Breast Cancer Research and Treatment (2019) 175:59–75
https://doi.org/10.1007/s10549-019-05147-6

PRECLINICAL STUDY

An eight-lncRNA signature predicts survival of breast cancer patients: a comprehensive study based on weighted gene co-expression network analysis and competing endogenous RNA network

Min Sun[1,2] · Di Wu[2] · Ke Zhou[1] · Heng Li[1] · Xingrui Gong[2] · Qiong Wei[2] · Mengyu Du[2] · Peijie Lei[3] · Jin Zha[2] · Hongrui Zhu[2] · Xinsheng Gu[4] · Dong Huang[1]

Received: 1 October 2018 / Accepted: 22 January 2019 / Published online: 4 February 2019

Abstract

Purpose To identify a lncRNA signature to predict survival of breast cancer (BRCA) patients.

Methods A total of 1222 BRCA case and control datasets were downloaded from the TCGA database. The weighted gene co-expression network analysis of differentially expressed mRNAs was performed to generate the modules associated with BRCA overall survival status and further construct a hub on competing endogenous RNA (ceRNA) network. LncRNA signatures for predicting survival of BRCA patients were generated using univariate survival analyses and a multivariate Cox hazard model analysis and validated and characterized for prognostic performance measured using receiver operating characteristic (ROC) curves.

Results A prognostic score model of eight lncRNAs signature was identified as Prognostic score = $(0.121 \times \mathrm{EXP}_{\mathrm{AC007731.1}}) + (0.108 \times \mathrm{EXP}_{\mathrm{AL513123.1}}) + (0.105 \times \mathrm{EXP}_{\mathrm{C10orf126}}) + (0.065 \times \mathrm{EXP}_{\mathrm{WT1\text{-}AS}}) + (-0.126 \times \mathrm{EXP}_{\mathrm{ADAMTS9\text{-}AS1}}) + (-0.130 \times \mathrm{EXP}_{\mathrm{SRGAP3\text{-}AS2}}) + (0.116 \times \mathrm{EXP}_{\mathrm{TLR8\text{-}AS1}}) + (0.060 \times \mathrm{EXP}_{\mathrm{HOTAIR}})$ with median score 1.088. Higher scores predicted higher risk. The lncRNAs signature was an independent prognostic factor associated with overall survival. The area under the ROC curves (AUC) of the signature was 0.979, 0.844, 0.99 and 0.997 by logistic regression, support vector machine, decision tree and random forest models, respectively, and the AUCs in predicting 1- to 10-year survival were between 0.656 and 0.748 in the test dataset from TCGA database.

Conclusions The eight-lncRNA signature could serve as an independent biomarker for prediction of overall survival of BRCA. The lncRNA-miRNA-mRNA ceRNA network is a good tool to identify lncRNAs that is correlated with overall survival of BRCA.

Keywords Breast cancer · The cancer genome atlas · Competing endogenous RNA network · Prognostic signature · Weighted gene co-expression network analysis

Abbreviations

lncRNAs	Long noncoding RNAs
ceRNA	Competing endogenous RNA
BRCA	Breast cancer

Min Sun and Di Wu have contributed equally to this work.

Electronic supplementary material The online version of this article (https://doi.org/10.1007/s10549-019-05147-6) contains supplementary material, which is available to authorized users.

✉ Xinsheng Gu
gu.xinsheng@gmail.com

✉ Dong Huang
hd_814@sohu.com

[1] Department of General Surgery, Taihe Hospital, Hubei University of Medicine, Shiyan 442000, China

[2] Department of Anesthesiology, Institute of Anesthesiology, Taihe Hospital, Hubei University of Medicine, Shiyan 442000, China

[3] The First Clinical School, Hubei University of Medicine, Shiyan 442000, China

[4] College of Basic Medical Sciences, Hubei University of Medicine, Shiyan 442000, China

TCGA	The cancer genome atlas
DEmRNAs	Differentially expressed mRNAs
DEmiRNAs	Differentially expressed miRNAs
DElncRNAs	Differentially expressed lncRNAs
WGCNA	Weighted gene co-expression network analysis
OS	Overall survival
miRNA	microRNAs
MREs	microRNA response element
PCC	Pearson's correlation coefficient
PPI	Protein–protein interaction network
GO	Gene ontology
KEGG	Kyoto encyclopedia of genes and genomes
AIC	Akaike information criterion
ROC	Receiver operating characteristic
AUC	The area under the respective ROC curves
EXP	Expression
ER	Estrogen receptor
RCC	Renal cell carcinoma
HR	Hazard ratio
PCA	Principal component analysis

Introduction

Breast cancer (BRCA) is the most common and life-threatening cancer diagnosed among women worldwide [1]. A number of individual and signature molecular markers have been used in diagnosis and prognosis of BRCA, such as uPA and PAI-1 whose decreased levels are associated with lower risk of cancer recurrence, and Ki67, cyclin D, cyclin E, p27 and p21 overexpression indicate uncontrolled proliferation [2]. Oncotype DX of 21-gene panel predicts 10-year distant recurrence in patients with estrogen receptor-positive, axillary lymph node-negative BRCA [3]. MammaPrint, a 70-gene expression profile prognosis signature, predicts recurrence for hormone receptor-positive, lymph node-negative, human epidermal growth factor receptor type 2-negative early stage BRCA in Japanese patients and selects postmenopausal women for adjuvant chemotherapy [4–7]. Novel markers including microRNA and patient-specific subpathway activities signatures predict lymph node metastasis and prognosis in BRCA [8–10]. Hedgehog-mesenchyme gene signature identifies bimodal prognosis in luminal and basal BRCA sub-types [11]. As BRCAs are notoriously heterogeneous, multiparameter signatures are more appealing for cancer prognosis than a single biomarker that usually cannot accurately predict cancer progression.

A competing endogenous RNA (ceRNA) hypothesis has been proposed and developed to address how RNAs in a cell form a regulatory network comprised of microRNAs, messenger RNAs (mRNAs), transcribed pseudogenes (tpRNAs), long noncoding RNAs (lncRNAs) and circular RNAs (crRNAs). These RNAs interact with each other through microRNA response elements (MREs). MREs are sequences in mRNAs, tpRNAs, lncRNAs and crRNAs and can be recognized and bound by certain microRNAs. Through MREs, lncRNAs act as "sponges" for microRNAs, resulting in changes in the microRNAs-regulated mRNA levels. The lncRNA-miRNA-mRNA interactions form a ceRNA network which are involved in regulation of cell activity and function [12–15]. Several studies have shown that lncRNA-miRNA-mRNA ceRNA networks were involved in regulating BRCA progression [16–20]. However, understanding of the association of lncRNA-miRNA-mRNA ceRNA networks and their components with the prognosis of BRCAs remain very limited.

In this study, we identified a ceRNA network associated with overall survival of BRCA based on The Cancer Genome Atlas (TCGA) database and the weighted gene co-expression network analysis (WGCNA). TCGA database contains genomic sequence, expression, methylation and copy number variation data on over 11,000 individuals who represent over 30 different types of cancers [21, 22]. WGCNA is an algorithm to investigate the scale-free gene modules of co-expression and highly correlated genes and clinical traits and identify candidate biomarkers [23, 24]. We further identified and characterized an lncRNA signature for prediction of overall survival for BRCA patients. Our finding may provide a clue to develop a promising prognostic tool for overall survival of BRCA.

Materials and methods

The workflow of analysis and the datasets

The workflow of analysis is illustrated in Fig. 1. The mRNASeq gene expression data, miRNASeq data and clinical information of a total of 1222 cases (1101 BRCA cases, 104 normal cases and 17 cases with replicated data) were downloaded from the TCGA database using the Data-Transfer Tool. A total of 836 cases with complete clinical information and 251 cases with complete prognostic information were included in the train dataset and the test dataset, respectively.

Analysis of differential expressed mRNAs, miRNAs and lncRNAs (DEmRNAs, DEmiRNAs and DElncRNAs)

The DEmRNAs, DEmiRNAs and DElncRNAs lists were generated by the "edgeR" package after normalization and data filter with thresholds $|\log2 \text{ (fold change)}| > 2$ and adjusted P value < 0.01 by comparing the BRCA group with the normal group. The data with expression levels close to

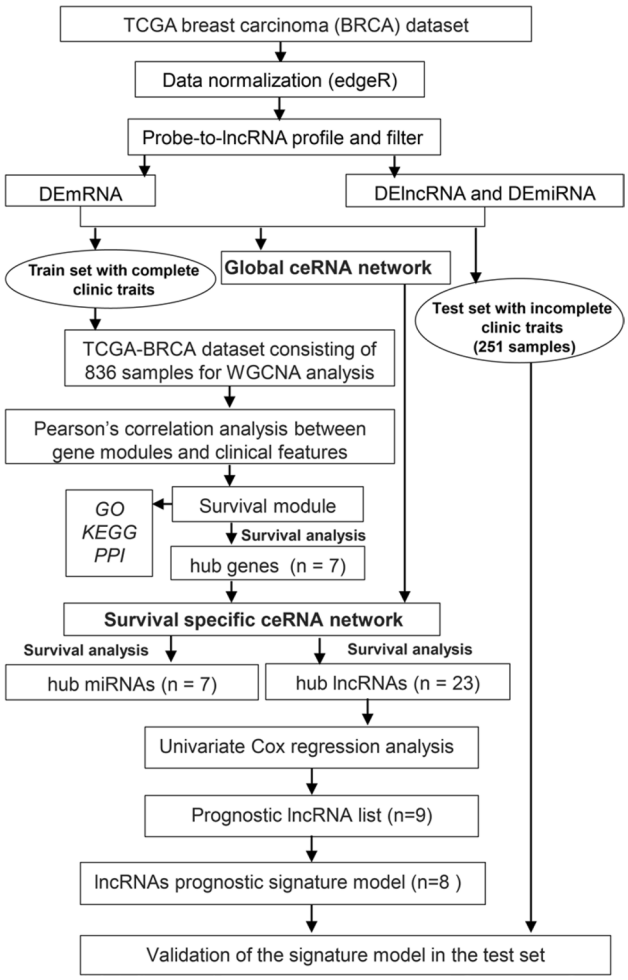


Fig. 1 Workflow for analysis of the TCGA-BRCA datasets

zero were excluded. All genes were annotated with the GENCODE database.

Construction of weighted gene co-expression network of DEmRNAs and identification and functional characterization of modules associated with overall survival status

Weighted gene co-expression network of DEmRNAs was constructed using the WGCNA package [23]. The network modules were generated using the topological overlapping measurement with a power cutoff threshold of 3 and a module size cutoff ≤ 50. The modules were then used to analyze their correlation with clinical traits by Pearson's correlation test. $P < 0.05$ was considered significant.

The function of the module genes was annotated using the "clusterProfiler" package of Bioconductor with FDR < 0.05 [25]. The overall survival modules were functionally annotated by Gene Ontology (GO) biological processes [26] and Kyoto Encyclopedia of Genes and Genomes (KEGG) pathways [27]. Protein interactions among DEmRNAs in a module were identified, and a PPI network (minimum required interaction score > 0.4) was constructed using the STRING online tools [28]. Networks were visualized with the Cytoscape software [29].

Construction of the global ceRNA network and the hub ceRNA network associated with BRCA

The global ceRNA network was constructed by retrieving miRNA-mRNA pairs from miRTarBase [30] and lncRNA–miRNA interactions from the miRcode database (http://www.mircode.org/index.php) using DEmRNAs, DEmiRNAs and DElncRNAs.

