Int J Clin Exp Med 2018;11(8):8139-8145
www.ijcem.com /ISSN:1940-5901/IJCEM0073914

Original Article

Circulating lncRNAs acting as diagnosis fingerprints for predicting triple negative breast cancer

Qiwei Du[2*], Yue Yang[3*], Xuandong Kong[2], Fang Lan[2], Jiancheng Sun[2], Haote Zhu[2], Yixiu Ni[2], Anping Pan[1]

[1]Department of Oncology, Hangzhou Cancer Hospital, Hangzhou, Zhejiang Province, China; [2]Department of Breast Surgery, Zhejiang University of Traditional Chinese Medicine, Hangzhou, Zhejiang Province, China; [3]Department of Plastic Surgery, First Affiliated Hospital of Zhejiang Chinese Medicine University, Hangzhou, Zhejiang Province, China. [*]Equal contributors.

Received October 8, 2017; Accepted May 24, 2018; Epub August 15, 2018; Published August 30, 2018

Abstract: The long non-coding RNAs (lncRNAs) has been identified as non-invasive markers in both diagnosis and predicting the prognosis of human malignant tumors. Here we mainly focused on the circulating expressed lncRNAs in triple negative breast cancer patients (TNBC) to screen potential lncRNAs as diagnosis fingerprint. All the studies focusing on the lncRNA in TNBC, we found 14 lncRNAs whom has been identified with significant increased level in TNBC tumor tissues. According to previously report, we enrolled 14 lncRNAs whom has been identified with significant increased level in TNBC tumor tissues, as candidate based on their aberrant expression in TNBC patients. The screening phase revealed that ANRIL, SOX2OT and ANRASSF1 presenting with increased level ($p<0.05$) in 20 paired TNBC patients, non-TNBC patients and controls. Further validation set also confirmed the increasing level of these three lncRNAs by Quantitative real-time PCR (qRT-PCR). The risk score analysis presenting by receiver operating characteristic curve (ROC) indicating ANRIL, SOX2OT and ANRASSF1 could distinguish TNBC patients from the non-TNBC patients. Among this, ANRIL provided the highest diagnostic performance with an area under ROC curve value (AUC) of 0.918. Further combination with the three factors indicated a higher power (AUC, 0.959; sensitivity, 92.8%; specificity, 90.5%). In conclusion, circulating lncRNAs ANRIL, SOX2OT and ANRASSF1 might act as fingerprints for predicting triple negative breast cancer.

Keywords: TNBC, long non-coding RNAs, AUC, risk score, fingerprints

Introduction

Breast cancer is the leading cause of cancer mortality among women worldwide. The incidence of breast cancer has been increasing by 3% per year in China, which has threatened the health of women and created a great burden on society [1, 2]. One of the most aggressive breast cancer subtypes is triple-negative breast cancer (TNBC), which lacks estrogen receptor (ER) and progesterone receptor (PR) expression and human epidermal growth factor receptor 2 (HER2) amplification [3, 4]. A TNBC represents approximately 10-20% of all breast cancers and has a larger tumor size, higher grade, more positive lymph nodes, and poorer prognosis than other subtypes of breast cancer [4]. Because TNBC patients lack effective therapeutic targets, such as tamoxifen in ER-positive BC and Herceptin in HER2-positive BC, they often have significantly lower rates of disease-free survival (DFS), overall survival (OS) and 5-year survival [5, 6]. The early diagnosis and the timely treatment might increase the prognosis of TNBC patients.

LncRNAs, which have been a focus of study recently, are segments of RNA that are more than 200 nucleotides in length with little translation capacity [7]. Recent studies have indicated that altered expression levels of lncRNAs are important regulatory molecules at every level of cellular physiology, and the alterations of which are related to multiple diseases, including cancers [8]. LncRNAs were reported as a biomarker for diagnosis multiple diseases [9, 10]; Researchers have recently identified that circulation long non-coding RNA (lncRNA) might act as biomarker for the early diagnosis of multiple human cancers such as hepatocel-

Table 1. Clinicopathological features analysis of Triple Negative Breast Cancer (TNBC) patients, Non-TNBC patients and healthy control

