

An 8-lncRNA signature predicts the survival of triple-negative breast cancer patients without germline BRCA1/2 mutation

Minling Liu

The Seventh Affiliated Hospital, Sun Yat-sen University

Wei Dai

Mengyuan Zhu

The Seventh Affiliated Hospital, Sun Yat-sen University

Xueying Li

The Seventh Affiliated Hospital, Sun Yat-sen University

Shan Haung

The Seventh Affiliated Hospital, Sun Yat-sen University

Lei Li

The Seventh Affiliated Hospital, Sun Yat-sen University

SHUO FANG (✉ fangsh9@mail.sysu.edu.cn)

The Seventh Affiliated Hospital, Sun Yat-sen University

Article

Keywords: long non-coding RNA, triple-negative breast cancer, germline BRCA1/2 mutation, overall survival

Posted Date: March 1st, 2021

DOI: <https://doi.org/10.21203/rs.3.rs-269884/v1>

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

[Read Full License](#)

Version of Record: A version of this preprint was published at Medical Research on September 30th, 2021. See the published version at [https://doi.org/10.6913/MRHK.202109_3\(3\).0003](https://doi.org/10.6913/MRHK.202109_3(3).0003).

Abstract

TNBC with germline BRCA1/2 mutation (gBRCAm) have higher sensitivity to DNA damaging agents including platinum-based chemotherapy and PARP inhibitors. But the treatment of TNBC without gBRCAm remains challenging. This study aimed to develop a long non-coding RNA (lncRNA) signature of TNBC patients without gBRCAm to improve risk stratification and optimize individualized treatment.

98 TNBC patients without gBRCAm were acquired from The Cancer Genome Atlas (TCGA) database. The univariable Cox regression analysis and LASSO Cox regression model were applied to establish an lncRNA signature. Then Kaplan–Meier survival curve and time-dependent ROC curve were used to validate the prognostic ability of the signature. The qPCR assay was performed to confirm the expressions and clinicopathological correlations of two potential lncRNAs HAGLROS and TONSL-AS1 in 30 paired samples. We developed an 8-lncRNA signature in the training cohort including HAGLROS, AL139002.1, AL391244.2, AP000696.1, AL391056.1, AL513304.1, TONSL-AS1 and AL031008.1. Patients with higher risk scores showed significantly worse overall survival compared to those with lower risk scores ($P=0.00018$ and $P=0.0068$ respectively). 30 paired specimens of TNBC without gBRCAm in our center showed that two potential lncRNAs HAGLROS and TONSL-AS1 were found frequently overexpressed, and significantly associated with tumor grade and invasion.

Introduction

Breast cancer (BC) is the most common cancer and the leading cause of cancer-related death among women worldwide^{1,2}. Triple-negative breast cancer (TNBC) is a particular aggressive type of BC, defining as the lack of estrogen receptor (ER) and progesterone receptor (PR) expression as well as human epidermal growth factor receptor 2 (HER2) amplification, accounting for approximately 15-20% of all breast cancers. The treatment of TNBC mainly relies on a combination of surgery, radiotherapy, and chemotherapy. Recently, several targeted therapies have been approved for TNBC, including the poly(ADP-ribose) polymerase(PARP) inhibitors olaparib³ and talazoparib⁴ for TNBC with germline BRCA1/2 mutation(gBRCAm) and the checkpoint inhibitor, atezolizumab⁵ in combination with nab-paclitaxel for programmed death-ligand 1 (PD-L1) positive advanced TNBC. But the TNBC is still characterized by poorer prognosis compared with other types of breast cancers⁶⁻⁸.

TNBC is considered a single clinical entity but molecular profiling has revealed an unexpectedly high level of heterogeneity, which result in different treatments sensitivity and outcomes among different subtypes. TNBC with gBRCAm have higher sensitivity to DNA damaging agents including platinum-based chemotherapy and PARP inhibitors^{3,4,9,10}. And they also benefit more from immunotherapy due to their genomic instability or relatively high mutational load. But the treatment of those TNBC patients without gBRCAm remains challenging. Therefore, developing potential therapeutic or prognostic biomarkers are urgently needed for TNBC without gBRCAm.

Long non-coding RNAs (lncRNAs) are a category of RNA transcripts with a length of > 200 nucleotides well known for their limited protein-coding potential. They comprise a heterogeneous class of intergenic transcripts, enhancer RNAs, and sense or antisense transcripts that overlap other genes¹¹. lncRNAs have been proposed to carry out diverse functions, including transcriptional regulation in cis-acting or in trans-acting manner, specific interactions with other cellular factors, namely proteins, DNA, and other RNA molecules¹². lncRNAs have been reported to play an important role in breast cancer pathogenesis, development, and metastasis. For instance, STAR1 is overexpressed in breast cancer and significantly associated with gene modulation activity in the interferon signaling pathway during breast tumorigenesis¹³. LINC00673 promotes the proliferation of breast cancer cells via the miR-515-5p/MARK4/Hippo signaling pathway¹⁴. LINC00261 reduces proliferation and migration of breast cancer cells via the NME1-EMT pathway¹⁵.

lncRNAs are also implicated in the development and progression of TNBC. CCAT2 promotes oncogenesis in triple-negative breast cancer by regulating stemness of cancer cells¹⁶. LINC00993 suppressed TNBC growth and its higher expression indicated better outcome¹⁷. lncRNA FAM83H-AS1 promotes triple-negative breast cancer progression by regulating the miR-136-5p/metadherin axis¹⁸. DCST1-AS1 promotes cell proliferation and metastasis in TNBC by forming a positive regulatory loop with miR-873-5p and MYC¹⁹.

In recent years, accumulating evidence suggests that lncRNAs could be promising prognostic biomarkers in breast cancer including TNBC²⁰. However, the studies about lncRNAs and TNBC without gBRCAm are very few by far. Here, we aimed to develop and validate a new lncRNA signature to predict the prognosis of TNBC patients without gBRCAm, using the clinical data and RNA-Sequencing (RNA-Seq) data acquired from the Cancer Genome Atlas (TCGA) database(<http://cancergenome.nih.gov/>). We hope our findings may provide orientations for the management of TNBC without gBRCAm and then improve their prognosis in the future.

Results

Demographic and clinical characteristics of TNBC patients without gBRCAm

A total of 98 TNBC patients without gBRCAm were enrolled in our study (Supplementary Table S1). The detailed demographic and clinical characteristics of these patients are summarized in Table 1. As shown in Table 1, the vast majority of patients (78.57%) were younger than 65 years old, and most of them (65.31%) were postmenopausal women. Almost all patients were in early or local advanced stage (TNM I&II&III; 98.98%). 60% TNBC patients were randomly selected as the training cohort (N = 59), and all patient (N = 98) as the validation cohort (also called the entire cohort). The demographic and clinical characteristics of the patients were similar between those two cohorts.

