Additional file 1: Method 1-4; Table S1-S21; Figure S1-S29.

Contents

Method 1

Source of public data

Genome-wide, level 3 processed, Infinium Human Methylation 450K array and 850K array were obtained from The Cancer Genome Atlas (TCGA) and Gene Expression Omnibus (GEO). Illumina normalized 450K methylation array of tissue samples from 30 different tumor types were downloaded from UCSC Xena (https://xena.ucsc.edu/public/), including 774 breast cancer (BC) tissues, 97 adjacent tissues, and 3,655 tissue samples from 29 other tumor types. Illumina normalized 450K methylation data from 859 tissue samples obtained from patients with BC were downloaded from GEO (GSE52865 [1], GSE66695, GSE69914 [2], GSE60185 [3]). Illumina normalized 450K methylation data from 911 samples of white blood cells (WBC) were downloaded from GEO (GSE40279 [4] and GSE51032). Additionally, 11 Illumina normalized 850K methylation datasets were obtained from GEO, involving tissue samples of 8 other common tumor types (GSE136380 [5], GSE136583 [5], GSE114989 [6], GSE159898 [7], GSE149282 [8], GSE200484, GSE192967, GSE161651 [9], GSE171142 [10], and GSE146003 [11]) and 144 samples of WBC (GSE140038 [12]). Illumina normalized 450K methylation data from 119 BC tissues were obtained from GSE72251, with survival information. All the samples included in the study were female samples without preoperative chemotherapy or radiotherapy. The methylation level for each CpG was represented as a beta value (β), which is a ratio of intensities between the methylated (M) allele and the sum of methylated (M) and unmethylated (U), and it was expressed as a continuous variable that ranged from 0 (no methylation) to 1 (full methylation). Details of the sample sources and sample sizes can be found in Table S1.

Preanalytical processing

Tissue sample collection and DNA extraction

For BC patients, tissue samples were collected during the surgery and immediately snap-frozen in liquid nitrogen and then stored at -80 ℃. Genomic DNA was extracted from the fresh-frozen tumor tissues with the QIAamp DNA Mini Kit (QIAGEN, Germany).

Blood sample collection and processing

For BC patients, breast benign tumor patients, healthy controls, and other type tumor patients, 8-10 mL whole blood samples were collected prior to any treatment using Ardent Cell-Free DNA blood collection tubes. The following steps were immediately performed (<4h): (i) centrifugation at 800g for 10 minutes at room temperature, (ii) separation of plasma (1750 μL aliquot) from buffy coat (500 μL aliquot) carefully, (iii) plasma was then centrifugated at 16,000 g for 10 minutes at 4 ° C to remove any remaining cells and subsequent collection of plasma phase in 1700 μL aliquot. (iv) Plasma and buffy coat were immediately stored at -80°C until assay. For each subject, approximately 3 aliquots of plasma were obtained.

DNA extraction and bisulfite conversion

Circulating cell-free (cfDNA) was extracted using the QIAamp® Circulating Nucleic Acid Kit (QIAGEN®, Germany) from 4-5 mL plasma, according to the manufacturer's instructions. cfDNA concentration was measured using a Qubit3.0 fluorometer (Thermo Fisher Scientific) before bisulfite treatment.

Genomic DNA was extracted from 200 μL of buffy coat using QIAamp® DNA Mini kit (QIAGEN®,

Germany), according to the manufacturer's instructions. DNAs were quantified using NanoDrop 2000 (Thermo Scientific, Wilmington, DE, United States) and then immediately stored at -20°C.

The extracted cfDNA samples and genomic DNAs were bisulfite converted using the EZ DNA Methylation-Gold™ Kit (Zymo Research, United States), according to the manufacturer's guidelines.

Method 2

Evaluation and validation of the diagnostic effect of BC tissue-specific methylation markers in 450k methylation data in public datasets

For candidate BC-specific methylation CpG sites, TCGA data of 774 BC and 97 tumor-adjacent tissue samples were used as the training set. Four 450K methylation datasets (GSE52865, GSE66695, GSE69914, and GSE60185) were utilized as external validation sets. To evaluate the diagnostic efficacy of each CpG site, the receiver operating characteristic curves (ROC) were used to obtain the area under the ROC curve (AUC) in the TCGA. According to the maximum Youden index, the evaluation indexes were obtained: the best cut-off value, sensitivity, and specificity.