To identify the hub ceRNA network associated with survival status, the genes common to both the genes set in the survival-associated WGCNA module from DEmRNAs and the DEmiRNAs target gene set were identified. These genes were considered to be hub genes. To construct the hub ceRNA network, the hub genes were used to retrieve miRNAs, lncRNAs and associated interactions from the global ceRNA network to form a sub-ceRNA network which was considered the hub ceRNA network. These miRNAs and lncRNAs in this network were defined as hub miRNA and hub lncRNA, respectively. The association of these hub genes, miRNAs and hub lncRNAs with survival time was analyzed using Kaplan–Meier curve.

Identification, characterization and validation of lncRNA signature predicting survival of BRCA patients

The hub lncRNAs were used to perform univariate survival analyses, and the lncRNAs with P value < 0.05 were selected for further analysis. A stepwise model selection by the Akaike information criterion (AIC) was then performed to avoid over-fitting. Finally, a multivariate Cox hazard model analysis was performed to generate an lncRNAs prognostic signature model, which calculated prognostic score as follows: Prognostic score = $\sum(C \times EXP_{lncRNA})$. EXP was the FPKM value of the lncRNA, and C was the regression coefficient for the corresponding lncRNA in multivariate Cox hazard model analysis. The median prognostic score of the training dataset was used to differentiate high-risk group and low-risk group. Higher scores predicted higher risk.

The lncRNA signature was examined for its association with patient survival. The prognostic performance was measured using receiver operating characteristic (ROC) curves by comparing the area under the respective ROC curves (AUC) calculated by logistic regression, support vector machine, decision tree and random forest methods. All reported P values were two-sided. All analyses were carried out via the R/BioConductor (version 3.2.3),

survival curves and ROCs were generated by ggplot2, survival and survivalROC packages.

Results

Identifying DEmRNAs, DEmiRNAs and DElncRNAs in BRCA tissues

We identified a total of 2,139 DEmRNAs with 1,351 (63.16%) upregulated and 788 (36.84%) (Fig. 2a, d) and 86 DEmiRNAs with 67 (77.91%) upregulated and 19 (22.09%) downregulated (Fig. 2b, e). We identified a total of 1062 DElncRNAs (Fig. 2c, f). Ten mRNAs, miRNAs and lncRNAs with highest fold changes are shown in Table 1.

Identification and functional characterization of modules associated with overall survival status

We constructed the co-expression network of 2139 DEmRNAs with WGCNA to identify functionally related or similar genes. The distribution of connections among genes indicated it was a scale-free network (Supplementary Fig. S1A-D) and then we analyzed it using methods for a scale-free network. We generated the modules by dynamic tree cut (Fig. 3a). After merging highly similar modules, we generated fifteen modules ranging from 46 to 485 genes in size. The non-coexpressed genes were included in the "grey" module which was not analyzed further (Fig. 3b, c). We analyzed the association of the fifteen modules with clinical traits. The results showed that the turquoise module was correlated significantly with survival status (PCC = 0.072, $P = 0.04$) (Fig. 3d). The turquoise module contained 485 genes (Fig. 3e, Supplementary Table 1) which were natural candidates for overall survival biomarkers of BRCA.

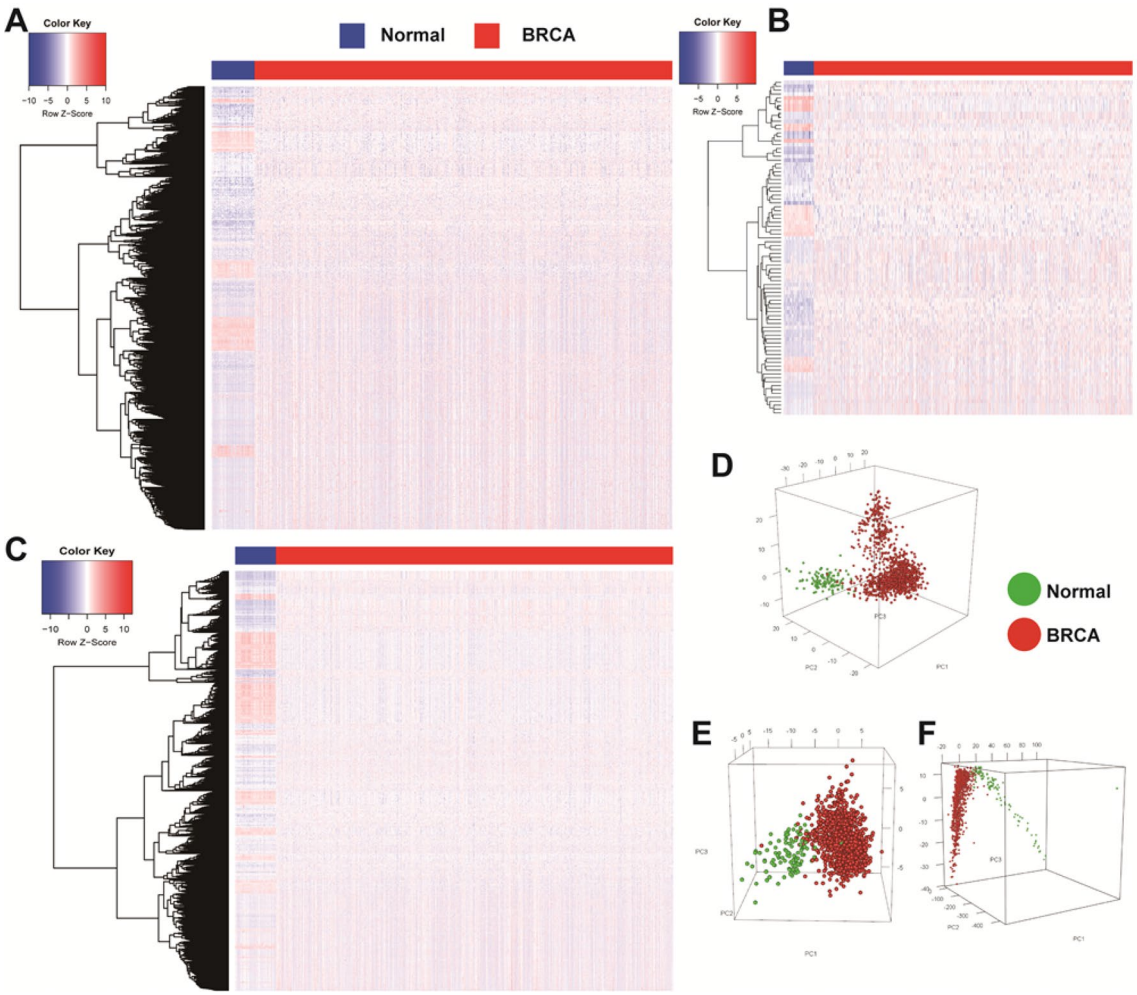


Fig. 2 Identification of DEmRNAs, DEmiRNAs and DElncRNAs in the TCGA-BRCA tissues. Heat map and two-way hierarchical clustering of DEmRNAs (**a**), DEmiRNA (**b**) and DElncRNAS (**c**), and PCA plotting of DEmRNAs (**d**), DEmiRNA (**e**) and DElncRNAS (**f**) were generated

Table 1 Top 10 DEmRNAs, DEmiRNAs and DElncRNAs

Gene symbol	logFC	Adjusted *P* value	Classification
Top 10 DEmRNAs			
MMP11	6.230500727	3.20E−162	Up
COL10A1	7.102468524	9.79E−162	Up
NEK2	4.21931208	2.45E−157	Up
KIF4A	3.811320912	2.79E−135	Up
PKMYT1	3.969878821	2.83E−135	Up
CKM	−8.26134	< 0.001	Down
ACTA1	−6.96741	< 0.001	Down
MYLPF	−6.92063	< 0.001	Down
PYGM	−6.90898	< 0.001	Down
SLN	−6.42667	< 0.001	Down
Top 10 DEmiRNAs			
hsa-miR-21	2.214203053	1.17E−121	Up
hsa-miR-96	3.351163524	8.13E−107	Up
hsa-miR-183	2.994455514	1.68E−98	Up
hsa-miR-141	2.291868915	4.28E−79	Up
hsa-miR-592	4.517167084	6.40E−79	Up
hsa-miR-133a-1	−6.577507556	0	Down
hsa-miR-133a-2	−6.534233186	0	Down
hsa-miR-1-2	−5.725915614	0	Down
hsa-miR-486-1	−4.527467291	5.77E−296	Down
hsa-miR-486-2	−4.518886465	2.29E−293	Down
Top 10 DElncRNAs			
LINC01614	5.826506552	3.9906E−129	Up
LINC01705	5.81357573	2.00264E−92	Up
LINC00922	4.758978121	2.21487E−92	Up
LEF1-AS1	2.593177109	1.93192E−89	Up
C6orf99	2.928494588	9.60826E−89	Up
AP001528.2	−3.07895	6.87E−267	Down
LINC02202	−3.03487	1.43E−259	Down
LINC01537	−3.44855	2.53E−237	Down
TRHDE-AS1	−5.37866	5.08E−228	Down
AL031316.1	−4.24909	1.98E−222	Down

DEmRNAs differentially expressed mRNAs, *DemiRNAs* differentially expressed microRNAs, *DelncRNAs* differentially expressed long non-coding RNAs, *FC* fold change

We performed functional enrichment analysis of the turquoise module. The results showed that cell cycle was the most enriched GO term and KEGG pathway (Fig. 4a, b). We constructed and visualized the PPI network with the genes in the turquoise module (Fig. 4c). We identified the hub genes based on the gene set of the turquoise module and the target gene set of DEmiRNAs. The results showed a total of seven genes (CCNB1, SHCBP1, KPNA2, CCNE2, SFRP1, ELAVL2 and CCL20) were hub genes (Fig. 4d).

Construction and characterization of the ceRNA network associated with survival status

We first constructed the global ceRNA network using DEmRNAs, DEmiRNAs and DElncRNAs. This global ceRNA network comprised of 94 lncRNAs, 18 miRNAs, 27 mRNAs and 379 edges (Supplementary Fig. S2).