Table 1
Demographic and clinical characteristics of TNBC patients without gBRCAm

Clinical factors	Training cohort (N = 59)	Validation cohort (N = 98)
Age		
<65years	47(79.66%)	77(78.57%)
≥ 65 years	12(20.34%)	21(21.43%)
menopause_status		
Pre	14(23.73%)	22(22.45%)
post	38(64.41%)	64(65.31%)
Peri	2(3.39%)	5(5.10%)
NA	5(8.47%)	7(7.14%)
Stage		
I	8(13.56%)	17(17.35%)
II	26(52.54%)	58(59.18%)
III	19(32.20%)	22(22.45%)
NA	1(1.69%)	1(1.02%)
Survival status		
Alive	47(79.66%)	83(84.69%)
Dead	12(20.34%)	15(15.31%)

Generate Prognostic LNCRNA Signature From The Training Cohort

Using the univariate Cox regression analysis method, a set of 173 prognostic lncRNAs was identified in the training cohort ($P < 0.05$). A LASSO Cox regression model was further applied to the top 20 lncRNAs to generate a prognostic signature (Table 2). As a result, we recognized a new 8-lncRNA signature that was highly associated with OS in TNBC patients without gBRCAm (Fig. 1A, 1B). As shown in Table 3, those 8 lncRNAs were all with positive coefficients, which meant they were all correlated with poor survival. Based on the expression of these 8 lncRNAs for OS prediction, we established a risk-score formula: Risk score = $(0.12701757 \times \text{expression level of HAGLROS}) + (0.205120473 \times \text{expression level of AL139002.1}) + (0.217051763 \times \text{expression level of AL391244.2}) + (0.257176862 \times \text{expression level of AP000696.1}) +$

(0.453822184* expression level of AL391056.1)+ (0.57841541* expression level of AL513304.1)+ (0.625591281*expression level of TONSL.AS1)+ (2.872258676* expression level of AL031008.1).

Table 2
Top 20 prognosis-related lncRNAs obtained based on univariate
Cox regression analysis in the training cohort (N = 59)

lncRNA	HR (95% CI)	P value
TONSL-AS1	5.95 (2.07–17.11)	0.000941
AL031008.1	9213.64(39.52-2147857.65)	0.001031
AL359878.1	300.67 (8.68-10418.13)	0.001608
LIVAR	42.89 (3.84-478.84)	0.002264
AL391056.1	44.48(3.72-531.59)	0.002714
AL391244.2	1.60 (1.17–2.17)	0.002922
AP000696.1	7.55 (1.95–29.18)	0.003378
AL513304.1	3.46 (1.50–7.96)	0.003591
THORLNC	3.35 (1.47–7.63)	0.004065
AC010655.2	9.37 (2.02–43.54)	0.004308
AC105450.1	28.58 (2.86-285.97)	0.004329
NDUFB2-AS1	46.23 (3.20-667.49)	0.00489
AC105219.2	1.51 (1.13–2.01)	0.004957
AC134682.1	7.99 (1.86–34.33)	0.005174
AL135791.1	3.13 (1.40–6.99)	0.005371
LNCOC1	18.85 (2.37-149.68)	0.005467
WAC-AS1	1.14 (1.04–1.24)	0.005609
AL596094.1	1.86 (1.20–2.90)	0.005721
AL139002.1	1.34 (1.09–1.65)	0.005886
HAGLROS	1.44 (1.11–1.86)	0.005931
Abbreviations: lncRNA, long non-coding RNA; HR = hazard ratio.		

Table 3
8 lncRNAs significantly associated with the overall survival

gene symbol	coefficient
HAGLROS	0.12701757
AL139002.1	0.205120473
AL391244.2	0.217051763
AP000696.1	0.257176862
AL391056.1	0.453822184
AL513304.1	0.57841541
TONSL.AS1	0.625591281
AL031008.1	2.872258676

Investigate the prognostic ability of the 8-lncRNA signature in the training cohort

We worked out the 8-lncRNA signature risk score for every patient in the training cohort. Using the median risk score as the cut-off point, the patients were categorized into a low risk group (N = 30) and high-risk group (N = 29). The Kaplan-Meier survival curve analysis showed that the overall survival rate of the high-risk group was lower, and the difference between the two groups was statistically significant ($P = 0.00018$, Fig. 2A). The prognostic ability of the 8-lncRNA signature was also evaluated by calculating the AUC of the time-dependent ROC curve. The ROC curve can be used to assess the specificity and sensitivity of the model (AUC > 0.7 indicates that the model has good sensitivity). The higher the AUC, the better is the prediction performance of the signature. For 1, 5, 8 years survival times, the AUC of the 8-lncRNA signature in the training cohort were 1.000, 1.000 and 0.908 respectively (Fig. 3A).

Validation the prognostic ability of the 8-lncRNA signature in the validation cohort

In order to confirm the power of the 8-lncRNA signature in predicting the OS of TNBC patients without gBRCAm, we validated our results in the entire cohort. By utilizing the same classification method, patients were classified into a high-risk group (N = 49) and a low risk group (N = 49). Consistent with previous findings, patients in the high-risk group revealed significantly worse OS compared to the low-risk group ($P = 0.0068$, Fig. 2B). And for 1, 5, 8 years survival times, the AUC of the 8-lncRNA signature in the entire cohort were 0.785, 0.790 and 0.892 respectively (Fig. 3B). It indicated that the prognostic ability 8-lncRNA signature is highly sensitive and specific, and also time-dependent.

Functional Enrichment Analysis Of Lncrna-related MRNAs

Based on the mRNA expression data from the TCGA database, 531 mRNAs were found to be closely related 8-lncRNA signature using the Pearson correlation with $|COR| > 0.3$ and $P < 0.05$ as the cutoff (Supplementary Table S2). The top ten related mRNA were CALML6, BRICD5, ADCK5, C10orf143, GLI4, FBXL6, KIFC2, CDCP2, CCDC154, TSTA3 with $COR \geq 0.46$. The functions of those mRNAs were analysis in the Metascape. The top 20 clusters of significantly enriched terms are shown in Fig. 4. The most significant term was GO:0034660 (ncRNA metabolic process), which included 26 genes: CDK9, BRF1, POLR2J, SNAPC4, RRP1, CPSF4, RRS1, BOP1, PTC1, CPSF1, SIRT7, SRRT, YBEY, EXOSC4, DDX4, DDX56, INTS11, OSGEP, DUS1L, TUT1, MRM1, ELL3, PUS1, CTU1, TRMT61A, TYW3. The top five GO biological process included ncRNA metabolic process, mRNA 3'-end processing, snRNA metabolic process, regulation of DNA repair, oxidoreduction coenzyme metabolic process. The KEGG analysis revealed that these genes mainly involved in mRNA surveillance pathway, Sulfur relay system, Phototransduction, ABC transporters. More details about the enriched terms and genes were shown in Supplementary Table S3. **Overexpression of lncRNAs TONSL-AS1 and HAGLROS in TNBC without gBRCAm and its clinical significance**

Based on our above findings and previous research results, we thought two lncRNAs TONSL-AS1 and HAGLROS may be potential biomarkers for TNBC patients without gBRCAm. qRT-PCR results showed that overexpression of TONSL-AS1 and HAGLROS (defined as a greater than 10-fold increase in tumor tissue) were observed in 10/30 (33.3% for TONSL-AS1) and 16/30 (53.3% for HAGLROS) of the primary breast tumors (Fig. 5; $P < 0.0001$). Clinicopathological association analysis found that the overexpression of TONSL-AS1 was significantly associated with advanced tumor stage, higher invasion rate and recurrence risk ($P < 0.05$, Table 4).

