Original Article

A nomogram based on CENPP expression for survival prediction in breast cancer

Heyan Chen, Shengyu Pu, Shibo Yu, Xiaoqin Liao, Jianjun He, Huimin Zhang^

Department of Breast Surgery, the First Affiliated Hospital of Xi'an Jiaotong University, Xi'an, China

Contributions: (I) Conception and design: H Chen, H Zhang; (II) Administrative support: H Zhang, J He; (III) Provision of study materials or patients: H Chen, S Yu; (IV) Collection and assembly of data: H Chen, S Pu, X Liao; (V) Data analysis and interpretation: H Chen, S Pu, X Liao; (VI) Manuscript writing: All authors; (VII) Final approval of manuscript: All authors.

Correspondence to: Huimin Zhang, Jianjun He. 277 West Yanta Road, Xi'an 710061, China. Email: huimin.zhang@xjtu.edu.cn; chinahjj@163.com.

Background: In recent years, it has been found that the expression of 17 centromere proteins (CENPs) was closely related to malignant tumors, however, the role of CENPs in breast cancer (BC) has not been fully investigated. This study intends to investigate the prognostic value of CENPs in BC and establish nomogram based on expression of CENPs to predict BC patients' prognosis.

Methods: A total of 800 BC patients with complete relevant data were included from the TCGA database and were further randomly divided into training set (N=480) and validation set (N=320). Univariate and multivariate Cox regression analysis were used to screen independent factors for overall survival (OS) prediction of BC patients in the training set. Then, the nomogram was established based on these independent predictors and further validated by receiver-operating characteristic (ROC) curves and calibration plots. The GEPIA and bcGenExMiner v4.4 databases were utilized to analyze mRNA expression of candidate gene in BC patients with different clinicopathological features, respectively.

Results: Multivariate Cox regression analysis showed that age, Her2 status, pathologic_T stage, pathologic_M stage and CENPP expression were of independent prognostic value for BC. CENPP was overexpressed in BC tissues (P<0.01) and lower expression of CENPP was associated with worse OS (P=0.005, HR =2.35; 95% CI: 1.30–4.23). We then established a nomogram based on those independent predictors, and the calibration curve demonstrated good fitness of the nomogram for OS prediction. In the training set, the AUCs of 3- and 5-year survival were 0.757 and 0.797, respectively. In the validation set, the AUCs of 3- and 5-year survival were 0.727 and 0.71, respectively.

Conclusions: Our study showed that CENPP was a novel prognostic factor for patients with BC, and the established nomogram could provide valuable information on prognostic prediction for patients with BC.

Keywords: Breast cancer; *CENPP* gene; overall survival; nomogram

Submitted Jan 15, 2021. Accepted for publication May 13, 2021.
doi: 10.21037/gs-21-30
View this article at: https://dx.doi.org/10.21037/gs-21-30

Introduction

Breast cancer (BC) is the most common malignancy and the leading cause of cancer related death among women worldwide (1). BC is a highly heterogeneous disease and its prognosis varies with different clinical stages, molecular subtypes and histologic types. Even in BC patients with same clinical stage, the histologic type or the molecular subtype, their prognosis is also sufficiently different, indicating the outcome variation cannot to be explained only by clinicopathological parameters. Therefore, it is of importance and far-

^ ORCID: 0000-0003-2636-7858.

reaching significance to explore the heterogeneity of gene expression in BC, search for appropriate molecular biomarkers and establish prognosis prediction models with combined information from both clinic and genetic data. At present, prognosis models based on multi-gene panel detection are increasingly utilized in the clinic to complement T, N, M and biomarker information, such as OncotypeDX, EndoPredict, PAM50 (Prosigna breast cancer prognosis markers), breast cancer index method, etc. (2). Unfortunately, due to the high cost, insufficient technology and poor reproducibility (3), the clinical application of these polygenic prediction models has been greatly limited. Therefore, it is necessary to find new and potential biological markers for BC and develop more practical and affordable prognostic assessment tools.

Constitutive centromere-associated network (CCAN) underlies the centromere specificity and stability of the kinetochore in mitosis of human cancer cells (4-8). To date, 17 members of CCAN have been identified in human, including CENPA/C/H/I/K/L/M/N/O/P/Q/R/S/T/U/W/X, and each of them was closely connected and interacted (9). Centromere protein abnormalities are essential for cancer development (10). In recent years, previous studies have shown that CENPA/H/U/I/O were associated with BC (11-16), lung cancer (17), bladder cancer (18), and gastric cancer (19). Regarding to BC, previous studies have reported that CENPK down-regulation in triple-negative breast cancer cells inhibited cell proliferation and invasion ability (11), down-regulation of CENPU and CENPH gene expression resulted in breast cancer cell proliferation inhibition by cell cycle arrest and apoptosis induction (12). The expression level of CENPA was higher in ER– tumors than in ER+ tumors, and it's an important independent prognostic indicator in ER+ BC patients who have not received systemic therapy (endocrine therapy or chemotherapy) (14). Thangavelu *et al.* found that CENPI mRNA and protein levels were significantly increased in ER+ tumors, and proved that CENPI overexpression promoted chromosomal instability in ER+ BC patients, leading to poor prognosis (16). The above studies on CENP-A/H/I/K/U indicated a vital role of CENPs played in the diagnosis and potential targeted therapy of BC patients. However, many members of this family have not been studied, which urged us to investigate the prognostic value of CENPs in BC.

Since members of CENPs family played a potential role in the occurrence and development of BC, and clinically, there's urgent need for a simple, economical and valuable unified model of BC prognosis prediction, the aim of this study was to explore the prognostic role of CENPs in BC, and establish a prognostic prediction model based on CENPs expression and prognostic clinicopathological parameters.

We present the following article in accordance with the TRIPOD reporting checklist (available at https://dx.doi.org/10.21037/gs-21-30).

Methods

Datasets source and patients selection

We downloaded the mRNA expression profile of CENPs in the TCGA Breast Cancer from the Xena system (https://xenabrowser.net/datapages/) for statistical analysis. In this study, we selected 1,215 BC samples with raw counts of RNAseq expression data of CENPs and corresponding clinicopathological features. Clinicopathological indicators, including age, gender, ER status, PR status, Her2 status, histological types, pathological stages, survival time and survival status, were included in this study. Cases with any of the above indicators missing were excluded from this study to ensure that included patients retained complete RNAseq expression data, clinicopathological characteristics, and prognostic information. Finally, 800 BC patients who met the inclusion criteria were screened out of 1,215 BC patients from the TCGA database. Additionally, since the expression of CENPC was not found on Xena platform, it was excluded from this study. The study was conducted in accordance with the Declaration of Helsinki (as revised in 2013).

In the Xena website, the gene expression was standardized and normalized through the FPKM-UQ (The upper quartile Fragments per Kilobase of transcript per Millionmapped reads) quantitative method provided by TCGA database. Then, in this study, the expression levels of CENPs were divided into high and low expression groups according to the median of mRNA expression profile.

Follow up

The clinical outcome endpoint of this study was overall survival (OS), which was defined as the time from the diagnosis of BC to death from any causes. Follow-up referred to the period from the diagnosis of BC to the

occurrence of an outcome event. The TCGA database recorded the survival status of the lost follow-up as "blank", which were excluded accordingly.

