Screening of differentially expressed genes and identification of NUF2 as a prognostic marker in breast cancer

WENJIE XU1, YIZHEN WANG1,2, YANAN WANG1, SHANMEI LV1,2, XIUPING XU1 and XUEJUN DONG1

1Department of Clinical Laboratory Center, Shaoxing Hospital, Zhejiang University School of Medicine, Shaoxing People's Hospital, 568 Zhongxing North Road, Shaoxing, Zhejiang 312000; 2Zhejiang Provincial Key Laboratory of Medical Genetics, Key Laboratory of Laboratory Medicine, Ministry of Education, School of Laboratory Medicine and Life Sciences, Wenzhou Medical University, Wenzhou, Zhejiang 325035, P.R. China

Received November 2, 2018; Accepted May 29, 2019

DOI: 10.3892/ijmm.2019.4239

Abstract. The aims of the present study were to screen differentially expressed genes (DEGs) in breast cancer (BC) and investigate NDC80 kinetochore complex component (NUF2) as a prognostic marker of BC in detail. A total of four BC microarray datasets, downloaded from the Gene Expression Omnibus (GEO) and The Cancer Genome Atlas (TCGA) databases, were used to screen DEGs. A total of 190 DEGs with the same expression trends were identified in the 4 datasets, including 65 upregulated and 125 downregulated DEGs. Functional and pathway enrichment analyses were performed using the Database for Annotation, Visualization and Integrated Discovery. The upregulated DEGs were enriched for 10 Gene Ontology (GO) terms and 7 pathways, and the downregulated DEGs were enriched for 10 GO terms and 10 pathways. A protein-protein interaction network containing 149 nodes and 930 edges was constructed using the Search Tool for the Retrieval of Interacting Genes, and 2 functional modules were identified using the MCODE plugin of Cytoscape. Based on an in-depth analysis of module 1 and literature mining, NUF2 was selected for further research. Oncomine database analysis and reverse transcription-quantitative PCR showed that NUF2 is significantly upregulated in BC tissues. In analyses of correlations between NUF2 and clinical pathological characteristics, NUF2 was significantly associated with the malignant features of BC. Using 5 additional datasets from GEO, it was demonstrated that NUF2 has a significant prognostic role in both ER-positive and ER-negative BC. A Gene Set Enrichment Analysis indicated that NUF2 may regulate breast carcinogenesis and progression via cell cycle-related pathways. The results of the present study demonstrated that NUF2 is overexpressed in BC and is significantly associated with its multiple pathological features and prognosis.

Introduction

Breast cancer (BC) is the most common female malignancy and the second leading cause of mortality in women worldwide (1). According to the World Health Organization in 2012, one-third of Asian women develop BC (2,3). Currently, BC treatment includes partial excision with or without radiotherapy and systemic therapies such as endocrine therapy, chemotherapy, molecular targeted therapy, and a combination of them (4). Although advanced therapeutic techniques based on surgery have considerably improved the survival of patients with BC and the five-year survival has increased from 75% in 1976 to 91% in 2017, high rates of metastasis and recurrence remain (1,5,6). Recently, molecular targeted therapy has been shown to play an important role in individualized treatment of BC. For instance, a monoclonal antibody against HER2, trastuzumab, has been demonstrated to improve survival of patients with BC; however, the prognosis remains poor (7,8). Therefore, to improve BC prognosis, effective therapeutic targets and prognostic biomarkers are needed.

NDC80 kinetochore complex component (NUF2), also known as CDC1A1, is a centromere-related protein (9). It regulates the binding of centromeres to spindle microtubules, participates in cell cycle regulation and has important roles in cell proliferation and apoptosis (10). NUF2 is overexpressed in a number of cancers, including lung cancer, cholangiocarcinoma, renal cell carcinoma and bladder cancer (11). Although the expression and prognostic significance of NUF2 in BC have been suggested (12,13), its precise role and underlying molecular mechanisms of action remain to be investigated.

In the present study, 4 mRNA microarray datasets were analyzed from the Gene Expression Omnibus (GEO) and The Cancer Genome Atlas (TCGA) databases to identify differentially expressed genes (DEGs) between BC tissues and normal breast tissues. The bioinformatics analysis and literature mining suggested that NUF2 is a key gene in the progression of BC. The expression of NUF2 in BC samples and its correlation with clinical pathological characteristics.
were then analyzed. In addition, the prognostic value of NUF2 was analyzed using individual and pooled methods. In a gene set enrichment analysis (GSEA), it was demonstrated that NUF2 might be involved in cell-cycle related pathways. The results of the present study suggest that NUF2 is a prognostic indicator of BC.

Materials and methods

Microarray data. GSE42568 (14), GSE45827 (15), GSE65194 (16) and TCGA BC microarray datasets, downloaded from GEO (17) and TCGA (18), were used to screen DEGs in BC. The TCGA dataset was used for analyses of clinical pathological characteristics associated with NUF2 in patients with BC. The following 5 additional BC microarray datasets were selected for prognostic analyses: GSE1456 (19), GSE22220 (20), NKI (21), GSE4299 (22) and GSE20685 (23).

To normalize mRNA levels, patients for each cohort were reclassified into four subsets (X1, X2, X3 and X4) based on the quartile for expression values. The datasets were then reclassified into a new dataset for a pooled analysis.

DEG identification. BC-related microarray data downloaded from the GEO and TCGA databases were processed using R software (version 3.4.3; https://cran.r-project.org/). DEGs between BC tissues and normal breast tissues were identified using the limma package in R. Fold-change (FC) values were calculated and the DEGs were further selected based on the following cutoff criteria: P<0.01 and log FC>2. Overlapping DEGs among the four datasets were identified using Funrich (version 3.1.3; http://www.funrich.org).

Functional and pathway enrichment analyses of DEGs. Gene Ontology (GO) is used to identify enriched functions of genes in three independent categories: Biological process (BP), molecular function (MF) and cellular component (CC) (24).