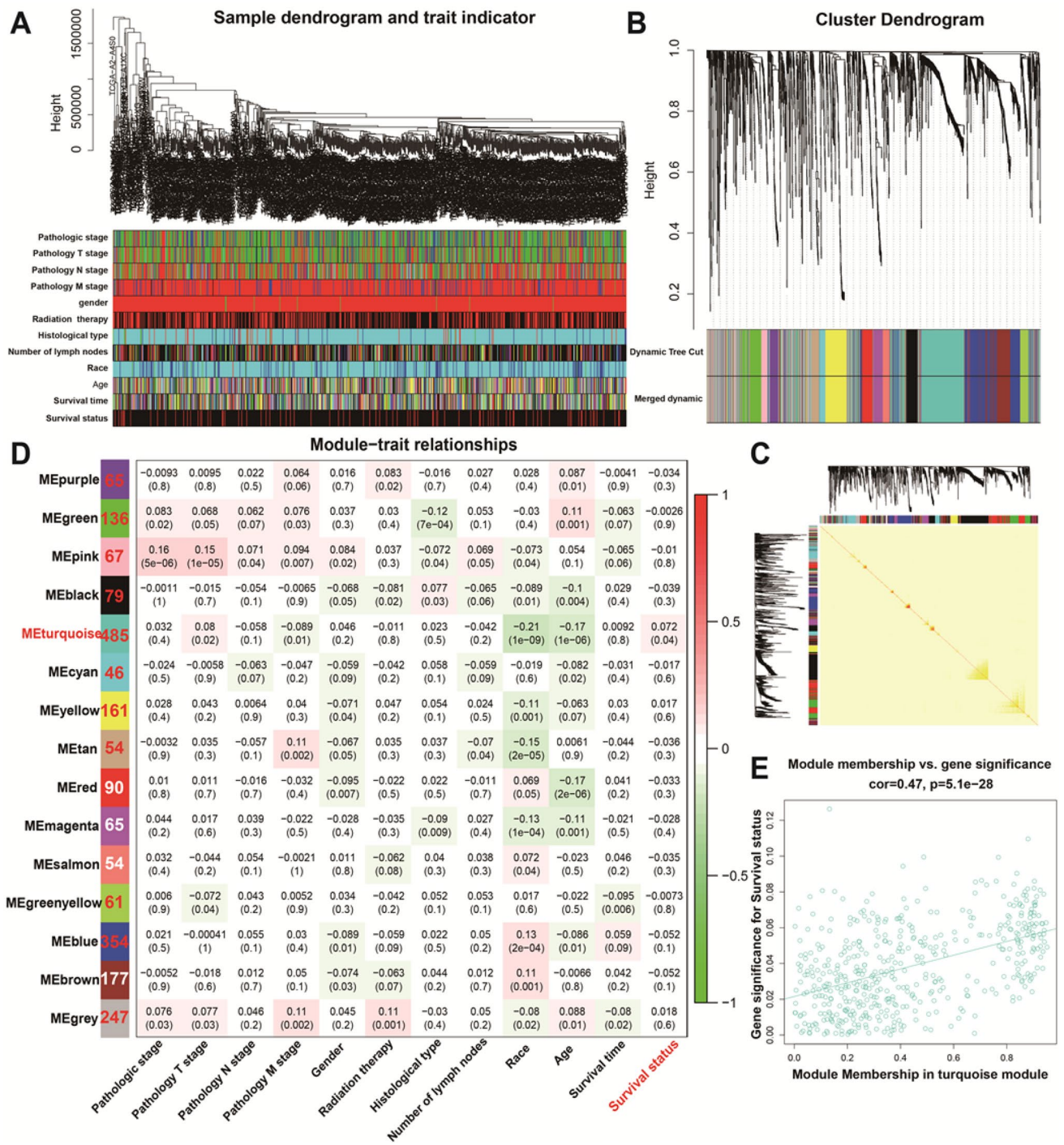
We generated the hub ceRNA network based on the global ceRNA network using the hub genes (Fig. 4e). The hub ceRNA network included 7 miRNAs, Hsa-miR-141, hsa-miR-144, hsa-miR-183, hsa-miR-200a, hsa-miR-206, hsa-miR-21, and hsa-miR-429, and 23 DElncRNAs (Table 2). We defined these miRNAs and lncRNAs as hub biomarkers (Fig. 4e).

We performed survival analysis of 23 DElncRNAs, 7 DEmiRNAs and 7 DEmRNAs. The results showed that up-regulation of 12 hub DElncRNAs, including WT1-AS, AL513123.1, FNDC1-IT1, LINC00200, AC007731.1, HOTAIR, UCA1, AC080037.1, TLR8-AS1, MUC19, LINC00461 and C10orf126, was associated with poor prognosis (Supplementary Fig. S3). Up-regulation of other 11 hub DElncRNAs predicated a longer patient survival time (Supplementary Fig. S4). MicroRNAs miR-144 and miR-429, but not miR-141, miR-183, miR-200a, miR-206, or miR-21, were associated with overall survival (Supplementary Fig. S5). The hub genes CCNB1, SHCBP1, KPNA2, CCNE2 and SFRP1, but not CCL20 and ELAVL2, were associated with overall survival (Supplementary Fig. S6–S7).

Identification and characterization of lncRNAs prognostic signature model

To define an lncRNA prognostic signature model, we performed univariate survival analyses of 23 hub lncRNAs and selected 9 lncRNAs with P value < 0.05 (Table 3). After a stepwise model selection by the Akaike information criterion (AIC) and multivariate Cox hazard model analysis, we generated a lncRNAs signature model, which calculated prognostic score: Prognostic score = $(0.121 \times \mathrm{EXP}_{\mathrm{AC007731.1}}) + (0.108 \times \mathrm{EXP}_{\mathrm{AL513123.1}}) + (0.105 \times \mathrm{EXP}_{\mathrm{C10orf126}}) + (0.065 \times \mathrm{EXP}_{\mathrm{WT1-AS}}) + (-0.126 \times \mathrm{EXP}_{\mathrm{ADAMTS9-AS1}}) + (-0.130 \times \mathrm{EXP}_{\mathrm{SRGAP3-AS2}}) + (0.116 \times \mathrm{EXP}_{\mathrm{TLR8-AS1}}) + (0.060 \times \mathrm{EXP}_{\mathrm{HOTAIR}})$.

We analyzed prognostic scores (risk scores) of patients from the training set ($n = 836$) using the lncRNA expression levels. The results showed that the median prognostic score was 1.088 and separated patients into either the high-risk group or the low-risk group (Fig. 5a). Patients with the high-risk scores exhibited poorer overall survival than those with the low-risk scores did (HR 2.607, $P < 0.001$) according to the Kaplan–Meier curve (Fig. 5b).



We evaluated the prognostic performance of the eight-lncRNA signature using ROC curves. The results showed that the AUC of the logistic regression, support vector machine, decision tree and random forest models were 0.999, 0.973, 0.731 and 0.898, respectively, for the train set (Fig. 5c), and the AUC was between 0.664 and 0.746 using survivalROC package (Fig. 5d–m), suggesting that the eight-lncRNAs signature could predict prognosis of BRCA.

We evaluated the independent prognostic performance of individual lncRNA in the eight-lncRNAs signature using a multivariate analysis. The results showed that AC007731.1, WT1-AS, ADAMTS9-AS1, SRGAP3-AS2, TLR8-AS1, HOTAIR were independent prognostic factors for overall survival (Table 4). This was consistent with Kaplan–Meier curves (Fig. 6, a collection from Supplementary Figs. S3-S4). Of these eight-lncRNAs, four were associated with high risk (AC007731.1, WT1-AS,

◂**Fig. 3** Network construction of the weighted coexpressed genes and their associations with clinical traits. **a** Hierarchical clustering tree of the TCGA-BRCA samples based on the train set. Dendrogram tips are labeled with the TCGA-BRCA unique name and experiment identifier. In the hierarchical dendrogram, lower branches correspond to higher co-expression (height = Euclidean distance). Identical colors in the eight bands below the dendrogram depict the TCGA-BRCA clinical traits. **b** Cluster dendrogram and color display of the co-expression network modules produced by average linkage hierarchical clustering of genes based on topological overlaps in the DEmiRNAs. Each branch in the dendrogram is a line that represents a single gene. Height indicates the Euclidean distance. Each color indicates a single module which contained genes with conservation closely in the dataset. **c** Heat map view of topological overlap of coexpressed genes in different modules. The heat map was generated from the topological overlap values between genes. The genes were grouped into modules labelled by a color code, which are given under the gene dendrogram on both sides. The topological overlap was high among genes of same module. **d** Module-trait relationships. The numbers represent R^2 values of Pearson correlations between the modules and traits. The numbers in the parentheses are corresponding *P* values. The background colors of the numbers represent the strength of the correlation. **e** Intra-modular analysis of gene significance for survival status and module membership of the genes in the turquoise module. *T* extent of the tumor, *N* extent of spread to the lymph nodes, *M* presence of metastasis

TLR8-AS1, HOTAIR, hazard ratio (HR) > 1) and two were associated with protective (ADAMTS9-AS1, SRGAP3-AS2, HR < 1) (Table 4).

We evaluated the prognostic performance of different clinical characteristics using the univariate Cox proportional hazards regression. The results showed that gender, pathologic stage, pathology T stage, pathology M stage, histological type and risk score were associated with overall survival ($P = 0.018$, $P < 0.001$, $P = 0.021$, $P = 0.013$, $P = 0.005$ and $P < 0.001$, respectively) (Table 5; Fig. 7). We further performed a multivariate model regression using the lncRNA signature, gender and histological type as covariates. The results showed that the lncRNAs signature and gender were independent prognostic factors associated with overall survival (HR 2.158, $P = 0.001$; HR = 3.963, $P = 0.020$, respectively) (Table 5).

Examination of the expression patterns of these eight signature lncRNAs in the low-risk group and the high-risk group showed that the levels of AC007731.1, AL513123.1, C10orf126, WT1-AS, TLR8-AS1 and HOTAIR were higher in the high-risk group than those of the low-risk group. The levels of ADAMTS9-AS1 and SRGAP3-AS2 were lower in the high-risk group than those of the low-risk group (Fig. 8).

Validation of prognostic performance of the lncRNAs signature in BRCA

To evaluate the performance of the eight-lncRNAs signature in prognosis of BRCA patients, we tested it with the test cohort of 251 patients. With the same model and cutoff point as those derived from the train cohort, the test cohort was classified into the high-risk group ($n = 149$) and low-risk group ($n = 102$) (Fig. 9a). Compared with the low-risk group, the high-risk group in the TCGA BRCA test cohort had significantly shorter overall survival (HR 2.640, 95% CI 1.729–4.033, log-rank test) (Fig. 9b). The AUC of the eight-lncRNAs signature logistic regression, support vector machine, decision tree and random forest models were 0.979, 0.844, 0.99 and 0.997, respectively, for the test set (Fig. 9c). The AUCs in predicting 1- to 10-year survival in BRCA were between 0.656 and 0.748 in the test set (Fig. 9d). The results supported that the eight-lncRNA signature could predict prognosis of BRCA.

Discussion

Molecular signatures are superior to single biomarker in prognosis of BRCA. To identify the gene signature associated with survival status of BRCA, we first constructed the global ceRNA network based on the lncRNA-miRNA-mRNA ceRNA network hypothesis and then constructed a hub lncRNA-miRNA-mRNA ceRNA network using seven hub genes (CCNB1, SHCBP1, KPNA2, CCNE2, SFRP1, ELAVL2 and CCL20). These seven hub genes become candidate biomarkers for overall survival of breast cancer because they belong to the turquoise module that is correlated significantly with survival status of BRCA. Our survival analysis using the training dataset from the TCGA database showed that CCNB1, SHCBP1, KPNA2, CCNE2 and SFRP1, but not CCL20 and ELAVL2, were associated with overall survival. These genes are shown to be involved in progression of BRCA. CCNB1 is a prognostic biomarker for ER + BRCA [31], SHCBP1 is over-expressed in BRCA and is important in the proliferation and apoptosis of the human malignant BRCA cell line [32], KPNA2 is a nuclear export protein that contributes to aberrant localization of key proteins and poor prognosis of BRCA [33], CCNE2 is associated with HER2 + BRCA trastuzumab resistance [34] and involved in BRCA oncogenesis and aggressiveness [35, 36], and SFRP1 is a tumor suppressor for BRCA [37, 38]. Therefore, CCNB1, SHCBP1, KPNA2, CCNE2 and SFRP1are likely prognostic factors of overall survival of breast cancer.

