

Distinct expression of CDCA3, CDCA5, and CDCA8 leads to shorter relapse free survival in breast cancer patient

Nam Nhut Phan^{1,2,*}, Chih-Yang Wang^{3,*}, Kuan-Lun Li¹, Chien-Fu Chen⁴, Chung-Chieh Chiao⁴, Han-Gang Yu⁵, Pung-Ling Huang^{1,6} and Yen-Chang Lin¹

¹Graduate Institute of Biotechnology, Chinese Culture University, Taipei, Taiwan

²NTT Institute of Hi-Technology, Nguyen Tat Thanh University, Ho Chi Minh City, Vietnam

³Department of Biochemistry and Molecular Biology, Institute of Basic Medical Sciences, College of Medicine, National Cheng Kung University, Tainan, Taiwan

⁴School of Chinese Medicine for Post-Baccalaureate, I-Shou University, Kaohsiung, Taiwan

⁵Department of Physiology and Pharmacology, West Virginia University, Morgantown, WV, USA

⁶Department of Horticulture & Landscape Architecture, National Taiwan University, Taipei, Taiwan

*These authors have contributed equally to this work

Correspondence to: Pung-Ling Huang, **email:** pungling@ntu.edu.tw
Yen-Chang Lin, **email:** lycnthu@gmail.com

Keywords: cell cycle division-associated (CDCA) protein; breast cancer; cell cycle; prognosis; bioinformatics

Received: October 16, 2017

Accepted: January 03, 2018

Published: January 09, 2018

Copyright: Phan et al. This is an open-access article distributed under the terms of the Creative Commons Attribution License 3.0 (CC BY 3.0), which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.

ABSTRACT

Breast cancer is a dangerous disease that results in high mortality rates for cancer patients. Many methods have been developed for the treatment and prevention of this disease. Determining the expression patterns of certain target genes in specific subtypes of breast cancer is important for developing new therapies for breast cancer. In the present study, we performed a holistic approach to screening the mRNA expression of six members of the cell division cycle-associated gene family (CDCA) with a focus on breast cancer using the OncoPrint and The Cancer Cell Line Encyclopedia (CCLE) databases. Furthermore, Gene Expression-Based Outcome for Breast Cancer Online (GOBO) was also used to deeply mine the expression of each CDCA gene in clinical breast cancer tissue and breast cancer cell lines. Finally, the mRNA expression of the CDCA genes as related to breast cancer patient survival were analyzed using a Kaplan-Meier plot. CDCA3, CDCA5, and CDCA8 mRNA expression levels were significantly higher than the control sample in both clinical tumor sample and cancer cell lines. These highly expressed genes in the tumors of breast cancer patients dramatically reduced patient survival. The interaction network of CDCA3, CDCA5, and CDCA8 with their co-expressed genes also revealed that CDCA3 expression was highly correlated with cell cycle related genes such as CCNB2, CDC20, CDKN3, and CCNB1. CDCA5 expression was correlated with BUB1 and TRIP13, while CDCA8 expression was correlated with BUB1 and CCNB1. Altogether, these findings suggested CDCA3, CDCA5, and CDCA8 could have a high potency as targeted breast cancer therapies.

INTRODUCTION

According to the WHO report, the top five cancers were lung (1.69 million), liver (788,000), colorectal (774,000), stomach (754,000), and breast (571,000) (WHO,

Fact Sheet, Feb 2017). Although the underlying mechanism of cancer development was extensively studied, breast cancer patient, particularly women, are still coping with low survival rate. Therefore, novel and effective therapeutic treatments and drugs development are very crucial.

Cell division is the critical process of life. Many studies have proven that a malfunction in the cell division process results in cancer [1–5]. The cell cycle division-associated (CDCA) protein family is comprised of eight members (CDCA1–8). Cell division cycle-associated protein 1 (CDCA1) is critical for nuclear division and microtubule stabilization [6]. The function of CDCA2 is binding to the protein phosphatase 1 γ (PP1 γ) and controlling the DNA damage response in the cell cycle [7, 8]. CDCA3 is known to regulate cell cycle progression, and its levels are controlled by transcription and protein degradation during the G1 checkpoint of the cell cycle [9]. CDCA4 is a cell-cycle regulator that is associated with the G1/S transition [10]. CDCA4 also modulates p53 expression levels [11]. CDCA5 is critical regulator of sister-chromatid cohesion and separation during cell division [12]. CDCA7 is activated in hematopoietic stem cell precursors in the mouse embryo and maintained thereafter in distinct undifferentiated hematopoietic populations. CDCA8 is an essential regulator of mitosis [13].

The purpose of this study was to systematically investigate the relationship between the mRNA expression of the CDCA family and the survival probability of breast cancer patients using the Oncomine database (www.oncomine.org), Gene expression-based Outcome for Breast cancer Online database (GOBO; <http://co.bmc.lu.se/gobo/gsa.pl>), the Cancer Cell Line Encyclopedia database (CCLE; www.portals.broadinstitute.org/ccle), and a Kaplan-Meier plot (www.kmplot.com).

RESULTS

Expression of CDCA gene family in breast cancer tissue

The expression of CDCA2, CDCA3, CDCA4, CDCA5, CDCA7, and CDCA8 in 20 types of cancer is depicted in Figure 1. CDCA genes were dramatically overexpressed in breast cancer tissue relative to normal type-matched tissue.

Data mining from The Cancer Genome Atlas (TCGA) dataset for breast cancer showed significant overexpression of CDCA2 in some subtypes of breast cancer, which were Erb-B2 Receptor Tyrosine Kinase 2/ Estrogen receptor/progesterone receptor (ERBB2/ER/PR) negative (triple-negative breast cancer [TNBC]), invasive ductal breast carcinoma, invasive breast carcinoma, and invasive lobular breast carcinoma. The invasive ductal breast carcinoma subtype had highest expression-level with an over five-fold higher expression in cancerous tissue compared to normal breast tissue (Figure 2A). Analysis of the TCGA breast dataset showed a high expression of CDCA3 in four subtypes of breast cancer, which were invasive ductal breast carcinoma, intraductal cribriform breast adenocarcinoma, invasive breast

carcinoma, and invasive lobular breast carcinoma, with the highest expression-level fold-change of 4.05-times higher expression in invasive ductal breast carcinoma tissue (Figure 2B). The CDCA4 mRNA expression level in breast cancer subtypes was not as high as that of the CDCA3 level. The highest CDCA4 fold-change was 2.25-times higher relative to the control type-matched tissue (Figure 2C). The CDCA5 mRNA expression level in breast cancer tissue was the highest in the invasive ductal breast carcinoma subtype, with a 5.5-fold change relative to the control tissue (Figure 2D). The CDCA7 mRNA was overexpressed in the TNBC subtype of breast cancer by more than 1.5 - fold relative to normal control sample (Figure 2E). The CDCA8 mRNA expression level was relatively high in four subtypes of breast cancer, which were male breast carcinoma, invasive ductal breast carcinoma, invasive lobular breast carcinoma, and invasive breast carcinoma, with all of these subtypes displaying more than a 3.5-fold change over the normal tissue (Figure 2F).

Expression of the CDCA gene family in the breast cancer cell line

We continued investigating the expression of CDCA2 in cancer cell data using the CCLE database. On a log₂ scale, the breast cancer cell line data from 58 datasets showed that CDCA2 expression was significantly up-regulated by an estimated eight-fold. The copy number of the CDCA2 gene in the different carcinoma cell lines is displayed in Figure 3A. The mRNA expression and copy number of CDCA3 in various cancer cell lines is displayed in Figures 3B and 4A. Neve dataset analysis of breast cancer cell lines showed the expression of CDCA3, and intensity percentile for each cell line [14]. The log₂ expression levels of the six subtypes of breast cancers with up-regulated CDCA3 in basal, luminal, TN, HER2 subtypes as illustrated in Figure 5. The basal subtype had the highest expression level compared to the basal and luminal-like subtypes, whereas TN had highest expression level relative to the HER2 and hormone receptor subtypes. The highest mRNA expression of CDCA3 in breast cancer tumors was in the basal breast cancer patients. The RFS of patients with a medium expression level of CDCA3 showed an association to all tumor subtypes. The number of breast cancer samples with a high, medium, and low expression level of CDCA3 was presented in the Figure 5. The expression of all samples in various breast cancer datasets was displayed. Forest plots of 751 cases of breast cancer showed a hazard ratio (HR) for different parameters, such as tumor size, age, tumor grade, node status, and ER-positive status (Figure 5).

From the CCLE analysis, CDCA4 expression levels in breast cancer cell lines were high. The copy number of this gene was also lower than that of other cancer types (Figure 3C). The analysis of the GOBO database in tumor mode

showed that log2 expression levels of CDCA4 were altered between cell lines, but not in a specific manner. In clinical subtypes, CDCA4 expression was the highest in TNBC

and the lowest in HER2. Basal subtype had the highest expression of CDCA4, while the lowest was in the luminal-like subtype of breast cancer (Supplementary Figure 1).