Kyoto Encyclopedia of Genes and Genomes (KEGG) was used to identify relevant pathways for the genes (25). GO BP and KEGG signaling pathway analyses of the DEGs were performed using the Database for Annotation Visualization and Integrated Discovery (DAVID) online tool (https://david.ncifcrf.gov/) (26) with P<0.05 as the threshold for significance.

Protein-protein interaction (PPI) network analysis. The Search Tool for the Retrieval of Interacting Genes (STRING; https://string-db.org/) was used to develop a PPI network. Using the STRING database, DEGs with a combined score ≥0.4 were chosen to construct the network, which was visualized using Cytoscape (version 3.6.1) (27).

Molecular Complex Detection (MCODE), a plugin for Cytoscape, was used to construct functional modules in the PPI network.

Gene set enrichment analysis (GSEA). A GSEA was conducted based on protocols obtained from the website (http://software.broadinstitute.org/gsea/index.jsp) and a previous study (28). GSEA (version 3.0) was run for the KEGG gene sets (c2.cp.kegg.v.6.0.symbols.gmt). The number of permutations was set to 1,000 and the phenotype labels were NUF2-high and NUF2-low. FDR <0.25 and NOM P<0.05 indicated statistical significance.

Oncomine analysis. Oncomine (https://www.oncomine.org/) is an online cancer microarray database, aiming to facilitate the discovery of novel biomarkers from genome-wide expression analyses. In the present study, the mRNA expression differences of NUF2 between BC and normal breast tissues were explored using the Oncomine database.

Patients and samples. BC and matched adjacent tissues were collected from the Pathology Department of Shaoxing People's Hospital (Shaoxing, China). Samples were obtained from 42 patients at initial diagnosis and were immediately frozen in liquid nitrogen. The present study was authorized by the Hospital Ethics Committee and informed consent was obtained from all patients.

Reverse transcription (RT)-quantitative (q)PCR. Total RNA was isolated from the BC and matched adjacent tissues using TRIzol (Invitrogen; Thermo Fisher Scientific, Inc., Waltham, MA, USA). The Nanodrop 2000 (Thermo Fisher Scientific, Inc.) was used to detect the purity and concentration of the total RNA. According to the manufacturer's protocol, RT-qPCR was performed using the LightCycler® 480 PCR apparatus (Roche Diagnostics, Basel, Switzerland) and the One Step SYBR® PrimeScript™ RT-PCR kit II (Takara Bio, Inc., Otsu, Japan).

Amplification was performed under the following conditions: 42°C for 5 min, 95°C for 10 sec; 40 cycles of 95°C for 5 sec and 60°C for 20 sec; and 65°C for 15 sec. The primers used were as follows: NUF2 forward primer 5'-TACCATTCAAGATTAGTACT-3' and reverse primer 5'-TAGAATATCAGCTCTCAAAG-3'; and β-actin forward primer 5'-CATGTACGTGTCTATCCAGG-3' and reverse primer 5'-CTCCCTTATGTACGGCAGAT-3'. The relative levels of NUF2 expression were evaluated by the 2^ΔΔcq (29) method using β-actin as the control.

Statistical analyses. All statistical analyses were performed using SPSS 20.0 (IBM Corps., Armonk, NY, USA). An independent t-test was used for analyzing the continuous data. The \( \chi^2 \) test and \( \chi^2 \) test with continuity correction were performed to analyze the association of NUF2 with clinical pathological characteristics. Bonferroni's post hoc test was used to analyze the clinical pathological characteristics between more than 2 groups. Survival curves were generated by the Kaplan-Meier method and significance was determined using the log-rank test. Bonferroni's post hoc test was used for pairwise comparisons.

Multivariable survival analysis was performed using the Cox proportional hazards regression model and significance was determined using the likelihood ratio test. \( P<0.05 \) was considered to indicate statistically significant differences, while for Bonferroni's test, \( P<0.05/N \) was considered to indicate statistically significant differences, where \( N=\)the number of pairwise comparisons.

Results

Identification of DEGs in BC. DEGs between the BC and normal breast tissues were screened using the GEO and TCGA databases. As shown in Fig. 1A, 1,702, 461, 600 and 337 DEGs were upregulated in the GSE45827, GSE42568, GSE65194, and TCGA datasets, and 613, 715, 264, and 337 dEGs were upregulated in the GSE45827, GSE42568, GSE65194, and TCGA datasets, and 613, 715, 264, and...
872 DEGs were downregulated, respectively (Fig. 1B). In total, 190 DEGs exhibited the same expression trends in all datasets, including 65 upregulated and 125 downregulated genes.

**Functional and pathway enrichment for the DEGs.** GO BP and KEGG signaling pathway analyses of the DEGs were performed using DAVID. The upregulated DEGs were mainly enriched for the BP terms cell division, mitotic nuclear division and G2/M transition of mitotic cell cycle (Fig. 2A), while downregulated DEGs were significantly associated with lipid metabolic process, cholesterol homeostasis, and glucose metabolic process (Fig. 2B). Additionally, seven KEGG pathways were identified for the upregulated genes, including the p53 signaling pathway, cell cycle and extracellular matrix (ECM)-receptor interaction (Fig. 2C). The peroxisome proliferator-activated receptor (PPAR) signaling pathway, AMP-activated protein kinase (AMPK) signaling pathway and proximal tubule bicarbonate reclamation were associated with the downregulated DEGs (Fig. 2D). The detailed results are presented in Table I.

**PPI network analysis and the selection of NUF2.** Protein interactions often play important roles in cancer progression. A PPI network analysis was performed using the STRING database and Cytoscape. The PPI network was constructed...
Table I. Significantly enriched GO biological process terms and KEGG pathways.

