Comprehensive Analysis of the Expression and Prognostic Significance of the CENP Family in Breast Cancer

Xueliang Liu, Yunjiang Liu

Breast Cancer Center, The Fourth Hospital of Hebei Medical University, Shijiazhuang, 050000, Hebei, People's Republic of China

Correspondence: Yunjiang Liu, Tel +86-13703297890, Email lyj818326@outlook.com

Background: Centromere proteins (CENPs) are a set of protein-coding genes involved in the transient assembly of the kinetochore which occurs during mitosis. This study intended to clarify the expression patterns, prognosis and potential mechanisms of CENPs in breast cancer (BC).

Methods: Coexpedia was used to screen GEO datasets and PubMed articles related to CENPs and BC. CENPs expressions, prognosis and alteration were analyzed by Oncomine, Ualcan and Kaplan Meier plotter and cBioPortal. The correlation and interaction of CENPs was performed by Breast Cancer Gene-Expression Miner, GeneMANIA and STRING portal. Gene Ontology (GO) and Kyoto Encyclopedia of Genes and Genomes (KEGG) enrichment analysis were conducted to clarify the functional roles of CENPs. CENPF, E, U, A, N, I, K, W, M, L were selected for further analysis.

Results: All CENPs were highly expressed in BC compared to normal tissue. High expression of CENPF, E, U, A, N, I, W, M, L and CENPF, E, U, A, N, I, M correlated with worse relapse free survival (RFS) and worse overall survival (OS), respectively. All of 10 CENPs indicated positive correlations and complex interactions between each other at mRNA expression and protein level. CENPs were enriched GO terms mainly in centromere complex assembly and KEGG terms in progesterone-mediated oocyte maturation, cell cycle and oocyte meiosis.

Conclusion: The 10 CENPs could be diagnostic biomarkers and all of them except CENPK can be used as prognosis biomarkers in BC. CENPs play an oncogenic role and may be the potential therapy targets of treatment for BC patients.

Keywords: centromere proteins, breast cancer, diagnosis, prognosis, bioinformatics analysis

Introduction
Breast cancer (BC) is the most commonly diagnosed cancer and a major contributor to cancer-related deaths among women worldwide. Nearly 10% of cases are due to inherited gene alterations, such as those in BRCA1 and BRCA2. BC is divided into four intrinsic subtypes: luminal A, luminal B, HER2-enriched and basal-like. Additionally, classical clinical prognostic biomarkers such as ER, PR, and HER-2 have played a critical role in the classification of patients who are likely to benefit from endocrine therapy or targeted therapy. However, due to tumor heterogeneity, the endocrine therapy, chemotherapy and target therapy resistance reduce the effect of treatment. Therefore, new potential biomarkers are needed to compensate for the limitations of such classical biomarkers. Several studies have already shown some approach or index may be able to predict the survivability of breast cancer patients based on their subtypes or the clinical parameters, such as the size of the tumor and the number of lymph nodes, and aim to identify new biomarkers. The new biomarkers could be used as prognostic indicators to effectively improve prognostic prediction and facilitate personalized treatment.

Centromere proteins (CENPs) comprise a protein family associated with the centromere-kinetochore complex and play a critical role in kinetochore function and chromosome segregation in mitosis. Centromeres recruit a protein complex named the kinetochore, which facilitates the biorientation of replicated chromosome pairs to a meiotic spindle...
structure. The first description of autoantibodies to CENPs was published in 1980. In the following years, 19 additional human CENP family members (CENPF, CENPE, CENPU, CENPA, CENPN, CENPL, CENPK, CENPM, CENPL, CENPB, CENPO, CENPP, CENPR (ITGB3BP), CENPH, CENPT, CENPC, CENPX) were identified to form a constitutive centromere-associated network (CCAN). Numerous studies have verified that the cross-talk of CENPs is involved in many molecular processes, such as mitosis and kinetochore assembly and disassembly. It has been previously reported that the inhibition of CENPE and CENPF can inhibit the proliferation of BC cell lines. Furthermore, high expression of CENPA and CENPI indicate a poor prognosis and predict relapse in ER-positive BC. However, despite these previously shown associations between CENPs and BC, the expression characteristics and diagnostic and prognostic implications of the entire CENP family have not been assessed in BC. Furthermore, the function, mechanism and prognostic significance of CENP family genes in BC remain unclear.