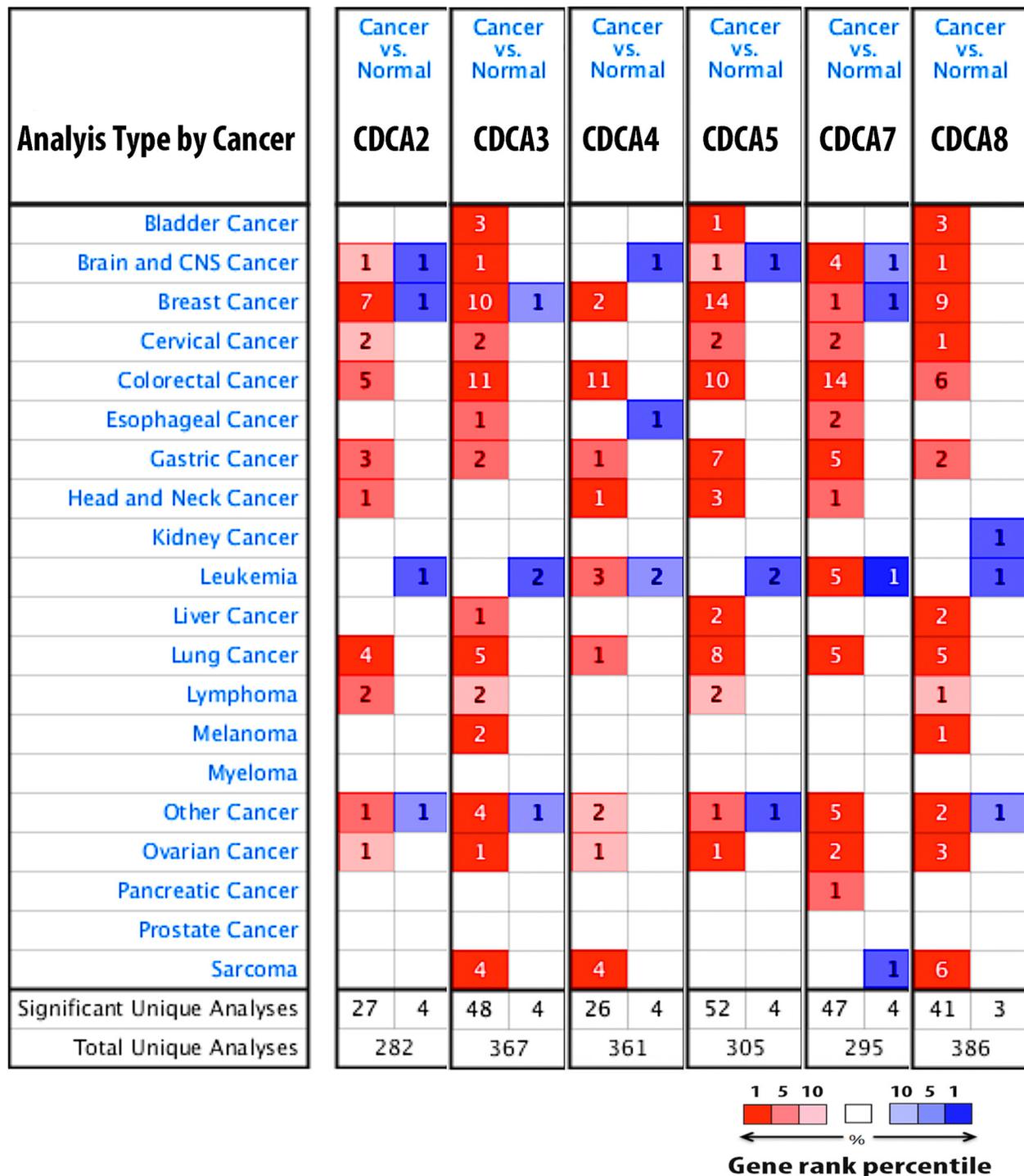
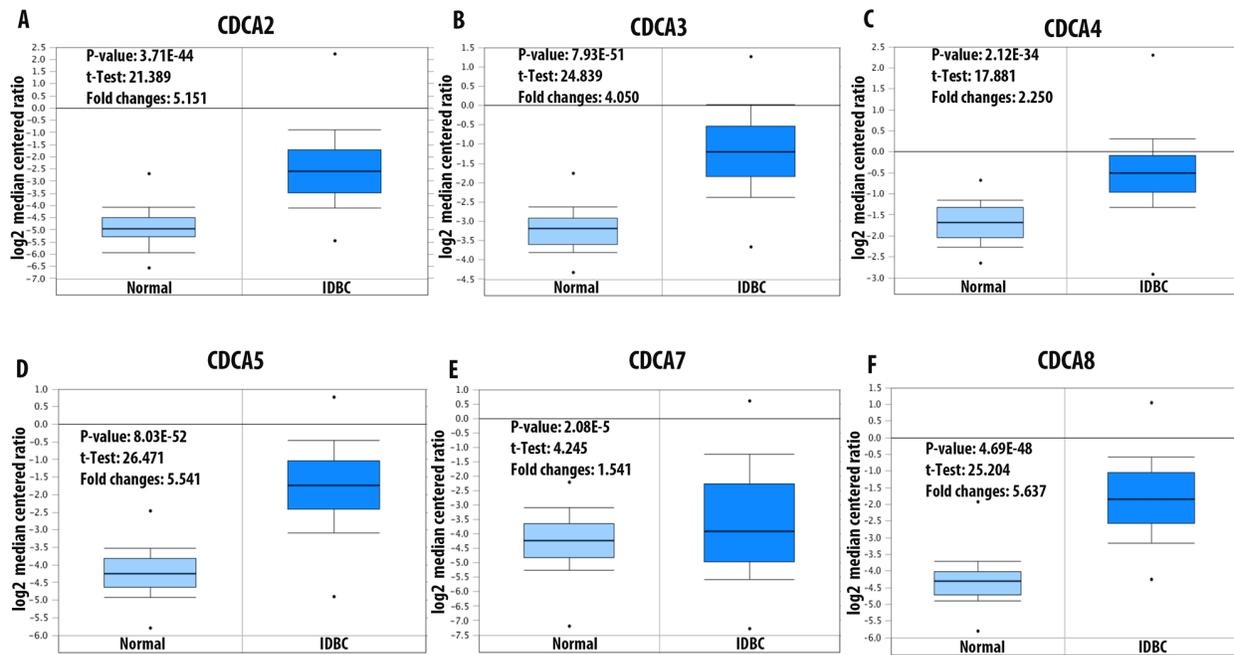


Figure 1: CDCA genes expression in 20 different types of cancer disease. Data compared mRNA expression of gene in cancer tissue relative to normal matched type tissue. Over and under expression of CDCA genes were displayed with color based manner. Red color represents for over expression of gene while blue color is for under expression of gene. Color transparency slightly shifted top 1% to top 10% in both over and under expression of gene. The number in each square denoted number of analyse(s) satisfy the threshold such as gene rank percentile (10%), p-value (10E-4), and fold change (1.5).



IDBC: Invasive Ductal Breast Carcinoma

Figure 2: mRNA expression of CDCA family members in invasive ductal breast carcinoma (IDBC). Log2 median centered ratio was used to express the fold changes between CDCA genes in IDBC tissue relative to normal matched type tissue. P-value, *t*-test, and fold changes information were displayed. Panel (A-F) represent for CDCA2, CDCA3, CDCA4, CDCA5, CDCA7, and CDCA8 mRNA expression in IDBC relative to control samples.

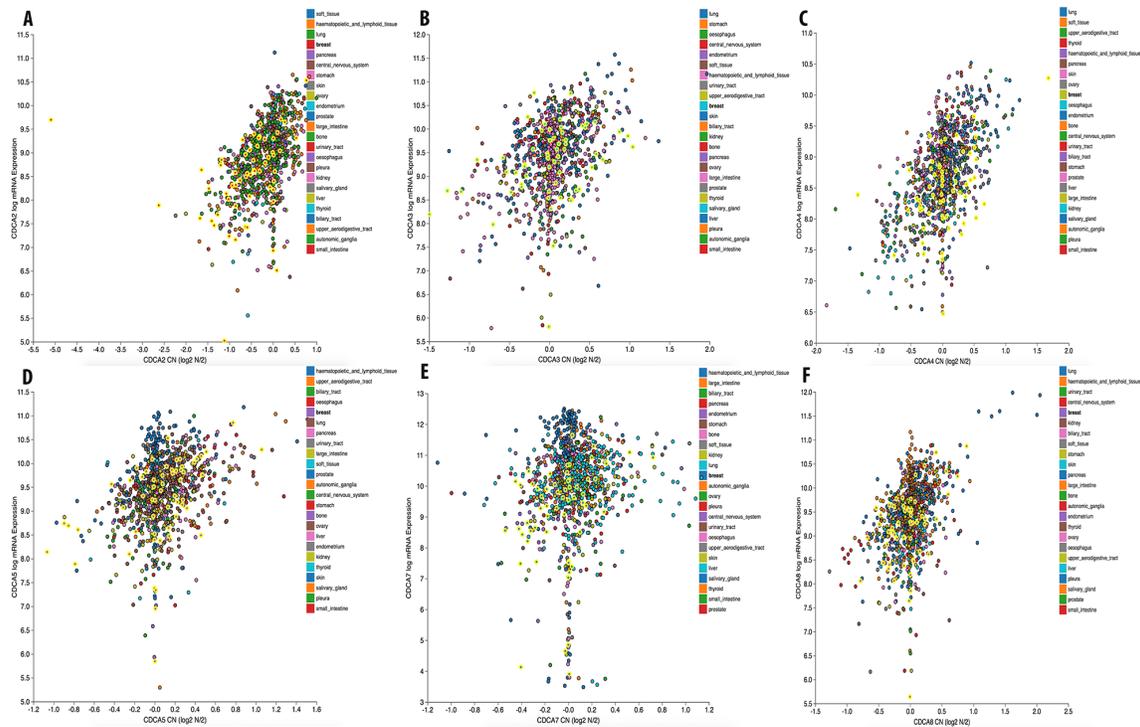


Figure 3: Gene expression and copy number of CDCA genes in various cancer cell lines. mRNA expression and copy number of CDCA genes (A), CDCA3 (B), CDCA4 (C), CDCA5 (D), CDCA7 (E), and CDCA8 (F) in breast cancer cell line were highlighted in bright yellow color circles.