### A. Upregulated

| Terms                              | Description                                                | Number of genes | P-value          |
|------------------------------------|------------------------------------------------------------|-----------------|-----------------|
| GO:0051301 Cell division           |                                                            | 18              | 1.02x10^{-14}   |
| GO:0007067 Mitotic nuclear division|                                                            | 16              | 1.90x10^{-14}   |
| GO:0000086 G2/M transition of mitotic cell cycle |                                | 9               | 4.82x10^{-06}   |
| GO:0031145 Anaphase-promoting complex-dependent catabolic process |                                | 7               | 5.40x10^{-07}   |
| GO:0042787 Protein ubiquitination involved in Ubiquitin-dependent protein catabolic process |                                | 8               | 1.83x10^{-06}   |
| GO:0007062 Sister chromatid cohesion |                                | 7               | 2.58x10^{-06}   |
| GO:0030574 Collagen catabolic process |                                | 6               | 4.40x10^{-06}   |
| GO:0008283 Cell proliferation      |                                                            | 9               | 7.20x10^{-05}   |
| GO:0051439 Regulation of ubiquitin-protein ligase activity involved in mitotic cell cycle |                                | 4               | 8.86x10^{-05}   |
| GO:0035978 Endodermal cell differentiation |                                | 4               | 1.45x10^{-04}   |

| KEGG pathways                      |                                                            |                |                 |
|------------------------------------|------------------------------------------------------------|----------------|-----------------|
| hsa04110 Cell cycle                |                                                            | 8              | 1.17x10^{-06}   |
| hsa04512 ECM-receptor interaction  |                                                            | 7              | 2.33x10^{-06}   |
| hsa04115 p53 signaling pathway     |                                                            | 5              | 0.000237        |
| hsa04114 Oocyte meiosis            |                                                            | 5              | 0.001499        |
| hsa04510 Focal adhesion            |                                                            | 5              | 0.014347        |
| hsa04151 PI3K-Akt signaling pathway |                                                            | 6              | 0.019962        |
| hsa05200 Pathways in cancer        |                                                            | 6              | 0.032832        |

### B. Downregulated

| Terms                              | Description                                                | Number of genes | P-value          |
|------------------------------------|------------------------------------------------------------|-----------------|-----------------|
| GO:0006629 Lipid metabolic process |                                                            | 9               | 1.32x10^{-05}   |
| GO:0019915 Lipid storage           |                                                            | 5               | 2.06x10^{-05}   |
| GO:0048662 Negative regulation of smooth muscle cell proliferation |                                | 5               | 4.49x10^{-05}   |
| GO:0042632 Cholesterol homeostasis |                                                            | 6               | 7.95x10^{-05}   |
| GO:0008217 Regulation of blood pressure |                                | 6               | 8.56x10^{-05}   |
| GO:0006006 Glucose metabolic process |                                | 6               | 9.90x10^{-05}   |
| GO:0019433 Triglyceride catabolic process |                                | 4               | 6.61x10^{-04}   |
| GO:0042593 Glucose homeostasis      |                                                            | 6               | 6.74x10^{-04}   |
| GO:0019432 Triglyceride biosynthetic process |                                | 4               | 7.44x10^{-04}   |
| GO:0001523 Retinoid metabolic process |                                | 5               | 8.33x10^{-04}   |

| KEGG pathways                      |                                                            |                |                 |
|------------------------------------|------------------------------------------------------------|----------------|-----------------|
| hsa03320 PPAR signaling pathway    |                                                            | 9              | 1.14x10^{-07}   |
| hsa04152 AMPK signaling pathway    |                                                            | 7              | 7.34x10^{-04}   |
| hsa04964 Proximal tubule bicarbonate reclamation |                                | 4              | 0.001072        |
| hsa04923 Regulation of lipolysis in adipocytes |                                | 5              | 0.001523        |
| hsa00982 Drug metabolism-cytochrome P450 |                                | 5              | 0.003116        |
| hsa04920 Adipocytokine signaling pathway |                                | 5              | 0.003462        |
| hsa00350 Tyrosine metabolism       |                                                            | 4              | 0.00367         |
| hsa00561 Glycerolipid metabolism   |                                                            | 4              | 0.014956        |
| hsa00980 Metabolism of xenobiotics by cytochrome P450 |                                | 4              | 0.028412        |
| hsa05205 Proteoglycans in cancer   |                                                            | 6              | 0.033006        |

KEGG, Kyoto Encyclopedia of genes and genomes; GO, Gene Ontology.
using 149 DEGs (57 upregulated and 92 downregulated DEGs) with combined scores ≥0.4, and contained 149 nodes and 930 edges (Fig. 3A). A total of two functional modules were identified using the MCODE plugin. Module 1 consisted of 35 nodes and 573 edges including NUF2, TOP2A, ASPM, and CCNB1 (Fig. 3B). Module 2 included 21 nodes and 104 edges including COL1A1, MMP1, MMP9, and LPL (Fig. 3C). Based on the degree of importance, module 1 was chosen for further analysis.

The 35 genes in module 1 were ranked based on log |FC| values in the TCGA database and selected the top 10 hub genes for further analysis. The expression levels of the 10 hub genes in the BC tissues were >10-fold (log |FC|≥3.42) increased compared with those in the normal breast tissues. Through literature mining, it was identified that UBE2C, ASPM, BIRC5, TOP2A, KIF20A, CEP55, TPX2, NEK2, and ANLN, but not NUF2, have been reported extensively in BC-related studies. Therefore, NUF2 was selected as the focus of subsequent analyses.

NUF2 expression in BC. The expression of NUF2 mRNA in the BC tissues was evaluated using Oncomine (https://www.oncomine.org/) (30). The results indicated that the NUF2 expression level is significantly increased in the BC tissues.
compared with in the normal breast tissues (P<0.01; Fig. 4A-C). To further verify these results, 42 pairs of BC tissues and adjacent tissues were analyzed by RT-qPCR. Consistent with the results of the database analysis, the expression of NUF2 mRNA in the BC tissues was significantly increased (P<0.001) compared with in the adjacent tissues (Fig. 4D).

**Association of NUF2 with clinical pathological characteristics and survival of patients with BC.** To further validate the clinical value of NUF2, the association between its expression and the clinical pathological characteristics of the 42 patients with BC recruited from Shaoxing People’s Hospital were assessed. The expression of NUF2 was only significantly associated with age (P<0.05). Using the data in the TCGA database (contains data on 1,090 patients with BC), NUF2 expression was found to be significantly associated with age (P<0.001), estrogen receptor (ER) status (P<0.001), progesterone receptor (PR) status (P<0.001), histological type (P<0.001), TNM stage (P<0.05), and molecular subtype (P<0.001). The results are shown in Tables II and III.

