

# Cell-free DNA in blood is a potential diagnostic biomarker of breast cancer

BO GONG<sup>1</sup>, JIANXIANG XUE<sup>3</sup>, JING YU<sup>1</sup>, HAICHUAN LI<sup>1</sup>, HEYU HU<sup>1</sup>, HSUEHHENG YEN<sup>2</sup>,  
JINYIN HU<sup>3</sup>, QIANGGANG DONG<sup>3</sup> and FUHUA CHEN<sup>3</sup>

<sup>1</sup>Department of Clinical Laboratory, Shanghai Changning Maternity and Infant Health Hospital, Shanghai 200051;

<sup>2</sup>Shanghai XP Biomed Ltd., Shanghai 200120; <sup>3</sup>Shanghai Cancer Institute, Shanghai 200032, P.R. China

Received November 11, 2011; Accepted January 13, 2012

DOI: 10.3892/ol.2012.576

**Abstract.** Breast cancer is a highly malignant disease in women. A convenient screening tool with high accuracy for early detection, not only in high-risk individuals but in the general population, is necessary. Two hundred breast cancer patients, 100 healthy controls and 100 hyperplasia patients were enrolled in this study. Samples were randomly assigned into training or testing cohorts. The receiver operating characteristic curve was used to explore the optimal concentration of cell-free DNA (*GAPDH*) in the training cohort and the cut-off point was validated in the testing cohort. The results showed that both in the training and testing cohorts, the overall accuracy of classification between cancer, healthy controls and hyperplasia was higher than 0.9. The sensitivity, specificity, positive predictive value and negative predictive value also reached 0.9, with the exception of the negative predictive value in the testing cohort. This study provides useful information on the use of concentration of free DNA for breast cancer detection. These findings need to be validated in a large prospective trial prior to clinical application.

## Introduction

Breast cancer is the most common type of cancer and is the second leading cause of cancer-related mortality in women (1). Early detection or diagnostic tools such as breast mammography or ultrasound may have the potential to detect breast cancer (2,3). However, these tools have limitations such as radiation exposure (4).

Apoptotic tumor cells may release DNA into the peripheral blood and this may hold possibilities for the early detection of cancer (5,6). Recent studies have shown that the concentration of free DNA in patients' blood has the potential to be a diagnostic marker for numerous cancers, including lung, colon and breast

cancer (6,7). The results of these studies demonstrated certain promising findings. However, lack of sufficient validation and small sample size has limited previous research.

In the present study, we utilized a case-control design and a split-sample approach to investigate the classification ability of free DNA concentration in breast cancer. A total of 200 breast cancer patients, 100 healthy controls, and 100 hyperplasia patients were enrolled in this study. The findings of this study may provide additional information regarding free DNA concentration in breast cancer and its clinical application.

## Materials and methods

**Patients and healthy controls.** We studied 200 patients with breast cancer at the Changning Maternity and Infant Health Hospital, China, from January 2008 until December 2009. The median age of the patients was 51.9 years (range, 25-74). Additionally, we enrolled 100 patients with hyperplasia whose median age was 41 years (range, 22-63) and 100 healthy women whose median age was 44 years (range, 24-64) as a control. Benign cancer cases were diagnosed by digital molybdenum target and malignant cancer cases were classified into four stages according to the 2003 World Health Organization (WHO) classification of tumors of the breast. This resulted in 8 patients of stage I, 86 patients of stage II, 93 patients of stage III and 13 patients of stage IV. Among these, 122 women had invasive ductal carcinoma, 17 had invasive lobular carcinoma, 19 had medullary carcinoma, 9 had invasive carcinoma, 5 had intraductal carcinoma, 13 had accompanying metastasis of the lymph nodes and 15 had other carcinoma including mucinous carcinoma and neuroendocrine carcinoma.

This study was approved by the institutional review boards of the hospital. Written informed consent was obtained from all patients. Their clinical characteristics are shown in Table I.