GEPIA and bcGenExMiner v4.4

GEPIA is a web-based tool to analyze the mRNA expression data of 8,587 normal and 9,736 tumor samples from the TCGA and the GTEx projects (20). In this study, we verified the mRNA expression of the candidate genes of CENPs family in BC via GEPIA (tumor *vs.* normal). bcGenExMiner v4.4 is a dataset of published annotated BC transcriptomic. The statistical analyses are divided into three modules: "expression", "prognosis" and "correlation" (21,22). The expression module could be utilized to compare the expression of candidate genes under different clinical features, such as receptor status (ER+ *vs.* ER−, PR+ *vs.* PR−, HER2+ *vs.* HER2− by IHC), nodal status, SBR, age, molecular subtypes and so on.

Statistical analysis

800 BC patients from the TCGA database were randomly divided into training set and validation set according to 3:2 using the "caret" package of R software (https://CRAN.R-project.org/package=caret). Chi-square test or Fisher's exact test was used to compare the distribution differences of classification variables between the two groups. Kaplan-Meier method was used to draw the survival curve and log-rank test was used to compare the survival difference between the two groups. Cox regression analysis was used to screen the independent prognostic factors via the "survival" (23) package of R software.

Based on the results of multivariate Cox analysis, nomogram was constructed using the "rms" (24) package of R software version 3.5.2 (https://www.r-project.org/). Then, receiver operating characteristic (ROC) curves and calibration plots were used to validate the performance of nomogram. The area under ROC curve (AUC) was used to evaluate the predicting ability of the model, and it ranged from 0 to 1.0, with 0 indicating discordance, 0.5 representing a random probability, while 1 indicating a perfect discrimination. The calibration plot was used to evaluate the accuracy of the nomogram. In a perfect calibration model, calibration plots (with 1,000 bootstrap resamples and 5-fold cross-validation) would fall on a 45-degree diagonal line.

All P values were two-sided and the level of significance was set at $P<0.05$.

Results

Clinicopathologic characteristics of the patients

The TCGA Breast Cancer (BRCA) dataset from Xena platform cataloged 1,215 BC patients. After excluding 415 BC patients with incomplete relevant information, 800 BC patients who met the eligible criteria were included, which were divided into a training set (N=480) and a validation set (N=320) randomly at the rate of 3:2. The clinicopathological characteristics of the training set and validation set were comparable (*Table 1*).

Independent prognostic factors of BC

Age, Her2 status, pathologic_T stage, pathologic_N stage, pathologic_M stage and CENPP expression were identified as predictive factors for OS of BC in the univariate analysis (*Table 2*), while except pathologic_N stage, all the other variables were further confirmed as independent predictive factors in the multivariate analysis (*Table 3*). The results showed that BC patients of 65–75 years old (P=0.015, HR =2.67; 95% CI: 1.21–5.86) and >75 years old (P<0.000, HR =3.63; 95% CI: 1.87–7.03) had worse OS than those <65 years old. In addition, Her2 positive patients (P=0.027, HR =2.04; 95% CI: 1.09–3.82) had worse OS than Her2 negative patients. BC patients with pathologic_T4 stage (P=0.003, HR =5.401; 95% CI: 1.78–16.38) and pathologic_M1 stage (P=0.040, HR =4.45; 95% CI: 0.07–18.47) had worse OS compared with pathologic_T1 stage and pathologic_M0 stage, respectively. Additionally, BC patients with expression of CENPP lower than the median (P=0.005, HR =2.35; 95% CI: 1.30–4.23) were significantly correlated with worse OS.

The expression pattern of CENPP in BC

Since CENPP expression was identified as the only independent prognostic factor in CENPs for BC patients, we analyzed its expression pattern through GEPIA and found that GENPP was overexpressed in BC tissues compared with normal tissues (P<0.01) (*Figure 1A*). Data from bcGenExMiner v4.4 showed that the CENPP mRNA level in Luminal A subtype ranked the highest among all 5 subtypes classified by PAM50 (P<0.0001) (*Figure 1B*). Further analysis showed that the expression of CENPP

Table 1 Clinicopathological characteristics of the study populations

Characteristics	Training set (N=480)		Validation set (N=320)		P value
	No.	%	No.	%	
Age, years					0.896
<65	334	69.6	220	68.8	
65–75	62	12.9	45	14.1	
>75	84	17.5	55	17.2	
Gender					0.784
Female	475	99.0	316	98.8	
Male	5	1.0	4	1.3	
ER					0.27
Negative	115	24.0	66	20.6	
Positive	365	76.0	254	79.4	
PR					0.602
Negative	163	34.0	103	32.2	
Positive	317	66.0	217	67.8	
Her2					0.864
Negative	370	77.1	245	76.6	
Positive	110	22.9	75	23.4	
Histological_type					0.418
IDC	351	73.1	234	73.1	
ILC	85	17.7	49	15.3	
Other	44	9.2	37	11.6	
Pathologic_T					0.323
T1	113	23.5	84	26.3	
T2	296	61.7	177	55.3	
T3	55	11.5	45	14.1	
T4	16	3.3	14	4.4	
Pathologic_N					0.98
N0	226	47.1	152	47.5	
N1	157	32.7	106	33.1	
N2	57	11.9	38	11.9	
N3	40	8.3	24	7.5	
Pathologic_M					0.223
M0	421	87.7	269	84.1	
M1	4	0.8	6	1.9	
Mx	55	11.5	45	14.1	

Table 1 (*continued*)

Table 1 (*continued*)

Characteristics	Training set (N=480)		Validation set (N=320)		P value
	No.	%	No.	%	
CENPA					0.164
Low	210	43.8	156	48.8	
High	270	56.3	164	51.3	
CENPH					0.078
Low	229	47.7	173	54.1	
High	251	52.3	147	45.9	
CENPI					0.541
Low	204	42.5	143	44.7	
High	276	57.5	177	55.3	
CENPK					0.073
Low	209	43.5	160	50.0	
High	271	56.5	160	50.0	
CENPL					0.795
Low	237	49.4	155	48.4	
High	243	50.6	165	51.6	
CENPM					0.209
Low	193	40.2	143	44.7	
High	287	59.8	177	55.3	
CENPN					0.118
Low	249	51.9	184	57.5	
High	231	48.1	136	42.5	
CENPO					0.371
Low	238	49.6	169	52.8	
High	242	50.4	151	47.2	
CENPP					0.885
Low	260	54.2	175	54.7	
High	220	45.8	145	45.3	
CENPQ					0.061
Low	233	48.5	177	55.3	
High	247	51.5	143	44.7	
CENPR					0.285
Low	238	49.6	171	53.4	
High	242	50.4	149	46.6	

Table 1 (*continued*)

Table 1 (*continued*)

Characteristics	Training set (N=480)		Validation set (N=320)		P value
	No.	%	No.	%	
CENPS					0.644
Low	238	49.6	164	51.3	
High	242	50.4	156	48.8	
CENPT					0.908
Low	262	54.6	176	55.0	
High	218	45.4	144	45.0	
CENPU					0.385
Low	255	53.1	180	56.3	
High	225	46.9	140	43.8	
CENPW					0.523
Low	259	54.0	180	56.3	
High	221	46.0	140	43.8	

ER, estrogen receptor; PR, progesterone receptor; Her2, human epidermal growth factor receptor 2; IDC, infiltrating ductal carcinoma; ILC, infiltrating lobular carcinoma; CENPA/H/I/K/L/M/N/O/P/Q/R/S/T/U/W, centromere protein A/H/I/K/L/M/N/O/P/Q/R/S/T/U/W.

was higher in ER+ or PR+ tumors (P<0.0001; P<0.0001), whereas lower in Her2+ tumors (P<0.0001) (*Figure 1C*,*D*,*E*). In order to investigate the correlation between CENPP expression and clinicopathological features, patients in the training set was divided into CENPP high and CENPP low groups by the median expression of CENPP. CENPP high expression group had higher ER+ and PR+ (P<0.001; P<0.001) ratio, and had a lower death rate compared to CENPP low expression group (P=0.003) (*Table 4*).