Furthermore, the clinical pathological characteristics that have multiple groups (>2) need a post hoc test to determine exactly what groups exhibit a difference. Therefore, Bonferroni’s post hoc test was used for pairwise comparison. The results showed that NUF2 expression is statistically different between TNM stage 1 and 2 (P<0.01), and tumor stage T1 and T2 (P<0.001). In terms of molecular subtype, NUF2 expression was also significantly different in all pairwise comparisons (P<0.001) except between luminal A and normal-like, and luminal B and basal-like. The detailed results are shown in Tables IV and V.

Furthermore, to elucidate the correlation between the expression of NUF2 and patient survival, 5 GEO datasets were used. Samples from each dataset were reclassified into four subsets (X1, X2, X3 and X4) according to the quartile of NUF2 expression. The X1 subset was set with the lowest expression as the reference to calculate the hazard ratio (HR). Each dataset was analyzed by Kaplan-Meier analysis and Cox proportional hazard analysis. High NUF2 expression was associated with shorter overall survival (OS) and progression-free survival (PFS) compared with low NUF2 expression in the GSE1456 dataset (Fig. 5A and B) and NKI dataset (Fig. 5C and D). Similar results were obtained in the GEO pooled analysis, as shown in Fig. 5E and F. Based on a further GEO pooled analysis, it was demonstrated that NUF2 expression levels are significantly associated with poor OS.
and PFS in both ER-positive (P<0.01; Fig. 6A and B) and ER-negative (P<0.01; Fig. 6C and D) BC, and the association is more obvious in ER-positive BC. The results of Cox proportional hazards analysis are shown in Table VII. In addition, the results of the Bonferroni’s post hoc tests used to compare the Kaplan-Meier survival curves corresponding to >2 groups are shown in Table VI. These results indicated that NUF2 might be a prognostic factor for BC.

**Discussion**

BC is the most common malignant tumor in women, accounting for 25% of female tumors (3). Despite advanced treatment techniques, BC remains the leading cause of death among women (31). Therefore, there is a pressing need for more effective molecular biomarkers to prevent, diagnose, and treat BC. With the development of microarray technology, hundreds of molecules have been found to be abnormally expressed during breast carcinogenesis and progression. The TCGA and GEO databases provide a large number of publicly available microarray datasets for biomarker identification.

In this study, 190 DEGs with the same expression trends were identified in four datasets, and a GO BP analysis showed that the upregulated DEGs were mainly enriched in the biological processes of cell division, mitotic nuclear division, and G2/M transition of mitotic cell cycle, while the downregulated DEGs
were related to lipid metabolic process, cholesterol homeostasis, and glucose metabolic process. Cell division and cell cycle are the basic processes in cell proliferation, and their abnormalities contribute to carcinogenesis and tumor progression (32). Furthermore, the activation of key regulators of lipid and cholesterol metabolism drives the estrogen-independent growth of invasive lobular breast carcinoma cells (33). A KEGG signaling pathway analysis of the DEGs in this study revealed

Table III. Association of NUF2 with clinical pathological characteristics of breast cancer patients derived from TCGA database.

| Pathological characteristics | Number of patients (%) | Low (%) | High (%) | P-value<sup>a</sup> |
|------------------------------|------------------------|---------|----------|---------------------|
| **Age**                     |                        |         |          |                     |
| <60                         | 579 (53.2)             | 255 (44.0) | 324 (66.0) | <0.001              |
| ≥60                         | 510 (46.8)             | 289 (56.7) | 221 (43.3) |                     |
| **HER2 status**             |                        |         |          | 0.296               |
| Positive                    | 90 (21.4)              | 39 (43.3) | 51 (56.7) |                     |
| Negative                    | 331 (78.6)             | 164 (49.5) | 167 (50.5) |                     |
| **ER status**               |                        |         |          | <0.001              |
| Positive                    | 803 (77.2)             | 467 (58.2) | 336 (41.8) |                     |
| Negative                    | 237 (22.8)             | 60 (25.3) | 177 (74.7) |                     |
| **PR status**               |                        |         |          | <0.001              |
| Positive                    | 694 (66.9)             | 415 (60.0) | 279 (40.0) |                     |
| Negative                    | 343 (33.1)             | 110 (32.1) | 233 (67.9) |                     |
| **Histology type**          |                        |         |          | <0.001              |
| IDC                         | 779 (79.3)             | 326 (41.8) | 453 (58.2) |                     |
| ILC                         | 203 (20.7)             | 156 (76.8) | 47 (23.2)  |                     |
| **TNM stage**               |                        |         |          | 0.032               |
| 1                           | 181 (17.0)             | 108 (59.7) | 73 (40.3)  |                     |
| 2                           | 619 (58.0)             | 293 (47.3) | 326 (52.7) |                     |
| 3                           | 247 (23.1)             | 119 (48.2) | 128 (51.8) |                     |
| 4                           | 20 (1.9)               | 10 (50.0)  | 10 (50.0)  |                     |
| **Tumor stage**             |                        |         |          | <0.001              |
| T1                          | 279 (25.7)             | 171 (61.3) | 108 (38.7) |                     |
| T2                          | 631 (58.0)             | 283 (44.8) | 348 (55.2) |                     |
| T3                          | 137 (12.6)             | 71 (51.8)  | 66 (48.2)  |                     |
| T4                          | 40 (3.7)               | 19 (47.5)  | 21 (52.5)  |                     |
| **Lymph node stage**        |                        |         |          | 0.095               |
| N0                          | 514 (48.0)             | 255 (49.6) | 259 (50.4) |                     |
| N1                          | 360 (33.6)             | 185 (51.4) | 175 (48.6) |                     |
| N2                          | 120 (11.2)             | 48 (40.0)  | 72 (60.0)  |                     |
| N3                          | 76 (7.2)               | 43 (56.6)  | 33 (43.4)  |                     |
| **Metastasis stage**        |                        |         |          | 0.82                |
| M0                          | 906 (97.6)             | 434 (47.9) | 472 (52.1) |                     |
| M1                          | 22 (2.4)               | 10 (45.5)  | 12 (54.5)  |                     |
| **Molecular subtype**       |                        |         |          | <0.001              |
| Luminal A                   | 419 (0.5)              | 305 (72.8) | 114 (27.2) |                     |
| Luminal B                   | 190 (0.23)             | 26 (13.7)  | 164 (86.3) |                     |
| HER2+                       | 67 (0.08)              | 26 (38.8)  | 41 (61.2)  |                     |
| Basal-like                  | 139 (0.16)             | 14 (10.1)  | 125 (89.9) |                     |
| Normal-like                 | 23 (0.03)              | 21 (91.3)  | 2 (8.7)    |                     |

