Supplementary Information

S1. Supplementary Methods

S2. Supplementary Tables

S3. Supplementary Figures

S1. Supplementary Methods

S1.1 Study Cohorts and Data Collection

SEER cohort

For the first cohort from the SEER database consisting of 18 population-based cancer registries, we selected patients diagnosed with invasive breast cancer between January 1, 2010 and December 31, 2014 (SEER provides HER2 status after 2010). We identified patients according to the following criteria: female, age 18-79, American Joint Committee on Cancer (AJCC) stages I–III, pathologically confirmed breast cancer (ICD-O-3 site code C50), diagnosis not obtained from a death certificate or autopsy, unilateral, known ER/PR/HER2 status, HER2 negative, known time of diagnosis, and breast cancer as the first cancer at diagnosis. ER-PR+HER2- cases were excluded. Data extraction was performed by SEER*Stat software v8.3.5 (http://seer.cancer.gov/seerstat/). Finally, we included 130,856 patients, which containing 13,084 ER+PR-HER2- cases (10.0%).

METABRIC cohort

METABRIC database is a Canada-UK Project which contains targeted sequencing data of 1,980 primary breast cancer samples[1]. Clinical and genomic data was downloaded from cbioportal (http://www.cbioportal.org/study?id=brca_metabric) on September 2, 2016. Though the maximum follow-up time is 351 months (Supplementary Fig. S1A-B), the follow-up time in our analysis was confined to 120 months since 10 years follow-up is enough. METABRIC database only supplied ER immunological histological chemistry (IHC) status. Thus, ER positive was defined as both “ER_IHC” and “ER_status” positive. PR negative was defined as “PR_status”

negative, and HER2 negative was defined as “HER2_status” negative after excluding “HER2_SNP6” gain. METABRIC database only had information about whether chemotherapy and hormone therapy were taken or not without detailed remedy. Genomic data included mRNA expression data (Illumina Human v3 microarray), copy number alteration (CNA) data and mutation data from targeted sequencing of 177 genes. A 1:1 pair match was taken to balance the distribution of age, stage and grade between “hormone therapy” patients and “no hormone therapy” patients.

TCGA cohort

Clinical data is publicly available released by TCGA and were downloaded in “nationwidechildrens.org_clinical_patient_brca” file from “https://tcga-data.nci.nih.gov/publications/tcga”. Eligible patients were as follow: female patients, stage I-III, breast malignancy on December 30, 2016. ER, PR and HER2 status were defined according to IHC staining and fluorescence in situ hybridization (FISH) results: HER2 negative was defined as HER2 IHC score 2+/1+/(0) and HER2 FISH status negative. Follow-up times and overall survival (OS) were updated from the follow-up tables on July 1, 2017. Genomic data, including TCGA Level 3 RNAseq Version 2 RSEM data, Level 3 WES data with tumor-specific mutations, somatic copy number alteration data, and Reverse Phase Protein Array data, and methylation (HM450) data from GDAC on December 30, 2016 (http://gdac.broadinstitute.org). The PAM50 classification of each tumor was downloaded from TCGA reference documents[2]. TCGA expression data, RSEM data, was downloaded from http://gdac.broadinstitute.org/ and transformed by log2(RSEM+1).

MDACC cohort

Three public neo-adjuvant geo datasets (GSE25066, GSE20194, GSE20271)[3-5] from MD Anderson Cancer Center (MDACC) were merged and re-normalized by frozen robust multi-array analysis (fRMA)[6]. ER, PR and HER2 status were defined according to IHC and FISH results. We selected 92 ER+PR-HER2- samples and extracted their microarray-based gene expression data. Probes for EGFR, KRT5 or GATA3 are 201983_s_at, 201820_at and 209603_at respectively.

FUSCC cohort

A prospective observational study cohort. A total of 245 consecutive operable patients treated in the Department of Breast Surgery at Fudan University Shanghai Cancer Center (FUSCC) from January 1, 2007 to December 31, 2014 were recruited according to the following criteria: (i) female patients diagnosed with unilateral disease; (ii) histologically confirmed invasive ductal carcinoma (IDC) or invasive lobular carcinoma (ILC) with the ER+PR-HER2- phenotype; and (iii) no metastatic loci at diagnosis. Exclusion criteria were as follow: (i) patients with breast carcinoma in situ and inflammatory breast cancer; (ii) patients who received any type of treatment before surgery. Pathological examination of tumor specimens was carried out in the Department of Pathology at FUSCC. The status of ER, PR and HER2 was reconfirmed by two experienced pathologists based on immunohistochemistry (IHC) and fluorescence in situ hybridization (FISH) [23-25]. The cutoff for ER-negative and PR-negative IHC status was less than 1% staining in the nuclei. HER2 status was considered negative when an IHC score was 0 or 1, or HER2 amplification was absent (ratio<2.2) by FISH analysis. If any disagreements arose during the evaluation of the IHC and FISH results, a third pathologist was

consulted. Follow-up for the patients was completed on March 1, 2018. The median length of follow-up was 49.9 months (interquartile range [IQR], 33.6 to 67.7 months). Recurrence-free survival (RFS) events included the following: the first recurrence of invasive disease at a local, regional, or distant site; contralateral breast cancer; and death from any cause . Patients without RFS events were censored at the last follow-up.

S1.2 Somatic Copy number Alterations (SCNAs)

Level 4 information of segmented CNA was downloaded in file "gdac.broadinstitute.org_BRCA-TP.CopyNumber_Gistic2.Level_4.2016012800.0.0" (http://gdac.broadinstitute.org) which contained CNA levels defined by GISTIC 2.0[7]. Genomic Identification of Significant Targets in Cancer (GISTIC 2.0) defines CNA calls as follow: -2 = homozygous deletion; -1 = hemizygous deletion; 0 = neutral / no change; 1 = gain; 2 = high level amplification. We defined copy number loss as homozygous deletion or hemizygous deletion.

S1.3 Methylation Level of PR

PR methylation information was extracted from "HM450" downloaded in file "gdac.broadinstitute.org_BRCA.Merge_methylation_humanmethylation450_jhu_usc_edu_Level_3_within_bioassay_data_set_function_data.Level_3.2016012800.0.0" which contains methylation (HM450) beta-values for genes in 885 cases of TCGA (http://gdac.broadinstitute.org). Among the multiple probes of PR, cg01671895 (promoter region) and cg27121959 (enhancer region) showed the most anti-relationship with PR mRNA expression by Pearson's correlation test.

S1.4 Pathway Analysis

GSEA analysis

GSEA software (GSEA 2.2.1) was downloaded from http://software.broadinstitute.org/gsea/) to analyze gene enrichment between groups with 13,310 gene sets downloaded from MSigDB[8]. Input gene expression value was transformed from log2 (RSEM+1). One thousand total permutations were used. The permutation type was set to “phenotype”.

Pathifier score

Pathifier is an algorithm that infers pathway deregulation scores for each tumor sample on the basis of expression data. The algorithm transforms gene-level information into pathway-level information, generating a compact and biologically relevant representation of each sample. Calculation procedures were progressed with R package “pathifier”[9].

S1.5 Detailed gene selection procedure (Figure A1)

a. Differentially expressed genes (DEGs) analysis.

DEGs between luminal-like and non-luminal-like tumors within ER+PR-HER2- breast cancer were calculated by limma test. There are 1,017 significant DEGs (|Fold change| >4, P<0.05), including 488 genes upregulated and 529 genes downregulated in non-luminal-like tumors. Of those DEGs, 23 genes from the PAM50 subtype signature were selected since they showed stable expression pattern across different cohorts. Besides, 7 genes reported to correlate with breast cancer were added based on current literatures. Thus, 30 genes were filtered out as candidate genes.

b. Feasibility of immunohistochemistry

To exclude genes that are not suitable for immunohistochemistry,

Pearson's correlation test between protein level and mRNA expression was operated. There were 15 genes with qualified coefficients included (Pearson's correlation coefficient >0.6). Furthermore, the immunohistochemistry data of each gene were queried from "The Human Protein Atlas" (https://www.proteinatlas.org). Five genes were excluded because of antibody staining mainly not consistent with RNA expression data. Thus 10 genes were left for further screening.

c. Logistic regression of candidate DEGs

Those ten genes were tested for their predictive ability by univariate logistic regression analysis in TCGA and MDACC cohort. Eight genes with significant regression coefficient in both two cohort were included ($P<0.05$), including two genes upregulated (*EGFR, KRT5*) and six genes (*GATA3, TFF1,TFF3, MAPT,SCUBE2* and *BCL2*) downregulated in non-luminal-like group.

Finally, three genes (*EGFR, KRT5 and GATA3*) with extensively reported biological significance and clinical feasibility in breast cancer were selected with priority.