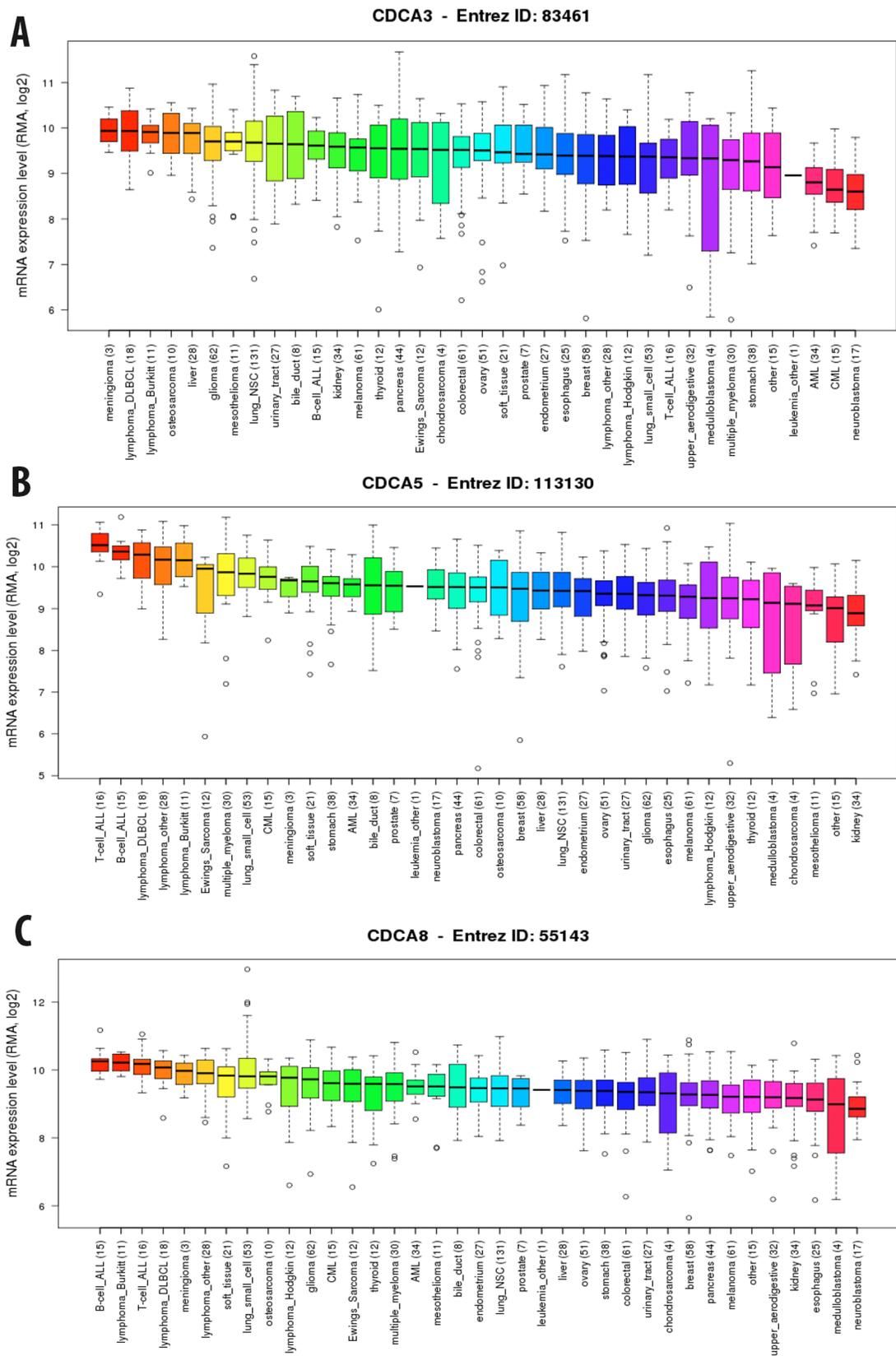


Figure 4: Expression of CDCA3 (A), CDCA5 (B), and CDCA8 (C) genes in breast cancer cell line from CCLE database. RMA, log2 was used to measure the expression of these three gene in breast cancer cell line.

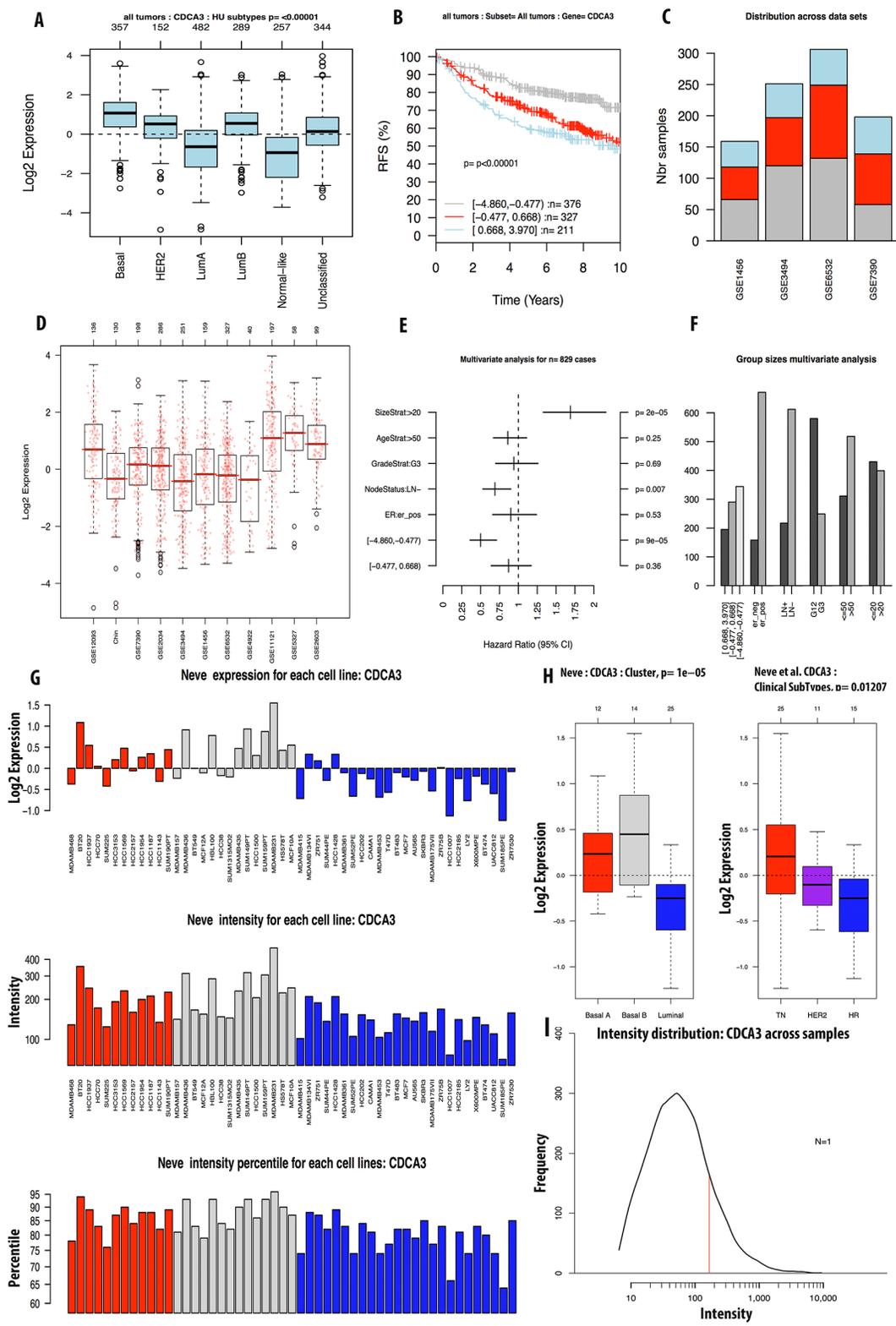


Figure 5: Analysis of CDCA3 expression in breast cancer tumor and cell line by GOBO database. Expression of CDCA3 in six subtypes of breast cancer tumors (A). RFS analysis of CDCA3 expression in tumor of patient survival with high expression in light blue, intermediate expression in red, and low expression in grey color (B). CDCA3 expression across datasets (C, D). Forest plot and bar chart of multivariate analysis on the expression of CDCA3 was displayed in (E, F). Expression of CDCA3 in breast cancer cell line using Neve et al dataset was display in (G-I).

