Cell Division Cycle Protein (CDC20), a Promising Prognostic Biomarker of Breast Cancer

Lin Cheng¹,²,*, Yu-Zhou Huang¹,*, Wei-Xian Chen², Liang Shi¹, Zhi Li¹, Xu Zhang¹, Xin-Yuan Dai¹, Ji-Fu Wei³,*, Qiang Ding¹,*

1. Jiangsu Breast Disease Center, the First Affiliated Hospital with Nanjing Medical University, 300 Guangzhou Road, Nanjing 210029, China
2. Department of Breast Surgery, the Affiliated Changzhou No. 2 People's Hospital of Nanjing Medical University, 29 Xinglong Lane, Changzhou 213003, China
3. Research Division of Clinical Pharmacology, the First Affiliated Hospital with Nanjing Medical University, 300 Guangzhou Road, Nanjing 210029, China

* These authors contributed equally to this work.

'Corresponding author: Dr. Qiang Ding. Address: Jiangsu Breast Disease Center, the First Affiliated Hospital with Nanjing Medical University, 300 Guangzhou Road, Nanjing 210029, China. Tel: 0086-25-83673123, Email: dinqiang@njmu.edu.cn

Dr. Ji-Fu Wei, Address: Research Division of Clinical Pharmacology, the First Affiliated Hospital with Nanjing Medical University, 300 Guangzhou Road, Nanjing 210029, China, Tel: 0086-25-83673123, E-mail: weijifu@hotmail.com

Key Words: CDC20, Breast cancer, Bioinformatics Analysis, Prognosis
Abbreviations

CDC20: Cell division cycle protein:
ER: Estrogen receptor,
PR: Progesterone receptor
HER-2: Epidermal growth factor receptor-2
SBR: Scarff-Bloom-Richardson
NPI: Nottingham prognostic index
TPX2: Targeting protein for Xenopus kinesin-like protein 2
RFS: Relapse-free survival
OS: Overall survival
TCGA: The Cancer Genome Atlas

Running title

CHENG et al: CDC20 COULD BE NOVEL PROGNOSTIC BIOMARKER OF BREAST CANCER
Abstract

CDC20 has been observed to be expressed higher in various kinds of human cancers and was associated with poor prognosis. However, studies on role of CDC20 in breast cancer are seldom reported till now, most of which are not systematical and conclusive. The study was performed to analyze the expression pattern, potential function, and distinct prognostic effect of CDC20 in breast cancer using several online databases including Oncomine, bc-GenExMiner, PrognoScan, and UCSC Xena. To verify the results from databases, we compared the mRNA CDC20 expression in breast cancer tissues and adjacent normal tissues of patients by real-time PCR. We found that CDC20 was expressed higher in different types of breast cancer, comparing to normal tissues. Moreover, the patients with a more advanced stage of breast cancer tended to express higher level CDC20. CDC20 was expressed higher in breast cancer tissues than normal tissues from patients in our hospital, consist with the results from databases. Estrogen receptor (ER) and progesterone receptor (PR) status were negatively correlated with CDC20 level. Conversely, Scarff–Bloom–Richardson (SBR) grade, Nottingham prognostic index (NPI), epidermal growth factor receptor-2 (HER-2) status, basal-like status, and triple-negative status were positively related to CDC20 expression in breast cancer patients in respect to normal individuals. Higher CDC20 expression correlated with worse survival. Finally, a positive correlation between CDC20 and Targeting protein for Xenopus kinesin-like protein 2 (TPX2) expression was revealed. CDC20 could be considered as a potential predictive indicator for prognosis of breast cancer with co-expressed TPX2 gene.
Introduction

Breast cancer is the most common malignant tumor and remains a major cause of death in women [1]. With the development of treatment, including surgery, chemotherapy, radiotherapy, endocrine therapy and target therapy, both disease-free survival and overall survival of breast cancer have been significantly improved. However, after systemic therapies, there are still some patients died of breast cancer, especially for advanced breast cancer. Breast cancer is related with inactivation of a large number of tumor suppressor genes and oncogenes [2]. As we know, biomarkers are reported as surrogates of the clinical features for predicting outcomes. It is important to identify more effective, sensitive, and specific biomarkers for the prognosis of patients with breast cancer [3].

The cell-division cycle consists of a series of complex processes which are regulated by numbers of cell cycle regulatory proteins [4]. The cell division cycle protein 20 (CDC20), acting as a regulatory protein, is a target molecule in the cell-cycle checkpoint [5]. It is also a key E3 ligase, which can activate adenomatous polyposis coli (APC) [5]. In addition to regulating cell cycle, recent evidence has demonstrated that CDC20 also plays an important role in carcinogenesis and cancer progression, having the potential to become a promising therapeutic target [6]. CDC20 has been observed expressed higher in different kinds of human cancers and was associated with poor prognosis such as, oral squamous cell carcinoma [7], gastric cancer [8], urothelial bladder cancer[9], colorectal cancer [10], lung cancer [11], and pancreatic cancer [12].