A total 23 hub lncRNAs were included in our hub ceRNA network (Fig. 4e; Table 2). Our survival analysis using the training dataset from the TCGA database showed that they are correlated with survival status of BRCA. Although studies that show the role of these lncRNAs in progression of BRCAs are very limited, several studies showed that lncRNA UCA1 promotes epithelial-mesenchymal transition (EMT) of BRCA cells [39]. HOTAIR is a key regulator of proliferation, colony formation, invasion and self-renewal capacity in BRCA stem cells [40], and it enhances

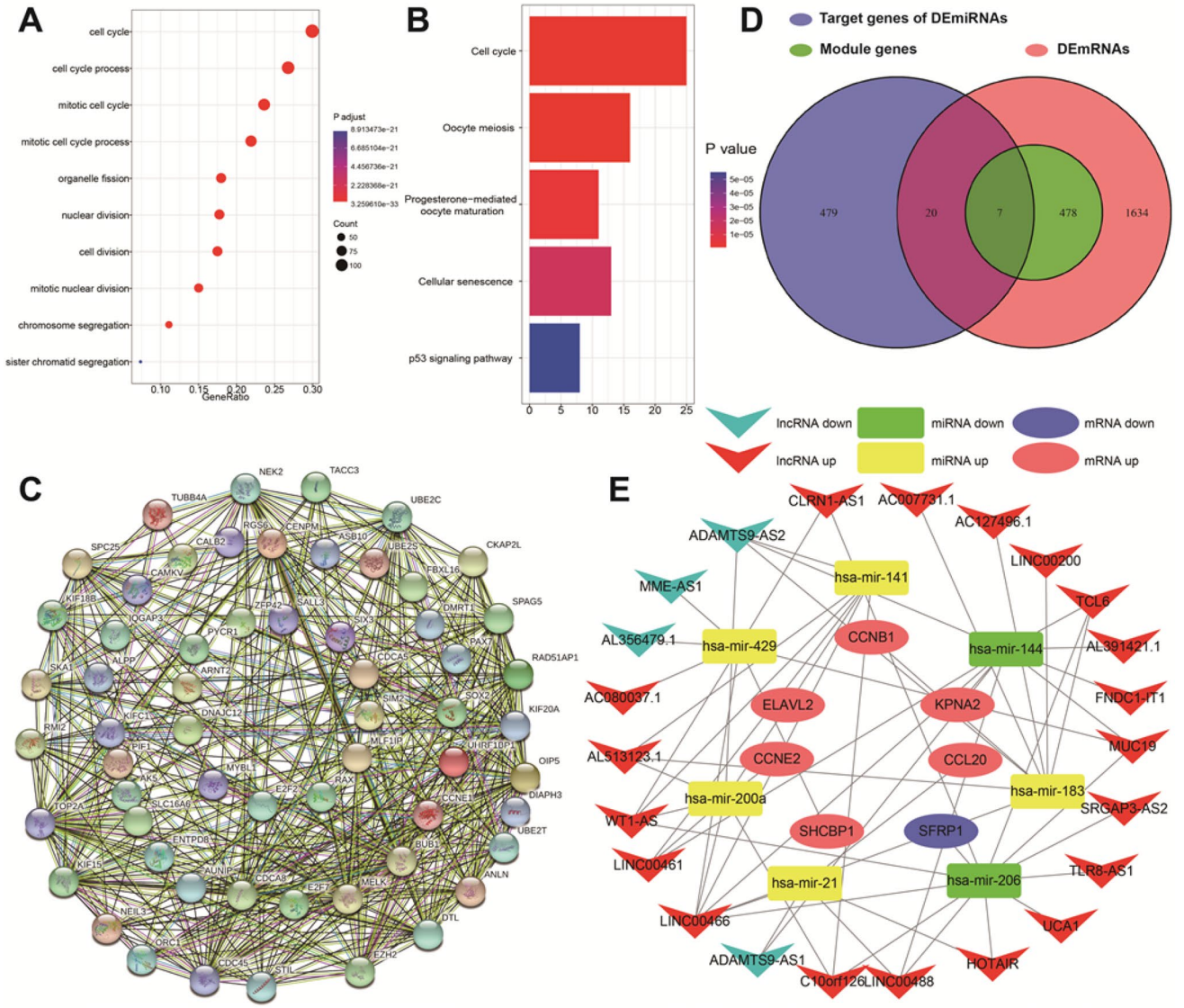


Fig. 4 Functional enrichment analysis, PPI, hub genes and ceRNA network of the module associated with survival status. The most significantly enriched GO terms (**a**) and KEGG pathways (**b**) of the genes in the turquoise module. **c** Protein–protein interaction network of the genes in the turquoise module. **d** Hub genes were identified using Venn diagram that showed the overlap of DEmRNAs, targeted genes of DEmiRNAs and turquoise module genes. **e** The lncRNA-miRNA-mRNA ceRNA network associated with survival status was generated based on the hub genes and the global ceRNA network

Table 2 The hub DElncRNAs, DEmiRNAs, and DEmRNAs preserved in ceRNA network

The type of RNAs	Gene symbols
DElncRNAs	WT1-AS, C10orf126, AL513123.1, LINC00466, ADAMTS9-AS2,LINC00461, TCL6, AL391421.1, MUC19, LINC00488, FNDC1-IT1, ADAMTS9-AS1, AC127496.1, LINC00200, AC007731.1, UCA1, HOTAIR, SRGAP3-AS2, TLR8-AS1, CLRN1-AS1, AL356479.1, AC080037.1, MME-AS1
DEmiRNAs	hsa-miR-141, hsa-miR-144, hsa-miR-183, hsa-miR-200a, hsa-miR-206, hsa-miR-21, hsa-miR-429
DEmRNAs	CCNB1(cyclin B1), SHCBP1(SHC SH2-domain binding protein 1), KPNA2(karyopherin alpha 2), CCNE2(cyclin E2), SFRP1(secreted frizzled-related protein 1), ELAVL2(ELAV like neuron-specific RNA binding protein 2), CCL20(chemokine C-C motif ligand 20)

DElncRNAs differentially expressed long noncoding RNAs, *DemiRNAs* differentially expressed microRNAs, *DEmRNAs* differentially expressed mRNAs, *ceRNA* competing endogenous RNA

Table 3 Prognostic value of the differentially expressed lncRNAs by univariate Cox regression analysis

LncRNA	Hazard ratio (95% CI)	*P* value	Z value	logHR	selogHR
AC007731.1*	1.205 (1.076–1.348)	0.001	3.242	0.186	0.057
AL513123.1*	1.188 (1.059–1.331)	0.003	2.948	0.172	0.058
C10orf126*	1.181 (1.06–1.316)	0.003	3.011	0.166	0.055
WT1-AS*	1.092 (1.026–1.161)	0.005	2.787	0.088	0.031
ADAMTS9-AS1*	0.867 (0.783–0.96)	0.006	−2.736	−0.143	0.052
SRGAP3-AS2*	0.886 (0.808–0.971)	0.01	−2.588	−0.121	0.047
TLR8-AS1*	1.154 (1.031–1.291)	0.013	2.494	0.143	0.057
HOTAIR*	1.068 (1.01–1.128)	0.02	2.327	0.065	0.028
AL391421.1*	0.942 (0.889–0.998)	0.042	−2.029	−0.06	0.03
LINC00200	1.093 (0.999–1.197)	0.053	1.933	0.089	0.046
MME-AS1	0.804 (0.643–1.005)	0.056	−1.914	−0.218	0.114
CLRN1-AS1	0.919 (0.833–1.014)	0.093	−1.681	−0.084	0.05
AL356479.1	0.949 (0.892–1.011)	0.104	−1.626	−0.052	0.032
LINC00488	0.853 (0.704–1.035)	0.107	−1.611	−0.159	0.098
AC127496.1	0.898 (0.787–1.024)	0.109	−1.604	−0.108	0.067
MUC19	1.039 (0.987–1.094)	0.144	1.462	0.038	0.026
LINC00466	0.936 (0.853–1.026)	0.16	−1.406	−0.066	0.047
AC080037.1	1.065 (0.974–1.165)	0.169	1.375	0.063	0.046
FNDC1-IT1	1.068 (0.959–1.189)	0.23	1.2	0.066	0.055
LINC00461	1.034 (0.974–1.098)	0.272	1.098	0.034	0.031
ADAMTS9-AS2	0.959 (0.863–1.066)	0.441	−0.771	−0.042	0.054
TCL6	1.027 (0.946–1.114)	0.528	0.63	0.026	0.042
UCA1	0.996 (0.926–1.072)	0.923	−0.097	−0.004	0.037

*$P < 0.05$

ER signaling and confers tamoxifen resistance in BRCA [41] and an independent prognostic marker of metastasis in estrogen receptor-positive primary BRCA [42]. It is likely that 23 hub DElncRNAs are promising individual independent prognostic biomarkers of overall survival for BRCA. Their roles in progression of BRCAs remain to be further investigated.

Our survival analysis showed that miR-144 and miR-429, but not miR-141, miR-183, miR-200a, miR-206 or miR-21 of the hub ceRNA network were significantly correlated with survival status of BRCA using the training dataset from the TCGA database. Published studies have shown that miR-144 suppresses proliferation, invasion and migration of BRCA cells [43, 44] and miR-429 inhibits migration and invasion of BRCA cells [45–47], suggesting that miR-144 and miR-429 are significantly involved in progression of BRCAs. Therefore, miR-144 and miR-429 could be prognostic factors of overall survival of BRCAs. According to the ceRNA network hypothesis, miRNA is the centric player that mediates mRNA and lncRNA levels and interactions through MREs [12–15]. In our hub ceRNA network, five mRNAs and all 23 lncRNAs but only two miRNAs are significantly correlated with survival status of BRCA. It seems that the miRNA-centric lncRNA-miRNA-mRNA ceRNA network is a good tool to identify lncRNAs that is correlated with overall survival of BRCA where the miRNAs are mediators and not necessary predictors for survival status.

Systematic analysis based on the lncRNA-miRNA-mRNA ceRNA network hypothesis has recently been applied to BRCAs. Distinct lncRNA-miRNA-mRNA patterns were discovered in BRCA molecular subtypes and might have a better prognostic value than that of the individual genes [16, 48, 49]. LINC0092 coexpressed with SFRP1 and RGMA and regulated by hsa-miR-449a and hsa-miR-452-5p and C2orf71 coexpressed with LINC00511 and regulated by hsa-miR-184 were correlated with better prognosis of ER(+) and ER(−) subtypes of BRCA [50]. miR-19a coexpressed with lncRNA-DLEU1 to co-regulate the expression of ESR1 [51]. Loss of PVT1 ceRNA activity contributes to human BRCA [52]. The over-expression levels of AKAP12, FOS, EMX2OS, MYCNOS, RP11-542B15.1, hsa-miR-9 and hsa-miR-183 predicted patients with poor outcome and high expression of hsa-miR-145, LINC00461, RP11-576D8.4 and RP11-496D24.2 have shown longer overall survival time [19]. Our hub ceRNA network identified five mRNAs, two miRNAs and 23 lncRNAs are promising prognostic biomarkers of overall survival for BRCA, providing a novel clue to further investigation of their prognostic value in subtype-specific overall survival of BRCA.