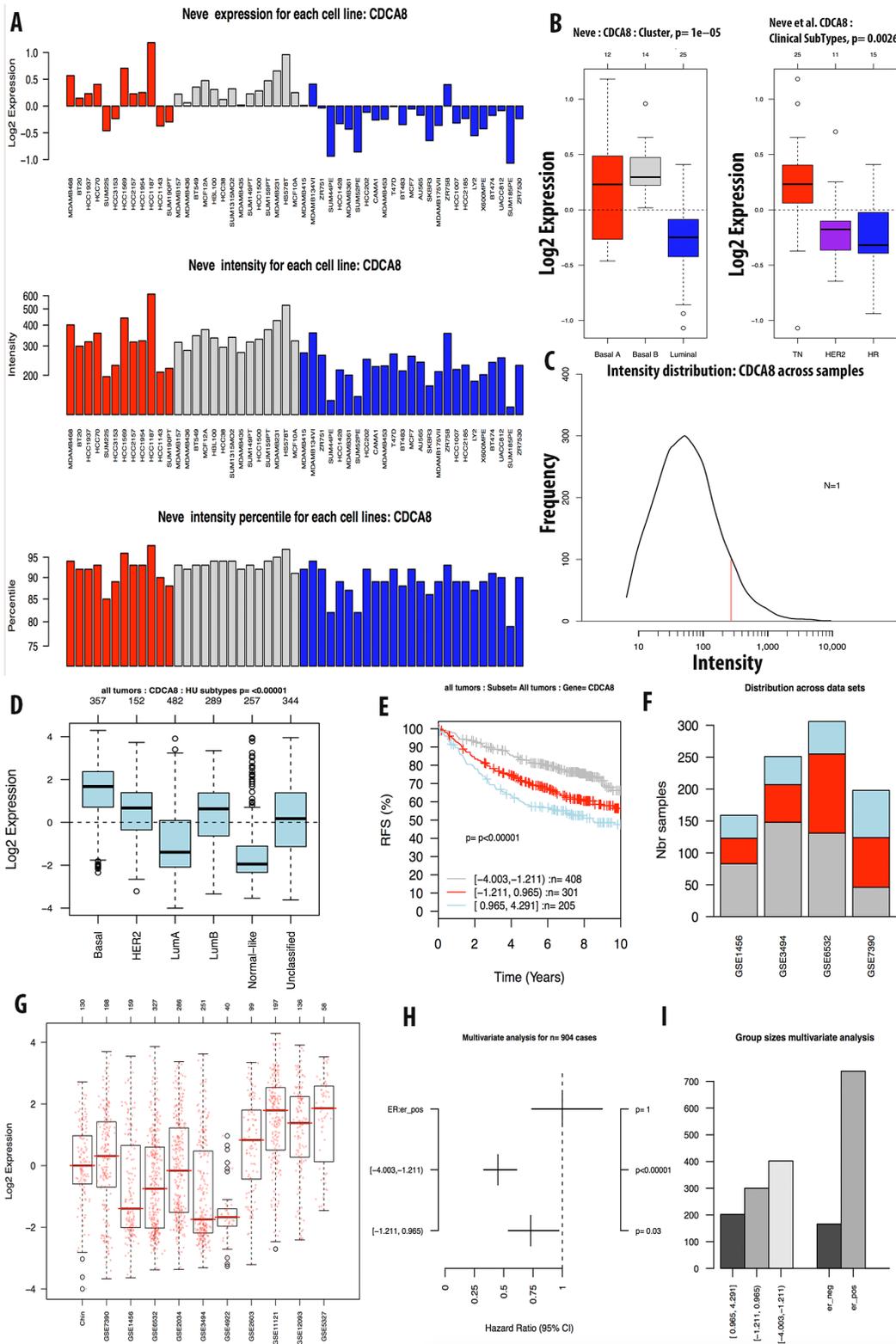


Figure 6: Analysis of CDCA8 expression in breast cancer tumor and cell line by GOBO database. Expression of CDCA8 in six subtypes of breast cancer tumors (A). RFS analysis of CDCA8 expression in tumor of patient survival with high expression in light blue, intermediate expression in red, and low expression in grey color (B). CDCA8 expression across datasets (C, D). Forest plot and bar chart of multivariate analysis on the expression of CDCA3 was displayed in (E, F). Expression of CDCA8 in breast cancer cell line using Neve et.al dataset was display in (G-I).

The copy number of CDCA5 was also high in breast and other cancers (Figure 3D). The CCLE analysis of CDCA5 in cancer cell lines revealed a high expression of this gene, at greater than nine-fold compared to normal control sample (Figure 4B).

The expression level of CDCA7 in various cancer cell lines was ten-fold greater to control sample with a high copy number (Figure 3E). The CCLE analysis for the CDCA8 expression in different cancer cell lines showed a high expression in breast and other types of cancers (Figure 4C). The copy number of CDCA8 was also high in breast and other cancers (Figure 3F). Further analysis of the expression of CDCA8 in tumors revealed a high expression level of this gene in the basal and TNBC subtypes. The expression, intensity, and intensity percentile of CDCA8 for each cell line was analyzed using the NEVE dataset, which is also displayed in Figure 6. The RFS of patients positively correlated with CDCA8 expression. The log₂ expression levels of CDCA8 in different datasets were displayed in Figure 6. The HR of the multivariate analysis for the 853 cases included tumor size, age, tumor grade, and node status of cancer patient, which are also illustrated (Figure 6).

Expression of CDCA genes and breast cancer patient overall survival

The overexpression of CDCA2 in breast cancer tissue was highly correlated with a poor prognosis for

breast cancer patients, with an HR of 1.36 (Figure 7A). A survival analysis of breast cancer patients with an overexpression of CDCA3 showed poor prognosis (HR = 1.59) (Figure 7B). CDCA4 expression was not significantly correlated with patient survival, with an HR of 1.14 (Figure 7C). This highly expressed of CDCA5 gene results in a poor prognosis value for breast cancer patients, with an HR of 1.46, and it significantly reduced patient survival over the 3-year and 5-years interval (Figure 7D). Analysis of CDCA7 expression and breast cancer patient survival revealed no significant correlation, with an HR of 1.35 (Figure 7E). The survival analysis of breast cancer patients with high CDCA8 expression levels showed a poor prognosis value, with an HR of 1.98 (Figure 7F).

Relapse-free survival analysis of breast cancer patients reveals a high correlation with CDCA3, CDCA5, and CDCA8

We further analyzed the overexpression of CDCA3, CDCA5, and CDCA8 with respect to patient RFS. In general, the high expression of these three genes resulted in a poor prognosis for the patient. A high expression of CDCA3 was highly correlated with patient survival under all breast cancer subtypes, with an HR of 1.59. In ER-positive, luminal A, and luminal B subtypes, the high expression of CDCA3 dramatically reduced the survival period for breast cancer patients. Interestingly,

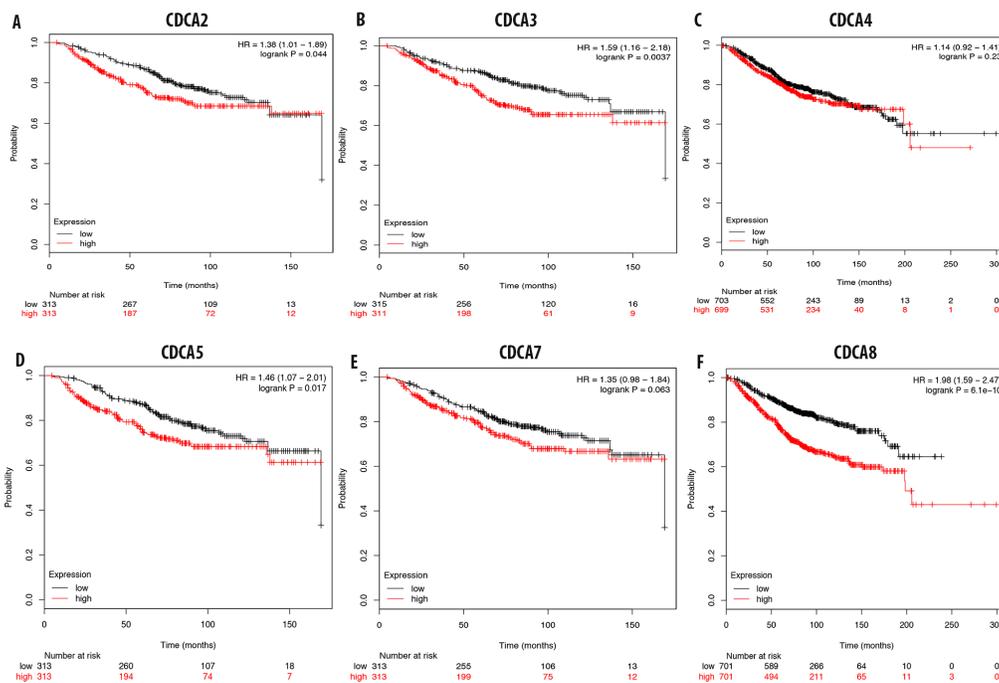


Figure 7: Correlation between CDCA genes (A-F) and overall survival of breast cancer patient. The over expression of CDCA2, CDCA3, CDCA5, and CDCA8 were highly correlated with patient survival with high HR which results in poor prognosis value. P-value<0.05 means significant different.

we found that patients undergoing chemotherapy treatment had a better prognosis relative to those who did not received this treatment, with HR values of 1.14 and 1.63, respectively. In the basal subtype, the high expression of CDCA3 was significantly correlated with a longer RFS, however in the HER2-positive subtype was not (Figure 8).

A high expression level of CDCA5 was correlated with a shorter RFS in breast cancer patients overall. With ER-positive, luminal A, and luminal B subtypes, a high expression of CDCA5 significantly shortened the RFS. Furthermore, patients without chemotherapy treatments could have a 2.85 times shorter RFS compared to the ones with chemotherapy (Figure 9).