Recently, CDC20 has been demonstrated to act as an oncogene in breast cancer progression [13], However, studies on role of CDC20 in breast cancer are seldom reported till now, most of which are not systematical and conclusive. Moreover, the CDC20 expression’s prognostic significance is also uncertain. Therefore, it is imperative to recognize that further study is necessary to determine the oncogenic role of CDC20 in breast tumorigenesis.

In the present study, we performed a deep bioinformatics analysis of the clinical parameters and survival data related to CDC20 in breast cancer patients using several online databases in order to evaluate the prognostic significance of CDC20 gene in breast cancer treatment. Moreover, we used 22 pairs of breast tissues from breast cancer patients in our hospital to compare the expression of CDC20 gene between the cancer tissues and normal tissues by real-time PCR.
Materials and methods

1. ONCOMINE data-mining analysis

ONCOMINE (www.oncomine.org), an online web-based cancer database for RNA and DNA sequences, was used to facilitate data-mining the transcriptional expressions of genes in 20 types of cancer [14]. Data used in this study from ONCOMINE was updated in April 2019. Transcriptional expressions of CDC20 in cancer samples were compared with those in normal individuals using Student’s t-test. Statistically significant values and fold change were demarcated as P-value ≤ 1E-4, and 2, respectively. Genes co-expressed with CDC20 were analyzed by using online Oncomine analysis tools.

2. UALCAN

UALCAN (http://ualcan.path.uab.edu/) is a user-friendly, interactive web resource for analyzing transcriptome data of cancers from The Cancer Genome Atlas (TCGA) [15]. Data from UALCAN was updated in March 2019. The mRNA expression of CDC20 between breast cancer and normal tissues as well as different stages cancer was also detected using the UALCAN web portal (TCGA level 3 data).

3. bc-GenExMiner v4.2

Breast Cancer Gene-Expression Miner v4.2 (bc-GenExMiner v4.2) is a statistical mining tool of incorporates 264 independent datasets and three classical mining functions: expression, prognosis, and correlation [16, 17]. Expression data were analyzed with clinical parameters such as age, nodal status, ER, PR, HER-2, Scarff-Bloom-Richardson (SBR) grade, and Nottingham prognostic index (NPI). Correlation of CDC20 and TPX2 was assessed using the correlation module.

3. PrognoScan

The PrognoScan (http://www.prognoscan.org/) is an online database for assessing the biological relationship between gene expression and survival data including overall survival, distant metastasis free survival, relapse free survival, and disease specific survival in breast cancer patients. The results base on a collection of publicly available cancer microarray datasets [18]. Data used in this study from PrognoScan was updated in April 2019. P-value, hazard ratio (HR), and 95% confidence intervals (CI) could be automatically calculated according to a certain gene expression. For statistical analysis and visualization, R packages (http://www.r-project.org) were used.
4. UCSC Xena

The UCSC Xena (http://xena.ucsc.edu/) is a popular genomics browser that provides visualization and integration for analyzing and viewing the public data hubs [19]. The heat map and correlation between CDC20 and TPX2 were generated by data mining in The Cancer Genome Atlas (TCGA) Breast Cancer using the UCSC Xena browser.

5. Breast tissue samples

22 pairs of breast tissue samples used in RT-PCR were obtained from the First Affiliated Hospital of Nanjing Medical University, China, between 2014 and 2016. The collection and use of the samples was reviewed and approved by the Institutional Ethics Committee of the First Affiliated Hospital of Nanjing Medical University.

6. RNA isolation and quantitative real-time PCR (RT-PCR) analysis

Total RNA was isolated using Trizol reagent (TaKaRa, Japan), and 1000 ng RNA was reverse-transcribed into cDNA using Primescript RT Reagent (TaKaRa, Japan). The RT-PCR was performed using FastStart Universal SYBR Green Master (Roche, Switzerland) in a RT-PCR instrument (Applied Biosystems, USA), and β-actin was used as endogenous control. The following PCR primers were used:

CDC20 forward, 5′-GCACAGTTCCGCTCGAGA3′
CDC20 reverse, 5′-CTGGATTTGCCAGGAGTTCGG3′
β-actin forward, 5′-GCTGTGCTATCCCTGTACGC3′
β-actin reverse, 5′-TGCCCTAGGCGAGCGGAAC3′

7. Statistical analysis

RT-PCR were repeated in triplicate, unless otherwise specified. The data were analyzed using the SPSS 20.0 software (Chicago, USA). We analyzed the statistical significance of the differences between groups using Students t-test, and a statistically significant difference was considered at the level of P < 0.05.
Results