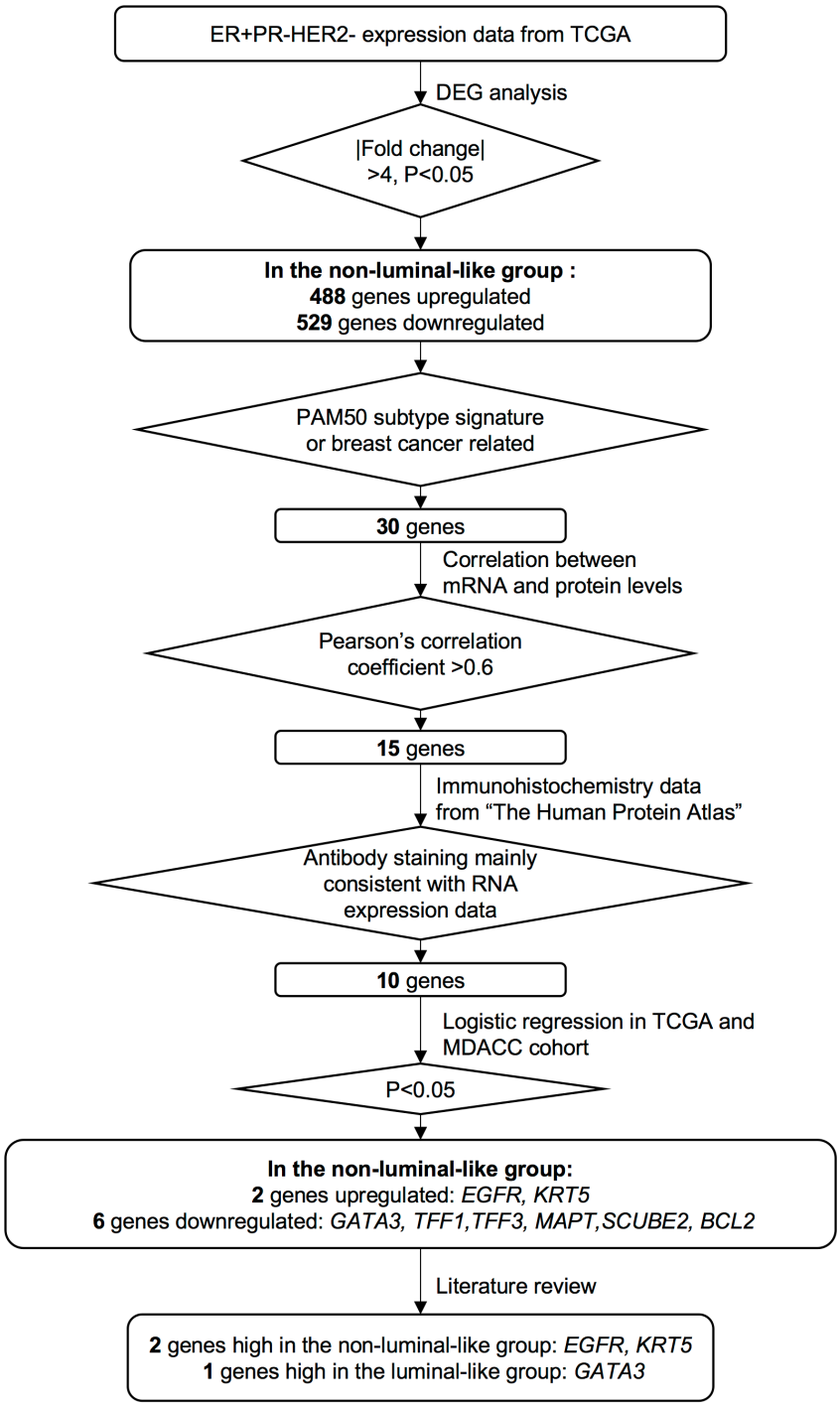


Figure A1. Flow chart of gene selection.

DEG: differentially expressed gene.

S2. Supplementary Tables

Supplementary Table S1. Clinicopathological characteristics of ER+PR-HER2- breast cancer from SEER, METABRIC, TCGA, MDACC and FUSCC cohort

		Cohort1: SEER	Cohort2: METABRIC	Cohort 3: TCGA	Cohort 4: MDACC	Cohort 5: FUSCC
		N=13,084 (%)	N=260 (%)	N= 66 (%)	N=92 (%)	N=245 (%)
Median Follow-up (IQR) (mo)		26 (11-41)	125.8 (75.3-194.0)	28.3 (16.4-55.9)	36.5 (22.5-52.4)	49.9 (33.6-67.7)
Age	18-49	2,375 (18.2)	21 (8.1)	12 (18.2)	40 (43.5)	53 (21.6)
	>=50	10,709 (81.9)	239 (91.9)	54 (81.8)	52 (56.5)	192 (78.4)
Race	White	9,827 (75.1)	-	50 (75.8)	-	-
	Black	1,883 (14.4)	-	7 (10.6)	-	-
	AS/AI/AP	1,278 (9.8)	-	2 (3.0)	-	100 (100.0)
	N/A	96 (0.7)	-	7 (10.6)	-	-
Histologic type	IDC	10,787 (82.4)	260 (100.0)	46 (69.7)	17 (18.5)	225 (91.8)
	ILC	1,747 (13.4)	0	13 (19.7)	0(0.0)	13 (5.3)
	Others and N/A	550 (4.2)	0	7 (10.6)	75 (81.5)	7 (2.9)
Grade	1	2,541 (19.4)	24 (9.2)	-	-	3 (1.2)
	2	4,814 (36.8)	113 (43.5)	-	-	135 (55.1)
	3	5,228 (40.0)	114 (43.9)	-	-	88 (35.9)
	Other|NA	501 (3.8)	9 (3.5)	-	-	19 (7.8)
T stage	T1	7,433 (56.8)	108 (41.5)	12 (18.2)	7 (7.6)	71 (29.0)
	T2	4,365 (33.4)	133 (51.2)	41 (62.1)	48 (52.2)	168 (68.6)
	T3-T4	1,269 (9.7)	18 (6.9)	13 (19.7)	37 (40.2)	6 (2.5)
	N/A	17 (0.1)	1 (0.4)	0 (0.0)	0 (0.0)	0 (0.0)

LN status	Negative	8,809 (67.3)	129 (49.6)	30 (45.5)	30 (32.6)	137 (55.9)
	Positive	4,268 (32.6)	131 (50.4)	36 (54.6)	62 (67.4)	108 (44.1)
	N/A	7 (0.1)	0 (0.0)	0 (0.0)	0 (0.0)	0 (0.0)
Stage	I	6,383 (48.8)	83 (31.9)	9 (13.6)	1 (1.1)	51 (20.8)
	II	4,878 (37.3)	153 (58.9)	39 (59.1)	46 (50.0)	143 (58.4)
	III	1,823 (13.9)	24 (9.2)	18 (27.3)	45 (48.9)	51 (20.8)
Chemotherapy	Yes	6,614 (50.6)	27 (10.4)	37 (56.1)	-	165 (67.4)
	No/ Unknown	6,470 (49.5)	233 (89.6)	29 (43.9)	-	80 (32.6)
Radiation	Yes	7,313 (55.9)	163 (62.7)	24 (36.4)	-	77 (31.4)
	No/Unknown	5,771 (44.1)	97 (37.3)	42 (63.6)	-	168 (68.6)
Endocrine therapy	Yes	-	206 (79.2)	41 (62.1)	-	200 (81.6)
	No/Unknown	-	54 (20.8)	25 (37.9)	-	45 (18.4)
Surgery	BCS	7,343 (56.1)	101 (38.9)	10 (15.2)	-	31 (12.7)
	Mastectomy	5,275 (40.3)	155 (59.6)	32 (48.5)	-	214 (87.4)
	Other|N/A	466 (3.6)	4 (1.5)	24 (36.4)	-	0 (0.0)

AS/AI/AP: Alaskan native/American Indian, and Asian/Pacific Islander, and others-unspecified; BCS: breast conserving surgery; ER: estrogen receptor; FUSCC: Fudan University Shanghai Cancer Center; HER2: human epidermal growth factor receptor 2; IDC: invasive ductal carcinoma; ILC: invasive lobular carcinoma; IQR: interquartile range; LN: lymph node; MDACC: MD Anderson Cancer Center; METABRIC: Molecular Taxonomy of Breast Cancer International Consortium; N/A: not available; PR: progesterone receptor; SEER: Surveillance, Epidemiology, and End Results; TCGA: the Cancer Genome Atlas.

Supplementary Table S2. Log-rank test P value between each two groups from SEER and METABRIC

	SEER cohort		METABRIC cohort	
	BCSS	**OS**	**10y-BCSS**	**10y-OS**
ER+PR-HER2- vs ER+PR+HER2-	<0.001	<0.001	<0.001	<0.001
ER+PR-HER2- vs TNBC	<0.001	<0.001	<0.05	0.241
ER+PR+HER2- vs TNBC	<0.001	<0.001	<0.001	<0.001
All	<0.001	<0.001	<0.001	<0.001

BCSS: breast cancer-specific survival; ER: estrogen receptor; HER2: human epidermal growth factor receptor 2; METABRIC: Molecular Taxonomy of Breast Cancer International Consortium; OS: overall survival; PR: progesterone receptor; SEER: Surveillance, Epidemiology, and End Results; TNBC: triple negative breast cancer.

Supplementary Table S3. Survival rate of each group from SEER and METABRIC

	SEER cohort		METABRIC cohort			
	5y-BCSS	5y-OS	5y-BCSS	10y-BCSS	5y-OS	10y-OS
ER+PR+HER2-	0.968	0.939	0.916	0.807	0.873	0.688
ER+PR-HER2-	0.906	0.873	0.838	0.701	0.786	0.577
TNBC	0.828	0.800	0.713	0.651	0.692	0.564
All	0.939	0.909	0.836	0.731	0.794	0.618

BCSS: breast cancer-specific survival; ER: estrogen receptor; HER2: Human Epidermal Growth Factor Receptor 2; METABRIC: Molecular Taxonomy of Breast Cancer International Consortium; OS: overall survival; PR: progesterone receptor; SEER Surveillance, Epidemiology, and End Results; TNBC: triple negative breast cancer.

Supplementary Table S4. Univariate and multivariate analysis by Cox proportional hazards models of overall survival in SEER and METABRIC cohorts

	SEER[a]		METABRIC[b]	
	Univariate			
	Hazard Ratio (95% CI)	P	Hazard Ratio (95% CI)	P
ER+PR+HER2-	1	-	1	-
ER+PR-HER2-	2.36 (2.18-2.55)	<.001	1.53 (1.20-1.94)	0.001
TNBC	4.39 (4.15-4.65)	<.001	1.81 (1.40-2.34)	<.001
	Multivariate			
	Hazard Ratio (95% CI)	P	Hazard Ratio (95% CI)	P
ER+PR+HER2-	1	-	1	-
ER+PR-HER2-	2.88 (2.62-3.18)	<.001	1.38 (1.09-1.76)	0.009
TNBC	5.58 (5.17-6.02)	<.001	1.81 (1.36-2.40)	<.001

ER: estrogen receptor; HR: hazard ratio; METABRIC: Molecular Taxonomy of Breast Cancer International Consortium; N/A: not available; PR: progesterone receptor; SEER: Surveillance, Epidemiology, and End Results; TNBC: triple negative breast cancer.

[a] Adjusted by age, race, stage, grade, histology, chemotherapy, and surgery.

[b] Adjusted by age, grade, stage, chemotherapy and surgery.

Supplementary Table S5. Mutation events in ER+PR-HER2-, ER+PR+HER2- and TNBC breast cancer from TCGA cohort

		ER+PR+HER2- (N=442)	%	ER+PR-HER2- (N=66)	%	TNBC (N=140)	%	P[a]	P[b]	P for all[c]
Mutation count (Median)		26		37		46		**0.007**[d]	-	**<0.001**[d]
MATH (Median)		37.1		38.8		43.8		0.241[d]	-	
TP53	Wild-type	367	83.0	46	69.7	35	25.0	**0.010**	**0.010**	**<0.001**
	Mutant	75	17.0	20	30.3	105	75.0			
PIK3CA	Wild-type	253	57.3	49	74.2	123	87.9	**0.009**	**0.009**	**<0.001**
	Mutant	189	42.7	17	25.8	17	12.1			
GATA3	Wild-type	374	84.6	56	84.9	137	97.9	0.961	0.867	**<0.001**
	Mutant	68	15.4	10	15.2	3	2.1			
MLL3	Wild-type	394	89.1	60	90.9	131	93.6	0.664	0.582	0.299
	Mutant	48	10.9	6	9.1	9	6.4			
CDH1	Wild-type	356	80.5	53	80.3	134	95.0	0.963	0.899	**<0.001**
	Mutant	86	19.5	13	19.7	7	5.0			

ER: estrogen receptor; HER2: human epidermal growth factor receptor 2; PR: progesterone receptor; MATH: mutant-allele tumor heterogeneity; TCGA: the Cancer Genome Atlas; TNBC: triple negative breast cancer.