Table 4
Clinicopathological correlation of TONSL-AS1 expression in Breast cancer

Features	Total	^a TONSL-AS1 expression		<i>P</i> value
		Normal expression	Overexpression	
Age				
≤ 60	23	15(65.2%)	8(34.8%)	0.571
≥60	7	5 (71.4%)	2 (28.6%)	
Invasion[†]				
Absent	17	17 (100.0%)	0 (0.0%)	0.000*
Present	13	3 (23.1%)	10 (76.9%)	
Tumor grade[†]				
I/II	22	18 (81.8%)	4 (18.2%)	0.007*
III/IV	8	2 (17.3%)	6 (82.7%)	
Recurrent[†]				
Without metastasis	27	20 (74.1%)	7 (25.9%)	0.030*
With metastasis	3	0 (0.0%)	3 (100.0%)	
[†] Breast cancer specimens demonstrating a 10-fold change higher expression level than matched NT samples are classified as "Overexpression" Group.				
*Statistical significance (<i>P</i> < 0.05) is shown in bold italic.				

And the overexpression of lncRNAs HAGLROS was significantly associated with advanced tumor stage, higher invasion rate (*P* < 0.05, Table 5).

Table 5
Clinicopathological correlation of HAGLROS expression in Breast cancer

Features	Total	^a HAGLROS expression		P value
		Normal expression	Overexpression	
Age				
≤ 60	23	11 (47.8%)	12 (52.2%)	0.581
>60	7	3 (42.9%)	4 (57.1%)	
Invasion[†]				
Absent	17	13 (76.5%)	4 (23.5%)	0.000*
Present	13	1 (7.7%)	12 (92.7%)	
Tumor grade[†]				
I/II	22	13 (59.1%)	9 (40.9%)	0.030*
III/IV	8	1 (12.5%)	7 (87.5%)	
Recurrent[†]				
Without metastasis	27	14 (51.9%)	13 (48.1%)	0.138
With metastasis	3	0 (0.0%)	3 (100.0%)	
[†] Breast cancer specimens demonstrating a 10-fold change higher expression level than matched NT samples are classified as "Overexpression" Group.				
[*] Statistical significance ($P < 0.05$) is shown in bold italic.				

Discussion

As mentioned above, unlike TNBC with gBRCAm, the treatment of TNBC patients without gBRCAm remains challenging. Identifying effective prognostic biomarkers which could help to stratify patients into different subgroups and guide individualized treatment is in great demand. In the other hand, lncRNAs have attracted increasing attention with the development of next-generation sequencing over the last decade. Accumulating research has demonstrated that that lncRNAs play an important role in the development and progression of the breast cancer, and different lncRNA signatures can predict the prognosis of breast cancer²¹⁻²⁷. However, the studies about lncRNAs and TNBC without gBRCAm are limited. In the present study, we identified 98 TNBC patients without gBRCAm from the TCGA database. In the training cohort, we finally obtained an 8-lncRNA signature using the univariable Cox regression

analysis and LASSO Cox regression model. Patients with higher 8-lncRNA signature risk scores showed worse OS compared to those with lower risk scores. High expression levels of HAGLROS, AL139002.1, AL391244.2, AP000696.1, AL391056.1, AL513304.1, *TONSL-AS1* and AL031008.1 was correlated with poor prognosis. The prognostic ability of the 8-lncRNA signature was validated by the time-dependent ROC curve both in the training cohort and the entire cohort.

Among these 8 lncRNAs, except for HAGLROS and *TONSL-AS1*, other 6 lncRNAs have not been investigated thoroughly and were discovered as prognostic signatures for the first time. HAGLROS is a lncRNA with a length of 699 bp, was reported to involve in the progression of various cancers^{28–35}, including lung cancer, colorectal cancer, lymphoma, ovarian cancer, osteosarcoma, nasopharyngeal carcinoma, intrahepatic cholangiocarcinoma. HAGLROS regulates apoptosis and autophagy via PI3K/Akt/mTOR signaling pathway^{28,30,36}, miR-5095/ATG12 axis³⁵ or miR-100/ATG5 axis^{30,37}. The apoptosis and autophagy regulated by HAGLROS contribute to cancer cells proliferation and chemoresistance³⁴. Patients with high expression levels of HAGLROS showed poor OS, which is consistent with our result. HAGLROS may serve as a potential therapeutic target for future treatment of TNBC patients without gBRCAm as mTOR inhibitors were available in clinical. Liu Y et al demonstrated that *TONSL-AS1* regulated miR-490-3p/CDK1 to affect ovarian epithelial carcinoma cell proliferation, and *TONSL-AS1* was upregulated in ovarian epithelial carcinoma, its high expression level was correlated with poor survival³⁸. Wang P et al demonstrated that *TONSL-AS1* regulated progression of gastric cancer via activating *TONSL*³⁹. Its high expression level was correlated with poor survival, which is also consistent with our result.

lncRNAs are non-coding transcripts, but they play important functional roles in modulating mRNA translation⁴⁰. In the present study, 8-lncRNA signature related mRNAs was identified using the Pearson correlation. The result showed that the top 10 related mRNA were CALML6, BRICD5, ADCK5, C10orf143, GLI4, FBXL6, KIFC2, CDCP2, CCDC154, TSTA3. CALML6 encoded RAS pathway related proteins, involving in calcium-mediated signaling and second-messenger-mediated signaling. BRICD5 mainly involved in cell proliferation pathway. ADCK5 is a member of an atypical kinase family and overexpressed in many carcinomas. Qiu M et al demonstrated that ADCK5 might regulate the expression of tumor oncogene PTTG1 by phosphorylating transcription factor SOX9, therefore enhancing the migration and invasion capabilities of lung cancer cells⁴¹. So the ADCK5-SOX9-PTTG1 pathway might also be a potential therapeutic target for TNBC without gBRCAm. GLI4, a zinc finger protein of unknown function, was localized to chromosome 8q24.3, distal to c-MYC⁴². Shi W et al have found that FBXL6 governs c-MYC to promote hepatocellular carcinoma through ubiquitination and stabilization of HSP90AA1⁴³. FBXL6-HSP90AA1-c-MYC axis might contribute to the oncogenesis of HCC. They proposed that inhibition of FBXL6 might be an effective therapeutic strategy for HCC treatment⁴³. *CCDC154* was reported to inhibit tumor cell growth⁴⁴. Sun Y et al showed that TSTA3 controls cell proliferation and invasion by regulating CXCR4 expression⁴⁵. TSTA3 was highly expressed in breast cancer cell and served as an independent prognostic factor for BC patients. C10orf143, KIFC2, CDCP2 were protein coding genes, however, the function of those genes in cancer have not been elucidated. Functional enrichment analysis revealed that

signature related mRNAs involved in RNA metabolic process and regulation of DNA repair terms, which were strongly associated with the proliferation and metastasis of tumor cells^{46,47}.

In the present study, we also found that lncRNAs HAGLROS and TONSL-AS1 were frequently overexpressed in primary triple-negative breast cancer tissues without gBRCAm, which was significantly associated with poor prognosis. These findings have been supported by the database analysis above, suggesting that lncRNAs HAGLROS and TONSL-AS1 plays an important oncogenic role in the development and progression of triple-negative breast cancer without gBRCAm.