Increased expression of CDC20 gene in breast cancer patients' tissues

Firstly, the expression of CDC20 gene in 20 types of cancer was measured and compared to normal tissues using the Oncomine online database (Figure 1A). We found that increased CDC20 (red) was observed in bladder cancer, brain and CNS cancer, cervical cancer, colorectal cancer, esophageal cancer, gastric cancer, head and neck cancer, liver cancer, lung cancer, lymphoma, ovarian cancer, pancreatic cancer, sarcoma, and especially breast cancer, whereas, decreased level of CDC20 (blue) was observed in leukemia and myeloma. Consistently, using UALCAN website, we also found that higher mRNA CDC20 was expressed in breast cancer tissues than in normal tissues (Figure 1B, P<0.05). Next, we focused on whether mRNA expression of CDC20 was related to cancer stage in individual patients. As shown in Figure 1C, the results indicated that patients with a more advanced stage of breast cancer tended to express higher levels CDC20.

To verify the results above, we further compared the mRNA CDC20 expression in breast cancer tissues and adjacent normal tissues of patients in our hospital and found that CDC20 was expressed higher in breast cancer tissues, consist with the results from databases (Figure 2, P<0.05).

Oncomine analysis also revealed that CDC20 was significantly expressed higher in medullary breast carcinoma, invasive ductal breast carcinoma, invasive lobular breast carcinoma, invasive breast carcinoma, invasive ductal and invasive lobular breast carcinoma, breast carcinoma, mucinous breast carcinoma, tubular breast carcinoma, intraductal cribriform breast adenocarcinoma, invasive ductal breast carcinoma, invasive breast carcinoma, mixed lobular and ductal breast carcinoma, invasive lobular breast carcinoma, ductal breast carcinoma with respect to normal individuals (Figure 3, Table 1).

CDC20 expression and clinical parameters of breast cancer patients

Using bc-GenExMiner v4.2 software, we implemented Welch’s test to compare the abnormal expression of CDC20 among different groups of patients according to clinical pathological features. For age criteria, CDC20 was significantly elevated in ≤51 year group with respect to >51 year group (Figure 4A, Table 2). As we know, the Scarff, Bloom & Richardson (SBR) is a histological grade that evaluates tubule formation, nuclear characteristics of pleomorphism, and mitotic index [20]. Based on tumor size, lymph node stage, and tumor grade, the Nottingham
Prognostic Index (NPI) is used to stratify patients into additional prognostic groups. The SBR grade and NPI index are two wonderful prognostic model for breast cancer [21]. More advanced SBR grade and NPI index were associated with higher CDC20 level (Figure 4B, C, Table 2). ER-positive or PR-positive breast cancer patients tended to express lower CDC20 gene compared with ER-negative or PR-negative patients (Figure 4D, E; Table 2). Patients with HER-2-negative status showed reduced expression of CDC20 than HER-2-positive patients (Figure 4F, Table 2). Regarding nodal status, there was no significant difference between positive and negative group (Figure 4G, Table 2). Moreover, CDC20 was significantly reduced in non-triple-negative and non-basal-like breast cancer patients compared to triple-negative and basal-like breast cancer patients (Figure 4H, I, Table 2).

**CDC20 expression and survival data of breast cancer patients**

Then, we investigated the prognostic value of CDC20 gene using the PrognoScan database. Breast cancer patients with lower expression of CDC20 (blue) significantly showed preferable distant metastasis free survival (Figure 5A, B, D, E, F, N, Table 3). Reduced CDC20 level (blue) was related to better relapse free survival (Figure 5C, H, M, Table 3) and the cases with increased CDC20 gene presented worse disease-free survival (Figure 5J, L, Table 3). Moreover, down-regulated CDC20 gene (blue) was strongly associated with better disease specific survival (Figure 5I, K, Table 3) and up-regulated CDC20 gene (red) was related to worse overall survival (Figure 5G, O, Table 3).

**Co-expression of CDC20 gene**

Finally, we investigated the co-expression of CDC20 gene using the Oncomine database. The co-expression profile of CDC20 was identified with a large cluster of 17,779 genes across 159 breast cancer samples (Figure 6A). Targeting protein for Xenopus kinesin-like protein 2 (TPX2) is a top correlated gene, which is a microtubule-associated protein and encoded by a gene located on human chromosome band 20q11.1 [22]. A positive correlation between CDC20 and TPX2 expression was revealed using bc-GenExMiner 4.2 (Figure 6C). Moreover, after analyzing breast cancer patient data in the TCGA database using the UCSC Xena web-based tool, again we confirmed a positive correlation between CDC20 and TPX2 expression, as shown in the heat map (Figure 6B, D). These results suggested that CDC20 might be closely related to the TPX2 signaling pathway in breast cancer.
Discussion

Cell division cycle 20 (CDC20) is a vital molecule which acts as an important role in the cell cycle and an activator of the anaphase-promoting complex (APC/C) [23]. Higher expression of CDC20 has been observed in a variety of human cancers and is correlated with poor prognosis [7, 24, 25]. However, the significance of CDC20 expression in the development and prognosis of breast cancer remains largely unclear. To the best of our knowledge, this is one of the first study to identify CDC20 as a potential predictive biomarker for prognosis of breast cancer using comprehensive bioinformatics analysis.