<sup>a</sup><sup>χ</sup><sup>2</sup> tests were used for comparisons between groups. TNM, tumor node metastasis; ER, estrogen receptor; PR, progesterone receptor; IDC, infiltrating ductal carcinoma; ILC, infiltrating lobular carcinoma; NUF2, NDC80 kinetochore complex component.
the importance of the cell cycle, ECM-receptor interaction, PPAR signaling pathway, and AMPK signaling pathway in BC. Previous studies have reported that ECM could regulate tissue homeostasis, and its dysregulation could promote tumor

Table IV. Comparison of clinical pathological characteristics of breast cancer patients from Shaoxing People's Hospital among multiple groups.

| Pathological characteristics | Pairwise comparisons (P-values) |
|-----------------------------|-------------------------------|
| TNM stage                   |                               |
| 1                           | N/A                           |
| 2                           | 0.556                         |
| 3                           | 0.620                         |
| Tumor stage                 |                               |
| T1                          | N/A                           |
| T2                          | 0.278                         |
| T3                          | 1.000                         |
| Lymph node stage            |                               |
| N0                          | N/A                           |
| N1                          | 1.000                         |
| N2                          | 1.000                         |
| N3                          | 1.000                         |

P<0.05/N was considered statistically significant, where N was the number of pairwise comparisons. TNM, tumor node metastasis.

Table V. Comparison of clinical pathological characteristics of breast cancer patients in The Cancer Genome Atlas database among multiple groups.

| Pathological characteristics | Pairwise comparisons (P-values) |
|-----------------------------|-------------------------------|
| TNM stage                   |                               |
| 1                           | N/A                           |
| 2                           | 0.004<sup>a</sup>             |
| 3                           | 0.019                         |
| 4                           | 0.405                         |
| Tumor stage                 |                               |
| T1                          | N/A                           |
| T2                          | 0.000<sup>a</sup>             |
| T3                          | 0.066                         |
| T4                          | 0.097                         |
| Lymph node stage            |                               |
| N0                          | N/A                           |
| N1                          | 0.605                         |
| N2                          | 0.058                         |
| N3                          | 0.257                         |
| Molecular subtype           |                               |
| Luminal A                   | N/A                           |
| Luminal B                   | 0.000<sup>a</sup>             |
| HER2+                       | 0.000<sup>a</sup>             |
| Basal-like                  | 0.000<sup>a</sup>             |
| Normal-like                 | 0.049                         |

P<0.05/N was considered statistically significant, where N was the number of pairwise comparisons. <sup>a</sup>P<0.05/N. TNM, tumor, node and metastasis.
Table VI. Comparison of Kaplan-Meier curves among multiple groups.

| Datasets          | Subsets | X1         | X2         | X3         | X4         |
|-------------------|---------|------------|------------|------------|------------|
|                   | GSE1456 OS | X1         | X2         | X3         | X4         |
|                   |         | N/A        | 0.088      | 0.008*     | 0.003*     |
|                   |         | X2         | 0.088      | N/A        | 0.356      | 0.253      |
|                   |         | X3         | 0.008*     | 0.356      | N/A        | 0.851      |
|                   |         | X4         | 0.003*     | 0.253      | 0.851      | N/A        |
|                   | GSE1456 PFS | X1         | X2         | X3         | X4         |
|                   |         | N/A        | 0.259      | 0.002*     | 0.004*     |
|                   |         | X2         | 0.259      | N/A        | 0.044      | 0.068      |
|                   |         | X3         | 0.002*     | 0.044      | N/A        | 0.800      |
|                   |         | X4         | 0.004*     | 0.068      | 0.800      | N/A        |
|                   | NKI OS  | X1         | X2         | X3         | X4         |
|                   |         | N/A        | 0.036      | 0.000*     | 0.000*     |
|                   |         | X2         | 0.036      | N/A        | 0.113      | 0.041      |
|                   |         | X3         | 0.000*     | 0.113      | N/A        | 0.663      |
|                   |         | X4         | 0.000*     | 0.041      | 0.663      | N/A        |
|                   | NKI PFS | X1         | X2         | X3         | X4         |
|                   |         | N/A        | 0.361      | 0.002*     | 0.003*     |
|                   |         | X2         | 0.361      | N/A        | 0.034      | 0.057      |
|                   |         | X3         | 0.002*     | 0.034      | N/A        | 0.840      |
|                   |         | X4         | 0.003*     | 0.057      | 0.840      | N/A        |
|                   | GEO pooled OS | X1         | X2         | X3         | X4         |
|                   |         | N/A        | 0.007*     | 0.000*     | 0.000*     |
|                   |         | X2         | 0.007*     | N/A        | 0.063      | 0.017      |
|                   |         | X3         | 0.000*     | 0.063      | N/A        | 0.630      |
|                   |         | X4         | 0.000*     | 0.017      | 0.630      | N/A        |
|                   | GEO pooled PFS | X1         | X2         | X3         | X4         |
|                   |         | N/A        | 0.027      | 0.000*     | 0.000*     |
|                   |         | X2         | 0.027      | N/A        | 0.010      | 0.002*     |
|                   |         | X3         | 0.000*     | 0.010      | N/A        | 0.639      |
|                   |         | X4         | 0.000*     | 0.002*     | 0.639      | N/A        |
|                   | GEO pooled ER(+) OS | X1         | X2         | X3         | X4         |
|                   |         | N/A        | 0.006*     | 0.000*     | 0.000*     |
|                   |         | X2         | 0.006*     | N/A        | 0.307      | 0.314      |
|                   |         | X3         | 0.000*     | 0.307      | N/A        | 0.925      |
|                   |         | X4         | 0.000*     | 0.314      | 0.925      | N/A        |
|                   | GEO pooled ER(+) PFS | X1         | X2         | X3         | X4         |
|                   |         | N/A        | 0.197      | 0.000*     | 0.001*     |
|                   |         | X2         | 0.197      | N/A        | 0.022      | 0.065      |
|                   |         | X3         | 0.000*     | 0.022      | N/A        | 0.730      |
|                   |         | X4         | 0.001*     | 0.065      | 0.730      | N/A        |
|                   | GEO pooled ER(-) OS | X1         | X2         | X3         | X4         |
|                   |         | N/A        | 0.495      | 0.031      | 0.016      |
|                   |         | X2         | 0.495      | N/A        | 0.079      | 0.028      |
|                   |         | X3         | 0.031      | 0.079      | N/A        | 0.762      |
|                   |         | X4         | 0.016      | 0.028      | 0.762      | N/A        |
|                   | GEO pooled ER(-) PFS | X1         | X2         | X3         | X4         |
|                   |         | N/A        | 0.679      | 0.029      | 0.055      |
|                   |         | X2         | 0.679      | N/A        | 0.095      | 0.106      |
|                   |         | X3         | 0.029      | 0.095      | N/A        | 0.834      |
|                   |         | X4         | 0.055      | 0.106      | 0.834      | N/A        |