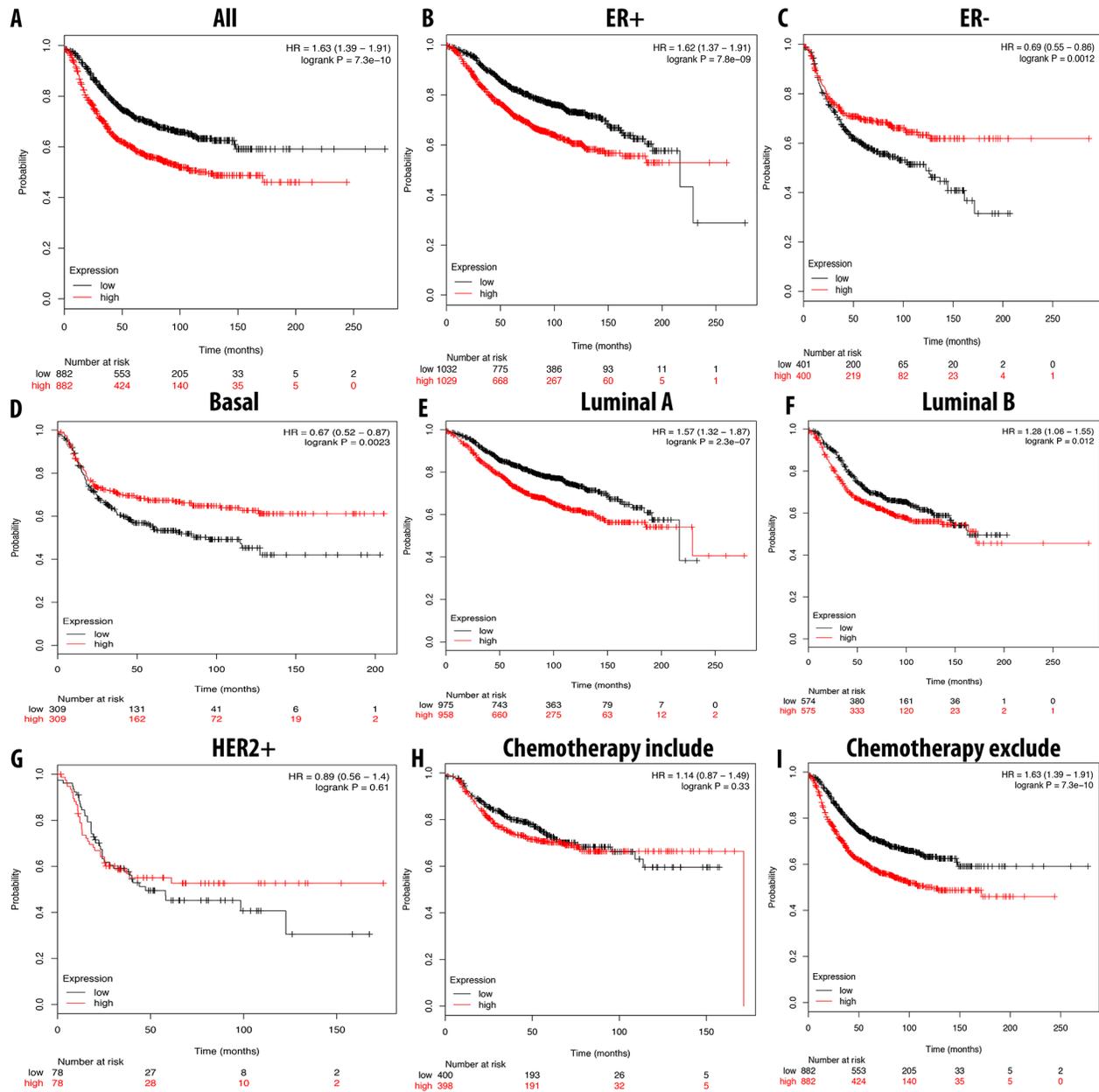


Figure 8: Correlation between expression of CDCA3 in RFS breast cancer patient (A-I). Breast cancer patient had significant shorter RFS with high expression of CDCA3 (A). High expression of CDCA3 was significantly associated with shorter RFS in ER positive but longer RFS in ER negative (B, C). In both luminal A and luminal B, high expression of CDCA3 was significantly indicated shorter value (E, F). High expression of CDCA3 was significantly linked to longer RFS in basal subtype but not in HER2 positive (D, G). Patient underwent chemotherapy had longer RFS relative to those who did not received chemotherapy treatment (H, I). P-value<0.05 means significant different.

In a similar expression pattern to CDCA3 and CDCA5, a high expression level of CDCA8 also correlated with a bad prognosis for breast cancer patients with a shorter RFS. In patients with luminal A, luminal B and ER-positive subtypes, a shorter RFS was highly correlated with CDCA8 overexpression. Patients undergoing chemotherapy treatment could prolong RFS, but not significantly (Figure 10).

Co-expression analysis of CDCA3, CDCA5, and CDCA8 revealed their expression networks in breast cancer

To further investigate the expression network for CDCA3, CDCA5, and CDCA8, co-expression analysis using clinical specimens was done with the Oncomine database. We found that CDCA3 expression was highly

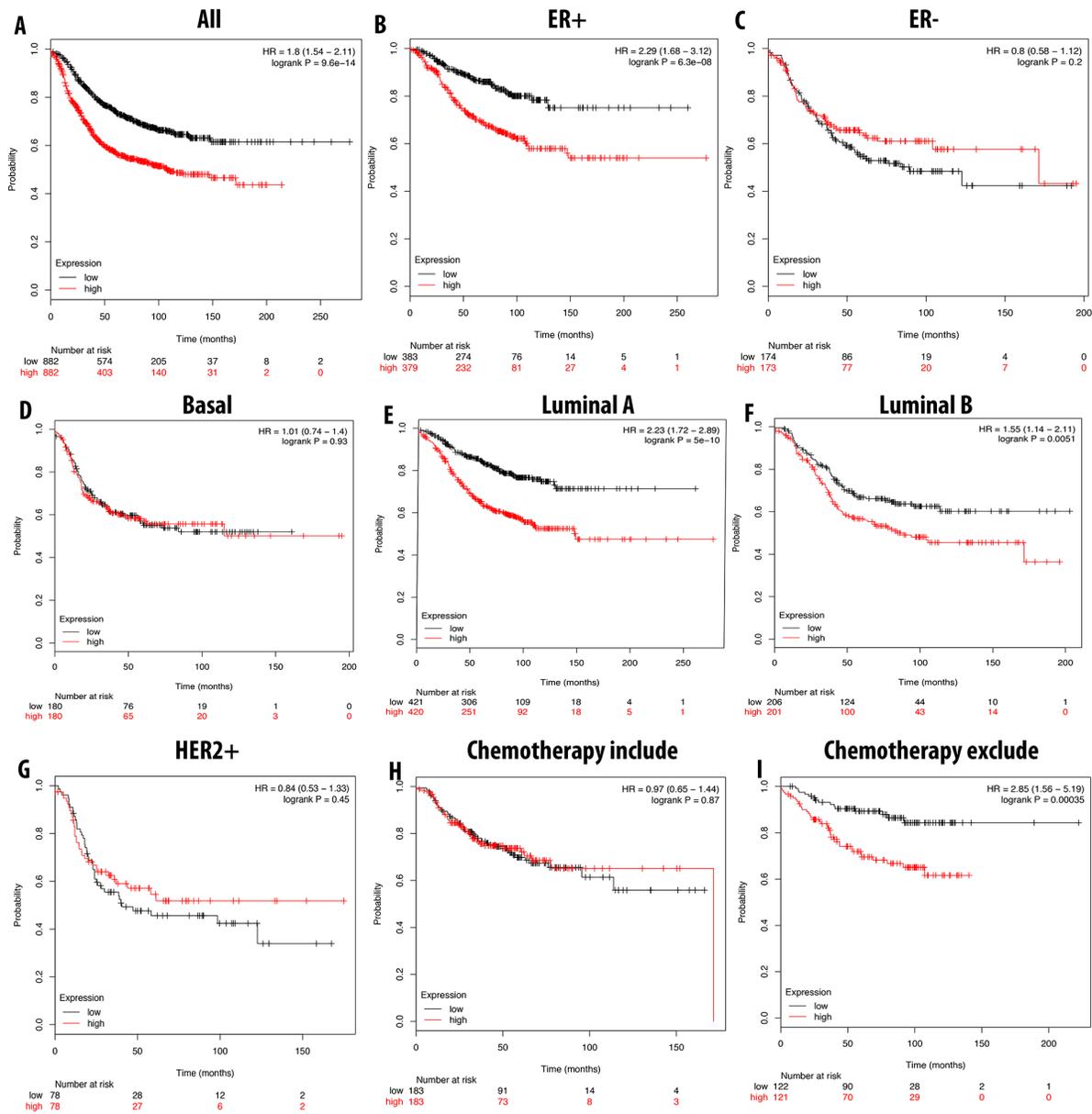


Figure 9: Correlation between expression of CDCA5 in RFS breast cancer patient (A-I). Breast cancer patient had significant shorter RFS with high expression of CDCA5 (A). High expression of CDCA5 was significantly associated with shorter RFS in ER positive but good prognosis in ER negative (B, C). In both luminal A and luminal B, high expression of CDCA5 was significantly indicated shorter value (E, F). High expression of CDCA3 was not significantly linked to longer RFS in HER2 positive but not in basal subtype (D, G). Patient underwent chemotherapy had longer RFS relative to those who did not received chemotherapy treatment (H, I). P-value<0.05 means significant different.

correlated with four genes, namely Cyclin B2 (CCNB2) ($R=0.89$), Cell Division Cycle 20 (CDC20) ($R=0.89$), Cyclin Dependent Kinase Inhibitor 3 (CDKN3) ($R=0.89$), and Cyclin B1 (CCNB1) ($R = 0.89$; Figure 11A). CDCA5 expression was correlated with two genes, namely budding uninhibited by benzimidazoles 1 (BUB1) ($R=0.89$) and Thyroid Hormone Receptor Interactor 13 (TRIP13)

($R = 0.89$; Figure 11B), while CDCA8 expression was correlated with BUB1 ($R = 0.928$) and CCNB1 ($R = 0.909$; Figure 11C).