In the present study, we performed a bioinformatics analysis of the clinical parameters and survival data related to CDC20 in breast cancer patients by pooling and analyzing several online tools. Comparing with normal tissues, Oncomine database revealed that CDC20 was expressed higher in different types of breast cancer including medullary breast carcinoma, invasive ductal breast carcinoma, invasive lobular breast carcinoma and so on. CDC20 was also expressed higher in breast cancer tissues compared to adjacent normal tissues of patients in our hospital, confirming the results from databases online. Moreover, we also found patients with a more advanced stage of breast cancer tended to express higher levels CDC20. Consistently, Yuan et al. reported that the mRNA and protein levels of CDC20 were significantly higher in breast cancer cells and high-grade primary breast cancer tissues [26].

For nodal status, there was no significant difference between positive and negative group. ER and PR status were negatively correlated with CDC20 level. Conversely, SBR grade, NPI index, HER-2 status, basal-like status, and triple-negative status were positively related to CDC20 expression in breast cancer patients in respect to normal individuals. It is generally known that breast cancer patients with ER or PR positive, HER-2 negative, non-basal-like or non-triple-negative status have a preferable outcome [27]. Therefore, these results indicated that lower expression of CDC20 may predict a better prognosis in breast cancer.

We further investigated the prognostic value of CDC20 in breast cancer using the PrognoScan database. These pooled results showed that higher CDC20 expression correlated with worse distant metastasis free survival, relapse free survival, disease free survival, disease specific survival and overall survival. These findings were in agreement with the notion of CDC20 as a tumor oncogene and a potential predictive biomarker for prognosis of breast cancer [13].
Finally, we checked the co-expression of CDC20 gene using the Oncomine, bc-GenExMiner, and UCSC Xena web-based tools and found that TPX2 was positively correlated with CDC20 expression. Targeting protein for Xenopus kinesin-like protein 2 (TPX2) is a microtubule-associated protein that is encoded by a gene located on human chromosome band 20q11.1[22]. Overexpression of TPX2 has been observed in lung cancer, hepatic cancer, colon cancer [28-30]. Moreover, TPX2 is also a marker of poor tumor prognosis in several cancers [31]. Both CDC20 and TPX2 were related to the process of cell cycle [32]. Using bioinformatics analysis, Zhang found that elevated mRNA Levels of CDC20 and TPX2 are associated with poor prognosis of lung adenocarcinoma [33]. These observations, along with our findings of CDC20 in survival data, provided evidence that CDC20 gene might promote tumor progress associated with TPX2 expression.

In conclusion, the study was performed to comprehensively analyze the expression pattern, potential function, and distinct prognostic effect of CDC20 in breast cancer by pooling all currently available data online. CDC20 was higher-expressed in different subtypes of breast cancer compared with normal tissues and was associated with several important clinical parameters. CDC20 could be considered as a potential predictive indicator for prognosis of breast cancer with co-expressed TPX2 gene. Over the past several decades, much research has focused on identifying new prognostic markers in order to make better clinical decisions and improve therapy and outcomes. More in-depth experiments are needed to validate the value of CDC20 for clinical decision-making in breast cancer.
Acknowledgements

This work was financially supported by Natural Science Foundation of China (81502294 and 81702591), the Project of Jiangsu Province Traditional Chinese Medicine Bureau (YB201952), Natural Science Foundation of Jiangsu Province (BK20170294) and Changzhou Sci&Tech Program (CJ20159044).

Conflict of interest

The authors declare no conflict of interest.

Author contribution

LC, JW and QD conceived and designed the experiments. LC, YH and WC performed the experiments. LS, ZL, XZ, and XD analyzed the data. LC wrote the manuscript.
References