**Preparation of serum and DNA extraction.** From each patient, 5 ml whole blood was collected in tubes prior to surgery or chemotherapy and immediately stored at 4°C. Serum and leukocyte preparations were performed within 4 h. The blood samples were centrifuged at 2,500 rpm for 5 min to obtain upper blood serum for circulating DNA extraction by in-house magnetic beads. The isolated DNA was dissolved in TE buffer at a volume of one-tenth of the original blood serum.

---

*Correspondence to:* Dr Bo Gong, Department of Clinical Laboratory, Shanghai Changning Maternity and Infant Health Hospital, No. 773, WuYi Road, Shanghai 200051, P.R. China  
E-mail: bogongsha@hotmail.com

**Key words:** breast cancer, free DNA, real-time PCR

Table I. Clinical characteristics of the study population.

Characteristics	Number (%)
Age (mean $\pm$ SD)	52.77 $\pm$ 9.86
Stage	
I	90 (45)
II	83 (41.5)
III	27 (13.5)
ER status	
Positive	114 (57)
Negative	86 (43)
Tumor size	
$\geq 2.5$ cm	113 (56.5)
$< 2.5$ cm	87 (43.5)

ER, estrogen receptor.

**Measurement of free DNA concentration.** The forward primer: 5'-GGAAGGTGAAGGTCGGAGTC-3', reverse primer: 5'-GAAGATGGTGATGGGATTTC-3' and probe: 5'-FAM-CAAGCTTCCCGTTCTCAGCC-TAMRA-3' were used to amplify the sequence of the *GAPDH* gene. The real-time PCR mixture comprised 20  $\mu$ l reaction mix, 10  $\mu$ l Premix Ex Taq (2X) (Takara Bio Inc., Shiga, Japan), 0.4  $\mu$ l PCR forward primer, 0.4  $\mu$ l reverse primer, 0.8  $\mu$ l probe, 0.4  $\mu$ l ROX (50X), 2  $\mu$ l serum DNA and 6  $\mu$ l ddH<sub>2</sub>O. Cycle conditions were as follows: preheating at 50°C for 5 min, denaturing at 94°C for 5 min (95°C for 20 sec; 60°C for 60 sec) at 40 cycles at 60°C detection. The result was expressed by copy number/reaction.

**Statistical analysis.** The 200 breast cancer patients, 100 healthy controls, and 100 hyperplasia patients, were randomly assigned to a training cohort or a testing cohort. In the training cohort, the receiver operating characteristic (ROC) curve was used to determine the optimal cut-off point of *GAPDH* concentration for separating breast cancer samples and healthy controls. The best cut-off point is the point on the curve with the minimum distance between the curve and the upper left corner in the ROC curve. To validate the predictive power of cut-off concentrations of *GAPDH*, the same cut-off point was applied in the testing cohort. Odds ratios were estimated using unconditional logistic regression. Analyses were carried out using SAS version 9.1 software (SAS Institute Inc.). Two-tailed tests were used and  $p < 0.05$  was considered to indicate a statistically significant result.

## Results

**Optimization of PCR assay conditions.** The DNA of breast cancer cells, MCF-7, (purchased from the Cell Institute of the Chinese Academy of Sciences) was extracted for PCR amplification and precise quantitation of *GAPDH* to set up the standard curve and to estimate the dynamic range of PCR assays. The results showed that the copy number and cycle number had apparent linear correlation, whereas the correla-

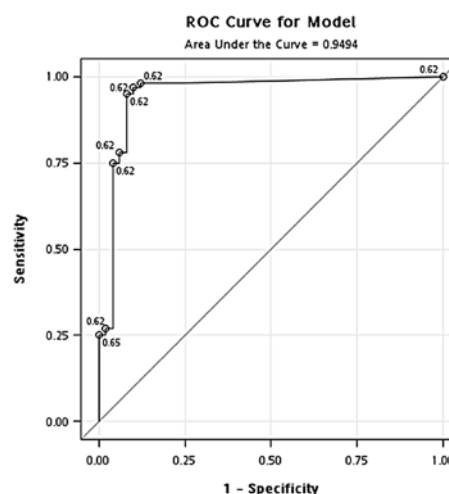


Figure 1. ROC curve showing the optimal cut-off concentration of *GAPDH* in the training cohort.