	TNBC	Non-TNBC	Control	*P* value
N	120	120	120	
Age Mean (SE) year[1]	41.31 (0.36)	41.77 (0.51)	42.01 (0.42)	0.22[a]
Metastasis[2]				
No	33	77	77	<0.001[b]
Yes	87	43	43	
Tumor Size (cm)[2]				0.19[b]
≤2 cm	76	66		
>2 cm	44	54		
Tumor grade[2]				0.43[b]
≤II	51	45		
>II	69	75		
Ki67, %[2]				<0.001[b]
≤14	42	74		
>14	78	46		

[a]Student t-test, [b]Chi-square test, [1]continuous variable, [2]categorical variable.

lular carcinoma [11, 12], gastric cancer [13, 14] and prostate cancer [15, 16]. However, the systematic investigation regarding circulating lncRNA in TNBC was little conducted. Here we enrolled 14 identified lncRNAs including TCONS_l2_00003938, HOTAIR, MALAT1, ANRIL, UCA1, ENST00000460164, ENST00000425295, HIF1A-AS2, C17orf76-AS1, CTC-338M12.3, XR_250621.1, SOX2OT, PTPRG-AS1 and ANRASSF1 which was different expression in the tumor tissues of patients with TNBC. We hypothesized that these TNBC-related lncRNAs might be released into the circulation during TNBC initiation and could be utilized to detect and monitor TNBC. We aimed to explore the potential lncRNA acting as fingerprint for the early identification of TNBC.

Materials and methods

Samples and screening phase

All plasma samples were obtained from patients diagnosed with TNBC between August 2014 and September 2016 at Zhejiang University of Traditional Chinese Medicine. Controls were from healthy volunteers without any health problems during their health check-ups at the Zhejiang University of Traditional Chinese Medicine. The research protocol was approved by the Institutional Ethics Committee of the Zhejiang University. Written informed consent was obtained from every participant. Peripheral blood samples of patients were collected before surgery. Blood samples were collected in a separate vacuum cube, followed by centrifugation at 3,000 rpm for 10 min. All samples were stored at -80°C until further analysis. The relevant clinical data of all patients were available. All patients were diagnosed by histological examination. The detailed information of patients and healthy controls was summarized in **Table 1**.

The screening phase was divided into training set and validation set. Twenty pair's samples including 20 patients with TNBC and 20 healthy controls were enrolled while the validation set contained two cohorts.

Training set

All candidates were tested in an independent cohort of 20 plasma samples obtained from patients. All the patients were diagnosed with BC pathologically at the Zhejiang University of Traditional Chinese Medicine. The expression levels of these candidates were analyzed in all samples, and comparative $2^{-\Delta\Delta Ct}$ method algorithms were used to analyze the difference between patients and health controls.

Validation set

A case-control study was designed to validate the obvious difference of relative expression levels of the selected potential biomarkers candidates in another two independent cohort including 100 TNBC patients, 100 non-TNBC patients and 100 healthy controls and was named as cohort I and II, respectively.

Quantitative real-time PCR (qRT-PCR)

Total RNAs were obtained from plasmic samples using TRIzol reagent as described by the manufacturer (Invitrogen Life Technologies Co, Carlsbad, CA, USA). For mRNA detection, total RNAs (500 ng) were reverse transcribed using the reverse transcription kit (Takara, Tokyo,