There were some limitations in this study. First, owing to the limited number of patients recruited in our study, patients were not divided randomly into training cohort and validation cohort. Second, the 8-lncRNA signature has not been validated in an independent cohort. Third, the underlying functions of these 8 lncRNAs have not been fully illuminated in this study.

Methods

Data collection and preprocessing

In the present study, we downloaded the clinical data and RNA-Seq data from the TCGA database. Of the 1098 breast cancer patients in TCGA, 825 patients without gBRCAm were identified by Kraya et al. using DNA-sequencing data⁴⁸. Of these 825 patients, 99 were diagnosed with TNBC. Then we excluded patients with an unknown survival time (N=1), leaving 98 TNBC patients without gBRCAm in our study. 60% patients (N=59) were randomly selected as the training cohort, and all patient (N=98) as the validation cohort.

Identification and validation of the prognostic lncRNA Signature

First, the univariate Cox regression model was applied to the training cohort to detect the prognostic lncRNAs. A set of lncRNAs which *P*-value was <0.05 were identified. The top 20 lncRNAs were then analyzed in a training cohort utilizing the LASSO Cox regression model analysis. A list of prognostic lncRNAs with related coefficients was obtained from the lncRNA expression profile and the patients' overall survival (OS) according to the best lambda value. Furthermore, the risk score of every patient was calculated according to the expression level of each prognostic lncRNA and its corresponding coefficient. The training cohort was assigned to a high-risk or low-risk group using the median risk score as the cut-off. The method of Kaplan–Meier and the log-rank test were utilized to evaluate the OS difference between the high or low-risk groups. Meanwhile, the prognostic lncRNA signature was validated in the validation cohort. Based on the median risk score, the validation cohort was also split into a high-risk or low-risk group and the OS difference between the two groups were also evaluated by the Kaplan–Meier survival curves. Time-dependent receiver operating characteristic (ROC) analysis was further used to assess the prognostic value.

Identification of lncRNA-related mRNAs

lncRNA signature related mRNAs were identified using the Pearson correlation with $|COR| > 0.3$ and $P < 0.05$ as the cutoff. Then the functional enrichment analysis was performed to predict the potential functions of these mRNAs.

Functional Enrichment Analysis

Metascape (<http://metascape.org/>) was used to perform functional enrichment analysis, including the Gene Ontology(GO) biological process and Kyoto Encyclopedia of Genes and Genomes (KEGG) Pathway analyses. All genes in the genome have been used as the enrichment background. Terms with a p-value < 0.01 , a minimum count of 3, and an enrichment factor > 1.5 (the enrichment factor is the ratio between the observed counts and the counts expected by chance) are collected and grouped into clusters based on their membership similarities. Kappa scores are used as the similarity metric when performing hierarchical clustering on the enriched terms, and sub-trees with a similarity of > 0.3 are considered a cluster. The most statistically significant term within a cluster is chosen to represent the cluster.

Further analysis of candidate lncRNAs in an independent cohort

Additionally, 30 TNBC samples and their paired adjacent normal tissues were collected consecutively at Affiliated Hospital of Guangdong Medical University. All enrolled patients were diagnosed with TNBC without gBRCAm. This study was reviewed and approved by the committee for ethical review of research involving human subjects at Guangdong Medical University. The expression levels of candidate lncRNAs were detected using quantitative real-time PCR assay. lncRNAs expression between tumor and normal specimens was compared with the aid of paired Student *t*-tests. The relationship between the expression of lncRNAs and clinicopathological factors were identified using Pearson's χ^2 test.

Quantitative real-time PCR (qRT-PCR)

For the quantitative real-time PCR, a SYBR Green PCR Kit (Applied Biosystems, Carlsbad, CA) was used to amplify the cDNA of paired breast cancer clinical specimens and performed under an ABI Prism 7900 System (Applied Biosystems Carlsbad, CA). 18srRNA was involved as the endogenous reference. In final analysis, all the data were analyzed using the ABI SDS v2.3 software (Applied Biosystems, Carlsbad, CA). The ABI SDS v2.3 software (Applied Biosystems, Carlsbad, CA) was used to analyze the all the data in final statistics. The relative expressions of lncRNAs were analyzed with formula $2^{-\Delta CT}$ ($\Delta CT = CT^{\text{target}} - CT^{18S}$) in paired clinical samples. Correspondingly, the average of non-tumor tissues value was defined as a criteria equal to 1.0, considered as normalization. For the estimation of lncRNA expression, the RiboTM mRNA/lncRNA qRT-PCR Starter Kit (Ribobio, China) was used for quantitative real-time PCR.

Statistical analysis

When a two-sided *P*-value was < 0.05 , the statistical analyses were defined as statistically significant. All statistical analyses were performed using the SPSS 16.0 or R software (R version 3.6.2).

References

- 1 Bray, F. *et al.* Global cancer statistics 2018: GLOBOCAN estimates of incidence and mortality worldwide for 36 cancers in 185 countries. *CA: a cancer journal for clinicians* **68**, 394-424, doi:10.3322/caac.21492 (2018).
- 2 Siegel, R. L., Miller, K. D. & Jemal, A. Cancer statistics, 2019. *CA: a cancer journal for clinicians* **69**, 7-34, doi:10.3322/caac.21551 (2019).
- 3 Robson, M. *et al.* Olaparib for Metastatic Breast Cancer in Patients with a Germline BRCA Mutation. *The New England journal of medicine* **377**, 523-533, doi:10.1056/NEJMoa1706450 (2017).
- 4 Litton, J. K. *et al.* Talazoparib in Patients with Advanced Breast Cancer and a Germline BRCA Mutation. *The New England journal of medicine* **379**, 753-763, doi:10.1056/NEJMoa1802905 (2018).
- 5 Schmid, P. *et al.* Atezolizumab and Nab-Paclitaxel in Advanced Triple-Negative Breast Cancer. *The New England journal of medicine* **379**, 2108-2121, doi:10.1056/NEJMoa1809615 (2018).
- 6 Foulkes, W. D., Smith, I. E. & Reis-Filho, J. S. Triple-negative breast cancer. *The New England journal of medicine* **363**, 1938-1948, doi:10.1056/NEJMra1001389 (2010).
- 7 Bosch, A., Eroles, P., Zaragoza, R., Viña, J. R. & Lluch, A. Triple-negative breast cancer: molecular features, pathogenesis, treatment and current lines of research. *Cancer treatment reviews* **36**, 206-215, doi:10.1016/j.ctrv.2009.12.002 (2010).
- 8 Dent, R. *et al.* Triple-negative breast cancer: clinical features and patterns of recurrence. *Clinical cancer research : an official journal of the American Association for Cancer Research* **13**, 4429-4434, doi:10.1158/1078-0432.ccr-06-3045 (2007).
- 9 Isakoff, S. J. *et al.* TBCRC009: A Multicenter Phase II Clinical Trial of Platinum Monotherapy With Biomarker Assessment in Metastatic Triple-Negative Breast Cancer. *Journal of clinical oncology : official journal of the American Society of Clinical Oncology* **33**, 1902-1909, doi:10.1200/jco.2014.57.6660 (2015).
- 10 Silver, D. P. *et al.* Efficacy of neoadjuvant Cisplatin in triple-negative breast cancer. *Journal of clinical oncology : official journal of the American Society of Clinical Oncology* **28**, 1145-1153, doi:10.1200/jco.2009.22.4725 (2010).
- 11 Derrien, T. *et al.* The GENCODE v7 catalog of human long noncoding RNAs: analysis of their gene structure, evolution, and expression. *Genome research* **22**, 1775-1789, doi:10.1101/gr.132159.111 (2012).
- 12 Ulitsky, I. & Bartel, D. P. lincRNAs: genomics, evolution, and mechanisms. *Cell* **154**, 26-46, doi:10.1016/j.cell.2013.06.020 (2013).