P<0.05/N was considered statistically significant, where N was the number of pairwise comparisons. *P<0.05/N. OS, overall survival; PFS, progression-free survival; ER, estrogen receptor.
progression by affecting adhesion, migration, differentiation, proliferation, and apoptosis of tumor cells (34,35). In addition, an increase in the rigidity of the ECM due to the local accumulation of crosslinked collagen matrix is associated with cancer progression (36). Yao et al (37) found that the PPAR signaling pathway is involved in breast carcinogenesis. Song et al (38) suggested that activation of the AMPK signaling pathway may be beneficial for the promotion of tumor necrosis factor-related apoptosis-inducing ligand-mediated anti-Bc treatment. A PPI network was constructed in this study and two functional modules were identified. According to the MCODE scores, which represent importance, module 1 was found to play a major role in the PPI network. By combining the log |Fc| values of the dEGs in the TCGA database and literature mining, NUF2 was selected for further research as a key gene in Bc.

Although Xiang et al (12) found that NUF2 is upregulated in Bc, using cDNA microarray data of Bc patients, further experiments have not been performed to verify this finding. In this study, by Oncomine analysis and RT-qPCR assay, it was verified that NUF2 is overexpressed in BC tissues, further confirming the results obtained by data mining. Shiraishi et al (39) found that NUF2 expression is significantly associated with prostate cancer recurrence, and patients with high NUF2 expression have significantly shorter survival times without biochemical recurrence. Hu et al (10) showed that the overexpression of NUF2 could be related to poor prognosis in pancreatic cancer. Zhang et al (13) found that NUF2 expression has prognostic values for Bc patients, by simply using the Bc-GenExMiner online analysis tool. However, further analysis has not been conducted. To this end, the present study analyzed the precise roles and underlying molecular mechanisms of NUF2 in Bc. By stratified analysis and pooled analysis of five GEO datasets, it was found that patients with BC and high NUF2 expression had worse prognosis than patients with low NUF2 expression in both ER-positive and ER-negative BC. Using clinical data for 42 patients, it was demonstrated that NUF2 expression was only associated with

Figure 5. Role of NUF2 in the prognosis of BC patients. (A) NUF2 expression was significantly associated with OS of BC patients in the GSE1456 dataset, (C) NKI dataset, and (E) GEO pooled dataset. NUF2 expression was significantly associated with PFS of BC patients in the (B) GSE1456 dataset, (D) NKI dataset and (F) GEO pooled dataset. GEO, gene expression omnibus; BC, breast cancer; OS, overall survival; PFS, progression free survival; ER, estrogen receptor; PR, progesterone receptor; NUF2, Ndc80 kinetochore complex component.
Figure 6. Prognostic value of NUF2 in ER-positive and ER-negative BC. Association of NUF2 expression with (A) OS and (B) PFS in ER-positive BC in the GEO pooled dataset. Association of NUF2 expression with (C) OS and (D) PFS in ER-negative BC in the GEO pooled dataset. GEO, gene expression omnibus; BC, breast cancer; OS, overall survival; PFS, progression-free survival; ER, estrogen receptor; PR, progesterone receptor; NUF2, Ndc80 kinetochore complex component.

Figure 7. Signaling pathways associated with NDC80 kinetochore complex component predicted by gene set enrichment analysis. (A) Cell cycle (P=0.000; FDR=0.000; ES=0.72). (B) DNA replication (P<0.001; FDR<0.001; ES=0.84). (C) p53 signaling pathway (P=0.000; FDR=0.004; ES=0.51). FDR, false discovery rate; KEGG, Kyoto Encyclopedia of genes and genomes; ES, enrichment score.

age. Small sample size, erroneous tissue sampling, RNA degradation, and fluctuating efficiency of RT-qPCR may affect the results of the analysis. This hypothesis can be tested through the following methods: Increasing sample size, determining the type of tissue with pathological examination, detecting RNA degradation by RNA electrophoresis, and verifying amplification efficiency of RT-qPCR by the standard curve method. The lack of additional experiments to test these possibilities is a limitation to the present study. Therefore, the relationship was further analyzed using clinical data for patients with BC in the TCGA database. The expression of NUF2 was positively correlated with tumor stage and negatively correlated with ER status, consistent with the results from a number of studies showing that advanced tumors and ER-negative tumors are probably related to poor survival (40-42), suggesting that NUF2 plays an important role in tumor progression and prognosis. To elucidate the molecular mechanisms by which NUF2 mediates breast carcinogenesis and progression, GSEA was performed in this study. The results revealed that NUF2 is involved in cell cycle-related pathways.