Prognostic values of CENPP in BC

Kaplan-Meier plotter showed that higher CENPP expression was associated with better OS in BC patients (P=0.0019) (*Figure 2A*). Regarding to histological types, we concluded that higher expression of CENPP was associated with better OS in IDC or ILC (P=0.0031; P=0.046) (*Figure 2B*,*C*). In addition, CENPP high expression indicated better OS in BC with ER+ or PR+ (P=0.0059; P=0.011), whereas no significant correlation with prognosis in ER– or PR– tumors (P=0.31; P=0.14) (*Figure 2D*,*E*,*F*,*G*). Moreover, regardless of Her2 status, higher expression of CENPP was associated with better OS (Her2+, P=0.017; Her2-, P=0.039) (*Figure 2H*,*I*).

Construction and validation of the nomogram for OS

Based on the results of the multivariate analysis, a nomogram was established with independent prognostic predictors for BC including age, Her2, pathologic_T stage, pathologic_M stage and CENPP expression (*Figure 3*). Different variables of each patient pointed to a score according to the top scale, and then all scores were summed up to get a total score. Based on the total score of the bottom scale, 3- and 5-year survival probabilities of BC could be evaluated. Next, the ROC curve was performed to evaluate the effectiveness of the nomogram (*Figure 4*). In the training set, the AUCs of 3- and 5-year survival prediction were 0.757 and 0.797 (*Figure 4A*,*B*), respectively. In the validation set, the AUC of 3- and 5-year survival prediction were 0.727 and 0.71 (*Figure 4C*,*D*), respectively. The calibration plot (*Figure 5A*,*B*,*C*,*D*) suggested that the nomogram was well calibrated. These results suggested that this nomogram displayed good accuracy in predicting both 3- and 5-year overall survival for patients with BC.

In order to explicit the necessity of including CENPP expression in the nomogram, we then established a new prognostic model with only four traditional clinicopathological features (age, Her2 status, pathological T stage and pathological M stage) included and CENPP

Table 2 Univariate analysis of the training set

Characteristics	Univariate analysis		
	HR	95% CI	P value
Age, years			
<65	1		
65–75	1.85	0.89–3.85	0.100
>75	2.87	1.59–5.17	0.000
Gender			
Female	1		
Male	0.00	0.00–Inf	0.996
ER			
Negative	1		
Positive	0.62	0.37–1.04	0.072
PR			
Negative	1		
Positive	0.72	0.44–1.17	0.185
Her2			
Negative	1		
Positive	2.31	1.32–4.04	0.003
Histological_type			
IDC	1		
ILC	0.83	0.42–1.65	0.595
Other	1.21	0.56–2.62	0.621
Pathologic_T			
T1	1		
T2	1.03	0.55–1.93	0.932
T3	1.49	0.68–3.29	0.320
T4	6.15	2.67–14.16	0.000
Pathologic_N			
N0	1		
N1	1.22	0.67–2.2	0.520
N2	1.91	0.92–3.95	0.080
N3	5.44	2.44–12.16	0.000
Pathologic_M			
M0	1		
M1	11.87	3.6–39.15	0.000
Mx	1.22	0.48–3.06	0.677

Table 2 (*continued*)

Table 2 (*continued*)

Characteristics	Univariate analysis		
	HR	95% CI	P value
CENPA			
High	1		
Low	1.51	0.92–2.48	0.102
CENPH			
High	1		
Low	1.51	0.92–2.48	0.106
CENPI			
High	1		
Low	1.30	0.77–0.80	0.296
CENPK			
High	1		
Low	1.11	0.68–1.82	0.670
CENPL			
High	1		
Low	1.16	0.71–1.90	0.555
CENPM			
High	1		
Low	1.40	0.86–2.30	0.177
CENPN			
High	1		
Low	1.26	0.77–2.08	0.356
CENPO			
High	1		
Low	1.37	0.83–2.24	0.219
CENPP			
High	1		
Low	2.20	1.28–3.80	0.005
CENPQ			
High	1		
Low	1.29	0.78–2.11	0.320
CENPR			
High	1		
Low	1.49	0.9–2.45	0.120

Table 2 (*continued*)

Table 2 (*continued*)

Characteristics	Univariate analysis		
	HR	95% CI	P value
CENPS			
High	1		
Low	1.43	0.85–2.38	0.176
CENPT			
High	1		
Low	0.95	0.58–1.55	0.830
CENPU			
High	1		
Low	1.39	0.85–2.30	0.200
CENPW			
High	1		
Low	1.39	0.84–2.30	0.201

ER, estrogen receptor; PR, progesterone receptor; Her2, human epidermal growth factor receptor 2; IDC, infiltrating ductal carcinoma; ILC, infiltrating lobular carcinoma; CENPA/H/I/K/L/M/N/O/P/Q/R/S/T/U/W, centromere protein A/H/I/K/L/M/N/O/P/Q/R/S/T/U/W.

expression excluded (Figure S1). The verification results showed that the AUC values of 3- and 5-year OS prediction in the training set were 0.676 and 0.677, respectively (Figure S2A,B). In the validation set, the AUC values for 3- and 5-year OS prediction were 0.646 and 0.615, respectively (Figure S2C,D). These results suggested that the model with CENPP expression has better performance than the model only with conventional clinicopathological features in predicting 3- and 5-year OS of BC patients.

Table 3 Multivariate analysis of the training set

Characteristics	Multivariate analysis		
	HR	95% CI	P value
Age, years			
<65	1		
65–75	2.67	1.21–5.86	0.015
>75	3.63	1.87–7.03	0.000
Her2			
Negative	1		
Positive	2.04	1.09–3.82	0.027
Pathologic_T			
T1	1		
T2	1.12	0.57–2.21	0.736
T3	2.13	0.91–5.04	0.083
T4	5.40	1.78–16.38	0.003
Pathologic_N			
N0	1		
N1	1.17	0.63–2.17	0.620
N2	1.10	0.47–2.58	0.820
N3	1.48	0.55–3.98	0.437
Pathologic_M			
M0	1		
M1	4.45	0.07–18.47	0.040
Mx	0.61	0.22–1.67	0.332
CENPP			
High	1		
Low	2.35	1.30–4.23	0.005

Her2, human epidermal growth factor receptor 2; CENPP, centromere protein P.