In this study, we aimed to comprehensively evaluate the expression status and diagnostic and prognostic value of CENPs in BC and assess their biological mechanisms by employing bioinformatics analysis tools.

Materials and Methods

Coexpedia Analysis
The coexpression network of CENPs in the Gene Expression Omnibus (GEO) database was analyzed using Coexpedia (http://www.coexpedia.org/), which was utilized to identify the CENPs related to breast cancer. We visualized the network of CENP genes automatically (14 co-expressed genes in all 19 queried genes) and the score for each gene was calculated as the summation of edge weights (log-likelihood score) to all connected genes in the network. Next, we screened 10 CENPs related to the medical subject headings (MeSH) terms “breast neoplasms”, which were second-ranked by calculating the contribution of their associated GEO series in the total sum of LLS scores.

Oncomine Analysis
The expression patterns of the CENP family in pan-cancer and normal samples were assessed with the Oncomine gene expression profile microarray datasets (https://www.oncomine.org/), which is a publicly accessible online cancer microarray database to facilitate the genome-wide expression analyses. Student’s t-test was used to compare the mRNA expression levels of CENP family genes between normal and cancer samples, and false discovery rates was used as a corrected measure of significance. The thresholds for P value, fold change and expressed gene rank were set at <0.01, 2 and the top 10%, respectively.

UALCAN Dataset Analysis
In order to analyze the expression of CENP family genes in BC tissues and normal breast tissues, we visualized it by UALCAN, which is a publicly available web portal (http://ualcan.path.uab.edu) that offers online analysis of data from The Cancer Genome Atlas (TCGA). The expression level of CENP members was normalized as transcripts per million reads, and differences with a P value of less than 0.01 according to Student’s t-test were considered to be significant.

The Kaplan–Meier Plotter Analysis
The prognostic value of CENP mRNA expression was evaluated using an online database, Kaplan–Meier Plotter (www.kmplot.com), which contains gene expression data and survival information of BC patients (http://kmplot.com/analysis/index.php?p=service&cancer=breast). To analyze the relapse-free survival (RFS) and overall survival (OS) of patients with BC, patient samples were split into two groups according to the median expression (high versus low expression) of each gene and assessed via Kaplan–Meier survival analysis; plots were drawn, and hazard ratios (HRs), 95% confidence intervals (CIs) and log-rank p values were calculated using this method.

cBioPortal Analysis
The breast invasive carcinoma dataset (The Cancer Genome Atlas, Firehouse legacy), which includes data from 1108 cases with pathology reports was selected for further analysis CENPs genetic alterations and the association between
genetic alterations and survival outcome in BC patients using cBioPortal. Genetic alterations in CENP family and their association with OS and DFS of breast cancer patients were visualized by Kaplan–Meier plots, and the Log rank test was performed to identify the significance of the diversity between the survival curves, and the differences with a P value of less than 0.05 were considered statistically significant.

**Breast Cancer Gene-Expression Miner (bc-GenExMiner), GeneMANIA, and STRING Analysis**

In order to explore the correlation of CENPs, we evaluated and visualized it by bc-GenEx-Miner v4.5, a statistical mining tool of published annotated BC transcriptomic data. The correlation between each pair of CENP genes was assessed by Pearson’s correlation with “correlation” module.

GeneMANIA and STRING were used to identify the interactions of CENPs at the gene and protein levels. GeneMANIA ([http://genemania.org](http://genemania.org)) is a Cytoscape plugin for identifying associations between genes within gene sets. STRING ([https://string-db.org/](https://string-db.org/)) is a database that searches for interactions between proteins and includes both physical interactions between proteins and functional correlations between proteins.

**Gene Ontology (GO) Annotation and Kyoto Encyclopedia of Genes and Genomes (KEGG) Pathway Enrichment Analysis**

To reveal the functions of CENPs, ClusterProfiler package in R software (version 4.0.2) was used to conduct GO enrichment and KEGG pathway enrichment analyses. The GO terms were grouped into three categories: biological process (BP), cellular component (CC) and molecular function (MF). An adjusted P < 0.05 was considered statistically significant.