[a] P value between ER+PR+HER2- and ER+PR-HER2- by chi-square test and Fisher's exact test if needed.

[b] P value between ER+PR+HER2- and ER+PR-HER2- by logistic regression model adjusted age, race, stage and histology.

[c] P value among all these three groups, chi-square test and Fisher's exact test if needed.

[d] Wilcoxon signed-rank test

Supplementary Table S6. Focal copy number amplification events within ER+HER2- breast cancer from TCGA

Amplification[a]	ER+PR+HER2- (N=433) (%)	ER+PR-HER2- (N=65) (%)	Gene in Region	Chi-square test (P-value)	Logistic model[b] (p-value)
Chr1p22.3	18 (4.2)	4 (6.2)		0.569	0.482
Chr1q21.3	94 (21.7)	17 (26.2)		0.469	0.494
Chr1q44	109 (25.2)	15 (23.1)		0.643	0.627
Chr3p25.1	5 (1.2)	3 (4.6)		0.068	0.100
Chr3q26.32	13 (3.0)	1 (1.5)		0.506	0.506
Chr4q13.3	8 (1.9)	0 (0.0)		0.125	-
Chr5p15.33	14 (3.2)	2 (3.1)		0.661	0.816
Chr6p23	12 (2.8)	2 (3.1)		0. 937	0.840
Chr6q21	6 (1.4)	5 (7.7)		0.005	**0.008**
Chr8p11.21	78 (18.0)	21 (32.3)	KAT6A	0.012	**0.007**
Chr8p11.23	63 (14.6)	16 (24.6)	ZNF703	0.038	**0.044**
Chr8q24.21	106 (24.5)	26 (40.0)	MYC	0.008	**0.009**
Chr10p15.1	7 (1.6)	2 (3.1)		0.084	0.444
Chr10q22.3	9 (2.1)	5 (7.7)		0.030	**0.022**
Chr11p13	10 (2.3)	1 (1.5)		0.865	0.625
Chr11q13.3	86 (19.9)	16 (24.6)		0.138	0.443
Chr11q14.1	41 (9.5)	6 (9.2)		0.951	0.991
Chr12p13.3	6 (1.4)	3 (4.6)		0.068	0.174
Chr12q15	23 (5.3)	4 (6.2)		0.780	0.754
Chr13q34	3 (0.7)	0 (0.0)		0.501	-
Chr14q21.1	9 (2.1)	2 (3.1)		0.610	0.668
Chr15q26.3	15 (3.5)	6 (9.2)	IFG1R	0.031	**0.016**

Chr17p11.2	9 (2.1)	3 (4.6)		0.214	0.258
Chr17q23.1	41 (9.5)	12 (18.5)	TUBD1	0.028	**0.019**
Chr19p13.12	2 (0.5)	2 (3.1)	NOTCH3	0.028	**0.035**
Chr19q13.42	17 (3.9)	0 (0.0)		0.149	-
Chr19q12	8 (1.9)	2 (3.1)		0.510	0.594
Chr20q13.2	46 (10.6)	9 (13.9)		0.588	0.382

ER: estrogen receptor; HER2: human epidermal growth factor receptor 2; PR: progesterone receptor; TCGA: the Cancer Genome Atlas.

[a] Amplification was defined as high level amplification, namely the amplitude threshold equals 2 (t>0.9). For more detailed information, please check the "all_lesions.conf_99.txt" from gistic 2.0 results of TCGA.

[b] Logistic regression model adjusted age, race, stage and histology.

Supplementary Table S7. Focal copy number deletion events within ER+HER2- breast cancer from TCGA

Deletion[a]	ER+PR+HER2- (N=433) (%)	ER+PR-HER2- (N=65) (%)	Genes in region	Chi-square test (P-value)	Logistic model[b] (p-value)
Chr1p36.13	158 (36.5)	27 (41.5)		0.432	0.457
Chr1p22.1	141 (32.6)	24 (36.9)		0.486	0.532
Chr2q37.3	106 (24.5)	17 (26.2)		0.771	0.847
Chr3p21.31	89 (20.6)	25 (38.5)		0.001	**0.001**
Chr4p16.3	92 (21.3)	21 (32.3)		0.047	0.057
Chr4q35.1	99 (22.9)	18 (27.7)		0.392	0.467
Chr5q11.2	46 (10.6)	18 (27.7)	TRIM23, CCNB1	<0.001	**<0.001**
Chr5q21.3	42 (9.7)	18 (27.7)	EFNA5	<0.001	**<0.001**
Chr6p25.3	88 (20.3)	9 (13.9)		0.219	0.260
Chr6q15	164 (37.9)	20 (30.8)		0.268	0.312
Chr6q27	149 (34.4)	27 (41.5)		0.262	0.119
Chr7p22.3	36 (8.3)	10 (15.4)		0.066	0.071
Chr7q36.1	63 (14.6)	12 (18.5)		0.411	0.435
Chr8p23.2	187 (43.2)	40 (61.5)	CSMD1, RNA5SP251	0.006	**0.005**
Chr8q11.21	56 (12.9)	11 (16.9)		0.379	0.351
Chr9p23	120 (27.7)	24 (36.9)		0.127	0.125
Chr9q21.3	123 (28.4)	26 (40.0)		0.057	0.060
Chr9q21.11	93 (21.5)	19 (29.2)		0.163	0.199
Chr9q34.2	86 (19.9)	18 (27.7)		0.148	0.164
Chr10q23.31	99 (22.9)	24 (36.9)	PTEN, SNORD74, KLLN	0.014	**0.019**
Chr10q26.3	91 (21.0)	21 (32.3)		0.042	0.056

Chr11p15.5	85 (19.6)	21 (32.3)		0.020	**0.026**
Chr11q13.2	105 (24.3)	14 (21.5)		0.633	0.574
Chr11q23.3	212 (49.0)	36 (55.4)		0.334	0.376
Chr11q25	188 (43.4)	34 (52.3)		0.179	0.192
Chr12p13.1	66 (15.2)	10 (15.4)		0.976	0.839
Chr12q23.1	48 (11.1)	12 (18.5)		0.088	0.117
Chr12q24.31	53 (12.2)	15 (23.1)	NCOR2	0.018	**0.006**
Chr13q14.2	178 (41.1)	31 (47.7)		0.316	0.397
Chr14q24.1	89 (20.5)	26 (40.0)	MLH3	0.001	**0.001**
Chr15q13.1	105 (24.3)	21 (32.3)		0.163	0.177
Chr16q24.3	329 (76.0)	37 (56.9)	TUBB3	0.001	**0.001**
Chr17p12	227 (52.4)	39 (60.0)		0.254	0.249
Chr17q21.31	95 (21.9)	27 (41.5)	BRCA1, MAPT	0.001	**0.001**
Chr18q23	119 (27.5)	17 (26.2)		0.823	0.769
Chr19p13.3	93 (21.5)	22 (33.9)		0.027	**0.028**
Chr19p13.32	55 (12.7)	15 (23.1)		0.025	**0.025**
Chr20p13	38 (8.8)	7 (10.8)		0.601	0.652
Chr21q11.2	75 (17.3)	16 (24.6)		0.156	0.194
Chr22q13.32	215 (49.7)	26 (40.0)		0.146	0.134

ER: estrogen receptor; HER2: human epidermal growth factor receptor 2; PR: progesterone receptor; TCGA: the Cancer Genome Atlas.

[a] Deletion was defined as hemizgous or homozygous deletion, namely the amplitude threshold equals -1 or -2. For more detailed information, please check the "all_lesions.conf_99.txt" from gistic 2.0 results of TCGA.

[b] Logistic regression model adjusted age, race, stage and histology.

Supplementary Table S8. Focal amplification CNA events from TCGA cohort

Amplification	ER+PR-HER2- %	ER+PR+HER2- %	TNBC %	Chi-square test (P value)
8p11.21	32.3	18.0	9.0	**<0.05**
8p11.23	24.6	14.6	6.6	**<0.05**
10q22.3	7.7	2.1	3.7	**<0.05**
17q23.1	18.5	9.5	4.4	**<0.05**

CNA: copy number alteration; ER: estrogen receptor; HER2: human epidermal growth factor receptor 2; PR: progesterone receptor; TCGA: the Cancer Genome Atlas; TNBC: triple negative breast cancer.

Supplementary Table S9. Gene level CNA events from TCGA and METABRIC cohort

	Chromosome band		TCGA			P-value[a]	METABRIC			P-value[a]
			ER+PR-HER2- (N=65) (%)	ER+PR+HER2- (N=433) (%)	TNBC (N=123) (%)		ER+PR-HER2- (N=260) (%)	ER+PR+HER2- (N=626) (%)	TNBC (N=169) (%)	
KAT6A	8p11.21	Del[b]	11 (16.9)	65 (15.0)	30 (24.3)	**0.011**	20 (7.8)	27 (4.4)	3 (2.0)	0.155
		Amp[c]	9 (13.9)	44 (10.2)	14 (11.4)		11 (4.1)	20 (3.2)	6 (3.5)	
ZNF703	8p11.23	Del	14 (21.5)	80 (18.5)	50 (40.6)	**<0.001**	13 (5.0)	43 (6.9)	7 (6.4)	**<0.001**
		Amp	14 (21.5)	59 (13.6)	9 (7.3)		43 (16.4)	36 (5.8)	3 (2.0)	
RPS6KB1	17q23.1	Loss	15 (23.1)	36 (8.31)	39 (31.7)	**<0.001**	5 (1.8)	14 (2.3)	5.7 (3.4)	**<0.001**
		Amp	12 (18.5)	34 (7.8)	8 (6.5)		15 (5.6)	18 (2.8)	0 (0.0)	

CNA: copy number alteration; ER: estrogen receptor; HER2: human epidermal growth factor receptor 2; METABRIC: Molecular Taxonomy of Breast Cancer International Consortium; PR: progesterone receptor; TCGA: the Cancer Genome Atlas; TNBC: triple negative breast cancer.

