Table S1 Clinicopathologic characteristics of the training cohort and validation cohort

Variable	Training cohort, n=9,616 (%)	5-yOS (%)	Validation cohort, n=675 (%)	5-yOS (%)	P
Age at diagnosis (years)					
≤50	2,578 (26.8)	91.0	316 (46.8)	92.0	0.000
>50	7,038 (73.2)	85.4	359 (53.2)	92.9	
Race					0.000
White	8,313 (86.4)	87.5	0 (0.0)	–	
Black	759 (7.9)	78.3	0 (0.0)	–	
Other	544 (5.7)	87.9	675 (100.0)	92.7	
Primary site					0.000
0	676 (7.0)	84.1	67 (9.9)	82.2	
1	3,594 (37.4)	87.6	316 (46.8)	80.6	
2	592 (6.2)	87.7	29 (4.3)	84.7	
3	395 (4.1)	86.3	31 (4.6)	83.3	
4	865 (8.9)	90.1	49 (7.3)	88.6	
5	1,867 (19.4)	87.2	109 (16.1)	90.1	
6	1,627 (17.0)	83.9	74 (11.0)	88.7	
Grade					0.000
Grade I	1648 (17.1)	91.0	76 (11.3)	92.1	
Grade II	5153 (53.6)	88.9	425 (62.9)	92.9	
Grade III + IV	2815 (29.3)	80.6	174 (25.8)	92.5	
T stage					0.000
0	3 (0)	66.7	3 (0.4)	100.0	
1	5,426 (56.4)	91.3	263 (39.0)	96.6	
2	3,266 (33.9)	83.9	373 (55.3)	92.2	
3	720 (7.5)	72.7	35 (5.2)	74.3	
4	201 (2.2)	61.1	1 (0.1)	100	
Radiation recode					0.000
Yes	4,610 (47.9)	89.2	258 (38.2)	91.5	
No	5,006 (52.1)	84.7	417 (61.8)	93.5	
Chemotherapy record					0.000
Yes	4,536 (47.2)	87.4	606 (89.8)	93.6	
No	5,080 (52.8)	86.4	69 (10.2)	86.0	
ER status recode					0.000
Positive	8,791 (91.4)	88.1	486 (72)	92.8	
Negative	825 (8.6)	73.7	189 (28)	92.2	
PR status recode					0.000
Positive	7,646 (79.5)	88.8	476 (70.5)	93.5	
Negative	1,970 (20.5)	79.3	199 (29.5)	91.4	

Primary site: 0, Nipple and Central portion of breast; 1, Upper-outer quadrant and Axillary tail of breast; 2, Lower-outer quadrant of breast; 3, Lower-inner quadrant of breast; 4, Upper-inner quadrant of breast; 5, Overlapping lesion of breast; 6, Breast, NOS. ER, estrogen receptor; PR, progesterone receptor.