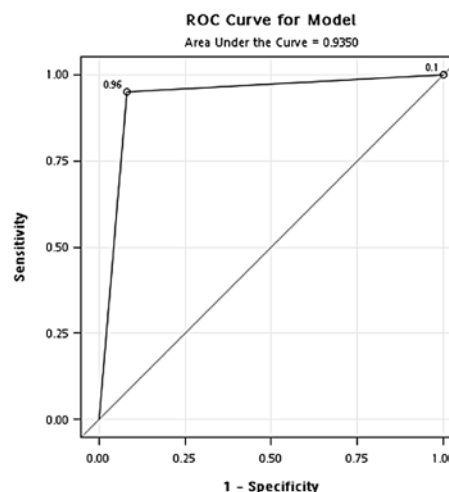


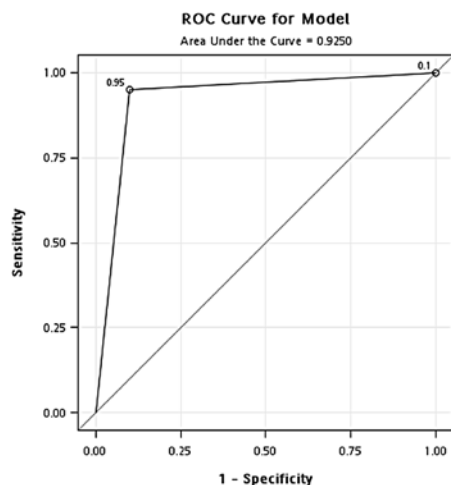
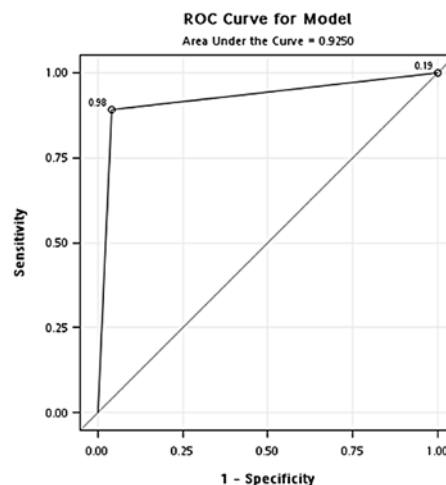
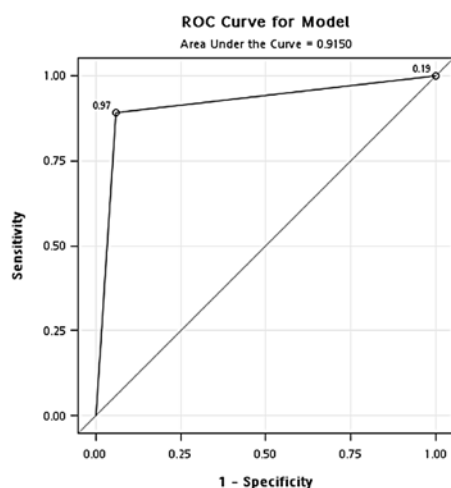
Figure 2. ROC curve analysis of the optimal concentration of *GAPDH* for classification of the breast cancer patients and healthy controls in the training cohort.

tion coefficient (R) was 0.999 when the DNA copy number was between  $10^4$  and  $10^{10}$ .