[a] Pearson's chi-square test

[b] Del = homozygous deletion / hemizygous deletion

[c] Amp = high level amplification

Supplementary Table S10. Univariate and multivariate analysis of ZNF703/RPS6KB1 amplification by Cox proportional hazards models in ER+HER2- group from METABRIC cohorts

	BCSS		OS			BCSS		OS	
	Hazard Ratio (95% CI)	P	Hazard Ratio (95% CI)	P		Hazard Ratio (95% CI)	P	Hazard Ratio (95% CI)	P
					Univariate				
ZNF703 no amp	1	-	1	-	**RPS6KB1 no amp**	1	-	1	-
ZNF703 amp	2.03 (1.34-3.06)	**0.001**	1.53 (1.07-2.17)	**0.019**	**RPS6KB1 amp**	1.91 (1.01-3.61)	**0.048**	1.34 (0.75-2.39)	0.320
					Multivariate[a]				
ZNF703 no amp	1	-	1	-	**RPS6KB1 no amp**	1	-	1	-
ZNF703 amp	1.93 (1.28-2.92)	**0.002**	1.45 (1.01-2.06)	**0.042**	**RPS6KB1 amp**	1.85 (0.97-3.51)	0.060	1.29 (0.72-2.31)	0.383

amp: amplification; BCSS: breast cancer-specific survival; ER: estrogen receptor; HER2: human epidermal growth factor receptor 2; HR: hazard ratio; METABRIC: Molecular Taxonomy of Breast Cancer International Consortium; OS: overall survival.

[a] Adjusted by age, grade, stage, chemotherapy and surgery.

Supplementary Table S11. Clinicopathologic characteristics of ER+HER2- breast cancer by ZNF703/RPS6KB1 amplification status in ER+HER2- group from METABRIC cohort

		ZNF703 amp	ZNF703 no amp	P	RPS6KB1 amp	RPS6KB1 no amp	P
		N= 79 (%)	N= 807 (%)		N= 32 (%)	N= 854 (%)	
Age	18-49	11 (13.9)	123 (15.2)	0.755	7 (21.9)	127 (14.9)	0.311
	>=50	68 (86.1)	54 (84.8)		25 (78.1)	727 (85.1)	
Histologic type	IDC	79 (100.0)	807 (100.0)	-	32 (100.0)	854 (100.0)	-
	ILC	0 (0.0)	0 (0.0)		0 (0.0)	0 (0.0)	
	Others and N/A	7 (10.6)	7 (10.6)		7 (10.6)	7 (10.6)	
Grade	1	3 (3.8)	99 (12.3)	**0.001**	1 (3.1)	101 (11.8)	**<0.001**
	2	31 (39.2)	411 (50.9)		6 (18.8)	436 (51.1)	
	3	42 (53.2)	259 (32.1)		25 (78.1)	276 (32.3)	
	Other|NA	3 (3.8)	38 (4.7)		0 (0.0)	41 (4.8)	
T stage	T1	31 (39.2)	374 (46.3)	0.243	13 (40.6)	374 (46.3)	0.856
	T2	42 (53.2)	401 (49.7)		18 (53.2)	401 (49.7)	
	T3-T4	6 (7.6)	31 (3.8)		6 (7.6)	31 (3.8)	
	N/A	0 (0.0)	1 (0.1)		0 (0.0)	1 (0.1)	
LN status	Negative	36 (45.6)	343 (42.5)	0.599	15 (46.9)	492 (57.6)	0.228
	Positive	43 (54.4)	464 (57.5)		17 (53.1)	362 (42.4)	
	N/A	0 (0.0)	0 (0.0)		0 (0.0)	0 (0.0)	

Stage	I	31 (39.2)	297 (36.8)	0.065	10 (31.3)	318 (37.2)	0.575
	II	39 (49.4)	468 (58.0)		21 (65.6)	486 (56.9)	
	III	9 (11.4)	42 (5.2)		1 (3.1)	50 (5.9)	
Chemotherapy	Yes	8 (10.1)	74 (9.2)	0.779	2 (6.3)	80 (9.4)	0.550
	No/ Unknown	71 (89.9)	733 (90.8)		30 (93.8)	774 (90.6)	
Radiation	Yes	27 (34.2)	301 (37.3)	0.583	20 (62.5)	538 (63.0)	0.954
	No/Unknown	52 (65.8)	506 (62.7)		12 (37.5)	316 (37.0)	
Endocrine therapy	Yes	62 (78.5)	568 (70.4)	0.130	28 (87.5)	602 (70.5)	**0.037**
	No/Unknown	17 (21.5)	239 (29.6)		4 (12.5)	252 (29.5)	
Surgery	BCS	30 (38.0)	361 (44.7)	0.472	13 (40.6)	378 (44.3)	0.792
	Mastectomy	48 (60.8)	440 (54.5)		19 (59.4)	469 (54.9)	
	Other|N/A	1 (1.27)	6 (0.7)		0 (0.0)	7 (0.8)	

amp: amplification; BCS: breast conserving surgery; ER: estrogen receptor; HER2: human epidermal growth factor receptor 2; IDC: invasive ductal carcinoma; ILC: invasive lobular carcinoma; IQR: interquartile range; LN: lymph node; METABRIC: Molecular Taxonomy of Breast Cancer International Consortium; N/A: not available

[a] Adjusted by age, grade, stage, chemotherapy and surgery.

Supplementary Table S12. Univariate and multivariate analysis of ZNF703/RPS6KB1 expression by Cox proportional hazards models in ER+HER2- group from METABRIC cohorts

	BCSS		OS			BCSS		OS	
	Hazard Ratio (95% CI)	P	Hazard Ratio (95% CI)	P		Hazard Ratio (95% CI)	P	Hazard Ratio (95% CI)	P
					Univariate				
ZNF703 high	1	-	1	-	**RPS6KB1 high**	1	-	1	-
ZNF703 low	1.41 (1.02-1.94)	**0.036**	1.23 (0.95-1.59)	0.112	**RPS6KB1 low**	1.49 (1.10-2.01)	**0.010**	1.42 (1.12-1.79)	**0.004**
					Multivariate[a]				
ZNF703 high	1	-	1	-	**RPS6KB1 high**	1	-	1	-
ZNF703 low	1.29 (0.94-1.79)	0.115	1.15 (0.89-1.49)	0.274	**RPS6KB1 low**	1.41 (1.04-1.91)	**0.027**	1.30 (1.03-1.65)	**0.028**

BCSS: breast cancer-specific survival; ER: estrogen receptor; HER2: human epidermal growth factor receptor 2; HR: hazard ratio; METABRIC: Molecular Taxonomy of Breast Cancer International Consortium. OS: overall survival.

[a] Adjusted by age, grade, stage, chemotherapy and surgery.

Supplementary Table S13. Clinicopathologic characteristics of ER+HER2- breast cancer by ZNF703/RPS6KB1 expression in ER+HER2- group from METABRIC cohort

		ZNF703 high	ZNF703 low	P	RPS6KB1 high	RPS6KB1 low	P
		N= 235 (%)	N= 638 (%)		N= 358 (%)	N= 500 (%)	
Age	18-49	35 (14.9)	97 (15.2)	0.910	42 (11.7)	89 (17.8)	**0.015**
	>=50	200 (85.1)	541 (84.8)		316 (86.3)	411 (82.2)	
Histologic type	IDC	235 (100.0)	638 (100.0)	-	32 (100.0)	854 (100.0)	-
	ILC	0 (0.0)	0 (0.0)		0 (0.0)	0 (0.0)	
	Others and N/A	7 (10.6)	7 (10.6)		7 (10.6)	7 (10.6)	
Grade	1	20 (8.5)	81 (12.7)	**<0.001**	34 (9.5)	66 (13.2)	**<0.001**
	2	98 (41.7)	337 (52.8)		152 (42.5)	274 (54.8)	
	3	107 (45.5)	190 (29.8)		157 (43.9)	138 (27.6)	
	Other|NA	10 (4.3)	30 (4.7)		15 (4.2)	22 (4.4)	
T stage	T1	75 (31.9)	248 (38.9)	0.131	117 (32.7)	195 (39.0)	0.119
	T2	141 (60.0)	358 (56.1)		222 (62.0)	273 (54.6)	
	T3-T4	19 (8.1)	32 (5.0)		19 (5.3)	32 (6.4)	
LN status	Negative	121 (51.5)	377 (59.1)	**0.044**	194 (54.2)	291 (58.2)	0.243
	Positive	114 (48.5)	261 (40.9)		164 (45.8)	209 (41.8)	
Stage	I	75 (31.9)	248 (38.9)	0.065	117 (32.7)	195 (39.0)	0.096
	II	141 (60.0)	358 (56.1)		222 (62.0)	273 (54.6)	

	III	19 (8.1)	32 (5.0)		19 (5.3)	32 (6.4)	
Chemotherapy	Yes	30 (12.8)	50 (7.8)	**0.025**	26 (7.3)	52 (10.4)	0.115
	No/ Unknown	205 (87.2)	588 (92.2)		332 (92.7)	448 (89.6)	
Radiation	Yes	155 (66.0)	395 (61.9)	0.272	226 (63.1)	319 (63.8)	0.840
	No/Unknown	80 (34.0)	243 (38.1)		132 (36.9)	181 (36.2)	
Endocrine therapy	Yes	183 (77.9)	438 (68.7)	**0.008**	264 (73.7)	349 (69.8)	0.207
	No/Unknown	52 (22.1)	200 (31.4)		94 (26.3)	151 (30.2)	
Surgery	BCS	95 (40.4)	289 (45.3)	0.302	148 (41.3)	233 (46.6)	0.122
	Mastectomy	139 (59.2)	343 (53.8)		209 (58.4)	262 (52.4)	
	Other|N/A	1 (0.4)	6 (0.9)		1 (0.3)	5 (1.0)	

BCS: breast conserving surgery; ER: estrogen receptor; HER2: human epidermal growth factor receptor 2; IDC: invasive ductal carcinoma; ILC: invasive lobular carcinoma; IQR: interquartile range; LN: lymph node; METABRIC: Molecular Taxonomy of Breast Cancer International Consortium; N/A: not available.