[1] L. Fan, K. Strasser-Weippl, J.-J. Li, J. St Louis, D.M. Finkelstein, K.-D. Yu, W.-Q. Chen, Z.-M. Shao, P.E. Goss, Breast cancer in China, The Lancet Oncology, 15 (2014) e279-e289.
[2] K. Rennstam, B. Baldetorp, S. Kytola, M. Tanner, J. Isola, Chromosomal rearrangements and oncogene amplification precede aneuploidization in the genetic evolution of breast cancer, Cancer research, 61 (2001) 1214-1219.
[3] M.J. Duffy, S. Walsh, E.W. McDermott, J. Crown, Biomarkers in Breast Cancer: Where Are We and Where Are We Going?, Adv Clin Chem, 71 (2015) 1-23.
[4] H. Yu, Cdc20: A WD40 Activator for a Cell Cycle Degradation Machine, Molecular Cell, 27 (2007) 3-16.
[5] J. Weinstein, F.W. Jacobsen, J. Hsu-Chen, T. Wu, L.G. Baum, A novel mammalian protein, p55CDC, present in dividing cells is associated with protein kinase activity and has homology to the Saccharomyces cerevisiae cell division cycle proteins Cdc20 and Cdc4, Molecular and cellular biology, 14 (1994) 3350-3363.
[6] L. Smolders, J.G. Teodoro, Targeting the anaphase promoting complex: common pathways for viral infection and cancer therapy, Expert opinion on therapeutic targets, 15 (2011) 767-780.
[7] I.M. Moura, M.L. Delgado, P.M. Silva, C.A. Lopes, J.B. do Amaral, L.S. Monteiro, H. Bousbaa, High Cdc20 expression is associated with poor prognosis in oral squamous cell carcinoma, Journal of oral pathology & medicine : official publication of the International Association of Oral Pathologists and the American Academy of Oral Pathology, 43 (2014) 225-231.
[8] Z.Y. Ding, H.R. Wu, J.M. Zhang, G.R. Huang, D.D. Ji, Expression characteristics of CDC20 in gastric cancer and its correlation with poor prognosis, International journal of clinical and experimental pathology, 7 (2014) 722-727.
[9] J.W. Choi, Y. Kim, J.H. Lee, Y.S. Kim, High expression of spindle assembly checkpoint proteins CDC20 and MAD2 is associated with poor prognosis in urothelial bladder cancer, Virchows Archiv : an international journal of pathology, 463 (2013) 681-687.
[10] W.J. Wu, K.S. Hu, D.S. Wang, Z.L. Zeng, D.S. Zhang, D.L. Chen, L. Bai, R.H. Xu, CDC20 overexpression predicts a poor prognosis for patients with colorectal cancer, Journal of translational medicine, 11 (2013) 142.
[11] T. Kato, Y. Daigo, M. Aragaki, K. Ishikawa, M. Sato, M. Kaji, Overexpression of CDC20 predicts poor prognosis in primary non-small cell lung cancer patients, Journal of surgical oncology, 106 (2012) 423-430.
[12] D.Z. Chang, Y. Ma, B. Ji, Y. Liu, P. Hwu, J.L. Abbruzzese, C. Logsdon, H. Wang, Increased CDC20 expression is associated with pancreatic ductal adenocarcinoma differentiation and progression, Journal of Hematology & Oncology, 5 (2012).
[13] H. Karra, H. Repo, I. Ahonen, E. Loyttyniemi, R. Pitkanen, M. Lintunen, T. Kuoppio, M. Soderstrom, P. Kronqvist, Cdc20 and securin overexpression predict short-term breast cancer survival, British journal of cancer, 110 (2014) 2905-2913.
[14] D.R. Rhodes, J. Yu, K. Shanker, N. Deshpande, R. Varambally, D. Ghosh, T. Barrette, A. Pandey, A.M. Chinnaian, ONCOMINE: a cancer microarray database and integrated data-mining platform, Neoplasia (New York, N.Y.), 6 (2004) 1-6.
[15] D.S. Chandrashekar, B. Bashel, S.A.H. Basubramanya, C.J. Creighton, I. Ponce-Rodriguez, B. Chakravarthi, S. Varambally, UALCAN: A Portal for Facilitating Tumor Subgroup Gene Expression and
Survival Analyses, Neoplasia, 19 (2017) 649-658.

[16] P. Jezequel, M. Campone, W. Gouraud, C. Guerin-Charbonnel, C. Leux, G. Ricolleau, L. Campion, bc-GenExMiner: an easy-to-use online platform for gene prognostic analyses in breast cancer, Breast cancer research and treatment, 131 (2012) 765-775.

[17] P. Jezequel, J.S. Frenel, L. Campion, C. Guerin-Charbonnel, W. Gouraud, G. Ricolleau, M. Campion, bc-GenExMiner 3.0: new mining module computes breast cancer gene expression correlation analyses, Database : the journal of biological databases and curation, 2013 (2013) bas060.

[18] H. Mizuno, K. Kitada, K. Nakai, A. Sarai, PrognoScan: a new database for meta-analysis of the prognostic value of genes, BMC medical genomics, 2 (2009) 18.

[19] J. Casper, A.S. Zweig, C. Villarreal, C. Tyner, M.L. Speir, K.R. Rosenbloom, B.J. Raney, C.M. Lee, B.T. Lee, D. Karolchik, A.S. Hinrichs, M. Haeussler, I. Fiddes, C. Eisenhart, M. Diekhans, H. Clawson, G.P. Barber, J. Armstrong, D. Haussler, W.J. Kent, The UCSC Genome Browser database: 2018 update, Nucleic acids research, 46 (2018) D762-d769.

[20] C. Bansal, M. Pujani, K.L. Sharma, A.N. Srivastava, U.S. Singh, Grading systems in the cytological diagnosis of breast cancer: a review, Journal of cancer research and therapeutics, 10 (2014) 839-845.