**Identification of optimal concentration of *GAPDH* in the blood.** The ROC curve was used to define the cut-off point for sample separation in the training cohort. The point with the minimum distance between the curve and the upper left corner was 471 ng/ml (Fig. 1). The identified optimal concentration was reanalyzed in the training cohort. In the classification of cancer and healthy controls, the results showed that the overall accuracy, sensitivity, specificity, positive predictive value and negative predictive value were 0.94, 0.95, 0.92, 0.96 and 0.90, respectively. In addition, the odds ratio was 218.5 [95% confidence interval (CI), 56.02-852.26] (Fig. 2 and Table II). In the classification of cancer and hyperplasia, the overall accuracy, sensitivity, specificity, positive predictive value and negative predictive value were 0.93, 0.95, 0.90, 0.95 and 0.90, respectively. In addition, the odds ratio was 171 (95% CI, 47.10-620.79) (Fig. 3 and Table II).

Table II. Accuracy indices of optimal *GAPDH* concentration for sample separation.

Index	Cancer vs. healthy controls		Cancer vs. hyperplasia	
	Training cohort	Testing cohort	Training cohort	Testing cohort
Sensitivity	0.95	0.89	0.95	0.89
Specificity	0.92	0.94	0.90	0.96
Positive predictive value	0.96	0.98	0.95	0.98
Negative predictive value	0.90	0.81	0.90	0.81
Overall accuracy	0.94	0.91	0.93	0.91

Figure 3. ROC curve analysis of the optimal concentration of *GAPDH* for classification of the breast cancer and hyperplasia patients in the training cohort.Figure 5. ROC curve analysis of the optimal concentration of *GAPDH* for classification of the breast cancer and hyperplasia patients in the testing cohort.Figure 4. ROC curve analysis of the optimal concentration of *GAPDH* for classification of the breast cancer patients and healthy controls in the testing cohort.

**Validation of optimal concentration of *GAPDH*.** To validate the optimal concentration of *GAPDH* identified in the training cohort, the same concentration of *GAPDH* (471 ng/ml) was used to classify samples in the testing cohort. In the classifi-

cation of cancer and healthy controls, the results showed that the overall accuracy, sensitivity, specificity, positive predictive value and negative predictive value were 0.91, 0.89, 0.94, 0.97 and 0.81, respectively. In addition, the odds ratio was 126.8 (95% CI, 33.71-476.69) (Fig. 4 and Table II). In the classification of cancer and hyperplasia, the overall accuracy, sensitivity, specificity, positive predictive value and negative predictive value were 0.91, 0.89, 0.96, 0.98 and 0.81, respectively. In addition, the odds ratio was 194.1 (95% CI, 41.34-911.70) (Fig. 5 and Table II).

## Discussion

In this study, samples were divided into training and testing cohorts and the optimal *GAPDH* concentration was identified in the training cohort. The optimal *GAPDH* concentration was then further validated in the testing cohort. The validation step in the testing cohort is the first advantage of this study. The second advantage is that we not only compared the free DNA concentration in breast cancer patients and healthy controls, but also compared it with that of hyperplasia patients. A comparison of healthy controls and hyperplasia patients revealed no significant difference. A possible reason may be that the free DNA is released from cancer cells and is not found in the hyperplasia samples or in healthy controls.

During the onset and progression of cancer, biomacromolecules, such as nucleic acid in the tumor cells, are released into the blood circulation. This tumor-derived free DNA (circulating nucleic acids) present in the small particles (microparticles) coated by the cell membrane and their molecular characteristics were mostly consistent with the primary tumor cells (8). Therefore, this free DNA has attracted widespread attention as non-invasive cancer research material (9). In recent years, molecular high-throughput methods have become widespread in free nucleic acid research (10-12) and these novel research methods have emerged as new potent and promising cancer molecular diagnostics strategies.

Developing this molecular detection of cancer based on free nucleic acids requires understanding of the basic data of such nucleic acid content in circulating cell-free DNA in cancer patients and the use of appropriate methods for purification. However, the lack of a large sample for clinical research still exists. In view of this fact, in the present study, whole blood samples were collected from 100 healthy women and 100 patients with benign breast disease and 200 patients with breast cancer and the free DNA content of these patients was analyzed by separation and quantitative PCR. The results showed that the free DNA of breast cancer patients was significantly increased compared to that of healthy women and women with benign breast disease.