- 13 Jeong, G. *et al.* A Kelch domain-containing KLHDC7B and a long non-coding RNA ST8SIA6-AS1 act oppositely on breast cancer cell proliferation via the interferon signaling pathway. *Scientific reports* **8**, 12922, doi:10.1038/s41598-018-31306-8 (2018).
- 14 Qiao, K. *et al.* LINC00673 is activated by YY1 and promotes the proliferation of breast cancer cells via the miR-515-5p/MARK4/Hippo signaling pathway. *Journal of experimental & clinical cancer research : CR* **38**, 418, doi:10.1186/s13046-019-1421-7 (2019).
- 15 Guo, G., Dai, S. & Chen, Q. Long Noncoding RNA LINC00261 Reduces Proliferation and Migration of Breast Cancer Cells via the NME1-EMT Pathway. *Cancer management and research* **12**, 3081-3089, doi:10.2147/cmar.s237197 (2020).
- 16 Xu, Z. *et al.* Long non-coding RNA CCAT2 promotes oncogenesis in triple-negative breast cancer by regulating stemness of cancer cells. *Pharmacological research* **152**, 104628, doi:10.1016/j.phrs.2020.104628 (2020).
- 17 Guo, S. *et al.* Novel Breast-Specific Long Non-coding RNA LINC00993 Acts as a Tumor Suppressor in Triple-Negative Breast Cancer. *Frontiers in oncology* **9**, 1325, doi:10.3389/fonc.2019.01325 (2019).
- 18 Han, C., Fu, Y., Zeng, N., Yin, J. & Li, Q. LncRNA FAM83H-AS1 promotes triple-negative breast cancer progression by regulating the miR-136-5p/metadherin axis. *Aging* **12**, 3594-3616, doi:10.18632/aging.102832 (2020).
- 19 Tang, L. *et al.* Long Noncoding RNA DCST1-AS1 Promotes Cell Proliferation and Metastasis in Triple-negative Breast Cancer by Forming a Positive Regulatory Loop with miR-873-5p and MYC. *Journal of Cancer* **11**, 311-323, doi:10.7150/jca.33982 (2020).
- 20 Tian, T. *et al.* Identification of long non-coding RNA signatures in triple-negative breast cancer. *Cancer cell international* **18**, 103, doi:10.1186/s12935-018-0598-8 (2018).
- 21 Sun, M. *et al.* 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. *Breast cancer research and treatment* **175**, 59-75, doi:10.1007/s10549-019-05147-6 (2019).
- 22 He, Y. *et al.* A prognostic 11 long noncoding RNA expression signature for breast invasive carcinoma. *Journal of cellular biochemistry* **120**, 16692-16702, doi:10.1002/jcb.28927 (2019).
- 23 Y, N. *et al.* Profiling the epigenetic interplay of lncRNA RUNXOR and oncogenic RUNX1 in breast cancer cells by gene in situ cis-activation. **9**, 1635-1649 (2019).
- 24 Lv, M. *et al.* LncRNAs as new biomarkers to differentiate triple negative breast cancer from non-triple negative breast cancer. *Oncotarget* **7**, 13047-13059, doi:10.18632/oncotarget.7509 (2016).

- 25 Fan, C. N., Ma, L. & Liu, N. Comprehensive analysis of novel three-long noncoding RNA signatures as a diagnostic and prognostic biomarkers of human triple-negative breast cancer. *Journal of cellular biochemistry* **120**, 3185-3196, doi:10.1002/jcb.27584 (2019).
- 26 Jiang, Y. Z. *et al.* Transcriptome Analysis of Triple-Negative Breast Cancer Reveals an Integrated mRNA-lncRNA Signature with Predictive and Prognostic Value. *Cancer research* **76**, 2105-2114, doi:10.1158/0008-5472.can-15-3284 (2016).
- 27 Meng, J., Li, P., Zhang, Q., Yang, Z. & Fu, S. A four-long non-coding RNA signature in predicting breast cancer survival. *Journal of experimental & clinical cancer research : CR* **33**, 84, doi:10.1186/s13046-014-0084-7 (2014).
- 28 Ma, J., Feng, J. & Zhou, X. Long non-coding RNA HAGLROS regulates lipid metabolism reprogramming in intrahepatic cholangiocarcinoma via the mTOR signaling pathway. *Experimental and molecular pathology* **115**, 104466, doi:10.1016/j.yexmp.2020.104466 (2020).
- 29 Wang, W. L., Yu, D. J. & Zhong, M. LncRNA HAGLROS accelerates the progression of lung carcinoma via sponging microRNA-152. *European review for medical and pharmacological sciences* **23**, 6531-6538, doi:10.26355/eurev_201908_18538 (2019).
- 30 Mu, G., Liu, Q., Wu, S., Xia, Y. & Fang, Q. Long noncoding RNA HAGLROS promotes the process of mantle cell lymphoma by regulating miR-100/ATG5 axis and involving in PI3K/AKT/mTOR signal. *Artificial cells, nanomedicine, and biotechnology* **47**, 3649-3656, doi:10.1080/21691401.2019.1645151 (2019).
- 31 Wu, P. F., Dai, Z. T., Liu, W. D., Zhao, Z. X. & Kong, Y. H. Elevated long noncoding RNA HAGLROS expression correlates with clinical progression and prognosis in osteosarcoma. *European review for medical and pharmacological sciences* **23**, 1428-1433, doi:10.26355/eurev_201902_17099 (2019).
- 32 Chen, Y. *et al.* HAGLROS is overexpressed and promotes non-small cell lung cancer migration and invasion. *Japanese journal of clinical oncology*, doi:10.1093/jjco/hyaa075 (2020).
- 33 Yang, M., Zhai, Z., Zhang, Y. & Wang, Y. Clinical significance and oncogene function of long noncoding RNA HAGLROS overexpression in ovarian cancer. *Archives of gynecology and obstetrics* **300**, 703-710, doi:10.1007/s00404-019-05218-5 (2019).
- 34 Bermúdez, M. *et al.* LncRNAs as Regulators of Autophagy and Drug Resistance in Colorectal Cancer. *Frontiers in oncology* **9**, 1008, doi:10.3389/fonc.2019.01008 (2019).
- 35 Wei, H. *et al.* Long noncoding RNA HAGLROS promotes cell proliferation, inhibits apoptosis and enhances autophagy via regulating miR-5095/ATG12 axis in hepatocellular carcinoma cells. *International immunopharmacology* **73**, 72-80, doi:10.1016/j.intimp.2019.04.049 (2019).