Table 2. Primers for Quantitative RT-PCR

Gene name		Sequence
TCONS_l2_00003938	Forward	GTGACCTGATGGAGACAGCC
	Reverse	CGTGAGGTCCCTTGTTGTGT
ENST00000460164	Forward	AAGTCTGCCCACATGTACCG
	Reverse	ACGTTGCAGGTGTAGGTCTG
ENST00000425295	Forward	CTCCCAGACCTCACGTCAAC
	Reverse	TCACCAGGCGGGTGTATCTA
MALAT1	Forward	GTCATAACCAGCCTGGCAGT
	Reverse	CGAAACATTGGCACACAGCA
HOTAIR	Forward	TTGCCCCCAGCAAGAATCAT
	Reverse	GTTCCGGAAATCAGGGCAGA
ANRIL	Forward	GCGCCGGACTAGGACTATTT
	Reverse	GCCAGGACGGAGATCAGATG
HIF1A-AS2	Forward	GTCTTTCCTCTCCGTCCCAG
	Reverse	CAGACACCCGATCTCCGTG
UCA1	Forward	CAACCCTTAAGCTCCTGGCA
	Reverse	GTGGCGGTCTGAATGGAGAA
C17orf76-AS1	Forward	TGGCAGCTGAGTATTGGAGC
	Reverse	CTGATGGCAGCTACTGGTCC
CTC-338M12.3	Forward	TGGCAGCTGAGTATTGGAGC
	Reverse	CTGATGGCAGCTACTGGTCC
XR_250621.1	Forward	AGCGTCAACTGCTTTTGTGT
	Reverse	AGGTAGAGCCGTTACTCCCG
SOX2OT	Forward	AACACCCTGATCTGGCATGG
	Reverse	ATATGGCTGTTGCCTGGCTT
PTPRG-AS1	Forward	GGGTAAGCGATCTACGCCC
	Reverse	GTGGTTGCCCTCCTTAGGTC
ANRASSF1	Forward	CGCGCAGAATTAGCCTCTCT
	Reverse	ACCCACTGAGATAGGTCGGG

Japan). The quality of RNA samples was assessed by a UV spectrophotometer (Bio-Rad, Hercules, CA, USA), and the 260/280 nm absorbance ratio of samples were limited to 1.8-2.0. QRT-PCR was performed by using ABI Prism 7900HT (Applied Biosystems, CA, USA). The detailed primer information has been added in **Table 2**.

Risk score analysis

Risk score analysis was performed to evaluate the associations between the concentrations of the plasma lncRNA expression levels. The upper 95% reference interval of each lncRNA value in controls or non-metastasis group was set as the threshold to code the expression level of the corresponding lncRNA for each sample as 0 and 1 in the training set. A risk score function (RSF) to predict TNBC group was defined according to a linear combination of the expression level for each lncRNA. For example, the RSF for sample i using information from three lncRNAs was: rsfi=$\sum$3j-1Wj.sij. In the above equation, sij is the risk score for lncRNA j on sample i, and Wj is the weight of the risk score of lncRNA j. The risk score of three lncRNAs was calculated using the weight by the regression coefficient that was derived from the univariate logistic regression analysis of each lncRNAs. Multivariate logistic regression was also conducted to identify the candidate lncRNAs for their contribution. Samples were ranked according to their RSF and then divided into a high-risk group, representing the TNBC patients, and a low-risk group, representing the predicted control individuals or non-metastasis patients. Frequency tables and ROC curves were then used to evaluate the diagnostic effects of the profiling and to find the appropriate cutoff point, and to validate the procedure and cutoffs in the next validation sample set.

Statistical analysis

Data were presented as mean (S.E.M.). The Student's t-test and Mann-Whitney unpaired test analysis of variance were used to evaluate statistical differences between patients and controls. Analysis of area under the ROC curve (AUC) was used to estimate the effectiveness of lncRNAs for prediction. Statistical analysis was performed using softwares of STATA 9.2 and GraphPad prism. In all cases, $P<0.05$ was considered to be significant. All *P* values were two-sided.

Results

Selection TNBC-associated lncRNAs in training set

We first detected 14 candidate lncRNAs in 20 plasma samples of TNBC patients, 20 plasma samples of non-TNBC patients and 20 healthy controls. As present in **Figure 1**, three lncRNA including ANRIL, SOX2OT and ANRASSF1 were identified as significant upregulation in plasma from TNBC patients which was consistence