The interaction network for CDCA3, CDCA5, and CDCA8 with their co-expressed genes was plotted using the STRING database for gene interaction (<https://string-db.org>) and GeneMANIA for gene networking and the

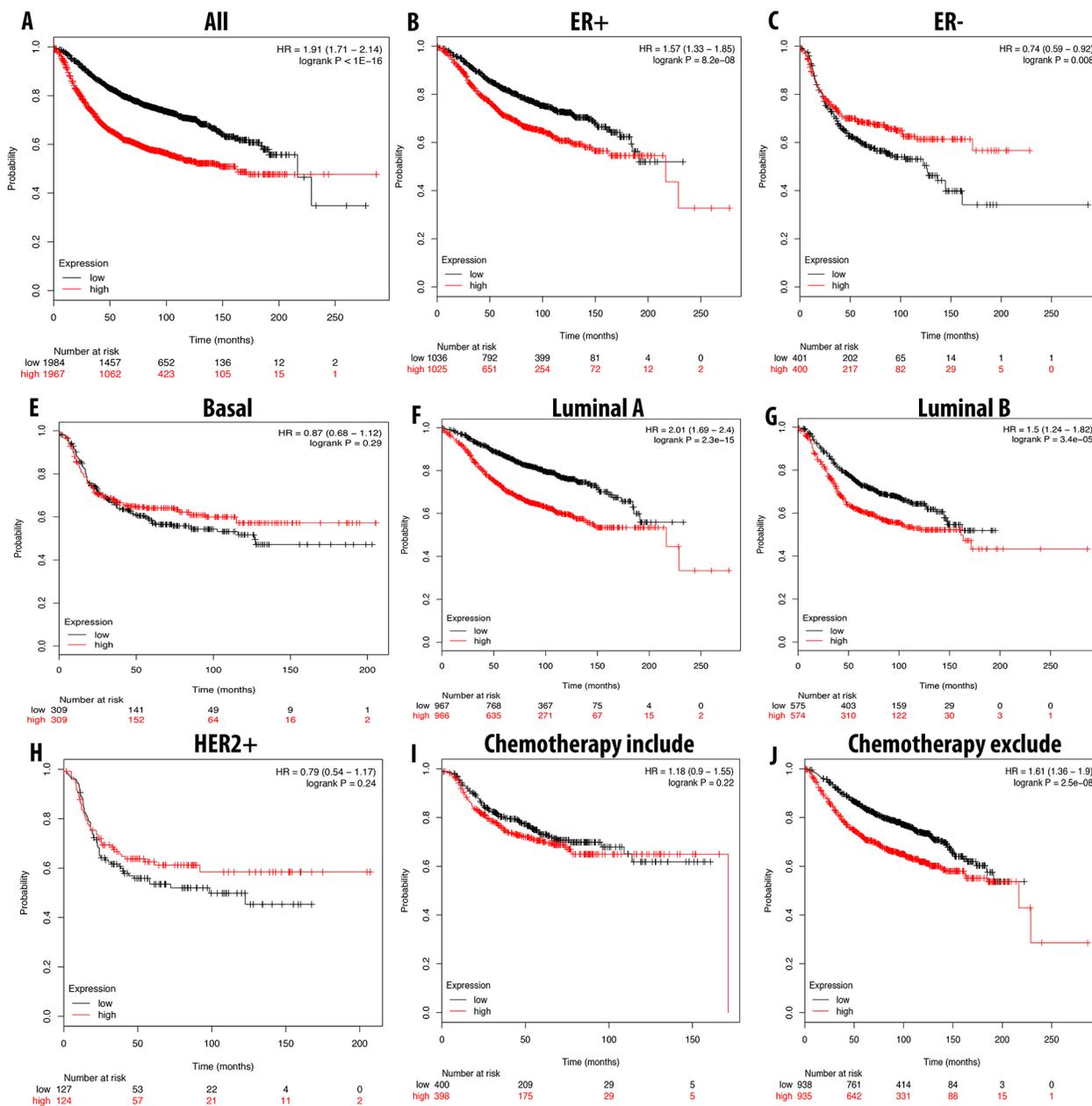


Figure 10: Correlation between expression of CDCA8 in RFS breast cancer patient (A-I). Breast cancer patient had significant shorter RFS with high expression of CDCA8 (A). High expression of CDCA8 was significantly associated with shorter RFS in ER positive but longer in ER negative (B, C). In both luminal A and luminal B, high expression of CDCA8 was significantly indicated shorter value (E, F). High expression of CDCA8 was significantly linked to longer RFS in basal subtype and HER2 positive (D, G). Patient underwent chemotherapy had longer RFS relative to those who did not received chemotherapy treatment (H, I). P-value<0.05 means significant different.

prediction of gene function. We established the functional protein interaction network for CDCA3, CDCA5, CDCA8, CCNB2, CDC20, CDKN3, BUB1, TRIP13, and CCNB1 by using known and predicted interactions (Figure 11D). The interaction of these genes was expanded to other related genes in their network through the use of the following additional traits: co-expression, physical interactions, pathways, shared protein domains, co-localization, and genetic interactions (Figure 11E).

DISCUSSION

Breast cancer has been found to be correlated with mutations and/or the overexpression of oncogenic genes. Finding new targets for breast cancer, particularly in specific subtypes, is extremely important for the prognosis and potential cure of this disease. In the present study, we found three members of the cell division cycle-associated gene family that had distinct mRNA expression in breast cancer tumors and cell lines, which are CDCA3, CDCA5, and CDCA8. The overexpression of these three genes correlated to the survival probability for a breast cancer patient in terms of the three-year and five-year survival periods. By analyzing the various tumor subtypes and cell lines of breast cancer, we found evidence for CDCA3, CDCA5, and CDCA8 involvement in breast cancer,

resulting in an overall poor prognosis. Further analysis of patient RFS with different subtypes of breast cancer revealed that patients with different intrinsic subtypes had poorer prognoses with a high HR.

CDCA3 plays an important role as a mitosis entry 1 trigger and controls cell cycle progression. Previously, CDCA3 was known to be involved in several types of cancer, such as prostate cancer, liver cancer, and oral squamous cell carcinoma [15–18]. In another bioinformatics study of 2158 full cancer transcriptomes from 163 diverse entities, CDCA3 was proven to be a novel gene involved in liver carcinogenesis [19]. In the current analysis, CDCA3 expression levels were high in invasive ductal breast carcinoma and were highly correlated with a low survival probability for breast cancer patients, leading to a poor prognosis. Our data revealed similar CDCA3 expression patterns to previous studies using whole transcription profiles of invasive ductal breast carcinoma obtained by either microarrays or RNA-sequence data. According to the previous studies, the overexpression of CDCA3 in invasive ductal breast carcinoma was likely to associate with oral carcinogenesis by decreasing the levels of cyclin-dependent kinase inhibitors, which resulted in cell cycle arrest at G1 [20, 21]. Altogether, the overexpression of CDCA3 is likely associated with cell cycle arrest at the G1 phase, a critical

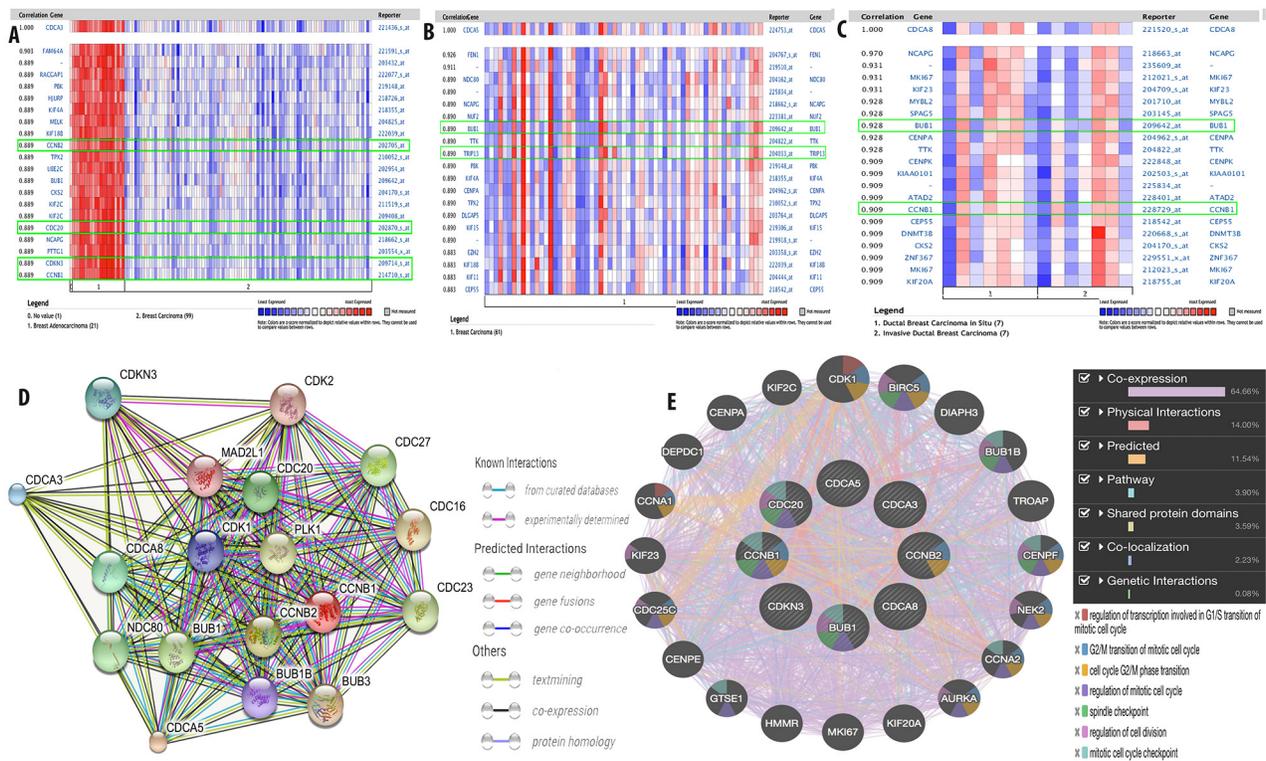


Figure 11: Co-expression analysis of CDCA3, CDCA5, and CDCA8 with other gene such as CCNB2, CDC20, CDKN3, CCNB1, BUB1, and TRIP13 in breast cancer (A-C). Functional proteins association network of CDCA3, CDCA5, CDCA8 with their co-expressed genes (D) and function (E) was done by STRING database and GENEMANIA.