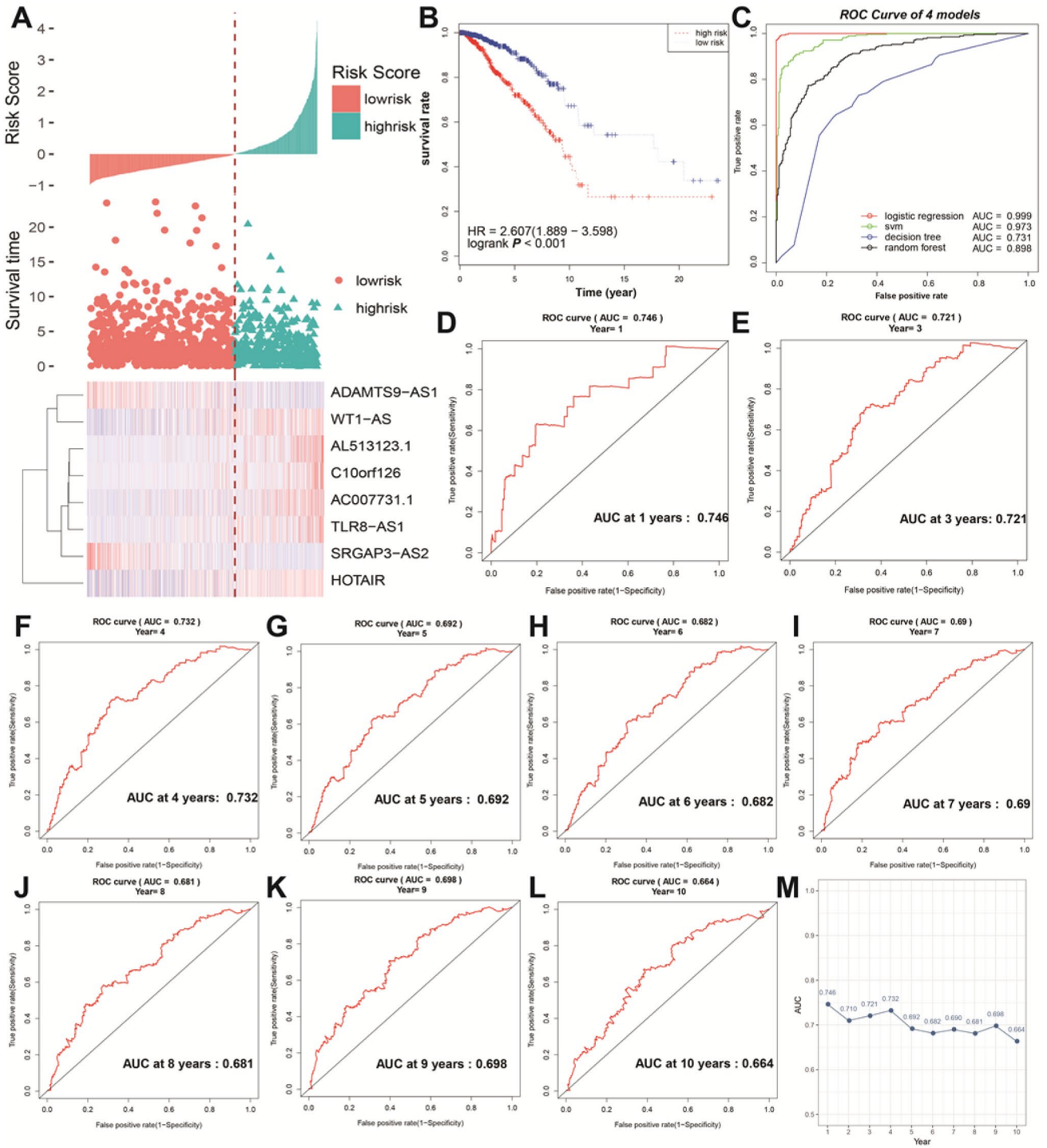


Fig. 5 Diagnostic efficiency of the eight-lncRNA signature in the TCGA-BRCA train cohort. **a** Risk score analysis of the eight-lncRNA signature of BRCA. Risk score of lncRNA signature (Top); duration of cases (Middle); low and high score groups for the 8 lncRNAs (Bottom). **b** Survival analysis of the high-risk group and the low-risk group using Kaplan–Meier curves. **c** The diagnostic efficiency of the eight-lncRNA signature. The AUC for the eight-lncRNAs signature logistic regression, support vector machine, decision tree and random forest models were calculated. **d–m** The diagnostic efficiency of the eight-lncRNA signature for survival time. ROC curves of the eight-lncRNA signature for predicting 1- to 10-year survival were analyzed

Table 4 Prognostic value of the differentially expressed lncRNAs by multivariate Cox regression analysis

LncRNA	Hazard ratio (95% CI)	*Z* value	*P* value
AC007731.1*	1.129 (1.004;1.268)	2.033	0.042
AL513123.1	1.114 (0.987;1.258)	1.743	0.081
C10orf126	1.111 (0.994;1.242)	1.856	0.063
WT1-AS*	1.067 (1.001;1.138)	1.966	0.049
ADAMTS9-AS1*	0.882 (0.792;0.982)	−2.298	0.022
SRGAP3-AS2*	0.878 (0.800;0.963)	−2.747	0.006
TLR8-AS1*	1.123 (1.002;1.260)	1.991	0.046
HOTAIR*	1.061 (1.003;1.123)	2.066	0.039

*$P < 0.05$

In the current study, we generated a prognostic score model of eight lncRNA signature. We showed that this eight lncRNA signature, together with gender, pathologic stage, pathology T stage, pathology M stage, histological type, were associated with overall survival using the univariate Cox proportional hazards regression. This signature was also an independent prognostic factors associated with overall survival according to a multivariate model regression. We further showed that the eight-lncRNA signature could predict overall survival using the test cohort of TCGA dataset based on survival analysis, logistic regression, support vector machine, decision tree and random forest models. Our

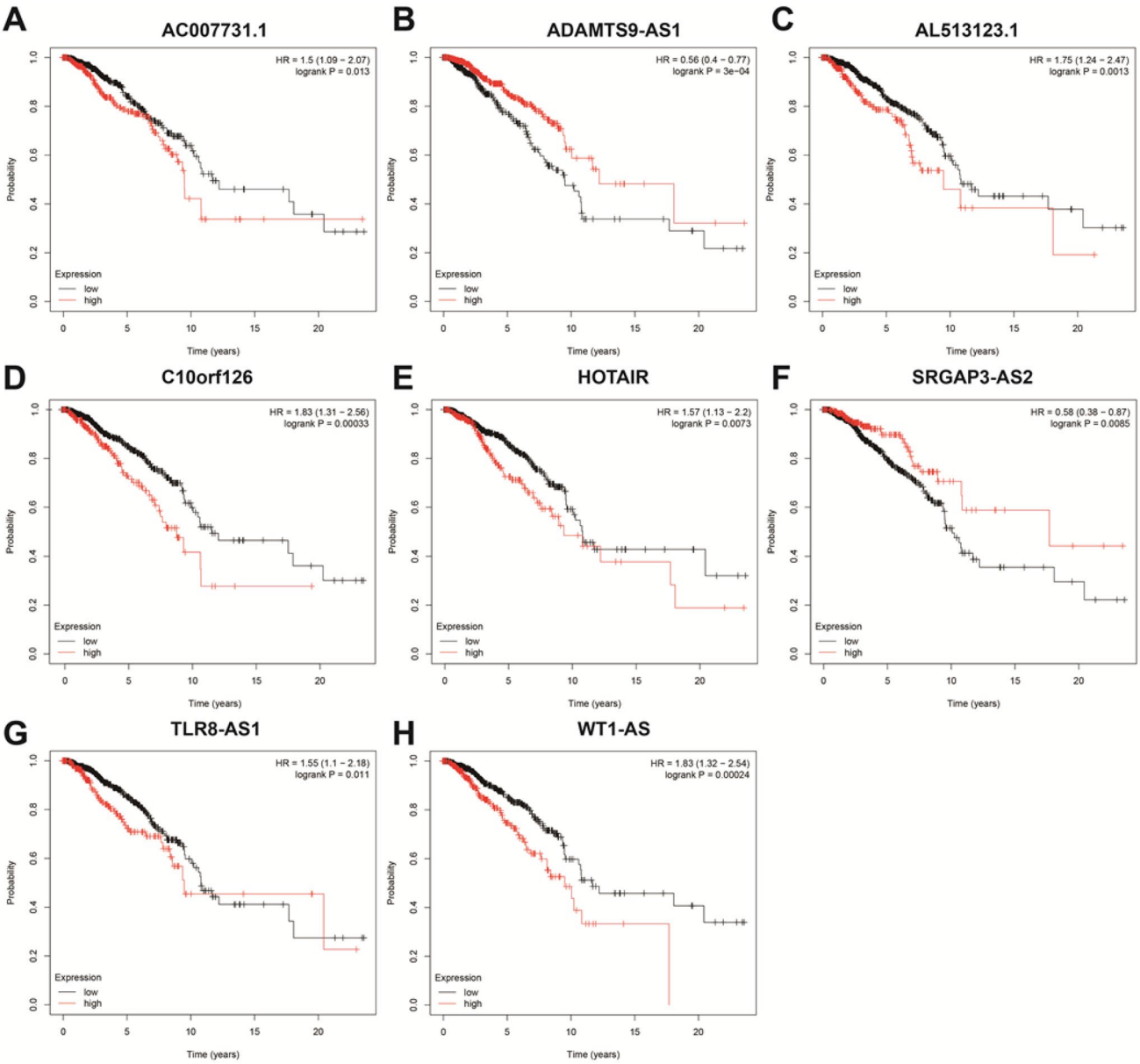


Fig. 6 Independent prognostic power (efficiency) of individual lncRNA in the eight-lncRNA signature. **a** AC007731.1, **b** ADAMTS9-AS1, **c** AL513123.1, **d** C10orf126, **e** HOTAIR, **f** SRGAP3-AS2, **g** TLR8-AS1, **h** WT1-AS. A multivariate analysis of the TCGA BRCA train set was performed to generate Kaplan–Meier curves. Horizontal axis, overall survival time. Vertical axis, overall survival

Table 5 Predictive values of related clinical features and risk score

	Univariate analysis[a]		Multivariate analysis[b]	
	HR (95% Cl)	*P* value	HR (95% CI)	*P* value
Age, years (≥median vs.<median)	1.062 (0.699;1.612)	0.778	–	–
Gender (male vs. female)	3.692 (1.158;11.770)	0.027	3.964 (1.242;12.653)	0.020
Race (white vs. others)	0.598 (0.188;1.899)	0.383	–	–
Pathologic stage (III and IV vs. I and II)	2.349 (1.542;3.577)	<0.001	–	–
Pathology T stage(III and IV vs. I and II)	1.735 (1.078;2.790)	0.023	–	–
Pathology N stage(II and III vs. I)	1.328 (0.869;2.030)	0.19	–	–
Pathology M stage (present vs. absent)	1.691 (1.114;2.568)	0.014	–	–
Radiation therapy (absent vs. present)	1.137 (0.745;1.736)	0.552	–	–
Histological type (infiltrating vs. others)	2.26 (1.255;4.072)	0.007	1.639 (0.878;3.056)	0.121
Lymph nodes (present vs. absent)	1.204 (0.758;1.914)	0.431	–	–
Risk score (≥median vs. <median)	2.459 (1.585;3.814)	<0.001	2.158 (1.355;3.436)	0.001