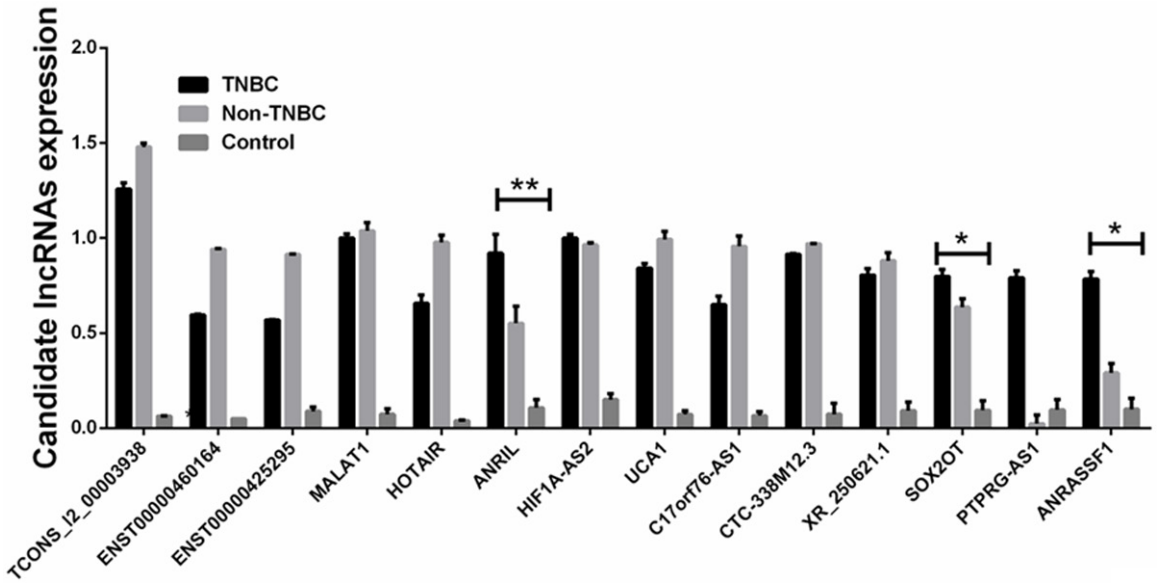


Figure 1. Relative expression of 14 candidate lncRNAs. Total 20 paired plasma from TNBC patients, 20 for, non-TNBC patients and 20 cancer-free controls were used in RT-qPCR analysis. Data was presented as mean ± SEM. Data was analyzed with student t test. *indicated $p<0.05$ and **indicated $p<0.01$.

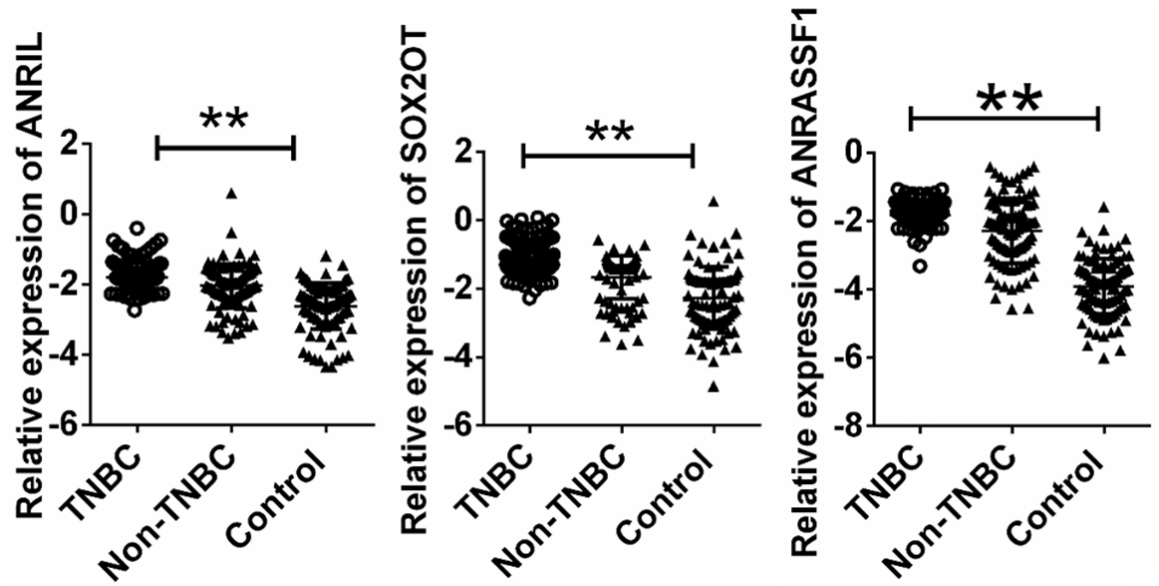


Figure 2. Validation of three different expressed lncRNA in cohort I. Plasma from 100 TNBC patients, 100 non-TNBC patients and 50 cancer-free controls were enrolled. Data were presented as plot of the median and range of log-transformed relative expression level and was analyzed with student t test. **indicated $p<0.01$.