Supplementary Table S14. Enriched pathways in ER+HER2- tumors with ZNF703 amplification in C2 sets (curated sets) by GSEA (NOM P<0.01)

Name	Size	NES	NOM p-val	FDR q-val
JAERVINEN_AMPLIFIED_IN_LARYNGEAL_CANCER	40	2.1463692	0	0.37563515
NIKOLSKY_BREAST_CANCER_8P12_P11_AMPLICON	55	2.0922172	0	0.26544356
AGUIRRE_PANCREATIC_CANCER_COPY_NUMBER_UP	293	2.0476024	0.003984064	0.2695679
BOYAULT_LIVER_CANCER_SUBCLASS_G3_UP	187	1.9883463	0.0041841	0.3507712
KEGG_RNA_DEGRADATION	57	1.9427629	0.007936508	0.43051714
NUNODA_RESPONSE_TO_DASATINIB_IMATINIB_UP	29	1.8810203	0	0.64612246
KEGG_HOMOLOGOUS_RECOMBINATION	26	1.8768625	0	0.5731757
REACTOME_G1_PHASE	35	1.8564644	0.008264462	0.60678667
NIKOLSKY_BREAST_CANCER_8Q12_Q22_AMPLICON	130	1.8510648	0.004608295	0.5066537
NIKOLSKY_MUTATED_AND_AMPLIFIED_IN_BREAST_CANCER	94	1.8481921	0	0.47224188
KEGG_CELL_CYCLE	118	1.8423429	0	0.45538518
MOHANKUMAR_TLX1_TARGETS_UP	405	1.8363556	0.008230452	0.44589967
REACTOME_AMINE_DERIVED_HORMONES	15	1.8243464	0	0.40285036
XU_RESPONSE_TO_TRETINOIN_AND_NSC682994_DN	15	1.8134319	0	0.37755236

UDAYAKUMAR_MED1_TARGETS_UP	132	1.8020736	0.003952569	0.3435485
LI_WILMS_TUMOR_VS_FETAL_KIDNEY_1_DN	161	1.7976847	0.007662835	0.32909077
WANG_METASTASIS_OF_BREAST_CANCER_ESR1_UP	22	1.7870872	0.004	0.33302078
ZEMBUTSU_SENSITIVITY_TO_METHOTREXATE	18	1.7779918	0.004524887	0.32640707
REACTOME_KINESINS	24	1.7708131	0.004048583	0.32511544
WHITFIELD_CELL_CYCLE_S	151	1.7688916	0.00390625	0.31013957
JEON_SMAD6_TARGETS_DN	18	1.7478148	0	0.3174453
DORMOY_ELAVL1_TARGETS	16	1.7305571	0	0.3334766
PID_MTOR_4PATHWAY	67	1.7143723	0.007692308	0.33521762
REACTOME_FACTORS_INVOLVED_IN_MEGAKARYOCYTE_DEVELOPMENT_AND_PLATELET_PRODUCTION	125	1.7010847	0.008658009	0.3610138
POMEROY_MEDULLOBLASTOMA_DESMOPLASIC_VS_CLASSIC_UP	60	1.6998909	0.007782101	0.35763603
REACTOME_G0_AND_EARLY_G1	23	1.6829524	0.004032258	0.37077963
ZEMBUTSU_SENSITIVITY_TO_CYCLOPHOSPHAMIDE	17	1.6320975	0.004	0.3982811
REACTOME_G1_S_SPECIFIC_TRANSCRIPTION	16	1.5360686	0	0.3465507
CAMPS_COLON_CANCER_COPY_NUMBER_UP	89	1.527604	0.007968128	0.34364656

ER: estrogen receptor; HER2: human epidermal growth factor receptor 2; GSEA: gene set enrichment analysis; ZNF703: zinc finger protein 703.

Supplementary Table S15. Enriched pathways in ER+HER2- tumors with ZNF703 amplification in C3 sets (motif sets) by GSEA (NOM P<0.01)

NAME	SIZE	NES	NOM p-val	FDR q-val
E2F_Q4_01	233	1.8727309	0.004	0.61495274
E2F_Q6_01	234	1.8433999	0.008032128	0.38300863
CCAGGTT_MIR490	63	1.8360833	0	0.2672649
E2F_03	237	1.8265266	0.004255319	0.22759774
TGCACGA_MIR517A_MIR517C	17	1.7943902	0.004149378	0.12564214
GABP_B	253	1.7539319	0.008097166	0.09625589
GGCKCATGS_UNKNOWN	66	1.6874338	0.004484305	0.13301794
CAGNWMCNNNGAC_UNKNOWN	84	1.6749204	0	0.12379758
ZF5_B	234	1.5816078	0.008474576	0.22577669

ER: estrogen receptor; HER2: human epidermal growth factor receptor 2; GSEA: gene set enrichment analysis; ZNF703: zinc finger protein 703.

Supplementary Table S16. Multivariate analysis of "three-marker" for RFS in FUSCC cohort

	HR	95% CI	P
Three-marker			
Luminal-like	1	-	-
Non-luminal-like	3.12	1.61-6.03	**0.001**
Chemotherapy	1.01	0.68-1.49	0.965
Radiation	1.06	0.71-1.59	0.767
LN stage	1.52	1.21-1.92	**0.000**
Tumor size	1.30	1.04-1.61	**0.018**
Grade	1.14	1.03-1.27	**0.016**
Age at diagnosis	1.00	0.98-1.03	0.783

CI: confidence interval; FUSCC: Fudan University Shanghai Cancer Center; HR: hazard ratio; LN: lymph node; RFS: recurrence-free survival.

Supplementary Table S17. Interaction test for RFS in FUSCC cohort

	HR (95% CI)	P
Hormone therapy (Insufficient vs Sufficient)	2.81 (1.62-4.86)	<0.001
Three-marker (Non-luminal vs Luminal)	2.78 (1.47-5.27)	0.002
Interaction	5.00 (2.21-11.29)	<0.001

CI: confidence interval; FUSCC: Fudan University Shanghai Cancer Center; HR: hazard ratio; RFS: recurrence-free survival.

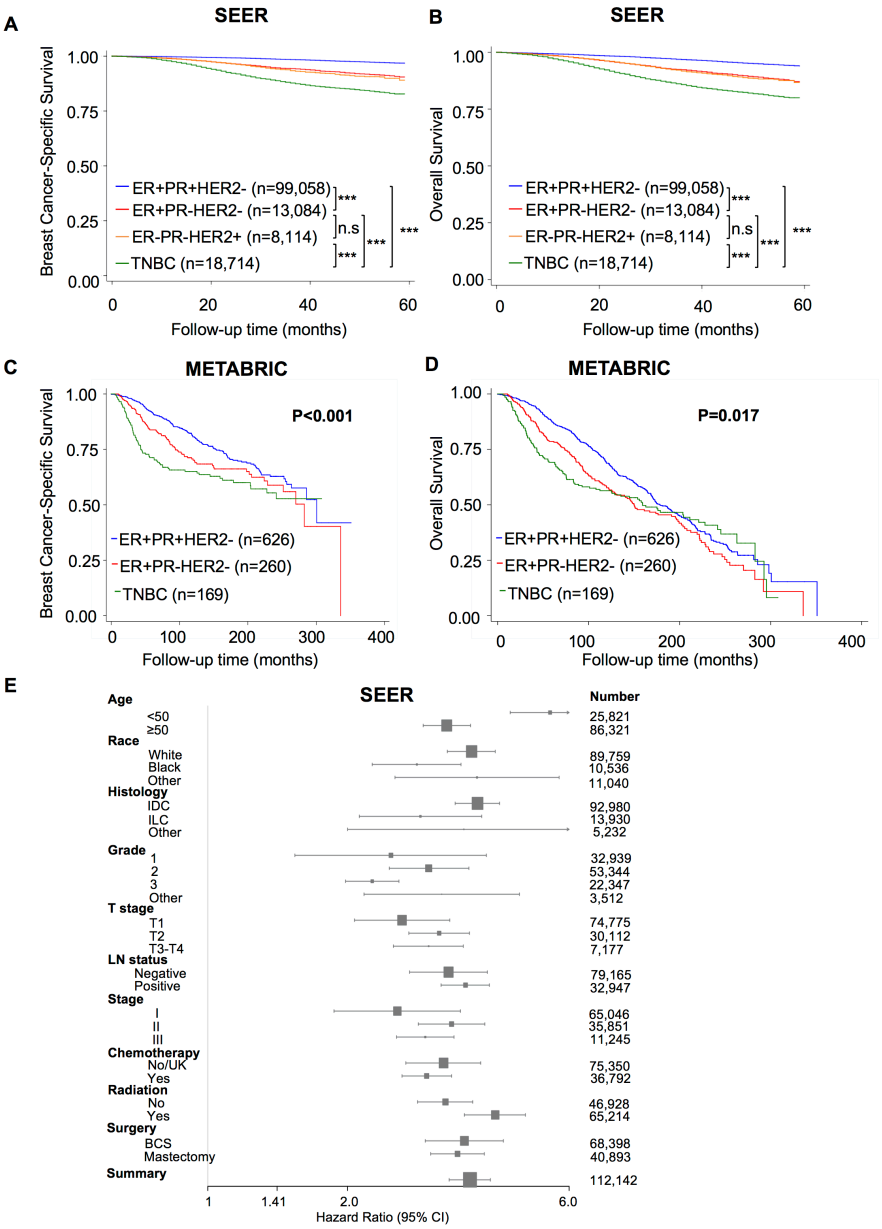
Supplementary Table S18. Treatment efficacy for luminal-like and non-luminal-like ER+PR-HER2- cases in FUSCC cohort

	Luminal-like (n=10)	Non-luminal-like (n=10)	P value
Median recurrence-free time during adjuvant hormone therapy (IQR)	32.0 (21.8-37.3)	12.0 (6.0-17.2)	**0.017**
Median progression-free time during salvage hormone therapy (IQR)	18.5 (9.0-27.8)	3.0 (2.5-4.5)	**0.034**
Median progression-free time during salvage chemotherapy (IQR)	6.0 (4.0-7.0)	4.5 (3.8-8.3)	0.724

ER: estrogen receptor; FUSCC: Fudan University Shanghai Cancer Center; HER2: human epidermal growth factor receptor 2; IQR: interquartile range; PR: progesterone receptor.