checkpoint for cell division, resulting in a chain reaction of descending processes that likely leads to the development of cancers. In addition, the survival analyses revealed that the expression of CDCA3 in many breast cancer subtypes highly correlated with bad prognoses. In short, CDCA3 could be considered as a potential target for these breast cancer subtypes.

CDCA5 is also considered as oncogene since its overexpression has been found in many types and subtypes of cancer [22–25]. CDCA5 plays a crucial role in DNA repair [22], and is involved in the process of sister-chromatid cohesion and separation [26]. A poor prognosis for non-small cell lung cancer was linked to CDCA5 overexpression [12]. In another study, CDCA5 overexpression was linked to G1-S transition malfunction in urinary bladder urothelial carcinoma [22]. Furthermore, a method for lung cancer and/or esophageal cancer treatment and prevention based on the overexpression of CDCA5 was developed and patented, proving the possible application of CDCA5 to cancer therapeutics [24]. In this study, we found CDCA5 overexpression dramatically decreased the survival probability of breast cancer patients to lower than 0.5 in the three-year survival rate. In addition, RFS patients with different subtypes of breast cancer, ER-positive, luminal A, and luminal B, had a poorer prognosis. On the whole, CDCA5 could be considered as a target for breast cancer, particularly invasive ductal breast carcinoma.

CDCA8, a regulator of cell mitosis, was shown to be associated with lung cancer when it was phosphorylated at four positions, Ser¹⁵⁴, Ser²¹⁹, Ser²⁷⁵, and Thr²⁷⁸, by aurora kinase B [27]. One meta-analysis using public microarray data and immunohistochemistry revealed that the overexpression of CDCA8 in breast cancer, especially TNBC, reduced patient survival [28]. Our data showed that CDCA8 had a high expression level in the male breast carcinoma, invasive lobular breast carcinoma, invasive ductal breast carcinoma, and invasive breast carcinoma subtypes. Moreover, highly expressed CDCA8 was also associated with an extremely low survival probability and a poor prognosis for patients with a probability lower than 0.4 at the five-year interval. RFS patients with ER-positive, luminal A, and luminal B subtypes had a poorer prognosis than the other subtypes. Thus, these data suggested the potential role of CDCA8 as a treatment target in these subtypes. In conclusion, three members of the cell division cycle-associated gene family, CDCA3, CDCA5, and CDCA8, displayed distinct overexpression in breast cancer in both tumors and cancer cell lines, and this overexpression was associated with a poor prognosis for the breast cancer patient with a low survival probability. These three genes could be considered as potential targets for breast cancer treatment and prevention.

MATERIALS AND METHODS

Oncomine database analysis

The CDCA family mRNA expression level was analyzed by the Oncomine database using public microarray as well as RNA-sequence database [29, 30]. This method has been clearly described in our previous studies [31–34]. Briefly, the names of the CDCA genes (CDCA2, CDCA3, CDCA4, CDCA5, CDCA7, and CDCA8) were keyed into the search box with the threshold for the p-value set to < 0.001, the fold change > 1.5, and the gene rank percentage < 10% when comparing cancerous tissue with normal type-matched tissue. Co-expression analysis was performed with clinical breast cancer samples. All calculations were set to the default settings, including the p-value, the two-sided *t*-test to compare the mean mRNA expression level between the cancer and control groups, and the multiple testing correction for the p-value to avoid the false discovery of genes with a small p-value that were not significant.

Gene expression-based outcome for breast cancer online (GOBO) database analysis

The mRNA expression level of the CDCA gene family, specifically in breast cancer, was analyzed using the GOBO database [35]. The GOBO database allows users to perform a gene set analysis in four modes, a tumor mode, a cell line mode, a co-expressed gene mode, and a sample prediction mode. In the present study, we applied the tumor and cell line modes for the analysis of our target genes. In the tumor mode, we used the default setting from GOBO for recurrence-free survival (RFS) as the endpoint of the Kaplan-Meier survival analysis and used the estrogen receptor (ER)-status for multivariate parameters. In the cell line mode, we used the breast cancer cell line from Neve et al. database [14].

The cancer cell line encyclopedia (CCLE) database analysis

Transcriptomic expression levels of the CDCA gene family across multiple types of cancer cell lines were analyzed using the CCLE database [36]. From the CCLE, two types of data could be acquired, namely gene expression and copy number in the datasets. Gene expression in the dataset provides information about an mRNA expression level in robust multichip average (RMA) log₂ form across all of the cancers from the CCLE database. Meanwhile, the gene expression and copy number in the datasets of genes from different types of cancer such as liver and lung cancer can also be acquired from CCLE database.

Kaplan-meier plot database analysis for survival probability

The correlation between mRNA expression levels of the CDCA gene family and the survival probability of breast cancer patients was analyzed using the Kaplan-Meier plot database as previously described [16]. In brief, we input the gene name of the CDCA family into the gene symbol search box and adjusted the survival type to RFS. We kept all of the default settings of the Kaplan-Meier plot database, such as the ER status, human epidermal growth factor receptor 2 (HER2) status, intrinsic subtype, and chemotherapy status, then plotted the Kaplan-Meier curve [37].

STRING database for functional protein association network

The Search Tool for the Retrieval of Interacting Genes (STRING) database (<https://string-db.org>) was used to create a network of the protein interactions for the interested genes [38]. We used the multiple protein to input data and chose *Homo sapiens* as the data source. All default settings were kept for further analysis.

GeneMANIA database for predicting the target gene function

Gene-association networks were made for CDCA3, CDCA5, and CDCA8 using GeneMANIA (<http://genemania.org>) [39]. The input genes were in stripped circles. The various proteins were colored based on their involvement in specific processes, such as “regulation of transcription involved in G1/S transition mitotic cell cycle”, “G2/M transition of mitotic cell cycle”, “regulation of mitotic cell cycle”, “spindle checkpoint”, “regulation of cell division”, and “mitotic cell cycle checkpoint”. The gene interaction network was created by co-expression, physical interaction, predicted interaction, pathway, shared protein domains, co-localization, and genetic interactions.

Abbreviations

Cell division cycle-associated protein: CDCA
Cancer Cell Line Encyclopedia Database: CCLE
Gene expression-based Outcome for Breast Cancer Online database: GOBO
The Cancer Genome Atlas: TCGA
Distant metastasis free survival: DMFS
Relapse free survival: RFS
Triple-negative breast cancer (sometimes abbreviated: TNBC)
Estrogen receptor: ERHormone receptor: HR
Human epidermal growth factor receptor 2: HER2

Author contributions

Conceptualization: NNP; CYW; KLL; CFC; CCC; PLH; HGY; YCL; Methodology: NNP; CYW; Software: NNP; CYW; Validation: NNP; CYW; CFC; CCC; Formal analysis: NNP; CYW Investigation: NNP; CYW; KLL; CFC; CCC; Resources: PLH; YCL; Data curation: NNP; CYW; KLL; CFC; CCC Writing-original draft: NNP; CYW; YCL; Writing – review and editing: NNP; CYW; CFC; CCC; PLH; HGY; YCL; Visualization: NNP; CYW; Supervision: PLH; YCL; Project administration: PLH; YCL; Funding acquisition: PLH; YCL; CYW

ACKNOWLEDGMENTS AND FUNDING

We thanks to National Science Council of the Executive Yuan (NSC 104-2320-B-034-003, NSC 105-2320-B-034-001 to YCL); Ministry of Science and Technology (104-2917-I-006-002 to CYW) for supporting this study. The funding bodies do not have role in study design, collection, analysis, interpretation and writing the manuscript.

DECLARATIONS

Ethics approval and consent to participate

not applicable

Consent for publication

not applicable

CONFLICTS OF INTEREST

The authors declare that they have no conflicts of interest.

REFERENCES

1. Preston-Martin S, Pike MC, Ross RK, Jones PA, Henderson BE. Increased cell division as a cause of human cancer. *Cancer Res.* 1990; 50: 7415-21.
2. Altieri DC. Survivin, versatile modulation of cell division and apoptosis in cancer. *Oncogene.* 2003; 22: 8581-9.
3. Vader G, Lens SM. The Aurora kinase family in cell division and cancer. *BiochimBiophys Acta.* 2008; 1786: 60-72.
4. Collins I, Garrett MD. Targeting the cell division cycle in cancer: CDK and cell cycle checkpoint kinase inhibitors. *Curr Opin Pharmacol.* 2005; 5: 366-73.
5. Együd L, Szent-Györgyi A. Cell division, SH, ketoaldehydes, and cancer. *Proc Natl Acad Sci U S A.* 1966; 55: 388-93.