Our study on the assessment of free DNA content is based on the copy number detection of the *GAPDH* gene. *GAPDH* is one of the most commonly used housekeeping genes in molecular biology. It is a member of the single-copy gene family in the human genome and is characterized by low gene amplification or deletion mutation probability in various types of cancer occurrence. Thus, *GAPDH* is a good candidate to quantify the free DNA content released from cancers and to reflect the tumor burden. In this study, the *GAPDH* gene copy number was regarded as an evaluation index and we found that 85% of the stage I and II breast cancer patients have 103 or more copies of the gene copy number, which is sufficient to meet the requirements of genetic diagnosis. These findings established the scientific foundations for the further development of specific new technologies of breast cancer genetic diagnosis.

Traditional screening tools such as breast mammography or ultrasound are useful in breast cancer screening and early detection. The combination of traditional screening tools and free DNA concentration may improve the accuracy of cancer detection and its use would not be limited to high-risk women.

In conclusion, the concentration of *GAPDH* may have potential for the early detection of breast cancer. However, its clinical application may need to be validated in a larger prospective trial.

## Acknowledgements

This study was sponsored by the Science and Technology Commission of Shanghai Municipality, grant number 08411966800. We are grateful for the support of Shanghai Changning Maternity and Infant Health Hospital.

## References

1. Jemal A, Siegel R, Xu J and Ward E: Cancer statistics. *CA Cancer J Clin*: caac.20073, 2010.
2. Nelson HD, Tyne K, Naik A, *et al*: Screening for breast cancer: U.S. preventive services task force recommendation statement. *Ann Intern Med* 151: 716-726, 2009.
3. Corsetti V, Houssami N, Ghirardi M, *et al*: Evidence of the effect of adjunct ultrasound screening in women with mammography-negative dense breasts: interval breast cancers at 1 year follow-up. *Eur J Cancer* 47: 1021-1026, 2011.
4. Tabár L, Vitak B, Chen TH-H, *et al*: Swedish two-county trial: impact of mammographic screening on breast cancer mortality during 3 decades. *Radiology* 260: 658-663, 2011.
5. Gahan PB and Swaminathan R: Circulating nucleic acids in plasma and serum. Recent developments. *Ann N Y Acad Sci* 1137: 1-6, 2008.
6. Schwarzenbach H, Hoon DSB and Pantel K: Cell-free nucleic acids as biomarkers in cancer patients. *Nat Rev Cancer* 11: 426-437, 2011.
7. Roth C, Pantel K, Muller V, *et al*: Apoptosis-related deregulation of proteolytic activities and high serum levels of circulating nucleosomes and DNA in blood correlate with breast cancer progression. *BMC Cancer* 11: 4, 2011.
8. Orozco AF and Lewis DE: Flow cytometric analysis of circulating microparticles in plasma. *Cytometry A* 77: 502-514, 2010.
9. Johnson PJ and Lo YMD: Plasma nucleic acids in the diagnosis and management of malignant disease. *Clin Chem* 48: 1186-1193, 2002.
10. Li J, Harris L, Mamon H, *et al*: Whole genome amplification of plasma-circulating DNA enables expanded screening for allelic imbalance in plasma. *J Mol Diagn* 8: 22-30, 2006.
11. Beck J, Urnovitz HB, Mitchell WM and Schütz E: Next generation sequencing of serum circulating nucleic acids from patients with invasive ductal breast cancer reveals differences to healthy and nonmalignant controls. *Mol Cancer Res* 8: 335-342, 2010.
12. Korshunova Y, Maloney RK, Lakey N, *et al*: Massively parallel bisulphite pyrosequencing reveals the molecular complexity of breast cancer-associated cytosine-methylation patterns obtained from tissue and serum DNA. *Genome Res* 18: 19-29, 2008.