[a]The data were subjected to Cox's proportional hazards regression model

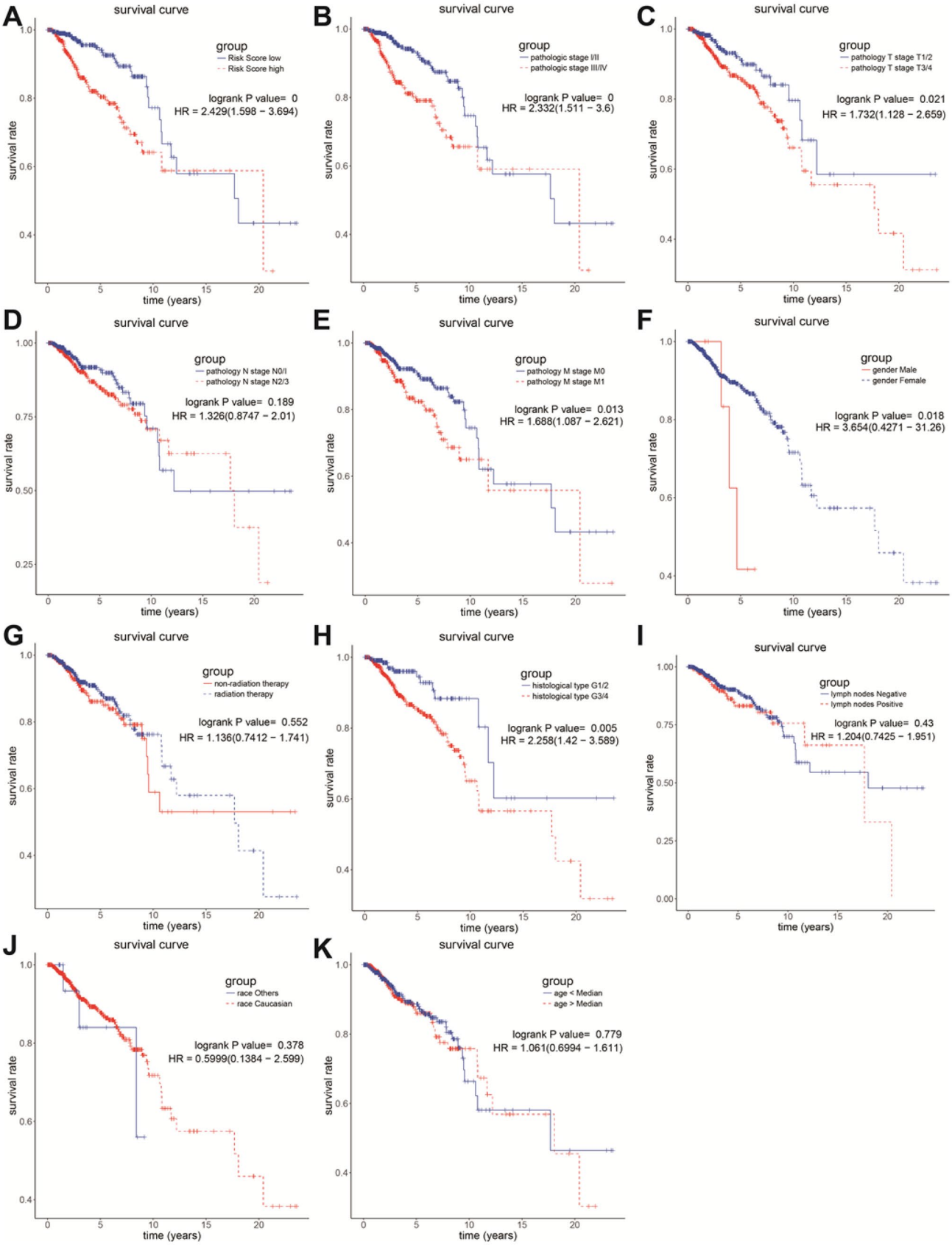
[b]Multivariate analysis used stepwise addition and removal of clinical covariates found to be associated with survival in univariate models ($P < 0.05$) and final models include only those covariates that were significantly associated with survival (Wald statistic, $P < 0.05$)

data supported that the eight lncRNAs signature is likely a promising and viable prognostic signature for survival status, adding to the signature pools of Oncotype DX [3], MammaPrint [4–7], microRNA signatures [8, 9], a functional patient-specific subpathway activities signature [10] and Hedgehog-mesenchyme gene signature [11]. In addition, the prognostic scores can be easily calculated according to the prognostic score equation and the levels of eight lncRNA in the signature which can be easily determined using quantitative real-time PCR or some other customized test methods. Risks can be predicted by the value of the prognostic scores. Higher prognostic scores predicted higher risks and poorer prognosis. Therefore, it is easy for oncologists to use prognostic score in practice.

In the eight-lncRNA risk-score signature, AC007731.1, AL513123.1, C10orf126, WT1-AS, TLR8-AS1 and HOTAIR were associated with high-risk (HR > 1), and the coefficients of ADAMTS9-AS1 and SRGAP3-AS2 are negative, denoting that high expression of ADAMTS9-AS1 and SRGAP3-AS2 is associated with protective effect (HR < 1). This is supported by that the levels of AC007731.1, AL513123.1, C10orf126, WT1-AS, TLR8-AS1 and HOTAIR were higher in the high-risk group than those of the low-risk group. The levels of ADAMTS9-AS1 and SRGAP3-AS2 were lower in the high-risk group than those of the low-risk group. Studies have showed that HOTAIR overexpression increases BRCA cell proliferation and mediates tamoxifen-resistant cell growth [41, 53, 54], probably mediated by NF-kappaB-HOTAIR axis driven a positive-feedback loop cascade during DNA damage response [55]. HOTAIR contributes to epithelial-mesenchymal transition mediated by Snail [56]. HOTAIR expression level in primary breast tumors was a robust predictor of eventual metastasis and death [57]. The function of other lncRNAs signature members in the progression of BRCA remains unclear although study showed that WT1-AS is involved in ovarian clear cell adenocarcinoma [58], gastric cancer [59], hepatocellular carcinoma [60], ADAMTS9-AS1 is involved in malignant epithelial ovarian cancer progression [61], correlated with bladder cancer patient survival [62], differentially expressed in colon adenocarcinoma [63], the development of ectoderm and epithelial cells [64], SRGAP3-AS2 was differentially expressed in human lung adenocarcinoma [65]. In the current study, we found for the first time that they are differentially expressed in BRCA and involved in BRCA prognosis, providing a clue to further investigation on their role in oncogenesis and progression of BRCA.

There are three main limitations in the current study. The first, our hub ceRNA network model might be changed with the update of the interaction databases of DEmRNA, DEmiRNA and DElncRNA. The second, we did not analyze the subtype-specific prognostic signature which has a much higher resolution in the risk stratification, leading to improved therapies and precision medicine for patients with BRCA [66]. The third, our model remains to be tested for validity in real samples. Some related prospective clinical studies in multiple oncology centers should be performed to

Fig. 7 Prognostic efficiency of the clinical features for OS of the patients in TCGA BRCA train set. **a** risk score, **b** pathologic stage, **c** pathology T stage, **d** pathology N stage, **e** pathology M stage, **f** gender, **g** radiation therapy, **h** histological type, **i** lymph nodes, **j** race, **k** age. The univariate Cox proportional hazards regression was performed. Horizontal axis, overall survival time. Vertical axis, overall survival


A
survival curve
group
Risk Score low
Risk Score high
logrank P value= 0
HR = 2.429(1.598 − 3.694)
survival rate
time (years)
B
survival curve
group
pathologic stage I/II
pathologic stage III/IV
logrank P value= 0
HR = 2.332(1.511 − 3.6)
survival rate
time (years)
C
survival curve
group
pathology T stage T1/2
pathology T stage T3/4
logrank P value= 0.021
HR = 1.732(1.128 − 2.659)
survival rate
time (years)
D
survival curve
group
pathology N stage N0/I
pathology N stage N2/3
logrank P value= 0.189
HR = 1.326(0.8747 − 2.01)
survival rate
time (years)
E
survival curve
group
pathology M stage M0
pathology M stage M1
logrank P value= 0.013
HR = 1.688(1.087 − 2.621)
survival rate
time (years)
F
survival curve
group
gender Male
gender Female
logrank P value= 0.018
HR = 3.654(0.4271 − 31.26)
survival rate
time (years)
G
survival curve
group
non-radiation therapy
radiation therapy
logrank P value= 0.552
HR = 1.136(0.7412 − 1.741)
survival rate
time (years)
H
survival curve
group
histological type G1/2
histological type G3/4
logrank P value= 0.005
HR = 2.258(1.42 − 3.589)
survival rate
time (years)
I
survival curve
group
lymph nodes Negative
lymph nodes Positive
logrank P value= 0.43
HR = 1.204(0.7425 − 1.951)
survival rate
time (years)
J
survival curve
group
race Others
race Caucasian
logrank P value= 0.378
HR = 0.5999(0.1384 − 2.599)
survival rate
time (years)
K
survival curve
group
age < Median
age > Median
logrank P value= 0.779
HR = 1.061(0.6994 − 1.611)
survival rate
time (years)

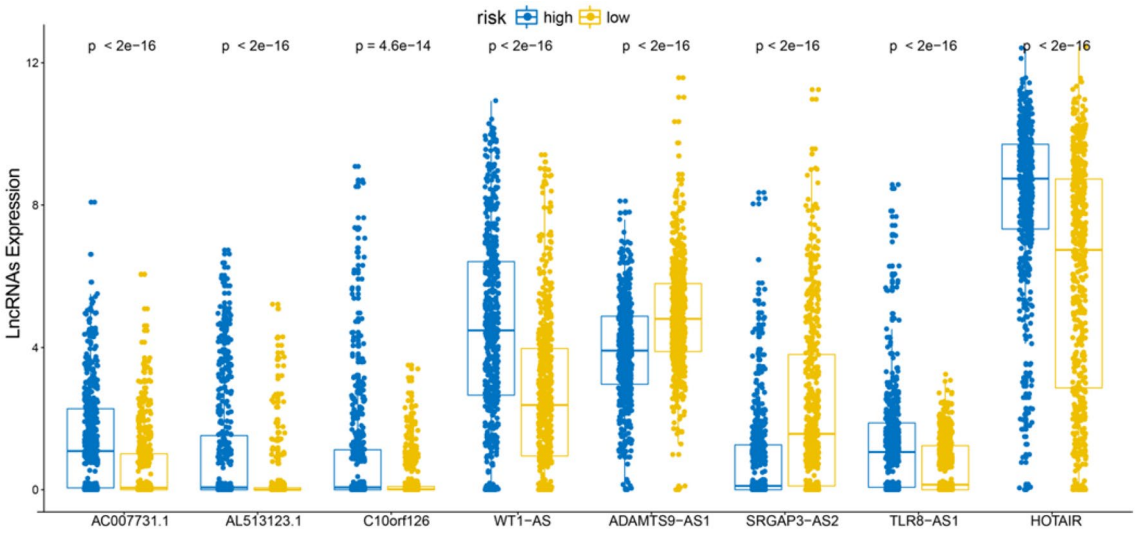


Fig. 8 Expression levels of the eight lncRNAs in the low-risk group and the high-risk group