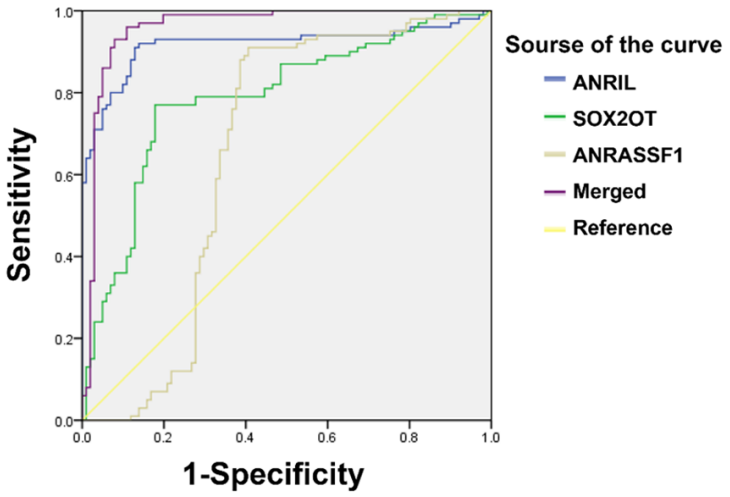
with the relative expression level in tumor tissues.

Validation of candidates lncRNAs in cohort I and II

We next examined the expression of these lncRNA levels through a larger individual samples including 100 cases (100 TNBC and 100 non-TNBC) and 50 controls. As shown in **Figure 2**, we found that ANRIL, SOX2OT and ANRASSF1 were consisted with the results in training set. The cohort I and II indicated samples obtained from different institutions.

Further experiment was conducted to explore the accuracy and specificity of these 3 lncRNA as a TNBC potential signature. Throughout the

A



B

Area Under the Curve

Test Result Variable(s)	Area	Std. Error[a]	Asymptotic Sig.[b]	Asymptotic 95% Confidence Interval	
				Lower Bound	Upper Bound
ANRIL	.918	.023	.000	.873	.962
SOX2OT	.788	.033	.000	.723	.852
ANRASSF1	.657	.042	.000	.573	.740
Merged	.959	.016	.000	.928	.990

a. Under the nonparametric assumption

b. Null hypothesis: true area = 0.5

Figure 3. ROC analysis of the three potential biomarkers for TNBC by using risk score analysis. The cutoff was 9.113. The sensitivity for ANRIL, SOX2OT, ANRASSF1 and the three merged were 0.873, 0.723, 0.573 and 0.928, respectively. The specificity for ANRIL, SOX2OT, ANRASSF1 and the three merged were 0.962, 0.852, 0.740 and 0.990, respectively.

multiphase testing and analysis, a profile of three lncRNAs might be considered to be the potential signature for the diagnosis of TNBC.

The risk score formula was applied to evaluate the diagnostic value of the three lncRNAs. Firstly, we divided the control group and case group in training set according to the upper 95% credibility interval (95% CI) in control group. The risk score was calculated based on the Logistic regression analysis. All plasma samples were then divided into a high-risk group which indicating the possible TNBC group, and a low-risk group, representing the predicted controls. We defined the cut-off value (9.113) as the maximal value of sensitivity + specificity. The positive predictive value (PPV) and negative predictive value (NPV) was calculated as 75% and 86% in the training set, respectively. We further applied the same value to calculate the risk score of samples in the validation sets, the PPV and NPV were 89%, 88%, respectively.

ROC curve analysis was used to evaluate the predicting diagnosis value of lncRNAs for TNBC. Areas under the ROC curves of the validated three lncRNAs signature were 0.918, 0.788 and 0.657, respectively While the combination of the three factors possessed a moderate ability for discrimination between HCC patients and controls with an area under the ROC curve of 0.959 (**Figure 3**).