Table S2 Subset comparison with the "other" race of the training cohort and validation cohort
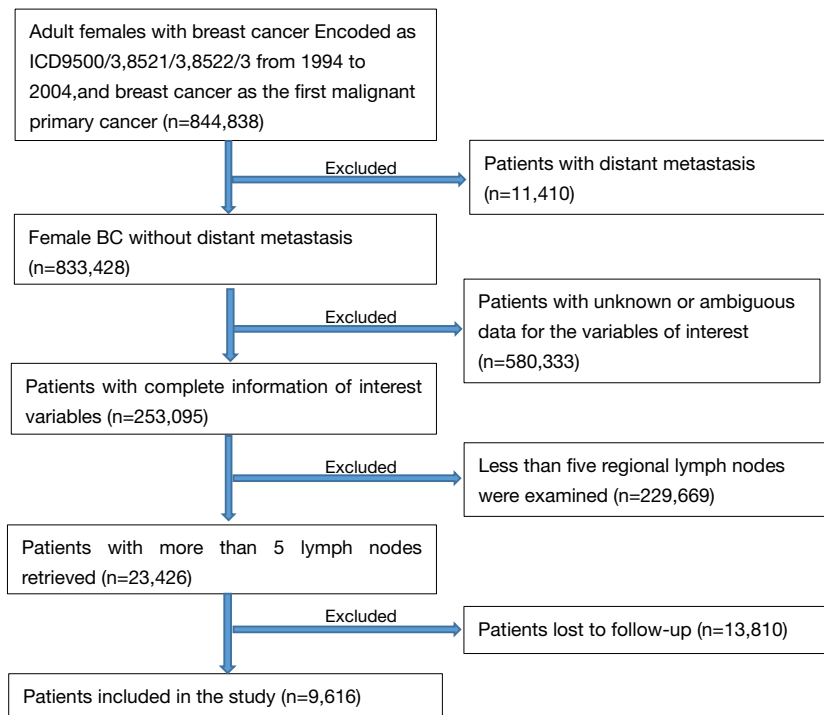
Variable	Other race in Training cohort, n=544 (%)	Other race in Validation cohort, n=675 (%)	P
Age at diagnosis (years)			
≤50	208 (38.2)	316 (46.8)	0.003
>50	336 (61.8)	359 (53.2)	
Primary site			0.006
0	46 (8.4)	67 (9.9)	
1	199 (36.6)	316 (46.8)	
2	29 (5.3)	29 (4.3)	
3	25 (4.6)	31 (4.6)	
4	58 (10.7)	49 (7.3)	
5	113 (20.7)	109 (16.1)	
6	74 (13.7)	74 (11.0)	
Grade			0.001
Grade I	88 (16.2)	76 (11.3)	
Grade II	288 (52.9)	425 (62.9)	
Grade III + IV	168 (30.9)	174 (25.8)	
T stage			0.000
0	0 (0)	3 (0.4)	
1	292 (53.7)	263 (39.0)	
2	193 (35.5)	373 (55.3)	
3	46 (8.5)	35 (5.2)	
4	13 (2.3)	1 (0.1)	
Radiation recode			0.000
Yes	271 (49.8)	258 (38.2)	
No	273 (50.2)	417 (61.8)	
Chemotherapy record			0.000
Yes	323 (59.4)	606 (89.8)	
No	221 (40.6)	69 (10.2)	
ER status recode			0.000
Positive	498 (91.5)	486 (72)	
Negative	46 (8.5)	189 (28)	
PR status recode			0.002
Positive	427 (78.5)	476 (70.5)	
Negative	117 (21.5)	199 (29.5)	

Primary site: 0, Nipple and Central portion of breast; 1, Upper-outer quadrant and Axillary tail of breast; 2, Lower-outer quadrant of breast; 3, Lower-inner quadrant of breast; 4, Upper-inner quadrant of breast; 5, Overlapping lesion of breast; 6, Breast, NOS. ER, estrogen receptor; PR, progesterone receptor.

Table S3 Univariate and multivariate Cox proportional hazards regression for OS

Variables	No. of patients	Univariate analysis	Multivariate analysis	P value
		P value	HR (95% CI)	
Age at diagnosis		0.000		0.000
≤50	2,578		Reference	
>50	7,038		0.510 (0.466–0.557)	
Race		0.000		0.000
White	8,313		Reference	
Black	759		1.116 (0.954–1.305)	0.172
Other	544		1.684 (1.399–2.026)	0.000
Primary site		0.000		0.012
0	676		Reference	
1	3,594		0.901 (0.783–1.038)	0.149
2	592		0.848 (0.772–0.932)	0.001
3	395		0.915 (0.783–1.070)	0.267
4	865		1.032 (0.869–1.224)	0.722
5	1,867		0.843 (0.735–0.967)	0.015
6	1,627		0.924 (0.831–1.028)	0.146
Grade		0.000		0.000
Grade I	1,648		Reference	
Grade II	5,153		0.674 (0.606–0.749)	0.000
Grade III + IV	2,815		0.794 (0.740–0.857)	0.000
T stage		0.000		0.000
0	3		Reference	
1	5,426		0.200 (0.028–1.433)	0.109
2	3,266		0.261 (0.218–0.312)	0.000
3	720		0.432 (0.361–0.516)	0.000
4	201		0.706 (0.578–0.862)	0.001
Radiation recode		0.000		0.000
Yes	4,610		Reference	
No	5,006		1.229 (1.148–1.315)	
Chemotherapy recode		0.000		0.000
Yes	4,536		Reference	
No	5,080		1.271 (1.182–1.366)	
ER status recode		0.000		0.001
Positive	8,791		Reference	
Negative	825		1.235 (1.096–1.391)	
PR status recode		0.000		0.000
Positive	7,646		Reference	
Negative	1,970		1.178 (1.082–1.283)	

Supplement table 3: Univariate and multivariate Cox proportional hazards regression for OS. Primary site: 0, Nipple and Central portion of breast; 1, Upper-outer quadrant and Axillary tail of breast; 2, Lower-outer quadrant of breast; 3, Lower-inner quadrant of breast; 4, Upper-inner quadrant of breast; 5, Overlapping lesion of breast;6, Breast, NOS. ER, estrogen receptor; PR, progesterone receptor.

Figure S1 Flowchart of the cohort selection process for the training set.