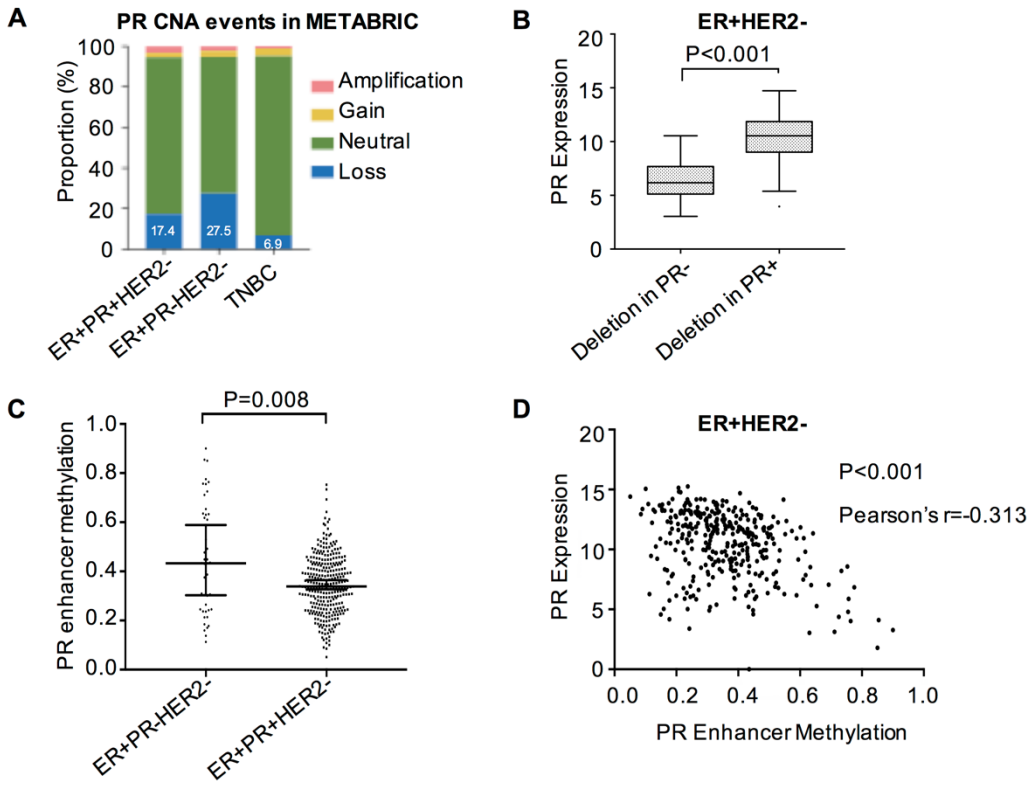
S3. Supplementary Figures

Supplementary Figure S1



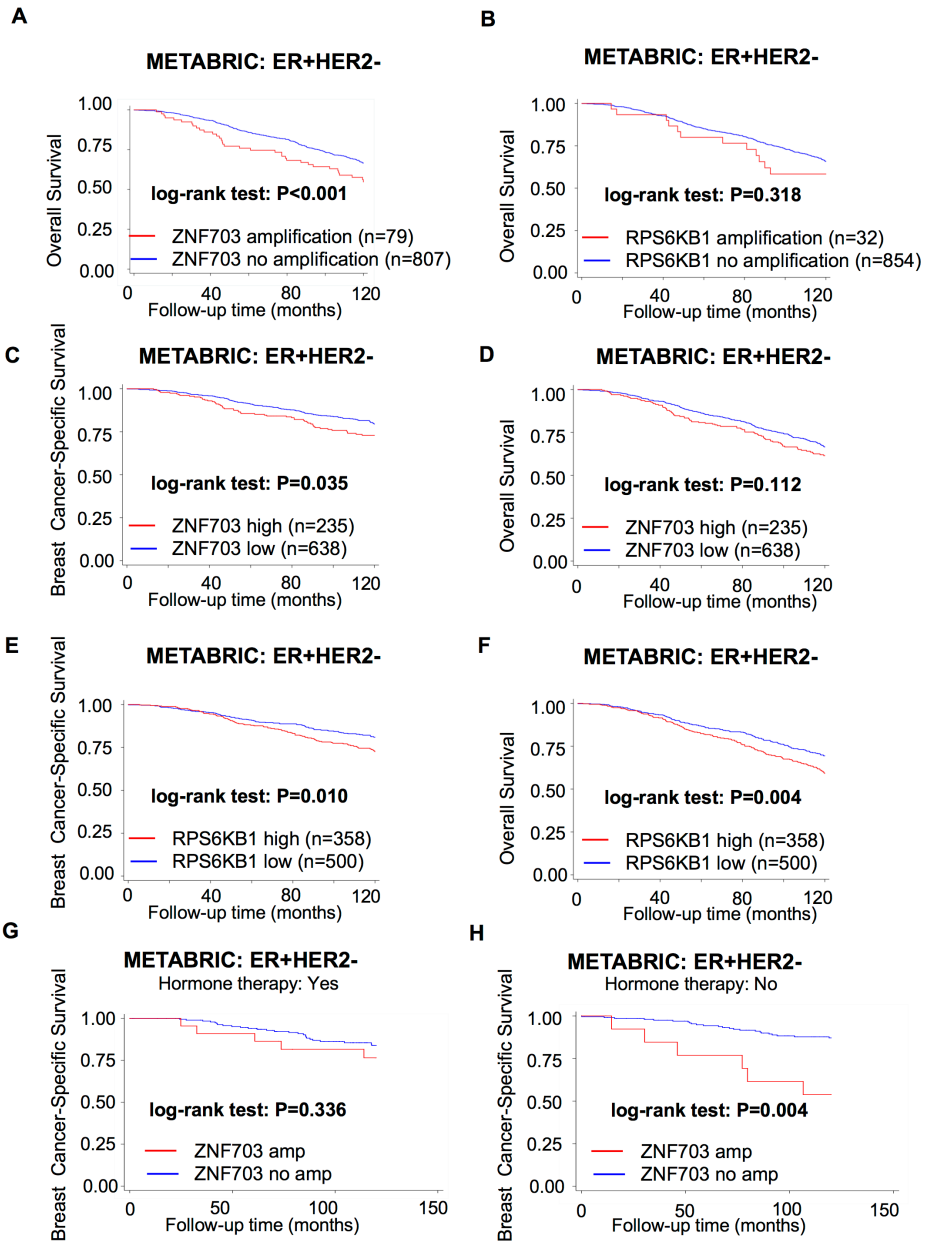
Supplementary Figure S1, (A) Breast cancer-specific survival or **(B)** Overall survival of each group in SEER cohort. **(C)** Breast cancer-specific survival or **(D)** Overall survival of each group in METABRIC cohort with long term follow-up. **(E)** Hazard ratio with 95% confidence interval (CI) of PR-negative in ER+HER2- tumors by subgroup analysis from SEER cohort. n.s: not significant.

Supplementary Figure S2



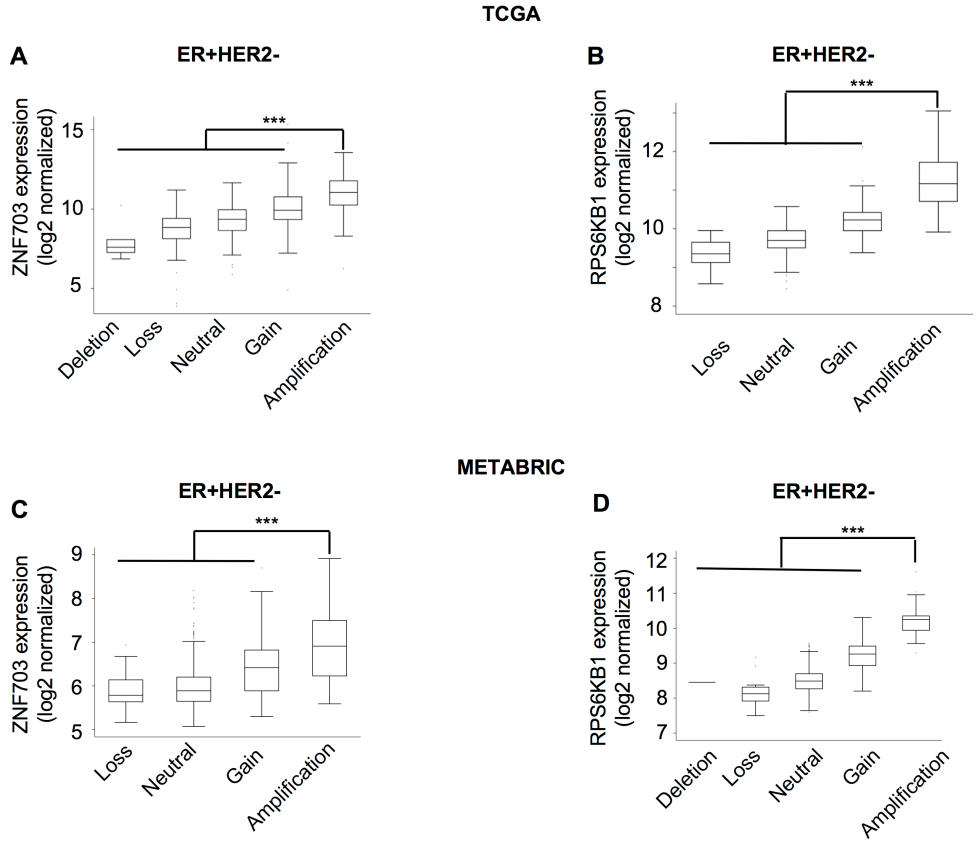
Supplementary Figure S2, (A) Progesterone receptor (PR) copy number alteration (CNA) status in METABRIC cohort. **(B)** PR expression in ER+HER2- tumors with PR copy number deletion in TCGA cohort. **(C)** PR enhancer methylation level in ER+PR-HER2- and ER+PR+HER2- tumors in TCGA cohort. **(D)** Correlation between PR expression and PR enhancer methylation level within ER+HER2- tumors in TCGA cohort.