Stability detection of lncRNAs in human plasma

We next amplified the three lncRNAs in five healthy controls, and detected the product of amplification by agarose electrophoresis. The bands presented in **Figure 4** indicated that both the three lncRNAs were detectable in human plasma. We further incubated human plasma obtained from three healthy controls at room temperature for 0 h, 12 h and 24 h and treated with frozen-thawing for 5 cycles. We found that neither the expression level of the three lncRNAs was alternated indicated that ANRIL, SOX2OT and ANRASSF1 was stably expressed and detectable in human plasma.

Discussion

Increasing evidence highlights that lncRNAs can serve as diagnostic biomarkers and therapeutic targets in solid tumors, including BC, however, their relative expression levels in various subtypes of human BC, particularly the TNBC subtype, remain unknown [17, 18]. ncRNA signatures of normal cancer tissues and metastases are used to classify different cancer types, indicating the potential of these lncRNAs as biomarkers for diagnosis, prognosis, and therapy [11, 12].

Thus far, ANRIL upregulation is considered the primary feature of many carcinomas, including BC [19]. Researchers has found ANRIL expres-

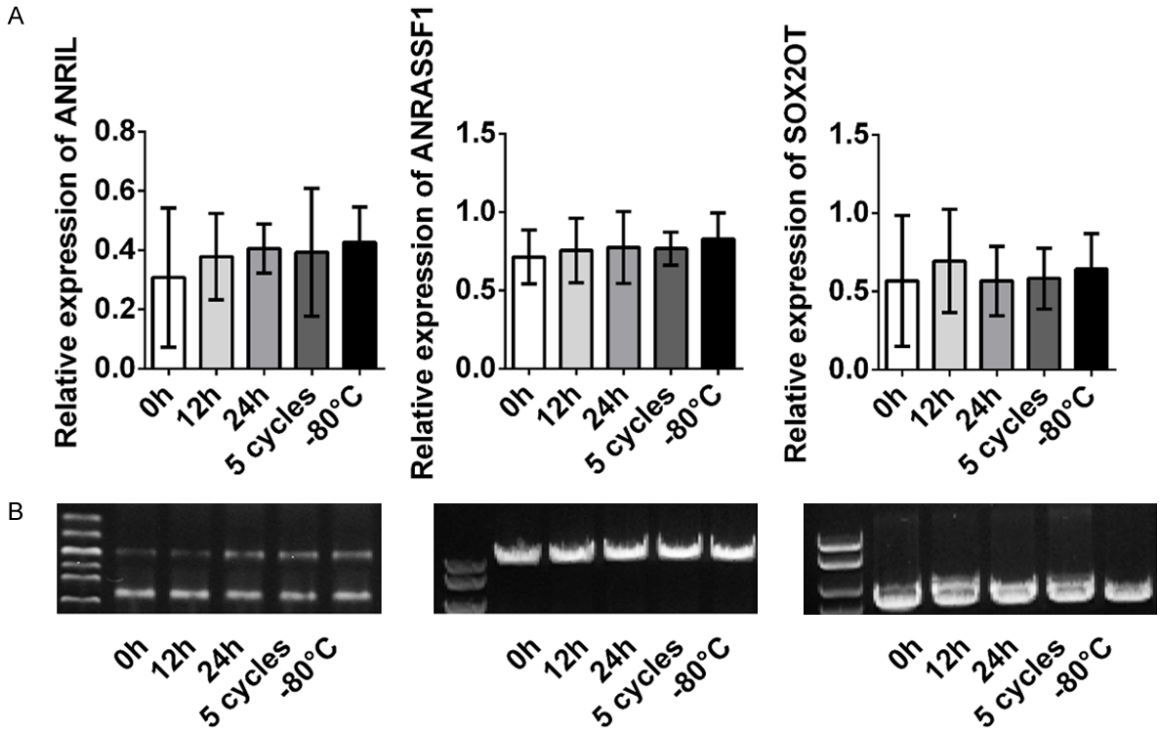


Figure 4. The products of the amplification fragment were detected by RT-PCR. Data was presented as mean ± SEM. The human plasma samples obtained from three healthy controls were incubated at room temperature for 0 h, 12 h and 24 h and treated with frozen-thawing for 5 cycles. A: The relative expression of lncRNAs by qRT-PCR. B: The band for AGE agarose gel electrophoresis.