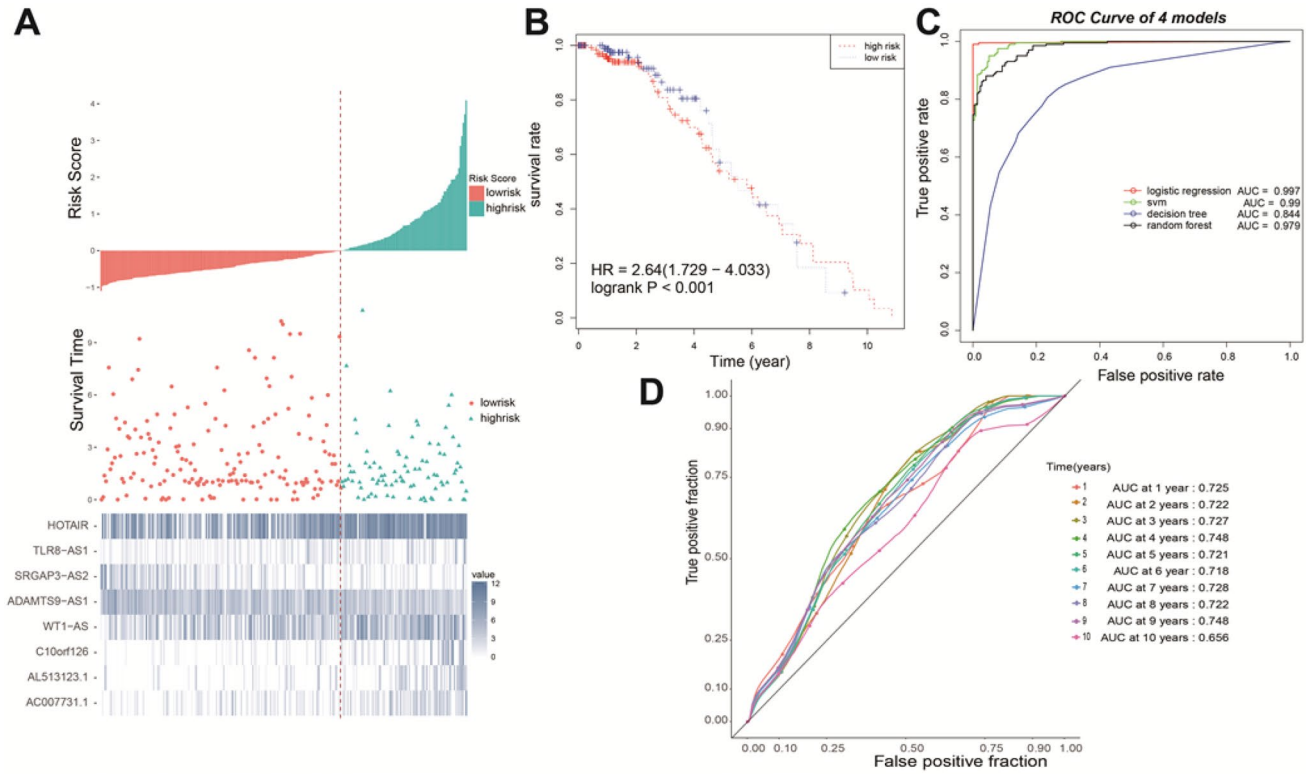


Fig. 9 Diagnostic performance of the eight-lncRNA signature in the TCGA-BRCA test cohort. **a** Risk score analysis of the eight-lncRNA signature of BRCA. Risk score of lncRNA signature (Top); duration of cases (Middle); low and high score groups for the 8 lncRNAs (Bottom). **b** Survival analysis of the high-risk group and the low-risk group using Kaplan–Meier curves. **c** The diagnostic efficiency of the eight-lncRNA signature. The AUC for the eight-lncRNAs signature logistic regression, support vector machine, decision tree and random forest models were calculated. **d** The diagnostic efficiency of the eight-lncRNA signature for survival time. ROC curves of the eight-lncRNA signature for predicting 1- to 10-year survival were analyzed

test the validity of the model in real samples. These limitations will be investigated in our studies in the future.

In summary, we identified a hub lncRNA-miRNA-mRNA ceRNA network comprising of the seven hub genes, seven hub miRNAs and 23 hub lncRNAs. All 23 lncRNAs are potentially associated with overall survival of BRCA. The lncRNA-miRNA-mRNA ceRNA network is a good tool to identify lncRNAs that is correlated with overall survival of BRCA. We identified and characterized an eight-lncRNA signature for prediction of overall survival for BRCA patients. Our finding may provide a clue to develop a promising prognostic tool for overall survival of BRCA.

Acknowledgements The authors gratefully acknowledge The Cancer Genome Atlas pilot project (established by NCI and NHGRI), which made the genomic data and clinical data of BRCA available.

Author contributions Min Sun, Dong Huang and Xinsheng Gu participated in research design. Min Sun, Di Wu, Mengyu Du and Jin Zha performed data analysis. Min Sun, Mengyu Du and Xinsheng Gu wrote or contributed to the writing of the manuscript.

Funding This research was supported by the Natural Science Foundation of Hubei Provincial Department of Education (Q20182105), Natural Science Foundation of Hubei Province of China (2016CFB530) and Faculty Development Foundation of Hubei University of Medicine (2014QDJZR01), Chen Xiao-ping Foundation for the development of science and technology of Hubei Provincial (CXP-JJH11800001-2018333) and Innovation and entrepreneurship training program (201810929009, 201810929068 and 201813249010).

Compliance with ethical standards

Conflict of interest The authors declare that they have no conflicts of interest in this work.

References

1. DeSantis CE, Ma J, Goding Sauer A, Newman LA, Jemal A (2017) Breast cancer statistics, 2017, racial disparity in mortality by state. CA 67(6):439–448
2. Li G, Hu J, Hu G (2017) Biomarker studies in early detection and prognosis of breast cancer. Adv Exp Med Biol 1026:27–39
3. Kaklamani V (2006) A genetic signature can predict prognosis and response to therapy in breast cancer: oncotype DX. Expert Rev Mol Diagn 6(6):803–809
4. Mook S, Schmidt MK, Weigelt B, Kreike B, Eekhout I, van de Vijver MJ, Glas AM, Floore A, Rutgers EJ, van 't Veer LJ (2010) The 70-gene prognosis signature predicts early metastasis in breast cancer patients between 55 and 70 years of age. Ann Oncol 21(4):717–722
5. Bueno-de-Mesquita JM, Linn SC, Keijzer R, Wesseling J, Nuyten DS, van Krimpen C, Meijers C, de Graaf PW, Bos MM, Hart AA et al (2009) Validation of 70-gene prognosis signature in node-negative breast cancer. Breast Cancer Res Treat 117(3):483–495
6. Weigelt B, Hu Z, He X, Livasy C, Carey LA, Ewend MG, Glas AM, Perou CM, van 't Veer LJ (2005) Molecular portraits and 70-gene prognosis signature are preserved throughout the metastatic process of breast cancer. Cancer Res 65(20):9155–9158
7. Kondo M, Hoshi SL, Ishiguro H, Toi M (2012) Economic evaluation of the 70-gene prognosis-signature (MammaPrint(R)) in hormone receptor-positive, lymph node-negative, human epidermal growth factor receptor type 2-negative early stage breast cancer in Japan. Breast Cancer Res Treat 133(2):759–768
8. Hironaka-Mitsuhashi A, Matsuzaki J, Takahashi RU, Yoshida M, Nezu Y, Yamamoto Y, Shiino S, Kinoshita T, Ushijima T, Hiraoka N et al (2017) A tissue microRNA signature that predicts the prognosis of breast cancer in young women. PLoS ONE 12(11):e0187638
9. Zhou J, Liu M, Chen Y, Xu S, Guo Y, Zhao L (2019) Cucurbitacin B suppresses proliferation of pancreatic cancer cells by ceRNA: Effect of miR-146b-5p and lncRNA-AFAP1-AS1. J Cell Physiol 234(4):4655–4667
10. Dianat-Moghadam H, Heydarifard M, Jahanban-Esfahlan R, Panahi Y, Hamishehkar H, Pouremamali F, Rahbarghazi R, Nouri M (2018) Cancer stem cells-emanated therapy resistance: implications for liposomal drug delivery systems. J Control Release 288:62–83
11. Torres-Garcia W, Domenech M (2017) Hedgehog-mesenchyme gene signature identifies bi-modal prognosis in luminal and basal breast cancer sub-types. Mol Biosyst 13(12):2615–2624
12. Salmena L, Poliseno L, Tay Y, Kats L, Pandolfi PP (2011) A ceRNA hypothesis: the Rosetta Stone of a hidden RNA language? Cell 146(3):353–358
13. Cesana M, Daley GQ (2013) Deciphering the rules of ceRNA networks. Proc Natl Acad Sci USA 110(18):7112–7113
14. An Y, Furber KL, Ji S (2017) Pseudogenes regulate parental gene expression via ceRNA network. J Cell Mol Med 21(1):185–192
15. Tay Y, Rinn J, Pandolfi PP (2014) The multilayered complexity of ceRNA crosstalk and competition. Nature 505(7483):344–352
16. Zhou S, Wang L, Yang Q, Liu H, Meng Q, Jiang L, Wang S, Jiang W (2018) Systematical analysis of lncRNA-mRNA competing endogenous RNA network in breast cancer subtypes. Breast Cancer Res Treat 169(2):267–275
17. Zheng L, Zhang Z, Zhang S, Guo Q, Zhang F, Gao L, Ni H, Guo X, Xiang C, Xi T (2018) RNA binding protein RNPC1 inhibits breast cancer cell metastasis via activating STARD13-correlated ceRNA network. Mol Pharm 15(6):2123–2132
18. Liu Y, Du Y, Hu X, Zhao L, Xia W (2018) Up-regulation of ceRNA TINCR by SP1 contributes to tumorigenesis in breast cancer. BMC Cancer 18(1):367
19. Yuan N, Zhang G, Bie F, Ma M, Ma Y, Jiang X, Wang Y, Hao X (2017) Integrative analysis of lncRNAs and miRNAs with coding RNAs associated with ceRNA crosstalk network in triple negative breast cancer. Onco Targets Ther 10:5883–5897
20. Li C, Zheng L, Xin Y, Tan Z, Zhang Y, Meng X, Wang Z, Xi T (2017) The competing endogenous RNA network of CYP4Z1 and pseudogene CYP4Z2P exerts an anti-apoptotic function in breast cancer. FEBS Lett 591(7):991–1000
21. Wang Z, Jensen MA, Zenklusen JC (2016) A practical guide to the cancer genome atlas (TCGA). Methods Mol Biol 1418:111–141
22. Tomczak K, Czerwinska P, Wiznerowicz M (2015) The cancer genome atlas (TCGA): an immeasurable source of knowledge. Contemp Oncol (Pozn) 19(1A):A68–A77
23. Langfelder P, Horvath S (2008) WGCNA: an R package for weighted correlation network analysis. BMC Bioinform 9:559
24. Zhang B, Horvath S (2005) A general framework for weighted gene co-expression network analysis. Stat Appl Genet Mol Biol 4:17
25. Yu G, Wang LG, Han Y, He QY (2012) clusterProfiler: an R package for comparing biological themes among gene clusters. OMICS 16(5):284–287
26. Harris MA, Clark J, Ireland A, Lomax J, Ashburner M, Foulger R, Eilbeck K, Lewis S, Marshall B, Mungall C et al (2004) The