Supplementary Figure S3



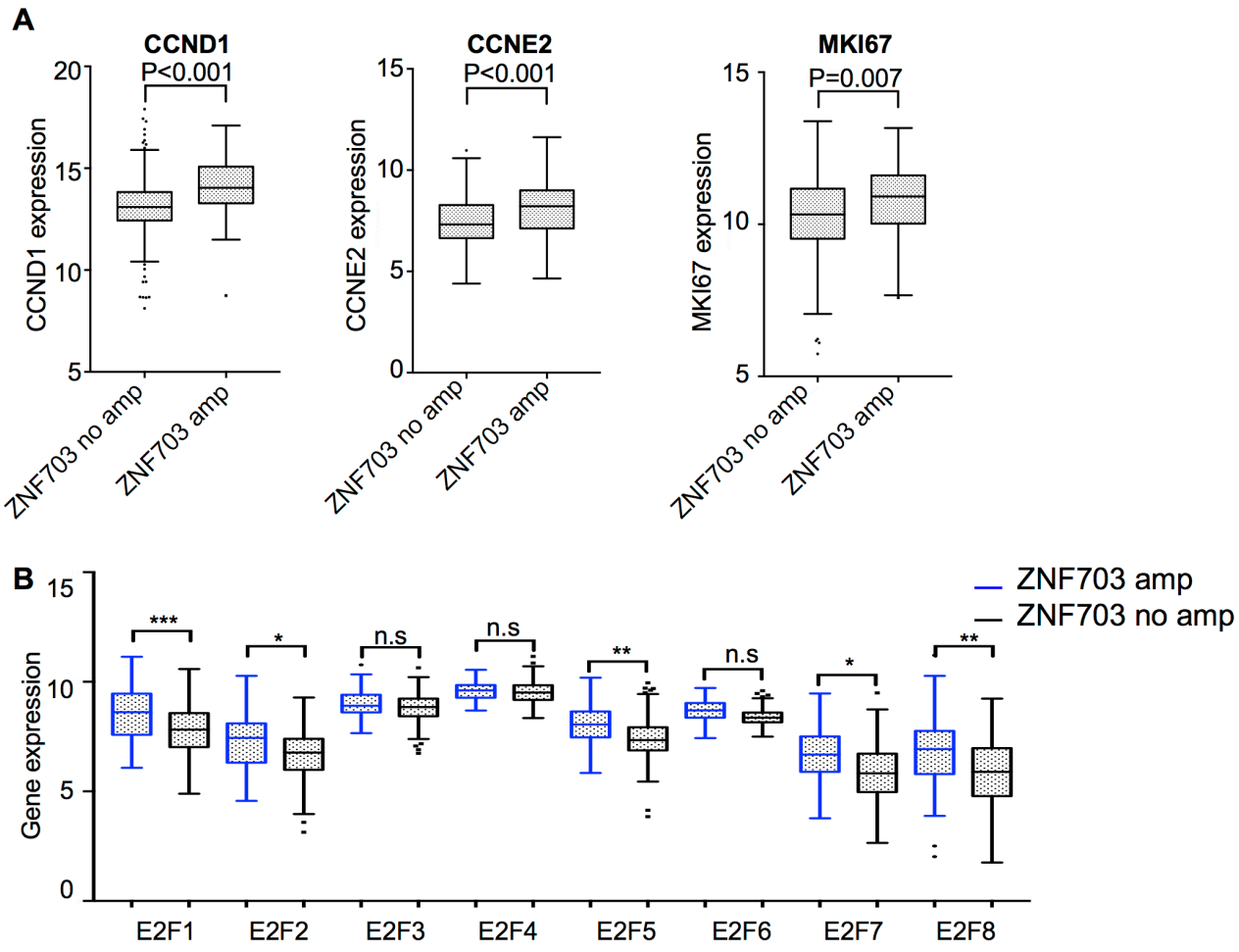
Supplementary Figure S3, Overall survival of ER+HER2- group in METABRIC cohort by **(A)** ZNF703 amplification or **(B)** RPS6KB1 amplification. **(C)** Breast cancer-specific survival (BCSS) or **(D)** Overall survival (OS) of ER+HER2- group in METABRIC cohort by ZNF703 expression. **(E)** BCSS or **(F)** OS of ER+HER2- group in METABRIC cohort by RPS6KB1 expression. **(G)** BCSS by ZNF703 amplification status in hormone therapy and **(H)** in no hormone therapy received ER+HER2- cases from METABRIC cohort. amp: amplification; no amp: not amplification.

Supplementary Figure S4



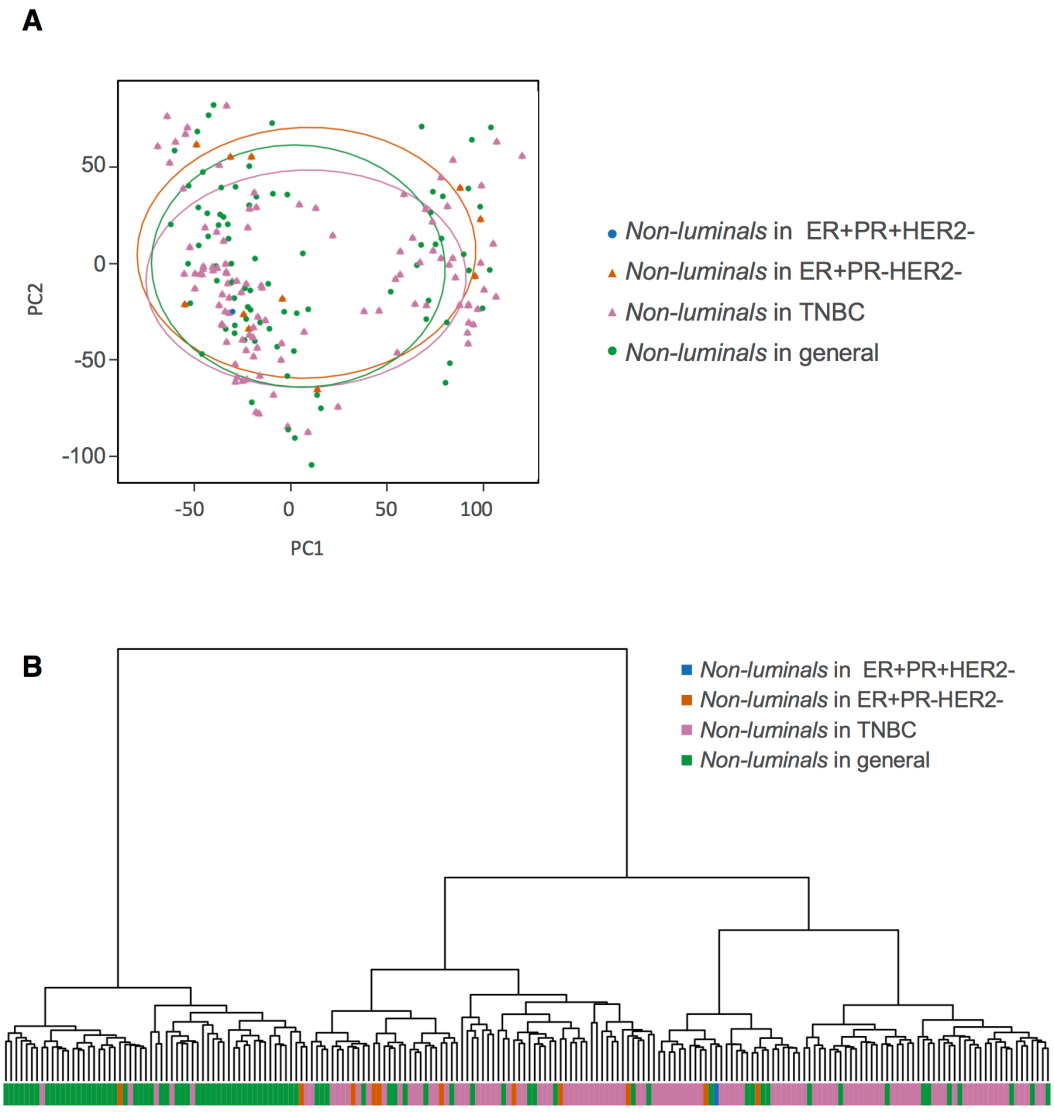
Supplementary Figure S4, Correlation between **(A)** ZNF703 or **(B)** RPS6KB1 gene expression level and copy number status within ER+HER2- tumors in TCGA cohort. Correlation between **(C)** ZNF703 or **(D)** RPS6KB1 gene expression level and copy number status within ER+HER2- tumors in METABRIC cohort. Kruskal-Wallis test: *: P<0.05; **: P<0.01; ***: P<0.001.

Supplementary Figure S5



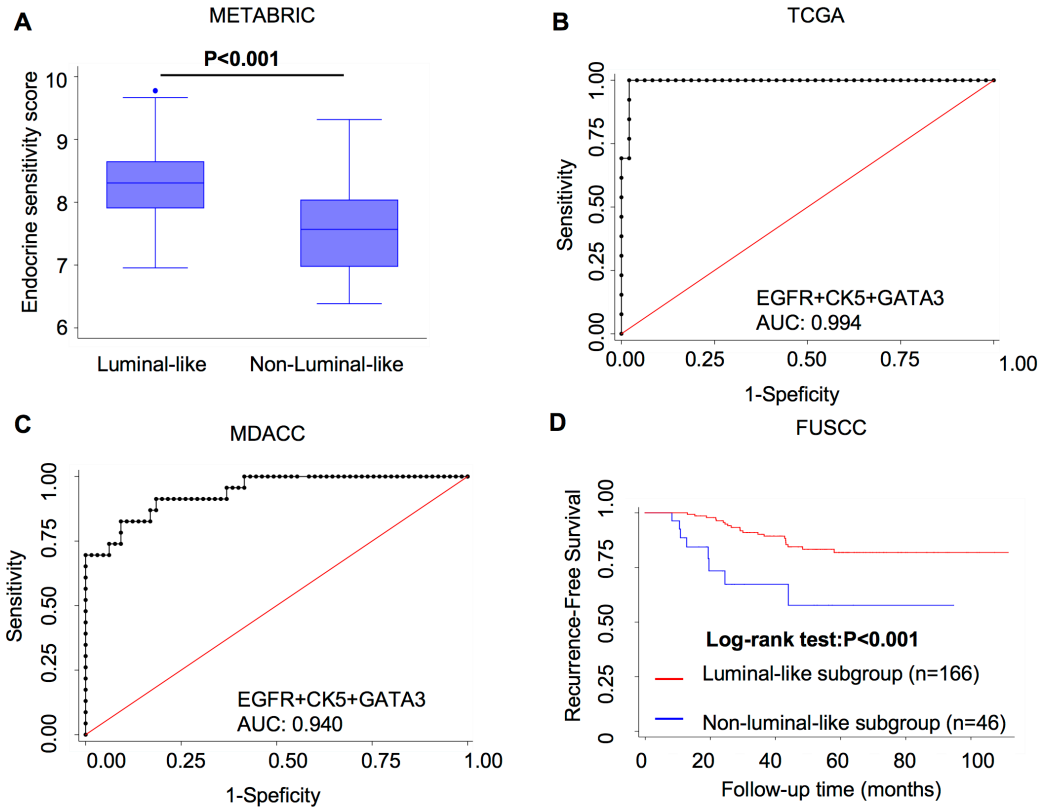
Supplementary Figure S5, *ZNF703* amplification correlated with cell-cycle progression via E2F regulation. **(A)** Expression levels of cell-cycle related genes (*CCND1*, *CCNE2*, *MKI67*) within ER+HER2- tumors in TCGA cohort. Mann-Whitney test was used. **(B)** Expression levels of E2F family genes within ER+HER2- tumors in TCGA cohort. Kruskal-Wallis test: *: P<0.05; **: P<0.01; ***: P<0.001; n.s: not significant.