Discussion

In this study, data of CENPs expression and clinicopathological features of BC patients were downloaded and analyzed from the TCGA database, which aimed to discover more biological genes and molecular indexes to accurately predict the prognosis of BC. Based on the Cox regression analysis, we identified that BC patients with older age, Her2 positivity, advanced T stage, advanced M stage, or lower expression of CENPP were accompanied by worse OS. Next, high expression of CENPP was proved to be associated with better OS in ER + BC or PR + BC, whereas regardless of Her2 status, higher expression of CENPP indicated better OS. Finally, we constructed a nomogram on the basis of CENPP expression as well as other independent predictors. The 3- and 5-year of AUCs in the training set were 0.757 and 0.797, and that of AUCs in the validation set were 0.727 and 0.71.

Cell mitosis is the process of transferring genetic information from the parent cell to the daughter cell. In the process of mitosis, CENPs not only provided energy

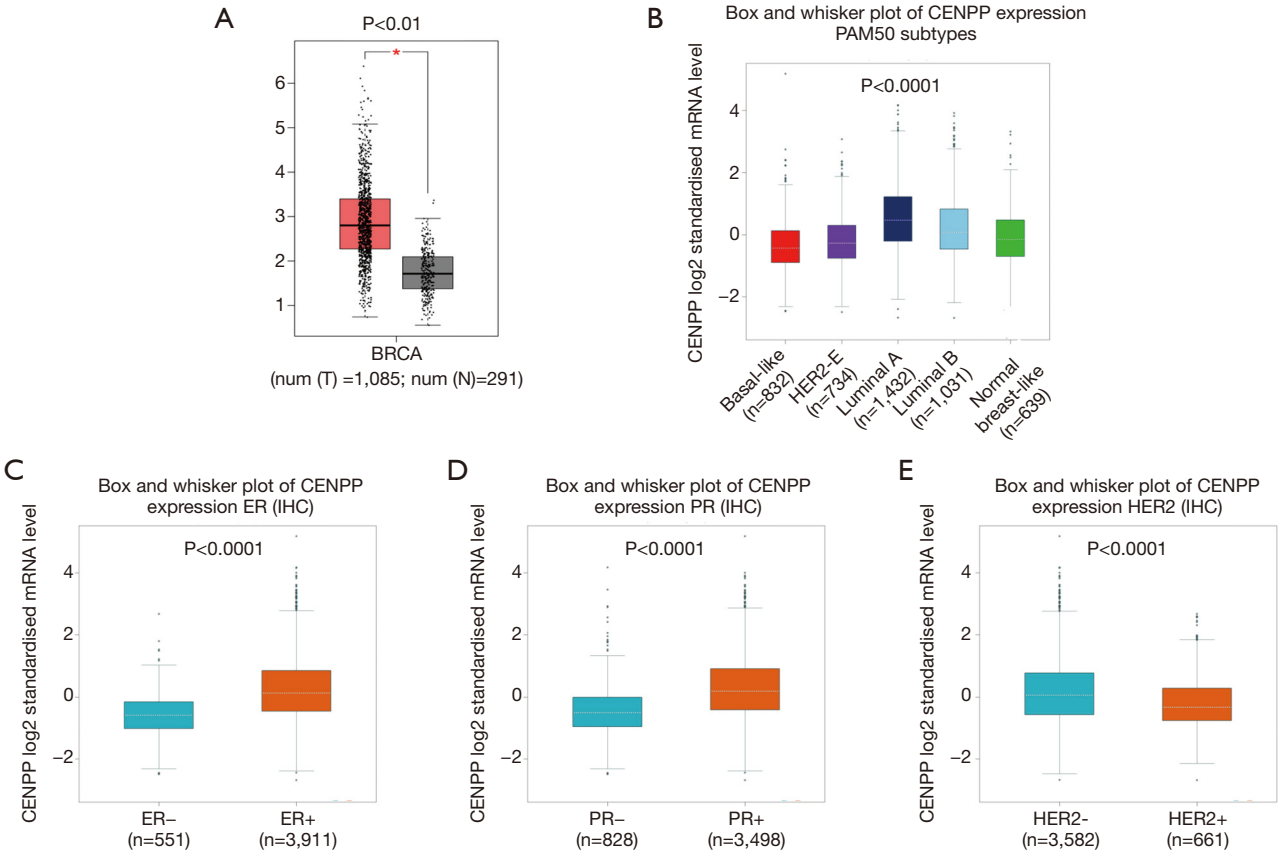

Figure 1 The expression level of CENPP in BC. (A) The expression level of CENPP between BC tissues and normal breast in GEPIA database. (B,C,D,E) The expression level of CENPP in different molecular subtypes (B), ER status (C), PR status (D) and Her2 status (E) (bcGenExMiner v4.4 database). *, P<0.01; ER, estrogen receptor; PR, progesterone receptor; Her2, human epidermal growth factor receptor; CENPP, centromere protein P.

Table 4 Correlation between clinical characteristics and expression of CENPP in the training set

Characteristics	High expression (N=220), n (%)	Low expression (N=260), n (%)	P value
Age, years			0.619
<65	150 (68.2)	184 (70.8)	
>75	32 (14.5)	30 (11.5)	
65–75	38 (17.3)	46 (17.7)	
Gender			0.837
Female	218 (99.1)	256 (98.5)	
Male	2 (0.9)	4 (1.5)	
ER			<0.001
Negative	33 (15.0)	82 (31.5)	
Positive	187 (85.0)	178 (68.5)	

Table 4 (*continued*)

Table 4 (*continued*)

Characteristics	High expression (N=220), n (%)	Low expression (N=260), n (%)	P value
PR			<0.001
Negative	50 (22.7)	113 (43.5)	
Positive	170 (77.3)	147 (56.5)	
Her2			0.676
Negative	172 (78.2)	198 (76.2)	
Positive	48 (21.8)	62 (23.8)	
Histological_type			0.08
IDC	151 (68.6)	200 (76.9)	
ILC	48 (21.8)	37 (14.2)	
Other	21 (9.5)	23 (8.8)	
Pathologic_T			0.189
T1	54 (24.5)	59 (22.7)	
T2	132 (60.0)	164 (63.1)	
T3	30 (13.6)	25 (9.6)	
T4	4 (1.8)	12 (4.6)	
Pathologic_N			0.548
N0	105 (47.7)	121 (46.5)	
N1	75 (34.1)	82 (31.5)	
N2	26 (11.8)	31 (11.9)	
N3	14 (6.4)	26 (10.0)	
Pathologic_M			0.179
M0	193 (87.7)	228 (87.7)	
M1	0 (0.0)	4 (1.5)	
Mx	27 (12.3)	28 (10.8)	
Status			0.003
Alive	202 (91.8)	214 (82.3)	
Dead	18 (8.2)	46 (17.7)	

ER, estrogen receptor; PR, progesterone receptor; Her2, human epidermal growth factor receptor 2; IDC, infiltrating ductal carcinoma; ILC, infiltrating lobular carcinoma; CENPP, centromere protein P.

for the separation of sister chromatids, but also served as a genomic information monitoring function. Once this process loses normal regulation or makes mistakes, it may induce the occurrence of malignant tumors (25,26). CENPs play a pivotal role in maintaining normal mitosis in cells. In this study, CENPP was identified as the only prognosis-related gene in CENPs for patients with BC. Some studies reported that CENPP was associated with mixed uterine carcinosarcoma and among its related pathways were mitotic metaphase, anaphase and signaling by G protein coupled receptor (GPCR) (4,27). However, its role in BC is unknown. To our knowledge, this is the