[21] A.H. Lee, I.O. Ellis, The Nottingham prognostic index for invasive carcinoma of the breast, Pathology oncology research : POR, 14 (2008) 113-115.

[22] G. Neumayer, C. Belzil, O.J. Gruss, M.D. Nguyen, TPX2: of spindle assembly, DNA damage response, and cancer, Cellular and molecular life sciences : CMLS, 71 (2014) 3027-3047.

[23] L. Chang, Z. Zhang, J. Yang, S.H. McLaughlin, D. Barford, Atomic structure of the APC/C and its mechanism of protein ubiquitination, Nature, 522 (2015) 450-454.

[24] Y. Kim, J.W. Choi, J.H. Lee, Y.S. Kim, MAD2 and CDC20 are upregulated in high-grade squamous intraepithelial lesions and squamous cell carcinomas of the uterine cervix, International journal of gynecological pathology : official journal of the International Society of Gynecological Pathologists, 33 (2014) 517-523.

[25] G. Marucci, L. Morandi, E. Magrini, A. Farnedi, E. Franceschi, R. Miglio, D. Calo, A. Pession, M.P. Foschini, V. Eusebi, Gene expression profiling in glioblastoma and immunohistochemical evaluation of IGFBP-2 and CDC20, Virchows Archiv : an international journal of pathology, 453 (2008) 599-609.

[26] B. Yuan, Y. Xu, J.H. Woo, Y. Wang, Y.K. Bae, D.S. Yoon, R.P. Wersto, E. Tully, K. Wilsbach, E. Gabrielson, Increased expression of mitotic checkpoint genes in breast cancer cells with chromosomal instability, Clinical cancer research : an official journal of the American Association for Cancer Research, 12 (2006) 405-410.

[27] S.P. Bagaria, P.S. Ray, M.S. Sim, X. Ye, J.M. Shamonki, X. Cui, A.E. Giuliano, Personalizing breast cancer staging by the inclusion of ER, PR, and HER2, JAMA surgery, 149 (2014) 125-129.

[28] Y. Ma, D. Lin, W. Sun, T. Xiao, J. Yuan, N. Han, S. Guo, X. Feng, K. Su, Y. Mao, S. Cheng, Y. Gao, Expression of targeting protein for xklp2 associated with both malignant transformation of respiratory epithelium and progression of squamous cell lung cancer, Clinical cancer research : an official journal of the American Association for Cancer Research, 12 (2006) 1121-1127.

[29] B. Liang, C. Jia, Y. Huang, H. He, J. Li, H. Liao, X. Liu, X. Liu, X. Bai, D. Yang, TPX2 Level Correlates with Hepatocellular Carcinoma Cell Proliferation, Apoptosis, and EMT, Digestive diseases and sciences, 60 (2015) 2360-2372.

[30] P. Wei, N. Zhang, Y. Xu, X. Li, D. Shi, Y. Wang, D. Li, S. Cai, TPX2 is a novel prognostic marker for the growth and metastasis of colon cancer, Journal of translational medicine, 11 (2013) 313.

[31] C. Tomii, M. Inokuchi, Y. Takagi, T. Ishikawa, S. Otsuki, H. Uetake, K. Kojima, T. Kawano, TPX2
expression is associated with poor survival in gastric cancer, World journal of surgical oncology, 15 (2017) 14.

[32] G. Neumayer, M.D. Nguyen, TPX2 impacts acetylation of histone H4 at lysine 16: implications for DNA damage response, PLoS One, 9 (2014) e110994.

[33] M.-Y. Zhang, X.-X. Liu, H. Li, R. Li, X. Liu, Y.-Q. Qu, Elevated mRNA Levels of AURKA, CDC20 and TPX2 are associated with poor prognosis of smoking related lung adenocarcinoma using bioinformatics analysis, International Journal of Medical Sciences, 15 (2018) 1676-1685.
**Figure legends**

**Figure 1.** Increased expression of CDC20 gene in breast cancer tissues in public databases. (A) Expression of CDC20 in 20 types of cancer vs. corresponding normal tissues using the Oncomine database with the threshold of fold change ≥2, P-value ≤1E-4, and gene rank ≥ top 10%. Red and blue respectively stand for the numbers of datasets with statistically significant (P<0.05) increased and decreased levels of CDC20 gene. (B) Higher mRNA CDC20 was expressed in breast cancer tissues than in normal tissues (P<0.05) using UALCAN website. (C) Patients with a more advanced stage of breast cancer tended to express higher levels CDC20.

**Figure 2.** Increased expression of CDC20 gene in breast cancer tissues from patients in our hospital. (A) CDC20 mRNA expression in 22 pairs of breast cancer and adjacent tissues. (B) Average expression level of RBMS2 mRNA in 22 pairs of breast cancer tissues and adjacent normal breast tissues. Breast cancer tissues had higher expression of CDC20 than adjacent breast tissues (P<0.05).