6. Harao M, Hirata S, Irie A, Senju S, Nakatsura T, Komori H, Ikuta Y, Yokomine K, Imai K, Inoue M. HLA-A2-restricted CTL epitopes of a novel lung cancer-associated cancer testis antigen, cell division cycle associated 1, can induce tumor-reactive CTL. *Int J Cancer*. 2008; 123: 2616-25.
7. Peng A, Lewellyn AL, Schiemann WP, Maller JL. Repo-man controls a protein phosphatase 1-dependent threshold for DNA damage checkpoint activation. *Curr Biol*. 2010; 20: 387-96.
8. Vagnarelli P. (2014). Repo-man at the intersection of chromatin remodelling, DNA repair, nuclear envelope organization, and cancer progression. *Cancer Biology and the Nuclear Envelope*: Springer, pp. 401-14.
9. Yoshida K. Cell-cycle-dependent regulation of the human and mouse Tome-1 promoters. *FEBS Lett*. 2005; 579: 1488-92.
10. Hayashi R, Goto Y, Ikeda R, Yokoyama KK, Yoshida K. CDCA4 is an E2F transcription factor family-induced nuclear factor that regulates E2F-dependent transcriptional activation and cell proliferation. *J Biol Chem*. 2006; 281: 35633-48.
11. Tategu M, Nakagawa H, Hayashi R, Yoshida K. Transcriptional co-factor CDCA4 participates in the regulation of JUN oncogene expression. *Biochimie*. 2008; 90: 1515-22.
12. Nguyen MH, Koinuma J, Ueda K, Ito T, Tsuchiya E, Nakamura Y, Daigo Y. Phosphorylation and activation of cell division cycle associated 5 by mitogen-activated protein kinase play a crucial role in human lung carcinogenesis. *Cancer Res*. 2010; 70: 5337-47.
13. Higuchi T, Uhlmann F. Cell cycle: passenger acrobatics. *Nature*. 2003; 426: 780-1.
14. Neve RM, Chin K, Fridlyand J, Yeh J, Baehner FL, Fevr T, Clark L, Bayani N, Coppe JP, Tong F. A collection of breast cancer cell lines for the study of functionally distinct cancer subtypes. *Cancer Cell*. 2006; 10: 515-27.
15. Ayad NG, Rankin S, Murakami M, Jebanathirajah J, Gygi S, Kirschner MW. Tome-1, a trigger of mitotic entry, is degraded during G1 via the APC. *Cell*. 2003; 113: 101-13.
16. Hu Q, Fu J, Luo B, Huang M, Guo W, Lin Y, Xie X, Xiao S. OY-TE5-1 may regulate the malignant behavior of liver cancer via NANOG, CD9, CCND2 and CDCA3: a bioinformatic analysis combine with RNAi and oligonucleotide microarray. *Oncol Rep*. 2015; 33: 1965-75.
17. Uchida F, Uzawa K, Kasamatsu A, Takatori H, Sakamoto Y, Ogawara K, Shiiba M, Tanzawa H, Bukawa H. Overexpression of cell cycle regulator CDCA3 promotes oral cancer progression by enhancing cell proliferation with prevention of G1 phase arrest. *BMC Cancer*. 2012; 12: 321.
18. Chen Y, Wu F, Parvez F, Ahmed A, Eunos M, McClintock TR, Patwary TI, Islam T, Ghosal AK, Islam S. Arsenic exposure from drinking water and QT-interval prolongation: results from the health effects of arsenic longitudinal study. *Environ Health Perspect*. 2013; 121: 427.
19. Itzel T, Scholz P, Maass T, Krupp M, Marquardt J, Strand S, Becker D, Staib F, Binder H, Wang X. Translating bioinformatics in oncology: guilty by profiling meta-analysis and identification of KIF18B and CDCA3 as novel driver genes in liver carcinogenesis. *Z Gastroenterol*. 2013; 51: P_4_26.
20. Arnutti P, Kotepui M, Asanprakit W, Punyarit P, Chavalitshewinkoon-Petmitr P, Harnroongroj T, Petmitr S. Determination of whole transcription profiles and specific pathways in invasive ductal breast carcinoma. *Int J Clin Exp Pathol*. 2013; 6: 1112-20.
21. Mercier I, Casimiro MC, Wang C, Rosenberg AL, Quong J, Minkeu A, Allen KG, Danilo C, Sotgia F, Bonuccelli G. Human breast cancer-associated fibroblasts (CAFs) show caveolin-1 down-regulation and RB tumor suppressor functional inactivation: implications for the response to hormonal therapy. *Cancer Biol Ther*. 2008; 7: 1212-25.
22. Chang IW, Lin VC, He HL, Hsu CT, Li CC, Wu WJ, Huang CN, Wu TF, Li CF. CDCA5 overexpression is an indicator of poor prognosis in patients with urothelial carcinomas of the upper urinary tract and urinary bladder. *Am J Transl Res*. 2015; 7: 710.
23. Cheng WY, Yang TH, Anastassiou D. Development of a prognostic model for breast cancer survival in an open challenge environment. *Sci Transl Med*. 2013; 5: 181ra50.
24. Nakamura Y, Daigo Y, Nakatsuru S. (2008). Cancer-related genes, cdca5, epha7, stk31 and wdhd1. Google Patents.
25. Ho DW, Kai AK, Ng IO. TCGA whole-transcriptome sequencing data reveals significantly dysregulated genes and signaling pathways in hepatocellular carcinoma. *Front Med*. 2015; 9: 322-30.
26. Zhang N, Pati D. Sororin is a master regulator of sister chromatid cohesion and separation. *Cell Cycle*. 2012; 11: 2073-83.
27. Hayama S, Daigo Y, Yamabuki T, Hirata D, Kato T, Miyamoto M, Ito T, Tsuchiya E, Kondo S, Nakamura Y. Phosphorylation and activation of cell division cycle associated 8 by aurora kinase B plays a significant role in human lung carcinogenesis. *Cancer Res*. 2007; 67: 4113-22.
28. Jiao D, Lu Z, Qiao J, Yan M, Cui S, Liu Z. Expression of CDCA8 correlates closely with FOXM1 in breast cancer: public microarray data analysis and immunohistochemical study. *Neoplasia*. 2014; 62: 464-9.
29. Rhodes DR, Yu J, Shanker K, Deshpande N, Varambally R, Ghosh D, Barrette T, Pander A, Chinnaiyan AM. ONCOMINE: a cancer microarray database and integrated data-mining platform. *Neoplasia*. 2004; 6: 1-6.
30. Rhodes DR, Kalyana-Sundaram S, Mahavisno V, Varambally R, Yu J, Briggs BB, Barrette TR, Anstet MJ, Kincead-Beal C, Kulkarni P. Oncomine 3.0: genes, pathways, and networks in a collection of 18,000 cancer gene expression profiles. *Neoplasia*. 2007; 9: 166-80.
31. Wang CY, Lai MD, Phan NN, Sun Z, Lin YC. Meta-analysis of public microarray datasets reveals voltage-gated calcium

- gene signatures in clinical cancer patients. *PLoS One*. 2015; 10: e0125766.
32. Wang CY, Shahi P, Huang JT, Phan NN, Sun Z, Lin YC, Lai MD, Werb Z. Systematic analysis of the achaete-scute complex-like gene signature in clinical cancer patients. *Mol Clin Oncol*. 2017; 6: 7-18.
 33. Phan NN, Wang CY, Chen CF, Sun Z, Lai MD, Lin YC. Voltage-gated calcium channels: novel targets for cancer therapy. *Oncol Lett*. 2017; 14: 2059-74.
 34. Phan NN, Huynh TT, Lin YC. Hyperpolarization-activated cyclic nucleotide-gated gene signatures and poor clinical outcome of cancer patient. *Transl Cancer Res*. 2017; 6: 698-708.
 35. Ringnér M, Fredlund E, Häkkinen J, Borg Å, Staaf J. GOBO: gene expression-based outcome for breast cancer online. *PLoS One*. 2011; 6: e17911.
 36. Barretina J, Caponigro G, Stransky N, Venkatesan K, Margolin AA, Kim S, Wilson CJ, Lehár J, Kryukov GV, Sonkin D. The Cancer Cell Line Encyclopedia enables predictive modelling of anticancer drug sensitivity. *Nature*. 2012; 483: 603-7.
 37. Györfy B, Lanczky A, Eklund AC, Denkert C, Budczies J, Li Q, Szallasi Z. An online survival analysis tool to rapidly assess the effect of 22,277 genes on breast cancer prognosis using microarray data of 1,809 patients. *Breast Cancer Res Treat*. 2010; 123: 725-31.
 38. Szklarczyk D, Morris JH, Cook H, Kuhn M, Wyder S, Simonovic M, Santos A, Doncheva NT, Roth A, Bork P. The STRING database in 2017: quality-controlled protein-protein association networks, made broadly accessible. *Nucleic Acids Res*. 2017; 45: D362-8.
 39. Warde-Farley D, Donaldson SL, Comes O, Zuberi K, Badrawi R, Chao P, Franz M, Grouios C, Kazi F, Lopes CT. The GeneMANIA prediction server: biological network integration for gene prioritization and predicting gene function. *Nucleic Acids Res*. 2010; 38: W214-20.