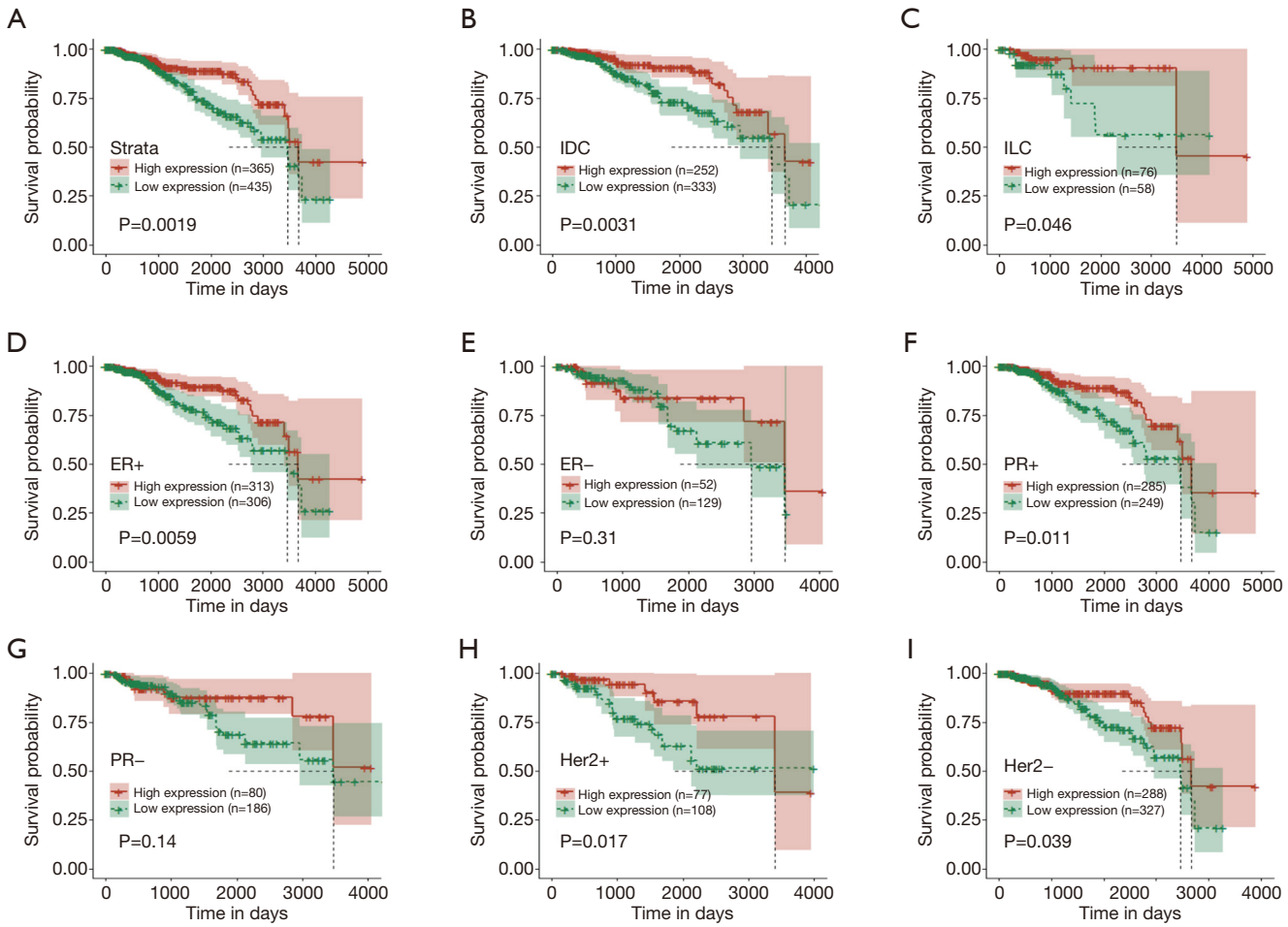

Figure 2 Kaplan-Meier survival curve of overall survival between the high-expression group and low-expression group of CENPP in BC. Kaplan-Meier curves were constructed to analyze the association of CENPP mRNA expression in the primary tumor with OS of all BC patients (A), IDC patients (B), ILC patients (C), ER+ patients (D), ER– patients (E), PR+ patients (F), PR– patients (G), Her2+ patients (H) and Her2– patients (I). IDC, infiltrating ductal carcinoma; ILC, infiltrating lobular carcinoma; ER, estrogen receptor; PR, progesterone receptor; Her2, human epidermal growth factor receptor; CENPP, centromere protein P.

first study to investigate the prognostic value of CENPP in BC.

Hormone receptor status plays a key role in the formation and development of BC. ER and PR status as an important biological indicator of choosing treatment schemes have been widely recognized and accepted in BC patients. Compared with ER– BC, the tumor differentiation of ER+ BC is better, the invasiveness is less, and the long-term survival rate is higher. According to the data of the American Registry of Cancer Research, 20% of patients with ER+ BC were PR– (28). Studies have shown that ER+PR– was a more invasive subtype of ER+ BC (29,30). The overall survival and disease-free survival of ER+PR– BC was lower than that of ER+PR+ BC. Purdie *et al.* believed that PR was an independent predictor of early breast cancer prognosis (31). Our study found that the mRNA level of CENPP was positively correlated with ER status and PR status, and higher mRNA expression of CENPP indicated better OS in BC with ER+ or PR+, which suggests a close relationship between CENPP and hormone receptor pathway and the underling mechanism requires further investigation. In addition, recent studies have shown that the overexpression of Her2 did not only indicate invasiveness and poor prognosis in BC, but also predict the sensitivity of BC to systemic treatment. The result of our study showed that no matter the Her2 status was positive or

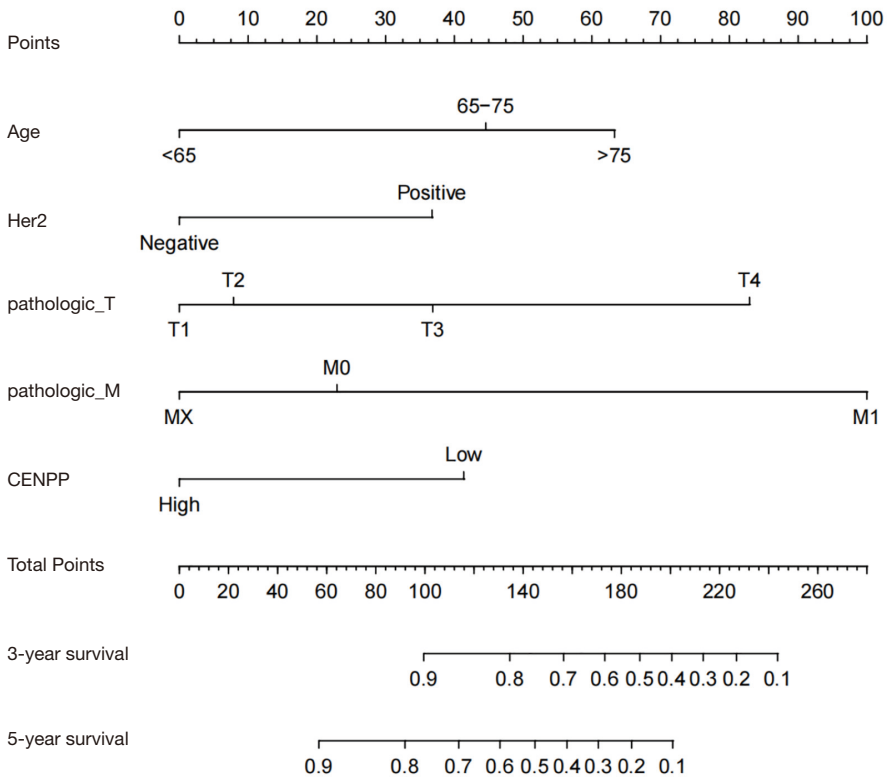

Figure 3 A nomogram to predict the prognosis of breast cancer. To use the nomogram, an individual patient's value is located on each variable axis, and a line is drawn upward to determine the number of points received for each variable value. The sum of these numbers is located on the Total Points axis, and a line is drawn downward to the survival axes to determine the likelihood of 3- or 5-year survival. Her2, human epidermal growth factor receptor; CENPP, centromere protein P.