gene ontology (GO) database and informatics resource. Nucleic Acids Res 32(Database issue):D258–D261
27. Kanehisa M, Goto S (2000) KEGG: kyoto encyclopedia of genes and genomes. Nucleic Acids Res 28(1):27–30
28. Szklarczyk D, Franceschini A, Wyder S, Forslund K, Heller D, Huerta-Cepas J, Simonovic M, Roth A, Santos A, Tsafou KP et al (2015) STRING v10: protein-protein interaction networks, integrated over the tree of life. Nucleic Acids Res 43:D447–D452
29. Cline MS, Smoot M, Cerami E, Kuchinsky A, Landys N, Workman C, Christmas R, Avila-Campilo I, Creech M, Gross B et al (2007) Integration of biological networks and gene expression data using cytoscape. Nat Protoc 2(10):2366–2382
30. Chou CH, Shrestha S, Yang CD, Chang NW, Lin YL, Liao KW, Huang WC, Sun TH, Tu SJ, Lee WH et al (2018) miRTarBase update 2018: a resource for experimentally validated microRNA-target interactions. Nucleic Acids Res 46(D1):D296–D302
31. Ding K, Li W, Zou Z, Zou X, Wang C (2014) CCNB1 is a prognostic biomarker for ER+ breast cancer. Med Hypotheses 83(3):359–364
32. Feng W, Li HC, Xu K, Chen YF, Pan LY, Mei Y, Cai H, Jiang YM, Chen T, Feng DX (2016) SHCBP1 is over-expressed in breast cancer and is important in the proliferation and apoptosis of the human malignant breast cancer cell line. Gene 587(1):91–97
33. Alshareeda AT, Negm OH, Green AR, Nolan CC, Tighe P, Albarakati N, Sultana R, Madhusudan S, Ellis IO, Rakha EA (2015) KPNA2 is a nuclear export protein that contributes to aberrant localisation of key proteins and poor prognosis of breast cancer. Br J Cancer 112(12):1929–1937
34. Tormo E, Adam-Artigues A, Ballester S, Pineda B, Zazo S, Gonzalez-Alonso P, Albanell J, Rovira A, Rojo F, Lluch A et al (2017) The role of miR-26a and miR-30b in HER2+ breast cancer trastuzumab resistance and regulation of the CCNE2 gene. Sci Rep 7:41309
35. Pegoraro S, Ros G, Ciani Y, Sgarra R, Piazza S, Manfioletti G (2015) A novel HMGA1-CCNE2-YAP axis regulates breast cancer aggressiveness. Oncotarget 6(22):19087–19101
36. Taghavi A, Akbari ME, Hashemi-Bahremani M, Nafissi N, Khalilnezhad A, Poorhosseini SM, Hashemi-Gorji F, Yassaee VR (2016) Gene expression profiling of the 8q22-24 position in human breast cancer: TSPYL5, MTDH, ATAD2 and CCNE2 genes are implicated in oncogenesis, while WISP1 and EXT1 genes may predict a risk of metastasis. Oncol Lett 12(5):3845–3855
37. Klopocki E, Kristiansen G, Wild PJ, Klaman I, Castanos-Velez E, Singer G, Stohr R, Simon R, Sauter G, Leibiger H et al (2004) Loss of SFRP1 is associated with breast cancer progression and poor prognosis in early stage tumors. Int J Oncol 25(3):641–649
38. Bernemann C, Hulsewig C, Ruckert C, Schafer S, Blumel L, Hempel G, Gotte M, Greve B, Barth PJ, Kiesel L et al (2014) Influence of secreted frizzled receptor protein 1 (SFRP1) on neoadjuvant chemotherapy in triple negative breast cancer does not rely on WNT signaling. Mol Cancer 13:174
39. Xiao C, Wu CH, Hu HZ (2016) LncRNA UCA1 promotes epithelial-mesenchymal transition (EMT) of breast cancer cells via enhancing Wnt/beta-catenin signaling pathway. Eur Rev Med Pharmacol Sci 20(13):2819–2824
40. Deng J, Yang M, Jiang R, An N, Wang X, Liu B (2017) Long non-coding RNA HOTAIR regulates the proliferation, Self-renewal capacity, tumor formation and migration of the cancer stem-like cell (CSC) subpopulation enriched from breast cancer cells. PLoS ONE 12(1):e0170860
41. Xue X, Yang YA, Zhang A, Fong KW, Kim J, Song B, Li S, Zhao JC, Yu J (2016) LncRNA HOTAIR enhances ER signaling and confers tamoxifen resistance in breast cancer. Oncogene 35(21):2746–2755
42. Sorensen KP, Thomassen M, Tan Q, Bak M, Cold S, Burton M, Larsen MJ, Kruse TA (2013) Long non-coding RNA HOTAIR is an independent prognostic marker of metastasis in estrogen receptor-positive primary breast cancer. Breast Cancer Res Treat 142(3):529–536
43. Pan Y, Zhang J, Fu H, Shen L (2016) miR-144 functions as a tumor suppressor in breast cancer through inhibiting ZEB1/2-mediated epithelial mesenchymal transition process. Onco Targets Ther 9:6247–6255
44. Yin Y, Cai J, Meng F, Sui C, Jiang Y (2018) MiR-144 suppresses proliferation, invasion, and migration of breast cancer cells through inhibiting CEP55. Cancer Biol Ther 19(4):306–315
45. Ye ZB, Ma G, Zhao YH, Xiao Y, Zhan Y, Jing C, Gao K, Liu ZH, Yu SJ (2015) miR-429 inhibits migration and invasion of breast cancer cells in vitro. Int J Oncol 46(2):531–538
46. Li D, Wang H, Song H, Xu H, Zhao B, Wu C, Hu J, Wu T, Xie D, Zhao J et al (2017) The microRNAs miR-200b-3p and miR-429-5p target the LIMK1/CFL1 pathway to inhibit growth and motility of breast cancer cells. Oncotarget 8(49):85276–85289
47. Wang C, Ju H, Shen C, Tong Z (2015) miR-429 mediates delta-tocotrienol-induced apoptosis in triple-negative breast cancer cells by targeting XIAP. Int J Clin Exp Med 8(9):15648–15656
48. Olgun G, Sahin O, Tastan O (2018) Discovering lncRNA mediated sponge interactions in breast cancer molecular subtypes. BMC Genom 19(1):650
49. Chen J, Xu J, Li Y, Zhang J, Chen H, Lu J, Wang Z, Zhao X, Xu K, Li X et al (2017) Competing endogenous RNA network analysis identifies critical genes among the different breast cancer subtypes. Oncotarget 8(6):10171–10184
50. Xiao B, Zhang W, Chen L, Hang J, Wang L, Zhang R, Liao Y, Chen J, Ma Q, Sun Z et al (2018) Analysis of the miRNA-mRNA-lncRNA network in human estrogen receptor-positive and estrogen receptor-negative breast cancer based on TCGA data. Gene 658:28–35
51. Wu Q, Guo L, Jiang F, Li L, Li Z, Chen F (2015) Analysis of the miRNA-mRNA-lncRNA networks in ER + and ER- breast cancer cell lines. J Cell Mol Med 19(12):2874–2887
52. Paci P, Colombo T, Farina L (2014) Computational analysis identifies a sponge interaction network between long non-coding RNAs and messenger RNAs in human breast cancer. BMC Syst Biol 8:83
53. Han L, Zhang HC, Li L, Li CX, Di X, Qu X (2018) Downregulation of long noncoding RNA HOTAIR and EZH2 Induces apoptosis and inhibits proliferation, invasion, and migration of human breast cancer cells. Cancer Biother Radiopharm 33(6):241–251
54. Zhao W, Geng D, Li S, Chen Z, Sun M (2018) LncRNA HOTAIR influences cell growth, migration, invasion, and apoptosis via the miR-20a-5p/HMGA2 axis in breast cancer. Cancer Med 7(3):842–855
55. Ozes AR, Miller DF, Ozes ON, Fang F, Liu Y, Matei D, Huang T, Nephew KP (2016) NF-kappaB-HOTAIR axis links DNA damage response, chemoresistance and cellular senescence in ovarian cancer. Oncogene 35(41):5350–5361
56. Battistelli C, Cicchini C, Santangelo L, Tramontano A, Grassi L, Gonzalez FJ, de Nonno V, Grassi G, Amicone L, Tripodi M (2017) The snail repressor recruits EZH2 to specific genomic sites through the enrollment of the lncRNA HOTAIR in epithelial-to-mesenchymal transition. Oncogene 36(7):942–955
57. Gupta RA, Shah N, Wang KC, Kim J, Horlings HM, Wong DJ, Tsai MC, Hung T, Argani P, Rinn JL et al (2010) Long non-coding RNA HOTAIR reprograms chromatin state to promote cancer metastasis. Nature 464(7291):1071–1076
58. Kaneuchi M, Sasaki M, Tanaka Y, Shiina H, Yamada H, Yamamoto R, Sakuragi N, Enokida H, Verma M, Dahiya R (2005) WT1 and WT1-AS genes are inactivated by promoter methylation in ovarian clear cell adenocarcinoma. Cancer 104(9):1924–1930
59. Du T, Zhang B, Zhang S, Jiang X, Zheng P, Li J, Yan M, Zhu Z, Liu B (2016) Decreased expression of long non-coding RNA

WT1-AS promotes cell proliferation and invasion in gastric cancer. Biochim Biophys Acta 1862(1):12–19
60. Lv L, Chen G, Zhou J, Li J, Gong J (2015) WT1-AS promotes cell apoptosis in hepatocellular carcinoma through down-regulating of WT1. J Exp Clin Cancer Res 34:119
61. Wang H, Fu Z, Dai C, Cao J, Liu X, Xu J, Lv M, Gu Y, Zhang J, Hua X et al (2016) LncRNAs expression profiling in normal ovary, benign ovarian cyst and malignant epithelial ovarian cancer. Sci Rep 6:38983
62. Zhu N, Hou J, Wu Y, Liu J, Li G, Zhao W, Ma G, Chen B, Song Y (2018) Integrated analysis of a competing endogenous RNA network reveals key lncRNAs as potential prognostic biomarkers for human bladder cancer. Medicine (Baltimore) 97(35):e11887
63. Xing Y, Zhao Z, Zhu Y, Zhao L, Zhu A, Piao D (2018) Comprehensive analysis of differential expression profiles of mRNAs and lncRNAs and identification of a 14-lncRNA prognostic signature for patients with colon adenocarcinoma. Oncol Rep 39(5):2365–2375
64. Li Z, Yao Q, Zhao S, Wang Y, Li Y, Wang Z (2017) Comprehensive analysis of differential co-expression patterns reveal transcriptional dysregulation mechanism and identify novel prognostic lncRNAs in esophageal squamous cell carcinoma. Onco Targets Ther 10:3095–3105
65. Yang Z, Li H, Wang Z, Yang Y, Niu J, Liu Y, Sun Z, Yin C (2018) Microarray expression profile of long non-coding RNAs in human lung adenocarcinoma. Thorac Cancer 9(10):1312–1322
66. Yu F, Quan F, Xu J, Zhang Y, Xie Y, Zhang J, Lan Y, Yuan H, Zhang H, Cheng S et al (2018) Breast cancer prognosis signature: linking risk stratification to disease subtypes. Brief Bioinform