**Results**

**Identification of the 10 CENPs Related to BC**

There were 14 genes in all 19 queried CENP genes were co-expressed and are shown in Figure 1. For further analyze the CENPs associated with BC, we filtered only 10 genes, and the other 9 genes were not included in the network. Interestingly, the 10 CENP genes related to BC were the top 10 log-likelihood score genes in the network (Table S1).

The 10 identified CENPs were CENPF, CENPE, CENPU, CENPA, CENPN, CENPI, CENPK, CENPW, CENPM, and CENPL (Figure 1), which are marked by thick blue borders and connection lines of the network.

**The mRNA Expression of CENPs in BC**

Next, the expression levels of the 10 CENPs in various cancer types and the corresponding normal tissues were assessed using the Oncomine database. The results showed that the expression of CENP genes was different in all types of cancer tissues and their matched normal tissues (Figure 2). CENPs were overexpressed in almost all cancers, especially BC, lung cancer, colorectal cancer, and sarcoma. Moreover, all the CENPs, especially CENPF, CENPE, CENPU, CENPA, CENPN, CENPM, and CENPL, were overexpressed in BC tissue compared to the corresponding normal tissues. CENPF, CENPE, CENPU, CENPA, and CENPM were the most commonly reported CENPs in breast cancer datasets.

The UALCAN database was used to verify the expression of CENPs in BC and matched normal tissues. The results showed that all of the top 10 CENPs identified with Coexpedia were overexpressed in BC samples compared to normal tissue samples (p < 0.0001 for all; Figure 3).

**The Prognostic Roles of CENP Genes in BC**

As shown by Kaplan–Meier Plotter, the overexpression of all CENPs was associated with worse RFS (p < 0.01), except that of CENPK (HR = 1.12, p = 0.13). Moreover, the p-values of the other 9 CENPs were well below 0.01 (Figure 4). Additionally, unfavorable OS in patients was associated with high CENPF (HR = 1.79, p=1.6E-09), CENPE (HR = 1.44, p = 0.00015), CENPU (HR = 1.64, p=2.6E-07), CENPA (HR = 1.57, p=2.8E-06), CENPN (HR = 1.47, p=1.8E-05), CENPI (HR = 1.25, p = 0.022) and CENPM (HR = 1.49, p=3.3E-05) expression. There was no significant difference in
OS among patients with high versus low expression of CENPK (HR = 1.11, p = 0.46), CENPW (HR = 1.2, p = 0.18), or CENPL (HR = 1.15, p = 0.31) (Figure 4). Interestingly, for CENPK, when the best cutoff instead of the median expression value was used to split the patients into two groups, the difference in RFS between groups was significant (p-value=0.0059) (Figure S1).

CENP Genetic Alterations in BC Patients
cBioPortal was used to analyze CENP genetic alterations in BC. The following alteration frequencies were identified in the TCGA provisional dataset: CENPF, 16%; CENPE, 5%; CENPU, 4%; CENPA, 4%; CENPN, 10%; CENPI, 4%; CENPK, 4%; CENPW, 5%; CENPM, 2.9%; CENPL, 14% (Figure 5A). The different types of genetic alterations of CENP genes are mainly in amplification, deep deletion, mRNA high and missense mutations (Figure 5A). Next, we assessed the difference in survival between patients with and without CENP genetic alterations. There was no significant difference between groups in terms of DFS (p = 0.248) or OS (p = 0.321) (Figure 5B and C).

Correlation and Interaction Analyses of CENPs in BC Patients
We calculated the correlations between CENP expression levels in BC based on bc-GenExMiner v4.5. Pearson’s correlation analysis indicated significant positive correlations between each pair of CENPs (all p-values <0.0001) (Figure 6A).

GeneMANIA was used to analyze the interactions of CENPs at the gene level. CENPU had the densest physical interaction network among CENPs but did not interact with CENPE and CEMPW (Figure 6B). Except for CEMPW, all the other CENPs shared at least one pathway. CENPA was the only CENP co-expressed with all other assessed CENPs. In addition, only CENPF and CENPE colocalized. Moreover, we used GeneMANIA to visualize the gene network of the 19 CENPs and the network genes was used to predict the function of CENPs by further GO and KEGG analysis.
We identified the interactions of CENPs at the protein level by using STRING. Most CENPs were shown to interact with each other according to results from text mining, experiments, databases, and coexpression analyses. Specifically, all CENPs except CENPW and CENPL were shown to interact with each other (Figure 6C).