sion was higher in nonsmall cell lung cancer tissues than adjacent nontumor tissues and was associated with high TNM stage and advanced lymph node metastasis. Similarly, ANRIL expression was higher in hepatocellular carcinoma tissues than in adjacent tumor-free tissues, and patients with high ANRIL expression showed significantly poor overall survival [20]. The lncRNA SOX2OT was also confirmed play important role in the pathogenesis of TNBC. For example, ectopic expression of SOX2OT led to an almost 20-fold increase in SOX2 expression, together with a reduced proliferation and increased breast cancer cell anchorage-independent growth. We propose that SOX2OT plays a key role in the induction and/or maintenance of SOX2 expression in breast cancer [21]. For ANRASSF1, the endogenous expression is higher in breast and prostate tumor cell lines compared with non-tumor. Researcher has found that the intronic long noncoding RNA ANRASSF1 could recruit PRC2 to the RASSF1A promoter, reducing the expression of RASSF1A and increasing cell proliferation [22].

In this study, 14 lncRNAS were enrolled as candidates based on the important function reported previously. These tissues different expressed lncRNA might present an alternated level in patients' bodily fluid. Research has speculated that the aberrant expressed circulation lncRNA might be derived from the tumor. Thus, we examined the expression of the 14 lncRNA in patients and healthy control. By employing a multiple screening phase including three cohorts, we found that only ANRIL, SOX2OT and ANRASSF1 were different expressed in TNBC with an increased level. In summary, we identified three lncRNAs, ANRIL, SOX2OT and ANRASSF1, as the potential biomarker for the tumorigenesis prediction of TNBC.

Disclosure of conflict of interest

None.

Address correspondence to: Anping Pan, Department of Oncology, Hangzhou Cancer Hospital, Yanguang Road 34#, Hangzhou 310002, Zhejiang

Province, China. Tel: +86-571-88122222; Fax: +86-571-88122222; E-mail: pananpingzj@yeah.net

References

[1] Polyak K. Heterogeneity in breast cancer. J Clin Invest 2011; 121: 3786-3788.

[2] Bosch A, Eroles P, Zaragoza R, Vina JR and Lluch A. Triple-negative breast cancer: molecular features, pathogenesis, treatment and current lines of research. Cancer Treat Rev 2010; 36: 206-215.

[3] Jiao Q, Wu A, Shao G, Peng H, Wang M, Ji S, Liu P and Zhang J. The latest progress in research on triple negative breast cancer (TNBC): risk factors, possible therapeutic targets and prognostic markers. J Thorac Dis 2014; 6: 1329-1335.

[4] Zhao J, Zhao D, Poage GM, Mazumdar A, Zhang Y, Hill JL, Hartman ZC, Savage MI, Mills GB and Brown PH. Death-associated protein kinase 1 promotes growth of p53-mutant cancers. J Clin Invest 2015; 125: 2707-2720.

[5] Bernardi R and Gianni L. Hallmarks of triple negative breast cancer emerging at last? Cell Res 2014; 24: 904-905.

[6] Horiuchi D, Camarda R, Zhou AY, Yau C, Momcilovic O, Balakrishnan S, Corella AN, Eyob H, Kessenbrock K, Lawson DA, Marsh LA, Anderton BN, Rohrberg J, Kunder R, Bazarov AV, Yaswen P, McManus MT, Rugo HS, Werb Z and Goga A. PIM1 kinase inhibition as a targeted therapy against triple-negative breast tumors with elevated MYC expression. Nat Med 2016; 22: 1321-1329.

[7] Fatica A and Bozzoni I. Long non-coding RNAs: new players in cell differentiation and development. Nat Rev Genet 2014; 15: 7-21.

[8] Batista PJ and Chang HY. Long noncoding RNAs: cellular address codes in development and disease. Cell 2013; 152: 1298-1307.