**Figure 3.** Box plot comparing CDC20 expression in normal individuals and breast cancer patients obtained from the Oncomine database. Analysis is shown for medullary breast carcinoma (A), invasive ductal breast carcinoma (B), invasive lobular breast carcinoma (C), invasive breast carcinoma (D), invasive ductal and invasive lobular breast carcinoma (E), breast carcinoma (F), mucinous breast carcinoma (G), tubular breast carcinoma (H), intraductal cribriform breast adenocarcinoma (I), invasive ductal breast carcinoma (J), invasive breast carcinoma (K), mixed lobular and ductal breast carcinoma (L), invasive lobular breast carcinoma (M), ductal breast carcinoma (N).

**Figure 4.** Box plot evaluating CDC20 expression among different groups of patients based on clinical parameters using the bc-GenExMiner 4.2 Analysis is shown for age (A), NPI index (B), SBR grade (C), ER (D), PR (E), HER-2 (F), nodal status (G), triple-negative status (H), and basal-like status (I).

**Figure 5.** Survive curve evaluating the prognostic value of CDC20 using the PrognoScan Analysis is shown for distant metastasis free survival (A, B, D, E, F, N), relapse free survival for different probe (C, H, M), disease free survival (J, L), disease specific survival (I, K), and overall survival (G, O). Red and blue stand for high and low expression of CDC20 gene, respectively.
Figure 6. Co-expression of CDC20 gene (A) CDC20 co-expression of genes analyzed using Oncomine. (B) Relationship between CDC20 and TPX2 in breast cancer analyzed using bc-GenExMiner v4.0. (C) Heat map of CDC20 expression and TPX2 mRNA expression across PAM50 breast cancer subtypes in the TCGA database, determined using UCSC Xena; D: Correlation between CDC20 and TPX2 mRNA expression in the TCGA database, determined using UCSC Xena.
A

Cancer Type

Significant Unique Analyses

Total Unique Analyses

Gene Rank Percentile (%)

1

5

10

15

Gene Rank Percentile (%)

1

5

10

15

B

Cancer Type

Bladder Cancer

Brain and CNS Cancer

Breast Cancer

Cervical Cancer

Colorectal Cancer

Esophageal Cancer

Gastric Cancer

Kidney Cancer

Leukemia

Lung Cancer

Melanoma

Malignant Tumor

Myeloma

Ovarian Cancer

Pancreatic Cancer

Prostate Cancer

Other Cancer

Sarcoma

C

Expression of CDC20 in BRCA based on individual cancer stages

Expression of CDC20 in BRCA based on Sample Types

Transcript per million

Transcript per million

Normal

Staged

Normal

Staged
Table 1. The significant changes of CDC20 expression in transcription level between different types of breast cancer and normal tissues (ONCOMINE database)

| Subtype of breast cancer                          | p-value   | Fold change | Rank (%) | Sample | Reference |
|--------------------------------------------------|-----------|-------------|----------|--------|-----------|
| Medullary Breast Carcinoma                       | 7.70E-21  | 14.001      | 1%       | 176    | 22522925  |
| Invasive Ductal Breast Carcinoma                 | 1.35E-115 | 5.535       | 1%       | 1700   | 22522925  |
| Invasive Lobular Breast Carcinoma                | 1.13E-42  | 3.205       | 2%       | 292    | 22522925  |
| Invasive Breast Carcinoma                        | 1.40E-7   | 4.392       | 2%       | 165    | 22522925  |
| Invasive Ductal and Invasive Lobular Breast Carcinoma | 3.68E-30  | 3.165       | 2%       | 234    | 22522925  |
| Breast Carcinoma                                 | 6.69E-6   | 3.634       | 3%       | 158    | 22522925  |
| Mucinous Breast Carcinoma                        | 1.58E-14  | 2.835       | 4%       | 190    | 22522925  |
| Tubular Breast Carcinoma                         | 6.13E-18  | 2.534       | 7%       | 211    | 22522925  |
| Intraductal Cribriform Breast Adenocarcinoma     | 1.69E-17  | 2.857       | 1%       | 64     | TCGA      |
| Invasive Ductal Breast Carcinoma                 | 2.56E-42  | 4.252       | 1%       | 450    | TCGA      |
| Invasive Breast Carcinoma                        | 5.80E-24  | 3.488       | 2%       | 137    | TCGA      |
| Mixed Lobular and Ductal Breast Carcinoma        | 1.97E-5   | 2.541       | 3%       | 68     | TCGA      |
| Invasive Lobular Breast Carcinoma                | 2.20E-12  | 2.731       | 3%       | 97     | TCGA      |
| Ductal Breast Carcinoma                          | 1.40E-6   | 21.440      | 4%       | 47     | 16473279  |
Table 2. The relationship between mRNA expression of CDC20 and clinicopathological parameters of breast carcinoma