In conclusion, the present data analysis and RT-qPCR validation indicated that NUF2 is highly expressed in BC and is associated with its multiple pathological features and prognosis.
NUF2 is therefore a potential therapeutic target and prognostic indicator of BC. However, this study had several limitations. First, only mRNA data were obtained from the databases and RT-qPCR, and this single-gene-level analysis cannot fully
capture the expression of NUF2 in BC. Second, experimental validation of the results was not performed. Therefore, further research is required to determine the roles of NUF2 in BC.

NUF2 is overexpressed in BC and is associated with its multiple pathological features. More importantly, NUF2 may play an important role in predicting the clinical outcomes of different BC subgroups.

Acknowledgments

Not applicable.

Funding

This study was supported by the Public Technology Research Project of Zhejiang Province (grant no. LGF18H200006) and the Medicines Health Technology Plan Project of Zhejiang Province (grant no. 2018PY073).

Availability of data and materials

The datasets analyzed during the current study are available in the TCGA (https://cancergenome.nih.gov/) and GEO (https://www.ncbi.nlm.nih.gov/geo/) databases.

Authors' contributions

WJX, YNW, and XJD participated in the study design. WJX, YNW, and XPD contributed to data collection and analysis. WJX, YZW, and SML were involved in the collection of samples and RT-qPCR. All authors were involved in the writing of the article. XJD critically reviewed the manuscript. All authors have read and approved the final manuscript.

Ethics approval and consent to participate

The present study was approved and supervised by the Ethics Committee of Shaoxing People's Hospital (Shaoxing, China). Written informed consent was obtained from all participants.

Patient consent for publication

Not applicable.

Competing interests

The authors declare that they have no competing interests.