**GO and KEGG Pathway Analysis of CENPs**

To study the functions of CENPs and neighboring genes of GeneMANIA network, we analyzed GO and KEGG pathways using R software. Analysis of significantly enriched GO terms indicated that CENPs were mainly localized in chromosomes, centromeric regions, kinetochores, condensed chromosomes, and centromeric regions and are primarily involved in centromere complex assembly, CENPA-containing nucleosome assembly, CENPA-containing chromatin organization, chromatin remodeling at the centromere, and DNA replication-independent nucleosome assembly, with additional roles in tubulin binding and microtubule binding (Figure 7A). KEGG pathway enrichment analysis showed that the genes were involved in progesterone-mediated oocyte maturation, the cell cycle and oocyte meiosis (Figure 7B).

**Discussion**

CENP overexpression has been reported in many cancers. Although the role of CENP genes in tumorigenesis and their poor prognostic implications in several cancers have been partially confirmed, further bioinformatics analysis of
This study is the first to explore the mRNA expression, prognostic value (for RFS and OS) and interactions of CENPs at the gene and protein levels in BC. Furthermore, our study performed a comprehensive analysis of the transcriptional expression and prognostic value of CENPs in BC. We hope that our findings will contribute to the available knowledge, identify therapy targets, and improve the prognosis of patients with BC.

Our results showed that the gene expression levels of all CENPs were elevated in BC tissues compared to normal tissues. This indicates that CENPs could play an important role in diagnosing BC. Survival analysis by Kaplan–Meier Plotter showed that the upregulation of CENPs, except CENPK, was positively associated with unfavorable RFS. However, CENPK could be a risk factor when we chose the best cutoff value rather than the median value. Hence, we cannot exclude CENPK as a potential biomarker for poor outcomes in BC. In addition, CENPK has been shown to exert a protumorigenic function by promoting G1 and G2/M cell cycle transition in the carcinogenesis of triple-negative breast cancer (TNBC). Furthermore, high expression of CENPF, CENPE, CENPU, CENPA, CENPN, CENPI and CENPM was significantly correlated with poor OS in BC patients in this study, which is consistent with clinical reports. In fact, O’Brien et al found that CENPF predicted a poor prognosis in BC. Upregulation of CENPU has been associated with a significantly lower OS rate in luminal BC, and high CENPU expression has been found to be an independent predictor of poor DFS. Pan et al proved by immunochemistry that high expression of CENPU is a predictor of worse DFS and OS in BC. High mRNA expression of CENPA is associated with local and metastatic relapse and poor patient outcomes in BC. CENPN was also found to be a predictor for both BC recurrence and worse OS among patients with a history of smoking. Overexpression of CENPI is a powerful independent biomarker for poor patient prognosis and survival in ER-positive BC patients. Although there are no direct reports on the prognostic value of CENPE, CENPK, CENPW, CENPM, and CENPL, they have all been shown to participate in the development of different tumor types and even have prognostic value in other cancers. CENPE and CENPK are associated with the proliferation of TNBC cells, and CENPK promotes lung adenocarcinoma cell proliferation, invasion, and metastasis and is associated with poor clinical outcomes. CENPW and CENPM are associated with hepatocellular carcinoma cell proliferation, migration, and invasion. Capra et al demonstrated that CENPL is associated with congenital infiltrating lipomatosis. Taken together, these findings indicate that all members of the CENP gene family could be considered possible biomarkers and therapeutic targets for BC.
Figure 4 The prognostic value (RFS (A) and OS (B)) of CENP mRNA levels in BC patients (Kaplan–Meier Plotter).
We aimed to assess the genetic alterations and carcinogenesis mechanisms of CENP family genes, and we found that the genetic alteration rates of CENP family genes in BC varied from 2.9% to 16% for individual genes. Furthermore, the results of Kaplan–Meier Plotter analysis and Log rank test showed no significant differences in OS and DFS in patients with and without CENP family gene alterations.

As revealed by the current study, significant positive correlations exist between CENPs at the expression level. Moreover, we showed by GeneMANIA and STRING analysis that CENPU is the most dense physical interaction network among CENPs.