negative, higher expression of CENPP was associated with better OS which indicated an inconsequential association between CENPP and Her2 pathway. Collectively, these findings suggested that CENPP was an effective prognostic predictor of BC and might also be a potential target for HR+ BC.

Currently, the nomogram has been developed and shown to be more accurate in predicting prognosis in some cancers than the conventional staging systems (32-34). This study attempted to establish a prognostic nomogram of BC and to determine whether the model can accurately predict survival of patients with BC. Age, Her2 status, pathologic_T stage, pathologic_M stage and CENPP expression were identified as predictive factors for OS of BC in the multivariate Cox analysis. This is the first study to set up a nomogram based on the CENPP expression and conventional prognosis predictors to predict OS in patients with BC. Through validation, the nomogram showed good performance in predicting survival, and its accuracy was supported by the ROC curves and the calibration curves. When a patient is diagnosed with BC and has obtained the above clinicopathological results, we can predict the clinical prognosis according to her own features and CENPP expression level. If the predicted prognosis is poor, intensive treatment might be recommended in the hope of gaining a better outcome.

There're some limitations of this study. Firstly, the demographic and clinical information provided by the TCGA database were not complete. For example, the database lacked detailed records like surgery, marital status and insurance status information. Different surgical approaches, marital status and insurance status may influence the outcome of BC patients. Secondly, this was a retrospective study, all the data of this study were obtained from publicly available databases. More prospective studies are needed to further confirm our conclusions.

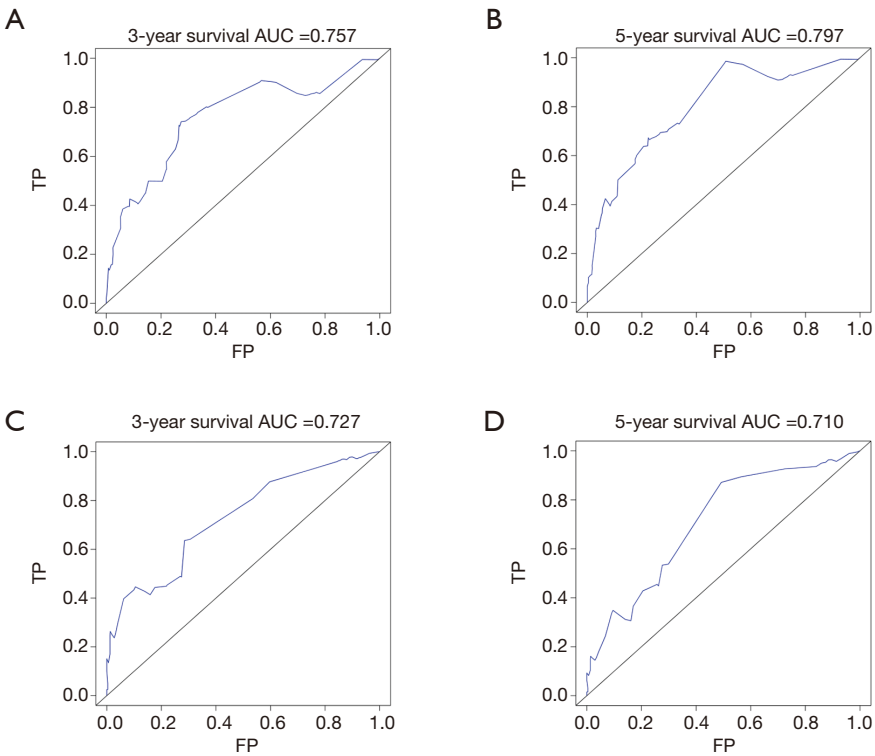

Figure 4 Validation of the nomogram by the ROC curves. The ROC curve for predicting patient survival at 3 years (A) and 5 years (B) in the training set, and at 3 years (C) and 5 years (D) in the validation set. ROC, receiver-operating characteristic; AUC, area under the curve. TP, true positive rate; FP, false positive rate.

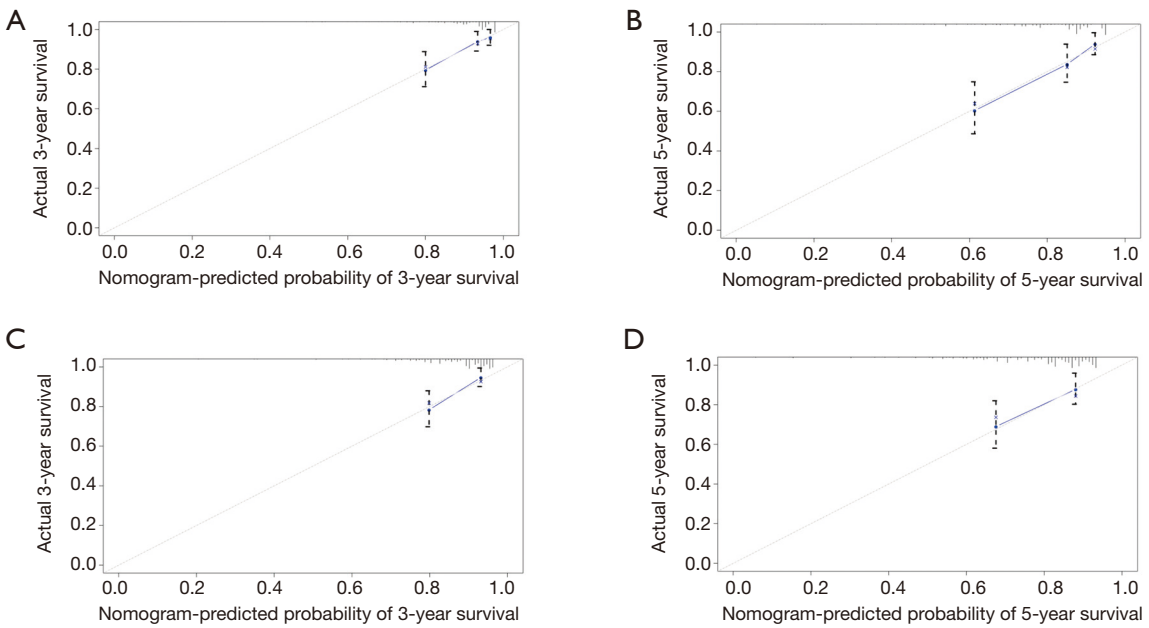

Figure 5 Validation of the nomogram by the calibration curves. The calibration curve for predicting patient survival at 3 years (A) and 5 years (B) in the training set, and at 3 years (C) and 5years (D) in the validation set. Nomogram-predicted probability of overall survival is plotted on the x-axis; actual overall survival is plotted on the y-axis.