- 36 Zhang, W., Zhang, Y. & Xi, S. Upregulation of lncRNA HAGLROS enhances the development of nasopharyngeal carcinoma via modulating miR-100/ATG14 axis-mediated PI3K/AKT/mTOR signals. *Artificial cells, nanomedicine, and biotechnology* **47**, 3043-3052, doi:10.1080/21691401.2019.1640233 (2019).
- 37 Zheng, Y., Tan, K. & Huang, H. Long noncoding RNA HAGLROS regulates apoptosis and autophagy in colorectal cancer cells via sponging miR-100 to target ATG5 expression. *Journal of cellular biochemistry* **120**, 3922-3933, doi:10.1002/jcb.27676 (2019).
- 38 Liu, Y., Li, L., Wang, X., Wang, P. & Wang, Z. LncRNA TONSL-AS1 regulates miR-490-3p/CDK1 to affect ovarian epithelial carcinoma cell proliferation. *Journal of ovarian research* **13**, 60, doi:10.1186/s13048-020-00657-0 (2020).
- 39 Wang, P. *et al.* A novel long non-coding RNA TONSL-AS1 regulates progression of gastric cancer via activating TONSL. *Experimental cell research* **382**, 111453, doi:10.1016/j.yexcr.2019.05.034 (2019).
- 40 Zhang, K. *et al.* Long non-coding RNAs as novel biomarkers for breast cancer invasion and metastasis. *Oncology letters* **14**, 1895-1904, doi:10.3892/ol.2017.6462 (2017).
- 41 Qiu, M. *et al.* aarF domain containing kinase 5 gene promotes invasion and migration of lung cancer cells through ADCK5-SOX9-PTTG1 pathway. *Experimental cell research* **392**, 112002, doi:10.1016/j.yexcr.2020.112002 (2020).
- 42 Kas, K., Wlodarska, I., Meyen, E., Van den Berghe, H. & Van de Ven, W. J. Assignment of the gene encoding human Krüppel-related zinc finger protein 4 (GLI4) to 8q24.3 by fluorescent in situ hybridization. *Cytogenetics and cell genetics* **72**, 297-298, doi:10.1159/000134207 (1996).
- 43 Shi, W. *et al.* FBXL6 governs c-MYC to promote hepatocellular carcinoma through ubiquitination and stabilization of HSP90AA1. *Cell communication and signaling : CCS* **18**, 100, doi:10.1186/s12964-020-00604-y (2020).
- 44 Liao, W. *et al.* Overexpression of a novel osteopetrosis-related gene CCDC154 suppresses cell proliferation by inducing G2/M arrest. *Cell cycle (Georgetown, Tex.)* **11**, 3270-3279, doi:10.4161/cc.21642 (2012).
- 45 Sun, Y. *et al.* Oncogenic potential of TSTA3 in breast cancer and its regulation by the tumor suppressors miR-125a-5p and miR-125b. *Tumour biology : the journal of the International Society for Oncodevelopmental Biology and Medicine* **37**, 4963-4972, doi:10.1007/s13277-015-4178-4 (2016).
- 46 Anastasiadou, E., Jacob, L. S. & Slack, F. J. Non-coding RNA networks in cancer. *Nature reviews. Cancer* **18**, 5-18, doi:10.1038/nrc.2017.99 (2018).
- 47 Jeggo, P. A., Pearl, L. H. & Carr, A. M. DNA repair, genome stability and cancer: a historical perspective. *Nature reviews. Cancer* **16**, 35-42, doi:10.1038/nrc.2015.4 (2016).

Figures

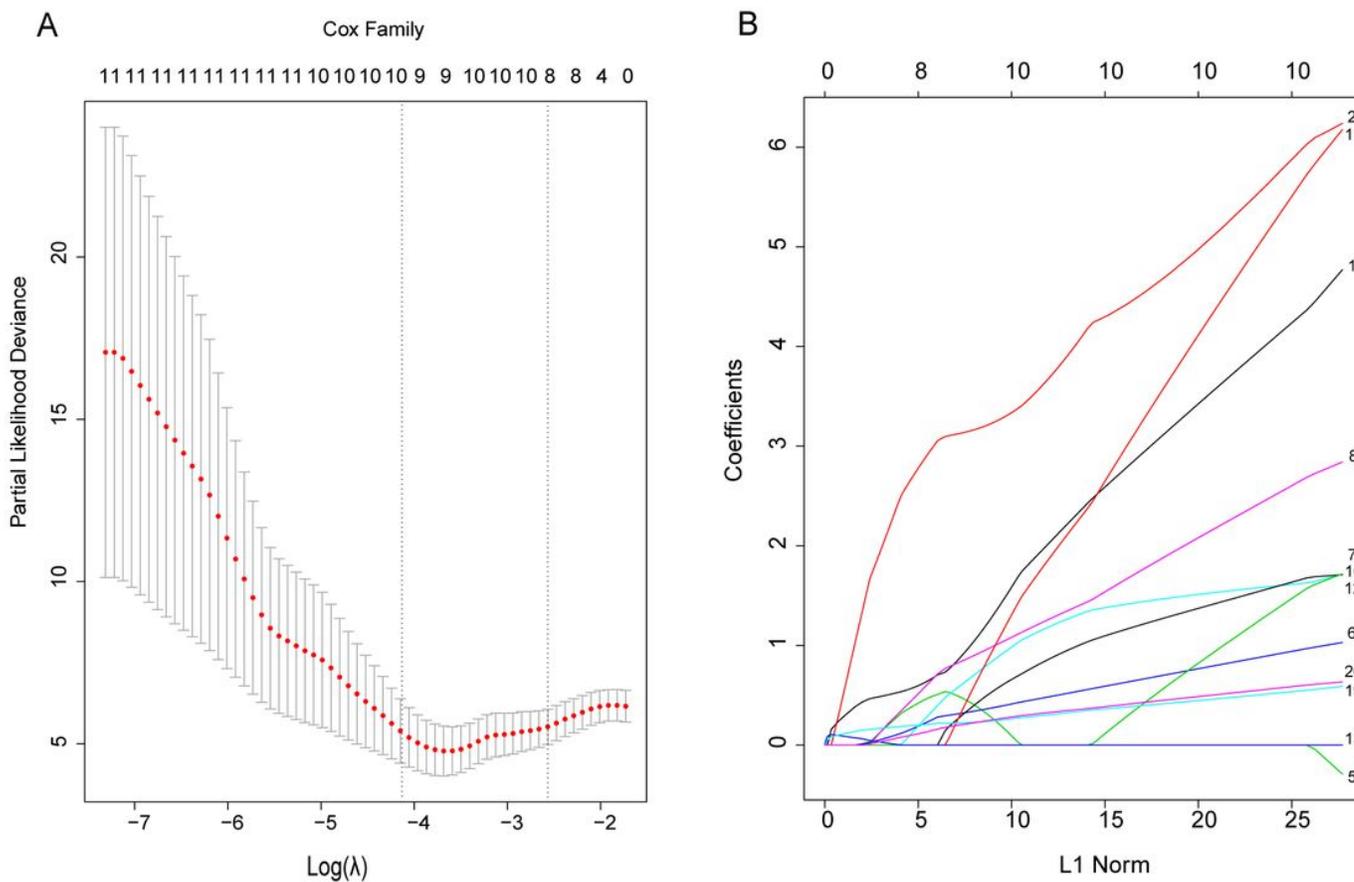


Figure 1

Construction of lncRNA signature by LASSO regression. (A) The tuning parameter selection plot of LASSO regression. The dotted lines represented the minimum and 1 standard error lambda for the optimal volume of variables. (B) The coefficient profiles of 20 prognostic lncRNAs.