We aimed to assess the genetic alterations and carcinogenesis mechanisms of CENP family genes, and we found that the genetic alteration rates of CENP family genes in BC varied from 2.9% to 16% for individual genes. Furthermore, the results of Kaplan–Meier Plotter analysis and Log rank test showed no significant differences in OS and DFS in patients with and without CENP family gene alterations.

As revealed by the current study, significant positive correlations exist between CENPs at the expression level. Moreover, we showed by GeneMANIA and STRING analysis that CENPU is the most dense physical interaction network among CENPs.
network among CENPs, which is consistent with Priyanka Singh’s report. In fact, CENPs such as CENPE and CENPU are components of the CENP complex that participate in chromosome segregation during mitosis. Hence, they may share the same pathway as other CENPs such as CENPA in the process of pathogenesis. These findings may partly verify the strong correlations among each of the CENPs; in addition, our GeneMANIA analysis showed that CENPA was the only CENP co-expressed with all others. Together, CENPs may function through cooperative mechanisms in the development of tumors including BC.

Our GO term analysis showed that CENP family genes contribute to centromere complex assembly, chromatin remodeling at the centromere and tubulin binding. KEGG pathway analysis showed that CENPs are significantly enriched in the progesterone-mediated oocyte maturation, cell cycle and oocyte meiosis pathways. In addition, CENPU has been reported to enhance angiogenesis by inhibiting the degradation of COX-2 in TNBC and to promote proliferation and migration through PI3K/AKT signaling in TNBC and lung cancer. CENPM has been identified to be associated with 5 genes at the mRNA expression level and one at the protein level in an ER-positive cell line, and it exerted its protumorigenic function by regulating the expression of cell cycle-associated proteins. Therefore, we hypothesized that CENP family genes induce tumorogenesis by regulating chromosome centromere checkpoint assembly and downstream genes to affect the division of tumor cells. However, even CENPs have similar biological characteristics, each CENP gene play its specific role in biological process. Our research has a limitation in that the enrichment analysis was based on CENP family gene set, which might weaken the biological process predictions of each CENP gene for breast cancer. In order to study their specific mechanisms in tumorogenesis, additional works and experiments need to be performed to validate the functions of them. Taken together, our findings offer new perspectives to study the role of CENPs and related signaling pathways for future research and clinical applications for BC patients.

Conclusions
In conclusion, our study illustrated that the expression level of CENPs is higher expressed in breast cancer compared to breast normal tissue. High expression of CENPF, E, U, A, N, I, W, M, L and CENPF, E, U, A, N, I, M correlated with worse relapse free survival (RFS) and worse overall survival (OS), respectively. It suggested that CENP family members could be new prognostic biomarkers for BC. They primarily involved in centromere complex assembly to participate in many signaling pathways including progesterone-mediated oocyte maturation. Therefore, the mechanisms of CENPs
need to be further studied based on our systematic analysis, and CENPs may become novel significant therapeutic targets in clinical practice.

**Abbreviations**

BC, breast cancer; CENPs, centromere proteins; DFS, disease-free survival; GO, Gene Ontology; GEO, Gene Expression Omnibus; KEGG, Kyoto Encyclopedia of Genes and Genomes; OS, overall survival; RFS, relapse-free survival; TNBC, triple-negative breast cancer.

**Data Sharing Statement**

Publicly available datasets were analyzed in this study. The datasets presented in this study can be found in online repositories.

**Ethics Approval and Informed Consent**

The study was approved by the medical ethics committee of the Fourth Hospital of Hebei Medical University (No. 209MEC065), and all experiments were conducted in accordance with the Declaration of Helsinki.

**Author Contributions**

All authors made a significant contribution to the work reported, whether that is in the conception, study design, execution, acquisition of data, analysis and interpretation, or in all these areas; took part in drafting, revising or critically reviewing the article; gave final approval of the version to be published; have agreed on the journal to which the article has been submitted; and agree to be accountable for all aspects of the work.

**Funding**

This work was supported by Natural Science Foundation of Hebei Province (No. H2020206210); Key Research and Development Foundation of Science and Technology Department of Hebei Province (No. 192777125D).

**Disclosure**

The authors declare that they have no competing interests in this work.

**References**