[9] Yu H, Xu Q, Liu F, Ye X, Wang J and Meng X. Identification and validation of long noncoding RNA biomarkers in human non-small-cell lung carcinomas. J Thorac Oncol 2015; 10: 645-654.

[10] Yang J, Lin J, Liu T, Chen T, Pan S, Huang W and Li S. Analysis of lncRNA expression profiles in non-small cell lung cancers (NSCLC) and their clinical subtypes. Lung Cancer 2014; 85: 110-115.

[11] Tang J, Zhuo H, Zhang X, Jiang R, Ji J, Deng L, Qian X, Zhang F and Sun B. A novel biomarker Linc00974 interacting with KRT19 promotes proliferation and metastasis in hepatocellular carcinoma. Cell Death Dis 2014; 5: e1549.

[12] Tang J, Jiang R, Deng L, Zhang X, Wang K and Sun B. Circulation long non-coding RNAs act as biomarkers for predicting tumorigenesis and metastasis in hepatocellular carcinoma. Oncotarget 2015; 6: 4505-4515.

[13] Dong L, Qi P, Xu MD, Ni SJ, Huang D, Xu QH, Weng WW, Tan C, Sheng WQ, Zhou XY and Du X. Circulating CUDR, LSINCT-5 and PTENP1 long noncoding RNAs in sera distinguish patients with gastric cancer from healthy controls. Int J Cancer 2015; 137: 1128-35.

[14] Arita T, Ichikawa D, Konishi H, Komatsu S, Shiozaki A, Shoda K, Kawaguchi T, Hirajima S, Nagata H, Kubota T, Fujiwara H, Okamoto K and Otsuji E. Circulating long non-coding RNAs in plasma of patients with gastric cancer. Anticancer Res 2013; 33: 3185-3193.

[15] Crea F, Watahiki A, Quagliata L, Xue H, Pikor L, Parolia A, Wang Y, Lin D, Lam WL, Farrar WL, Isogai T, Morant R, Castori-Eppenberger S, Chi KN and Helgason CD. Identification of a long non-coding RNA as a novel biomarker and potential therapeutic target for metastatic prostate cancer. Oncotarget 2014; 5: 764-774.

[16] Ren S, Wang F, Shen J, Sun Y, Xu W, Lu J, Wei M, Xu C, Wu C, Zhang Z, Gao X, Liu Z, Hou J and Huang J. Long non-coding RNA metastasis associated in lung adenocarcinoma transcript 1 derived miniRNA as a novel plasma-based biomarker for diagnosing prostate cancer. Eur J Cancer 2013; 49: 2949-2959.

[17] Zhang Z, Zhu Z, Watabe K, Zhang X, Bai C, Xu M, Wu F and Mo YY. Negative regulation of lncRNA GAS5 by miR-21. Cell Death Differ 2013; 20: 1558-1568.

[18] Huang J, Zhou N, Watabe K, Lu Z, Wu F, Xu M and Mo YY. Long non-coding RNA UCA1 promotes breast tumor growth by suppression of p27 (Kip1). Cell Death Dis 2014; 5: e1008.

[19] Liu M, Xing LQ and Liu YJ. A three-long noncoding RNA signature as a diagnostic biomarker for differentiating between triple-negative and non-triple-negative breast cancers. Medicine (Baltimore) 2017; 96: e6222.

[20] Boon RA, Jae N, Holdt L and Dimmeler S. Long noncoding RNAs: from clinical genetics to therapeutic targets? J Am Coll Cardiol 2016; 67: 1214-1226.

[21] Askarian-Amiri ME, Seyfoddin V, Smart CE, Wang J, Kim JE, Hansji H, Baguley BC, Finlay GJ and Leung EY. Emerging role of long non-coding RNA SOX2OT in SOX2 regulation in breast cancer. PLoS One 2014; 9: e102140.

[22] Beckedorff FC, Ayupe AC, Crocci-Souza R, Amaral MS, Nakaya HI, Soltys DT, Menck CF, Reis EM and Verjovski-Almeida S. The intronic long noncoding RNA ANRASSF1 recruits PRC2 to the RASSF1A promoter, reducing the expression of RASSF1A and increasing cell proliferation. PLoS Genet 2013; 9: e1003705.