Li HL, Miao GY, Wang JK and Zhang LJ: Upregulated long-non-coding RNA DLEU2 exon 9 expression was an independent indicator of unfavorable overall survival in patients with esophageal adenocarcinoma. *Biomed Pharmacother* 113: 108655, 2019.
32. Chen Z, Ju H, Yu S, Zhao T, Jing X, Li P, Jia J, Li N, Tan B and Li Y: Prader-Willi region non-protein coding RNA 1 suppressed gastric cancer growth as a competing endogenous RNA of *miR-425-5p*. *Clin Sci (Lond)* 132: 1003-1019, 2018.
33. Su H, Sun T, Wang H, Shi G, Zhang H, Sun F and Ye D: Decreased TCL6 expression is associated with poor prognosis in patients with clear cell renal cell carcinoma. *Oncotarget* 8: 5789-5799, 2017.
34. Kong X, Duan Y, Sang Y, Li Y, Zhang H, Liang Y, Liu Y, Zhang N and Yang Q: lncRNA-CDC6 promotes breast cancer progression and function as ceRNA to target CDC6 by sponging microRNA-215. *J Cell Physiol* 234: 9105-9117, 2019.
35. Xia F, Jin P, Ding Y, Li F and Shi C: GON4L drives nasopharyngeal carcinoma growth and proliferation through regulation of  $\beta$ -catenin/wnt signaling pathway. *Biomed Res* 28: 4348-4353, 2017.
36. Lam S, Lodder K, Teunisse AF, Rabelink MJ, Schutte M and Jochemsen AG: Role of Mdm4 in drug sensitivity of breast cancer cells. *Oncogene* 29: 2415-2426, 2010.
37. Liu H, Ma Y, He HW, Wang JP, Jiang JD and Shao RG: SLC9A3R1 stimulates autophagy via BECN1 stabilization in breast cancer cells. *Autophagy* 11: 2323-2334, 2015.
38. Li Q and Lozano G: Molecular pathways: Targeting Mdm2 and Mdm4 in cancer therapy. *Clin Cancer Res* 19: 34-41, 2013.
39. Busà R, Paronetto MP, Farini D, Pierantozzi E, Botti F, Angelini DF, Attisani F, Vespasiani G and Sette C: The RNA-binding protein Sam68 contributes to proliferation and survival of human prostate cancer cells. *Oncogene* 26: 4372-4382, 2007.
40. Chen S, Wu DD, Sang XB, Wang LL, Zong ZH, Sun KX, Liu BL and Zhao Y: The lncRNA HULC functions as an oncogene by targeting ATG7 and ITGB1 in epithelial ovarian carcinoma. *Cell Death Dis* 8: e3118, 2017.
41. Souchevnytskyi S: Proteomics of TGF-beta signaling and its impact on breast cancer. *Expert Rev Proteomics* 2: 925-935, 2005.
42. Chano T, Ikebuchi K, Ochi Y, Tameno H, Tomita Y, Jin Y, Inaji H, Ishitobi M, Teramoto K, Nishimura I, *et al*: RB1CC1 activates Rb1 pathway and inhibits proliferation and cologenic survival in human cancer. *PLoS One* 5: e11404, 2010.
43. Young LC, Hartig N, Muñoz-Alegre M, Osés-Prieto JA, Durdu S, Bender S, Vijayakumar V, Vietri Rudan M, Gewinner C, Henderson S, *et al*: An MRAS, SHOC2, and SCRIB complex coordinates ERK pathway activation with polarity and tumorigenic growth. *Mol Cell* 52: 679-692, 2013.
44. Mathias C, Zambalde EP, Rask P, Gradia DF and de Oliveira JC: Long non-coding RNAs differential expression in breast cancer subtypes: What do we know? *Clin Genet* 95: 558-568, 2019.
45. Klinge CM: Non-coding RNAs in breast cancer: Intracellular and intercellular communication. *Non-Coding RNA* 4: pii: E40, 2018.
46. Zheng Y, Xu Q, Liu M, Hu H, Xie Y, Zuo Z and Ren J: lncAR: A comprehensive resource for lncRNAs from cancer arrays. *Cancer Res* 79: 2076-2083, 2019.
47. Chen L, Sun F, Yang X, Jin Y, Shi M, Wang L, Shi Y, Zhan C and Wang Q: Correlation between RNA-Seq and microarrays results using TCGA data. *Gene* 628: 200-204, 2017.
48. Zheng S, Li M, Miao K and Xu H: SNHG1 contributes to proliferation and invasion by regulating miR-382 in breast cancer. *Cancer Manag Res* 11: 5589-5598, 2019.
49. Zhang YL, Li XB, Hou YX, Fang NZ, You JC and Zhou QH: The lncRNA XIST exhibits oncogenic properties via regulation of miR-449a and Bcl-2 in human non-small cell lung cancer. *Acta Pharmacol Sin* 38: 371-381, 2017.



This work is licensed under a Creative Commons Attribution-NonCommercial-NoDerivatives 4.0 International (CC BY-NC-ND 4.0) License.