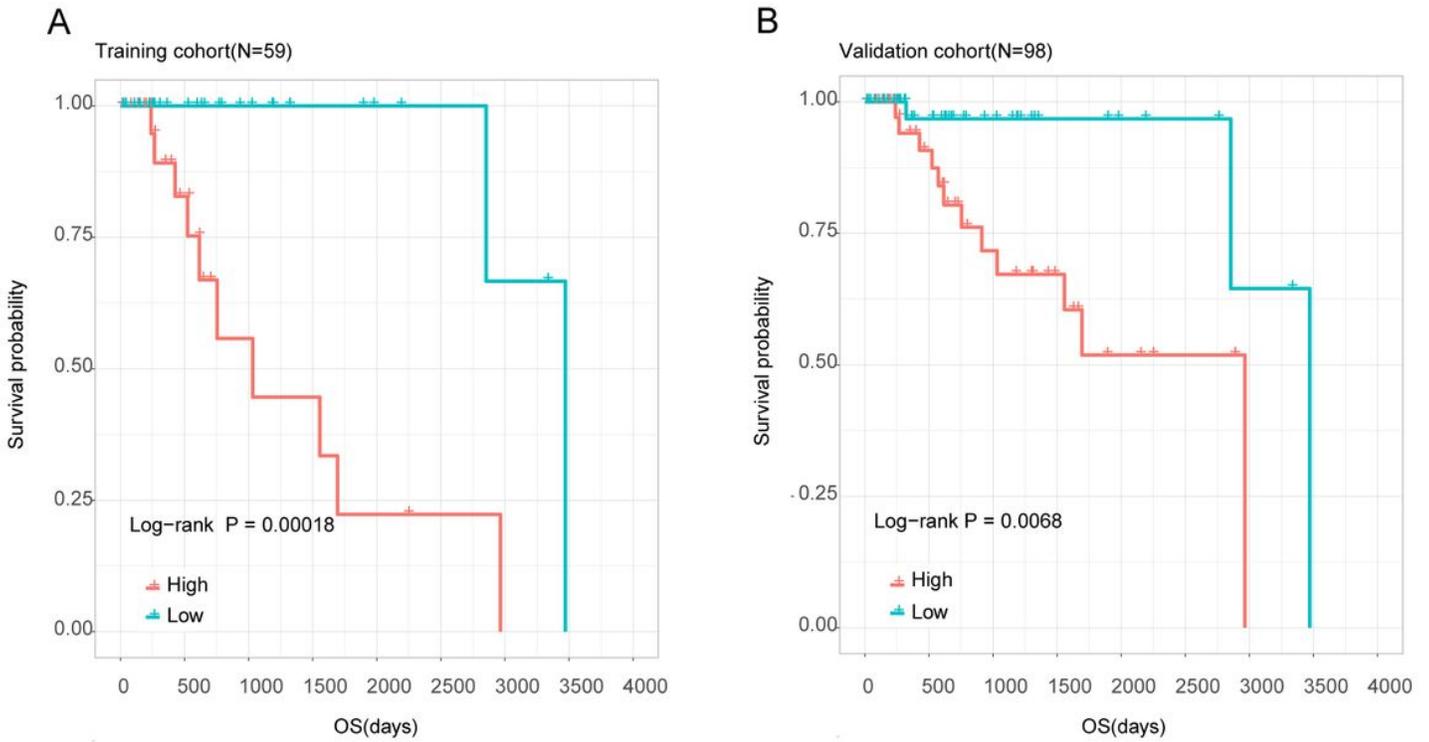


Figure 2

The Kaplan-Meier curve analysis showed the significant poorer OS of the high-risk group than the low-risk group both in the training cohort (A) and the entire cohort (B).

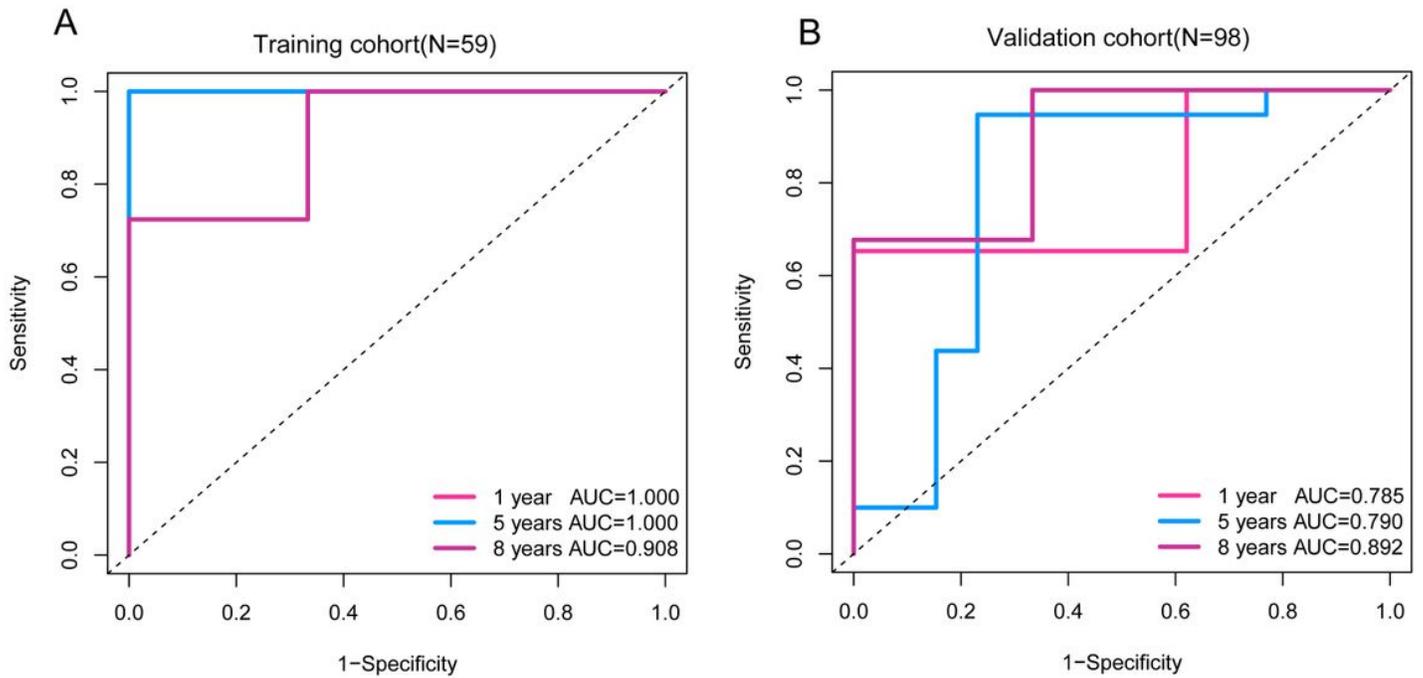


Figure 3

The result of the time-dependent ROC analysis in the training cohort(A)and the entire cohort(B).