Supplementary Figure S6



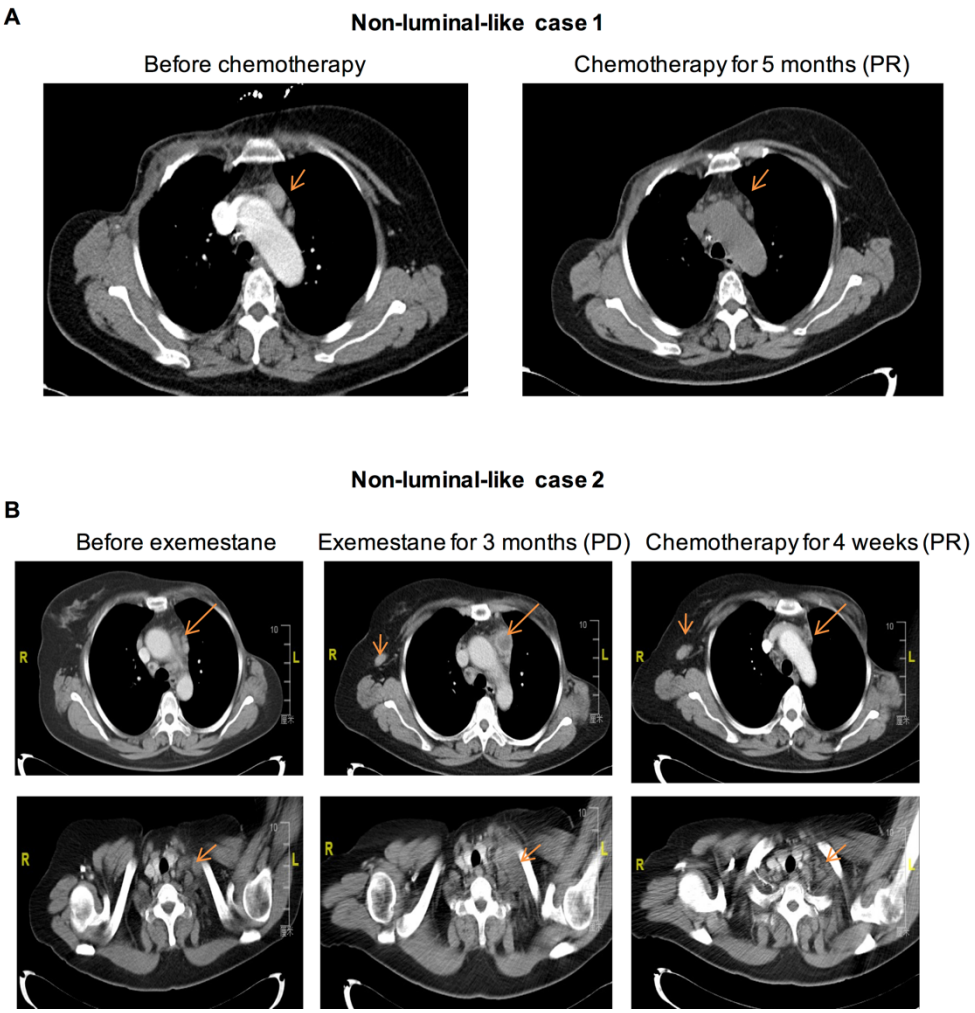
Supplementary Figure S6, Non-luminals in ER+PR-HER2- breast cancer clustered together with non-luminals in general **(A)** Principle component analysis (PCA) and **(B)** Hierarchical clustering of non-luminals from TCGA dataset.

Supplementary Figure S7



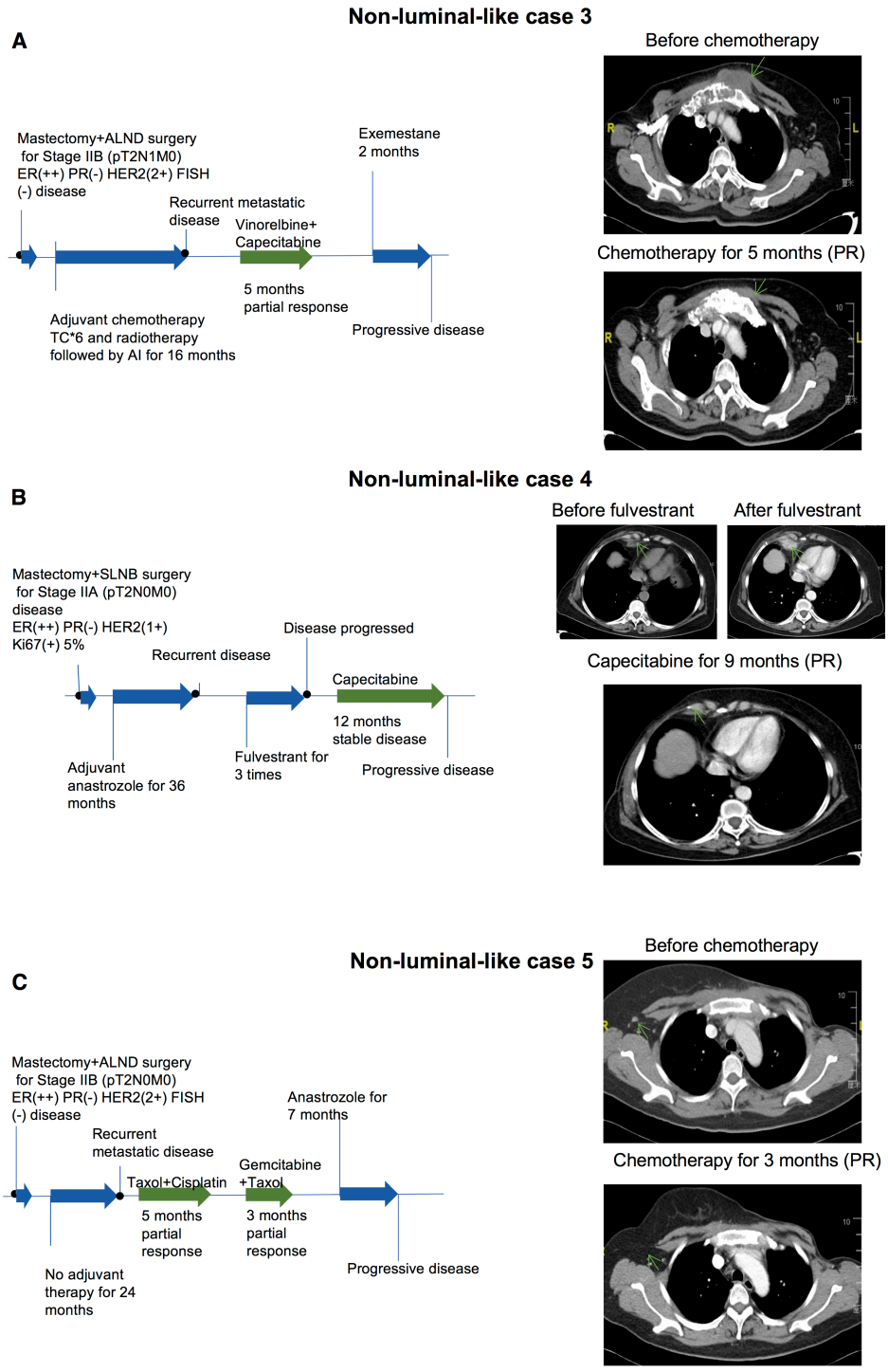
Supplementary Figure S7, **(A)** Endocrine sensitivity scores between luminal-like and non-luminal-like subgroups within ER+PR-HER2- tumors from METABRIC cohort. **(B)** Receiver operating characteristic (ROC) curve of three genes (CK5, EGFR, GATA3) in predicting non-luminal-like subtypes within ER+PR-HER2- tumors in TCGA cohort and **(C)** in MDACC cohort. Area under curve (AUC) was calculated. **(D)** Recurrence-free survival of luminal-like and non-luminal-like subgroups within ER+PR-HER2- tumors in FUSCC cohort.

Supplementary Figure S8



Supplementary Figure S8, Two non-luminal-like patients were identified by IHC-based three gene classifier. **(A)** Metastasis foci (orange arrow) shrunk back greatly after 5 months-chemotherapy in non-luminal-like case 1. **(B)** Metastasis foci (orange arrows) progressed quickly after 3 months-exemestane but shrunk greatly after 4 weeks-chemotherapy in non-luminal-like case 2. PR, partial response; PD, progressive disease.

Supplementary Figure S9



Supplementary Figure S9, Treatment procedure of another three non-luminal-like patients **(A)** Metastasis foci (green arrow) shrunk back greatly after 5 months-chemotherapy in non-luminal-like case 3. **(B)** Metastasis foci (green arrow) kept progressing after 3 times-fulvestrant but shrunk greatly and

kept stable during 9 months-chemotherapy in non-luminal-like case 4. **(C)** Metastasis foci (green arrow) kept shrunk greatly during 3 months-chemotherapy in non-luminal-like case 5. PR, partial response; PD, progressive disease.

References

1. Pereira B, Chin SF, Rueda OM, Vollan HK, Provenzano E, Bardwell HA, et al. The somatic mutation profiles of 2,433 breast cancers refines their genomic and transcriptomic landscapes. Nature communications. 2016; 7: 11479.
2. Ciriello G, Gatza ML, Beck AH, Wilkerson MD, Rhie SK, Pastore A, et al. Comprehensive Molecular Portraits of Invasive Lobular Breast Cancer. Cell. 2015; 163: 506-19.
3. Popovici V, Chen W, Gallas BG, Hatzis C, Shi W, Samuelson FW, et al. Effect of training-sample size and classification difficulty on the accuracy of genomic predictors. Breast Cancer Res. 2010; 12: R5.
4. Hatzis C, Pusztai L, Valero V, Booser DJ, Esserman L, Lluch A, et al. A genomic predictor of response and survival following taxane-anthracycline chemotherapy for invasive breast cancer. Jama. 2011; 305: 1873-81.
5. Tabchy A, Valero V, Vidaurre T, Lluch A, Gomez H, Martin M, et al. Evaluation of a 30-gene paclitaxel, fluorouracil, doxorubicin, and cyclophosphamide chemotherapy response predictor in a multicenter randomized trial in breast cancer. Clinical cancer research : an official journal of the American Association for Cancer Research. 2010; 16: 5351-61.
6. McCall MN, Bolstad BM, Irizarry RA. Frozen robust multiarray analysis (fRMA). Biostatistics. 2010; 11: 242-53.
7. Mermel CH, Schumacher SE, Hill B, Meyerson ML, Beroukhim R, Getz G. GISTIC2.0 facilitates sensitive and confident localization of the targets of focal somatic copy-number alteration in human cancers. Genome biology. 2011; 12: R41.
8. Subramanian A, Tamayo P, Mootha VK, Mukherjee S, Ebert BL, Gillette MA, et al. Gene set enrichment analysis: a knowledge-based approach for interpreting genome-wide expression profiles. Proc Natl Acad Sci U S A. 2005; 102: 15545-50.
9. Drier Y, Sheffer M, Domany E. Pathway-based personalized analysis of cancer. Proc Natl Acad Sci U S A. 2013; 110: 6388-93.