Conclusions

In this study, we identified that higher expression of CENPP was associated with better prognosis, and established a prognostic nomogram with good performance based on CENPP expression and clinicopathological features. Our study provided a novel method for clinical evaluation and a potential biomarker/target for BC which needs to be further validated in a prospective study and investigated in further basic research.

Acknowledgments

We would like to thank Ivan Chen for his help in polishing our paper.
Funding: This work was supported by Natural Science Foundation of Shaanxi Province, China (No. 2018JQ8004). We thank The Cancer Genome Atlas (TCGA) Database for sharing the large amount of data.

Footnote

Reporting Checklist: The authors have completed the TRIPOD reporting checklist. Available at https://dx.doi.org/10.21037/gs-21-30

Conflicts of Interest: All authors have completed the ICMJE uniform disclosure form (available at https://dx.doi.org/10.21037/gs-21-30). All authors report that this work was supported by Natural Science Foundation of Shaanxi Province, China (No. 2018JQ8004).

Ethical Statement: The authors are accountable for all aspects of the work in ensuring that questions related to the accuracy or integrity of any part of the work are appropriately investigated and resolved. The study was conducted in accordance with the Declaration of Helsinki (as revised in 2013).

References

1. Bray F, Ferlay J, Soerjomataram I, et al. Global cancer statistics 2018: GLOBOCAN estimates of incidence and mortality worldwide for 36 cancers in 185 countries. CA Cancer J Clin 2018;68:394-424.
2. Harris LN, Ismaila N, McShane LM, et al. Use of biomakers to guide decisions on adjuvant systematic therapy for women with early-stage invasive breast cancer: American Society of Clinical Oncology clinical practice guideline. J Clin Oncol 2016;34:1134-50.
3. Kwa M, Makris A, Esteva FJ. Clinical utility of gene-expression signatures in early stage breast cancer. Nat Rev Clin Oncol 2017;14:595-610.
4. Eskat A, Deng W, Hofmeister A, et al. Step-wise assembly, maturation and dynamic behavior of the human CENP-P/O/R/Q/U kinetochore sub-complex. PLoS One 2012;7:e44717.
5. Hoischen C, Yavas S, Wohland T, et al. CENP-C/H/I/K/M/T/W/N/L and hMis12 but not CENP-S/X participate in complex formation in the nucleoplasm of living human interphase cells outside centromeres. PLoS One 2018;13:e0192572.
6. Lera RF, Norman RX, Dumont M, et al. Plk1 protects kinetochore-centromere architecture against microtubule pulling forces. EMBO Rep 2019;20:e48711.
7. Logsdon GA, Gambogi CW, Liskovykh MA, et al. Human Artificial Chromosomes that Bypass Centromeric DNA. Cell 2019;178:624-39.e19.
8. Nechemia-Arbely Y, Miga KH, Shoshani O, et al. DNA replication acts as an error correction mechanism to maintain centromere identity by restricting CENP-A to centromeres. Nat Cell Biol 2019;21:743-54.
9. McKinley KL, Sekulic N, Guo LY, et al. The CENP-L-N Complex Forms a Critical Node in an Integrated Meshwork of Interactions at the Centromere-Kinetochore Interface. Mol Cell 2015;60:886-98.
10. Beh TT, Kalitsis P. The Role of Centromere Defects in Cancer. Prog Mol Subcell Biol 2017;56:541-54.
11. Komatsu M, Yoshimaru T, Matsuo T, et al. Molecular features of triple negative breast cancer cells by genome-wide gene expression profiling analysis. Int J Oncol 2013;42:478-506.
12. Liao WT, Feng Y, Li ML, et al. Overexpression of centromere protein H is significantly associated with breast cancer progression and overall patient survival. Chin J Cancer 2011;30:627-37.
13. Lin SY, Lv YB, Mao GX, et al. The effect of centromere

protein U silencing by lentiviral mediated RNA interference on the proliferation and apoptosis of breast cancer. Oncol Lett 2018;16:6721-8.
14. McGovern SL, Qi Y, Pusztai L, et al. Centromere protein-A, an essential centromere protein, is a prognostic marker for relapse in estrogen receptor-positive breast cancer. Breast Cancer Res 2012;14:R72.
15. Rajput AB, Hu N, Varma S, et al. Immunohistochemical Assessment of Expression of Centromere Protein-A (CENPA) in Human Invasive Breast Cancer. Cancers (Basel) 2011;3:4212-27.
16. Thangavelu PU, Lin CY, Vaidyanathan S, et al. Overexpression of the E2F target gene CENPI promotes chromosome instability and predicts poor prognosis in estrogen receptor-positive breast cancer. Oncotarget 2017;8:62167-82.
17. Wang X, Chen D, Gao J, et al. Centromere protein U expression promotes non-small-cell lung cancer cell proliferation through FOXM1 and predicts poor survival. Cancer Manag Res 2018;10:6971-84.
18. Wang S, Liu B, Zhang J, et al. Centromere protein U is a potential target for gene therapy of human bladder cancer. Oncol Rep 2017;38:735-44.
19. Cao Y, Xiong J, Li Z, et al. CENPO expression regulates gastric cancer cell proliferation and is associated with poor patient prognosis. Mol Med Rep 2019;20:3661-70.
20. Tang Z, Li C, Kang B, et al. GEPIA: a web server for cancer and normal gene expression profiling and interactive analyses. Nucleic Acids Res 2017;45:W98-W102.
21. Jezequel P, Campone M, Gouraud W, et al. bc-GenExMiner: an easy-to-use online platform for gene prognostic analyses in breast cancer. Breast Cancer Res Treat 2012;131:765-75.
22. Jezequel P, Frenel JS, Campion L, et al. bc-GenExMiner 3.0: new mining module computes breast cancer gene expression correlation analyses. Database (Oxford) 2013;2013:bas060.
23. Li JCA. Modeling survival data: Extending the Cox model. Sociol Method Res 2003;32:117-20.
24. Nunez E, Steyerberg EW, Nunez J. Regression modeling strategies. Rev Esp Cardiol 2011;64:501-7.
25. Beaver JA, Gustin JP, Yi KH, et al. PIK3CA and AKT1 mutations have distinct effects on sensitivity to targeted pathway inhibitors in an isogenic luminal breast cancer model system. Clin Cancer Res 2013;19:5413-22.
26. Lee H, Kim SJ, Jung KH, et al. A novel imidazopyridine PI3K inhibitor with anticancer activity in non-small cell lung cancer cells. Oncol Rep 2013;30:863-9.
27. Foltz DR, Jansen LE, Black BE, et al. The human CENP-A centromeric nucleosome-associated complex. Nat Cell Biol 2006;8:458-69.
28. Okada M, Cheeseman IM, Hori T, et al. The CENP-H-I complex is required for the efficient incorporation of newly synthesized CENP-A into centromeres. Nat Cell Biol 2006;8:446-57.
29. Cancello G, Maisonneuve P, Rotmensz N, et al. Progesterone receptor loss identifies Luminal B breast cancer subgroups at higher risk of relapse. Ann Oncol 2013;24:661-8.
30. Dunnwald LK, Rossing MA, Li CI. Hormone receptor status, tumor characteristics, and prognosis: a prospective cohort of breast cancer patients. Breast Cancer Res 2007;9:R6.
31. Purdie CA, Quinlan P, Jordan LB, et al. Progesterone receptor expression is an independent prognostic variable in early breast cancer: a population-based study. Br J Cancer 2014;110:565-72.
32. International Bladder Cancer Nomogram Consortium; Bochner BH, Kattan MW, et al. Postoperative nomogram predicting risk of recurrence after radical cystectomy for bladder cancer. J Clin Oncol 2006;24:3967-72.
33. Karakiewicz PI, Briganti A, Chun FK, et al. Multi-institutional validation of a new renal cancer-specific survival nomogram. J Clin Oncol 2007;25:1316-22.
34. Touijer K, Scardino PT. Nomograms for staging, prognosis, and predicting treatment outcomes. Cancer 2009;115:3107-11.