References

1. Siegel RL, Miller KD and Jemal A: Cancer Statistics, 2017. CA Cancer J Clin 67: 7-30, 2017.
2. Donepudi MS, Kondapalli K,amos SJ and Venkateshan P: Breast cancer statistics and markers. J Cancer Res Ther 10: 506-511, 2014.
3. Torre LA, Bray F, Siegel RL, Ferlay J, Lortet-Tieulent J and Jemal A: Global cancer statistics, 2012. CA Cancer J Clin 65: 87-108, 2015.
4. Tan W, Yang M, Yang H, Zhou F and Shen W: Predicting the response to neoadjuvant therapy for early-stage breast cancer: Tumor-, blood-, and imaging-related biomarkers. Cancer Manag Res 10: 4333-4347, 2018.
5. Salem Y, Yacov N, Propheta-Meirson O, Breitbart E and Mendel I: Newly characterized motile sperm domain-containing protein 2 promotes human breast cancer metastasis. Int J Cancer 144: 125-135, 2019.
6. Kinoshita MD, Nassar A, Siddiqui MT and Cohen C: Estrogen receptor (ER), progesterone receptor (PR), and HER2 expression pre- and post-neoadjuvant chemotherapy in primary breast carcinoma: A single institutional experience. Int J Clin Exp Pathol 5: 530-536, 2012.
7. Piao J, Chen L, Jin T, Xu M, Quan C and Lin Z: Pia1p affects breast cancer cell growth and represents a novel prognostic biomarker. Hum Pathol 73: 33-40, 2018.
8. Sandoo A, Kitas GD and Carmichael AR: Breast cancer therapy and cardiovascular risk: Focus on trastuzumab. Vasc Health Risk Manag 11: 223-228, 2015.
9. Nabetani A, Kojii T, Tsutsumi C, Haraguchi T and Hiraoka Y: A conserved protein, Nuf2, is implicated in connecting the centromere to the spindle during chromosome segregation: A link between the kinetochoore function and the spindle checkpoint. Chromosoma 110: 322-334, 2001.
10. Fu P, Chen X, Sun J, Bie P and Zhang LD: siRNA-mediated knockdown against NUF2 suppresses pancreatic cancer proliferation in vitro and in vivo. Biosci Rep 35: pii: e00170, 2015.
11. Fu HL and Shao L: Silencing of NUF2 inhibits proliferation of human osteosarcoma Saos-2 cells. Eur Rev Med Pharmacol Sci 20: 1071-1079, 2016.
12. Xiang YJ, Fu Q, Ma ZB, Gao DZ, Zhang Q, Li YY, Li L, Liu L, Ye CM, Yu ZG and Guo MM: Screening for candidate genes related to breast cancer with cDNA microarray analysis. Chronic Dis Transl Med 1: 65-72, 2015.
13. Zhang W, Mao J, Zhu W, Jain AK, Liu K, Brown JB and Karpen GH: Centromere and kinetochore gene misexpression predicts cancer patient survival and response to radiotherapy and chemotherapy. Nat Commun 7: 12619, 2016.
14. Clarke C, Madden SF, Doolan P, Ahern ST, Joyce H, O’Driscoll L, Gallagher WM, Hennessy BT, Moriarty M, Crowell J, et al: Correlating transcriptional networks to breast cancer survival: A large-scale coexpression analysis. Carcinogenesis 34: 2300-2308, 2013.
15. Grusso T, Miele N, Cardon M, Bourachot B, Kieffer Y, Devun F, Dubois T, Dutreix M, Vincent-Salomon A, Miller KM and Mechat-Grigoriou F: Chronic oxidative stress promotes H2AX protein degradation and enhances chemosensitivity in breast cancer patients. EMBO Mol Med 8: 527-549, 2016.
16. Maubant S, Tesson B, Maire V, Ye M, Rigaili G, Gentien D, Cruzaegui F, Tucker GC, Roman-Roman S and Dubois T: Transcriptome analysis of M1a-treated triple-negative breast cancer cells. PLoS One 10: e0122333, 2015.
17. Lee GY, Haverty PM, Li L, Kljavin NM, Bouron R, Lee J, Stern H, Modrusan Z, Seshaigari S, Zhang Z, et al: Comparative oncogenomics identifies PSMB4 and SHMT2 as potential cancer driver genes. Cancer Res 74: 3114-3126, 2014.
18. Tomczak K, Czerwinski P and Wizerowicz M: The cancer genome atlas (TCGA): An immeasurable source of knowledge. Contemp Oncol (Pozn) 19: A68-A77, 2015.
19. Pawitan Y, Bjohe J, Amler L, Borg AL, Egyhazi S, Hall P, Hans L, Holmberg L, Huang F, Klaar S, et al: Gene expression profiling spares early breast cancer patients from adjuvant therapy: Derived and validated in two population-based cohorts. Breast Cancer Res 7: R953-R964, 2005.
20. Buffa FM, Camps C, Winchester L, Snell CE, Gee HE, Sheldon H, Taylor M, Harris AL and Ragoussis J: microRNA-associated prognostic pathways and potential therapeutic targets identified by integrated mRNA and microRNA expression profiling in breast cancer. Cancer Res 71: 5635-5645, 2011.
21. van de Vijver MJ, He YD, van't Veer LJ, Dai H, Hart AA, Voskuil DW, Schreiber GJ, Peterse JL, Roberts C, Marton MJ, et al: A gene-expression signature as a predictor of survival in breast cancer. N Engl J Med 347: 1999-2009, 2002.
22. Jones SB, dePrimo SE, Whitehead ML and Brooks JD: Resveratrol-induced gene expression profiles in human prostate cancer cells. PLoS One 10: e0122333, 2015.
23. Kao KJ, Chang KM, Hsu HC and Huang AT: correlation activity in breast cancer patients. EMBO Mol Med 8: 527-549, 2016.
24. Ashburner M, Ball CA, Blake JA, Botstein D, Butler H, Cherry JM, Davis AP, Dolinski K, Dwight SS, Eppig JT, et al: Gene ontology: Tool for the unification of biology. The gene ontology consortium. Nat Genet 25: 25-29, 2000.
25. Kanehisa M and Goto S: KEGG: Kyoto encyclopedia of genes and genomes. Nucleic Acids Res 28: 27-30, 2000.
26. Dennis G Jr, Sherman BT, Hosack DA, Yang J, Gao W, Lane HC and Lempicki RA: DAVID: Database for annotation, visualization, and integrated discovery. Genome Biol 4: P3, 2003.
27. Doerks T, Copley RR, Schultz J, Ponting CP and Bork P: Systematic identification of novel protein domain families associated with nuclear functions. Genome Res 12: 47-56, 2002.
28. Subramanian A, Tamayo P, Mootha VK, Mukherjee S, Ebert BL, Gillette MA, Paulovich A, Pomeroy SL, Golub TR, Lander ES and Mesirov JP: Gene set enrichment analysis: A knowledge-based approach for interpreting genome-wide expression profiles. Proc Natl Acad Sci USA 102: 15545-15550, 2005.
29. Livak KJ and Schmittgen TD: Analysis of relative gene expression data using real-time quantitative PCR and the 2(-Delta Delta Ct) method. Methods 25: 402-408, 2001.
30. Rhodes DR, Kalyana-Sundaram S, Mahavisno V, Varambally R, Yu J, Briggs BB, Barrette TR, Anstet MJ, Kincaid-Beal C, Kulkarni P, et al: Oncomine 3.0: Genes, pathways, and networks in a collection of 18,000 cancer gene expression profiles. Neoplasia 9: 166-180, 2007.
31. Xue D, Cheng P, Han M, Liu X, Xue L, Ye C, Wang K and Huang J: An integrated bioinformatic analysis to evaluate the role of KIF4A as a prognostic biomarker for breast cancer. OncoTargets Ther 11: 4755-4768, 2018.
32. Du T, Sikora MJ, Levine KM, Tasdemir N, Riggins RB, Wendell SG, Van Houten B and Oesterreich S: Key regulators of lipid metabolism drive endocrine resistance in invasive lobular breast cancer. Breast Cancer Res 20: 106, 2018.
33. Kaushik N, Kim S, Suh Y and Lee SJ: Proinvasive extracellular matrix remodeling for tumor progression. Arch Pharm Res 42: 40-47, 2019.
34. Walker C, Mojares E and Del Río Hernández A: Role of extracellular matrix in development and cancer progression. Int J Mol Sci 19: pii, E3028, 2018.
35. Tilghman RW, Cowan CR, Mih JD, Koryakina Y, Gioeli D, Slack-Davis JK, Blackman BR, Tschumperlin DJ and Parsons JT: Matrix rigidity regulates cancer cell growth and cellular phenotype. PLoS One 5: e12905, 2010.
36. Yao PL, Morales JL, Zhu B, Kang BH, Gonzalez FJ and Peters JM: Activation of peroxisome proliferator-activated receptor-β/δ (PPAR-β/δ) inhibits human breast cancer cell line tumorigenicity. Mol Cancer Ther 13: 1008-1017, 2014.
37. Song W, Yan CY, Zhou QQ and Zhen LL: Galangin potentiates human breast cancer to apoptosis induced by TRAIL through activating AMPK. Biomed Pharmacother 89: 845-856, 2017.
38. Shiraishi T, Terada N, Zeng Y, Suyama T, Luo J, Trock B, Kulkarni P and Getzenberg RH: Cancer/testis antigens as potential predictors of biochemical recurrence of prostate cancer following radical prostatectomy. J Transl Med 9: 153, 2011.
39. Bagaria SP, Ray PS, Sim MS, Ye X, Shamoni JK, Cui X and Giuliano AE: Personalizing breast cancer staging by the inclusion of ER, PR, and HER2. JAMA Surg 149: 125-129, 2014.
40. Ryu JM, Choe HJ, Kim I, Lee SK, Yu J, Kim JE, Kang BI, Lee JE, Nam SJ and Kim SW: Only estrogen receptor ‘positive’ is not enough to predict the prognosis of breast cancer. Breast Cancer Res Treat 172: 627-636, 2018.
41. Dunnwald LK, Rossing MA and Li CI: Hormone receptor status, tumor characteristics, and prognosis: A prospective cohort of breast cancer patients. Breast Cancer Res 9: R6, 2007.

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