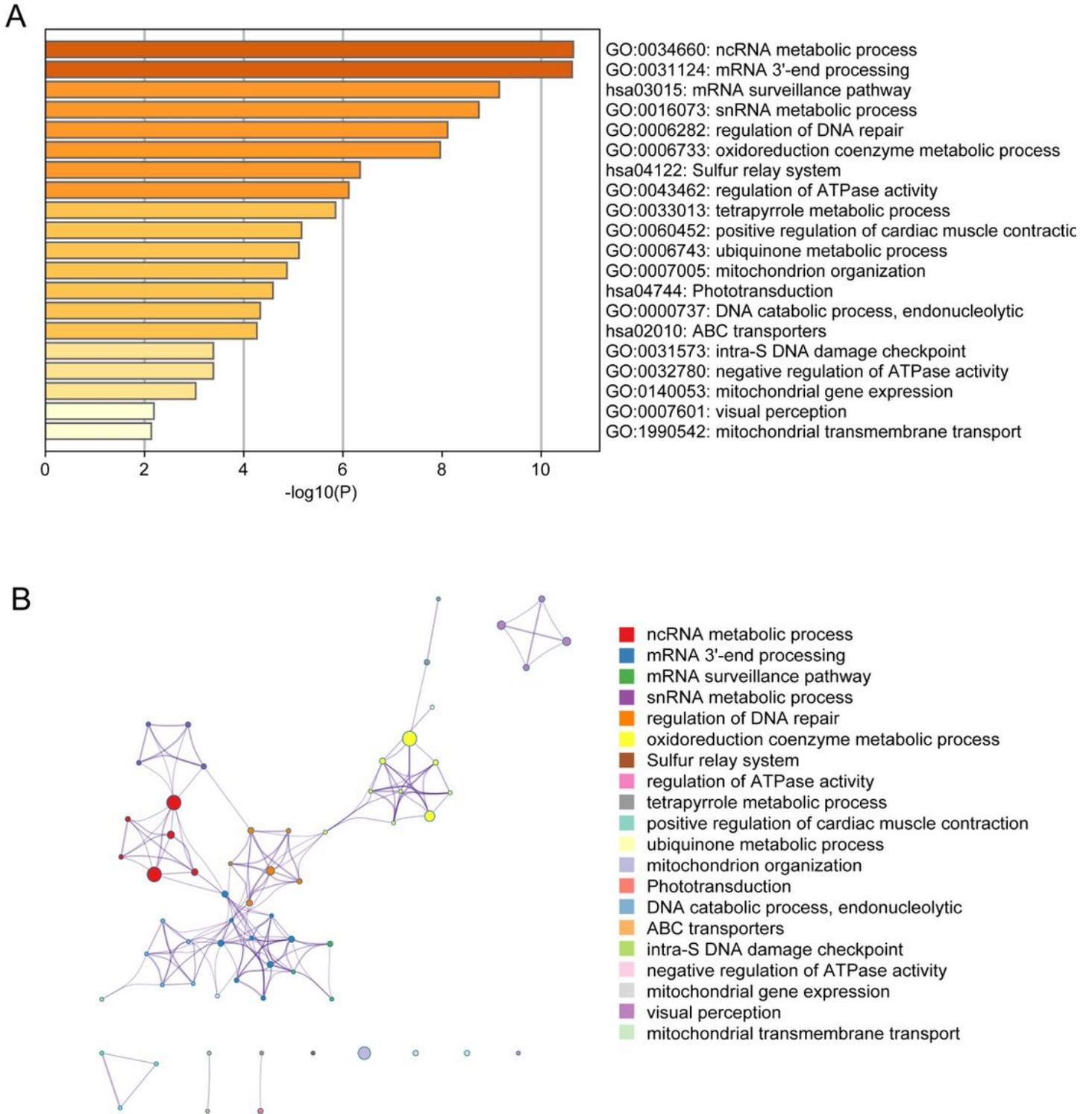


Figure 4

Functional enrichment results: the heatmap of top 20 clusters, colored by P-values. The smaller the P-value is the deeper the color is (A). The networks of the top 20 clusters. Each node represents an enriched term and is colored by its cluster ID, terms with a similarity > 0.3 are connected by edges. The thicker edge displayed, the higher the similarity is (B).

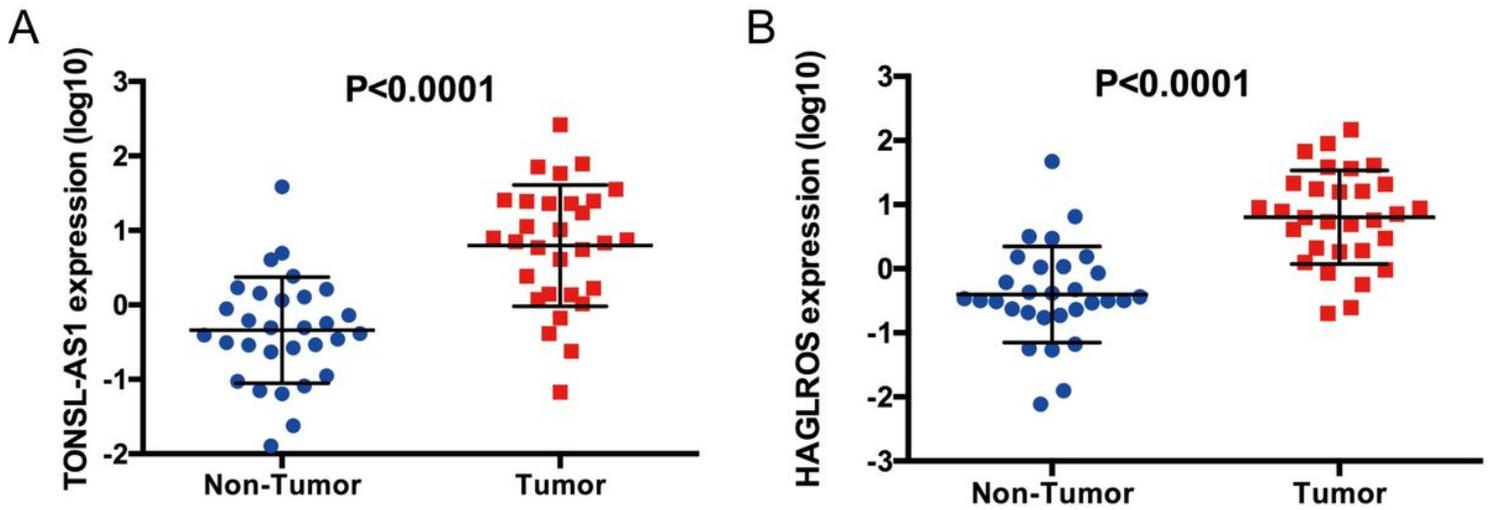


Figure 5

Overexpression of lncRNAs HAGLROS (A) and TONSL-AS1 (B) were frequently detected in primary breast cancers by qPCR. $P < 0.0001$.