Cite this article as: Chen H, Pu S, Yu S, Liao X, He J, Zhang H. A nomogram based on CENPP expression for survival prediction in breast cancer. Gland Surg 2021;10(6):1874-1888. doi: 10.21037/gs-21-30

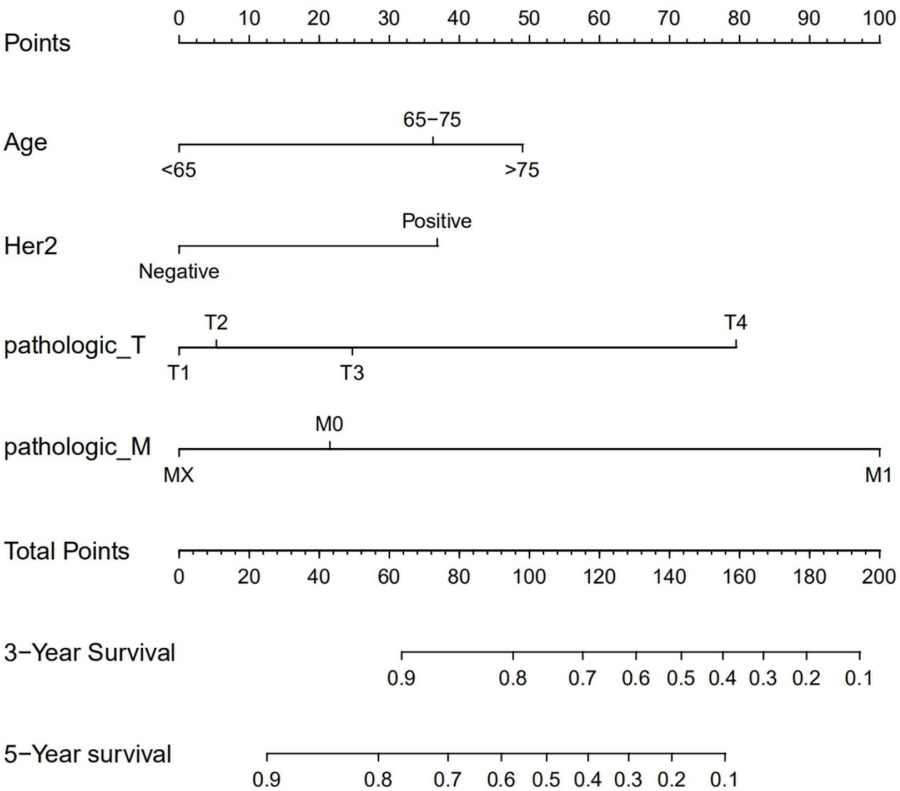

Figure S1 A nomogram (excluding CENPP) to predict the prognosis of breast cancer. To use the nomogram, an individual patient's value is located on each variable axis, and a line is drawn upward to determine the number of points received for each variable value. The sum of these numbers is located on the Total Points axis, and a line is drawn downward to the survival axes to determine the likelihood of 3- or 5-year survival. Her2, human epidermal growth factor receptor; CENPP, centromere protein P.

https://dx.doi.org/10.21037/gs-21-30

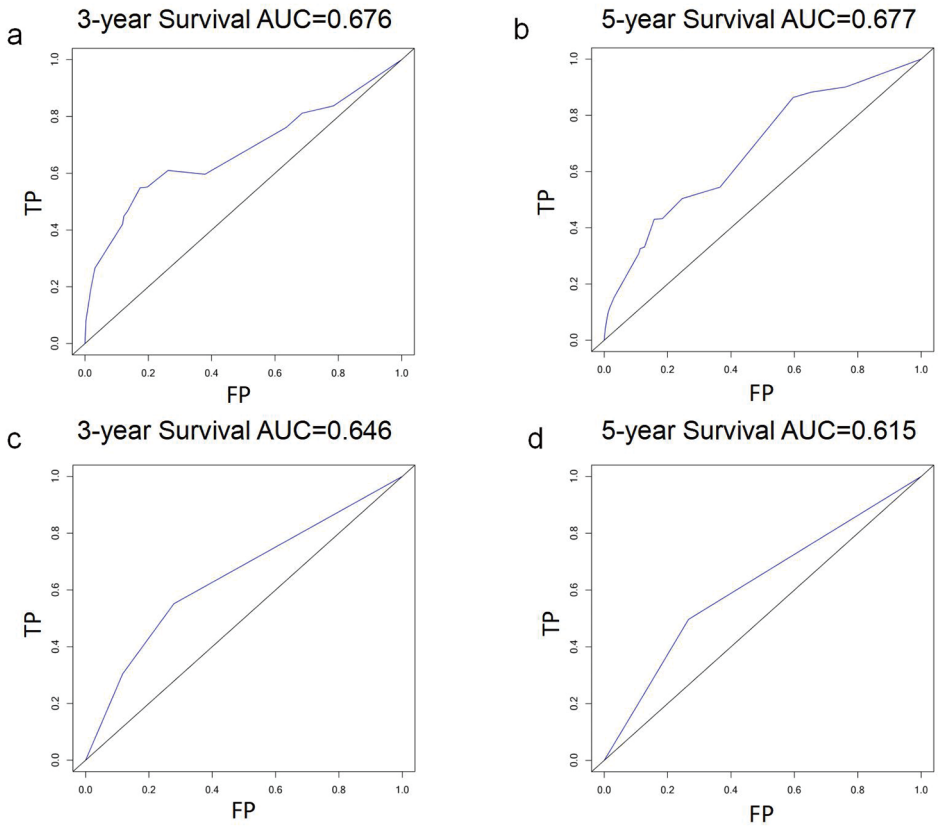

Figure S2 Validation of the nomogram (excluding CENPP) by the ROC curves. The ROC curve for predicting patient survival at 3 years (A) and 5 years (B) in the training set, and at 3 years (C) and 5 years (D) in the validation set. ROC, receiver-operating characteristic; AUC, area under the curve. TP, true positive rate; FP, false positive rate. ROC, receiver-operating characteristic; AUC, area under the curve. TP, true positive rate; FP, false positive rate.

https://dx.doi.org/10.21037/gs-21-30