| Variables              | NO. Patients | CDC20 mRNA | p-value    |
|------------------------|--------------|------------|------------|
| **Age**                |              |            |            |
| ≤ 51                   | 1392         | ↑          | <0.0001    |
| > 51                   | 2210         |            |            |
| **Nodal Status**       |              |            |            |
| -                      | 2493         |            | 0.6675     |
| +                      | 1562         |            |            |
| **ER**                 |              |            |            |
| -                      | 1446         | ↑          | <0.0001    |
| +                      | 3766         |            |            |
| **PR**                 |              |            |            |
| -                      | 804          | ↑          | <0.0001    |
| +                      | 1249         |            |            |
| **HER2**               |              |            |            |
| -                      | 1409         |            | 0.0009     |
| +                      | 201          | ↑          |            |
| **SBR**                |              |            |            |
| 1                      | 546          |            | <0.0001    |
| 2                      | 1431         | ↑          |            |
| 3                      | 1317         | ↑          |            |
| **NPI**                |              |            |            |
| 1                      | 931          |            | <0.0001    |
| 2                      | 727          | ↑          |            |
| 3                      | 125          | ↑          |            |
| **Basal-like Status**  |              |            |            |
| Not                    | 3492         |            | <0.0001    |
| Basal-like             | 1060         | ↑          |            |
| **Triple-negative**    |              |            |            |
| Status                 |              |            |            |
| Not                    | 3857         |            |            |
| TNBC                   | 374          | ↑          | <0.0001    |
### Table 3. The association of CDC20 expression and the survival in breast cancer patients

| DATASET       | ENDPOINT                       | PROBE ID  | Location          | N   | COX P-VALUE | HR(95% CI)          |
|---------------|--------------------------------|-----------|-------------------|-----|-------------|---------------------|
| GSE7390       | Distant Metastasis Free Survival | 202870_s_at | Uppsala, Oxford   | 198 | 0.040131    | 1.23 [1.01 - 1.49]  |
| GSE9195       | Distant Metastasis Free Survival | 202870_s_at | GUYT              | 77  | 0.001697    | 3.49 [1.60 - 7.60]  |
| GSE9195       | Relapse Free Survival          | 202870_s_at | GUYT              | 77  | 0.002865    | 2.67 [1.40 - 5.10]  |
| GSE12093      | Distant Metastasis Free Survival | 202870_s_at | IO, NCI           | 136 | 0.018342    | 2.26 [1.15 - 4.43]  |
| GSE11121      | Distant Metastasis Free Survival | 202870_s_at | Mainz             | 200 | 0.001782    | 1.63 [1.20 - 2.21]  |
| GSE2034       | Distant Metastasis Free Survival | 202870_s_at | Rotterdam         | 286 | 0.006541    | 1.40 [1.10 - 1.79]  |
| GSE1456-GPL96 | Overall Survival               | 202870_s_at | Stockholm         | 159 | 0.000207    | 1.94 [1.37 - 2.74]  |
| GSE1456-GPL96 | Relapse Free Survival          | 202870_s_at | Stockholm         | 159 | 0.000323    | 1.90 [1.34 - 2.70]  |
| GSE1456-GPL96 | Disease Specific Survival      | 202870_s_at | Stockholm         | 159 | 0.000155    | 2.22 [1.47 - 3.36]  |
| GSE7378       | Disease Free Survival          | 202870_s_at | UCSF              | 54  | 0.008193    | 2.79 [1.30 - 5.98]  |
| GSE3494-GPL96 | Disease Specific Survival      | 202870_s_at | Uppsala           | 236 | 0.000439    | 1.77 [1.29 - 2.44]  |
| GSE4922-GPL96 | Disease Free Survival          | 202870_s_at | Uppsala           | 249 | 0.000182    | 1.63 [1.26 - 2.11]  |
| GSE2990       | Relapse Free Survival          | 202870_s_at | Uppsala, Oxford   | 62  | 0.014206    | 1.72 [1.11 - 2.65]  |
| GSE2990       | Distant Metastasis Free Survival | 202870_s_at | Uppsala, Oxford   | 54  | 0.008976    | 2.08 [1.20 - 3.59]  |
| GSE7390       | Overall Survival               | 202870_s_at | Uppsala, Oxford   | 198 | 0.027783    | 1.25 [1.03 - 1.54]  |
| Study   | Overall Survival | Disease Specific Survival | Duke | UCSF | Ratio (95% CI) | p-value (95% CI) |
|---------|-----------------|---------------------------|------|------|---------------|-----------------|
| GSE3143 | 38414_at        | 158                       | 0.010445 | 0.75 [0.57 - 0.98] |
| E-TABM-158 | 202870_s_at    | 117                       | 0.038478 | 0.010445 | 1.67 [1.12-2.41] | 0.75 [0.57 - 0.98] |