Correlation between methylation and expression of candidate methylation markers in BC and tumor-adjacent tissues in the TCGA dataset

Considering the potential functions of candidate CpG sites in gene regulation, we analyzed the correlation between the methylation levels of these CpG sites and the expression levels of the corresponding genes by Pearson or Spearman method in the TCGA dataset.

Analysis of the tissue and buffy coat DNA methylation level with quantitative methylation-specific PCR (qMSP) assay

qMSP assay was performed for comparing the methylation difference of candidate markers between BC and tumor-adjacent tissues, and WBC. All primers and probes were designed using ABI Primer Express software (version 3.0.1) or manually. They were synthesized by Sangon Biotech (Shanghai, China). For qMSP reactions, each target gene was analyzed together with a methylation-insensitive target gene of *β*-actin (*ACTB*) control assay in the same reaction using FAM and VIC-based probes, respectively. The primers and probes of *ACTB* were reported elsewhere[13] The LightCycler® 480 Probes Master Mix (2×) Kit was used for qPCR. The reactions were performed in a 20μL final volume system with 10 μL of 2× Master Mix, 0.5 μM of each primer, 0.1 μM of each probe, and 10 to 20 ng of bisulfite-converted DNA. PCR cycling conditions were as follows: pre-incubation at 95 °C for 10 min, followed by 40 cycles of 95 °C for 10s and 60 °C for 30s amplification, and cooling. The ΔCq value was used to represent the methylation level, ΔCq = Cqreference (VIC) – Cqtarget (FAM). Primer and probe sequences are listed in Table S2.

Evaluation and validation of the prognotic effect of BC tissue-specific methylation markers in TCGA and GSE72251 cohort

Among 774 TCGA BC tissues, a total of 754 BC tissues included DNA methylation data, overall survival (OS) time (> 30 days), and survival status. OS was defined as the time from surgery to death from any cause or the last follow-up visit. For GSE72251 cohort, 118 BC tissues with clinical and DNA methylation data were included, though without the stage information.

For candidate methylation CpG sites, the optimal cutoff values for predicting the OS in TCGA BC patients were identified by X-tile 3.6.1 software (Yale University, New Haven, CT, USA), classifying patients into low-risk and high-risk groups. The Kaplan-Meier and log-rank tests were utilized to generate the survival curves and compare the survival differences between the groups.Then, we fitted a univariate Cox proportional hazard model in 754 BC tissues, and CpG sites with $p < 0.05$.were obtained. Subsequently, we fitted a multivariate Cox proportional hazard model with that CpG site as the predictor variable, and age, subtype, and stage as covariates, and again, CpG sites with $p < 0.05$.were obtained. The methylation risk score (MRS) was the sum of the methylation level of CpG multiplied by their multivariate Cox regression coefficient. MRS was also divided into low-risk and high-risk groups by X-tile software in TCGA cohort. Then, univariate and multivariate Cox regression analysis were performed to explore whether the MRS could be an independent predictor of OS in BC patients from TCGA and GSE72251 cohorts. The time-dependent ROC curve was generated to evaluate the predictive value of methylation risk score (MRS) for overall survival (OS) by the R package "timeROC". The survival analysis was generated by R package "survival" and "survminer".

Method 3

Cell lines and cell culture

Human BC cell lines MDA-MB-231 and T47D were purchased from Procell (Wuhan, China) and cultured in DMEM and RPMI-1640 media (Procell), respectively, supplemented with 10% fetal bovine serum (FBS) (Procell) and 10 U/mL penicillin-streptomycin (Beyotime, China). Cells were incubated at 37°C with 5% CO_2 and passaged when they reached 80% confluence.

RNA preparation

Total RNA was extracted from MDA-MB-231 and T47D cells using the E.Z.N.A.® Total RNA Kit I (Omega Bio-Tek, USA). The concentration and quality of the isolated RNA were measured using a NanoDrop™ One/OneC spectrophotometer (Thermo Scientific). Complementary DNA (cDNA) was synthesized by reverse transcription 1 μg of RNA with the BeyoRT™ III First Strand cDNA Synthesis Kit (Beyotime).

Real time quantitative PCR

cDNA amplification was performed using THUNDERBIRD Next SYBR qPCR Mix (Toyobo, Japan) on a LightCycler 480 PCR system (Roche, Switzerland). The following primer pairs were used:

FAM126A: sense 5′-GTTGTGGAGGAATGGTTGTCA-3′, antisense 3′-GACAGGTTCTAGCAACTCACTTT-5′

GAPDH: sense 5′-GGAGCCGAGATCCCCTCCAAAAT-3′, antisense 5′-GGCTGTTGTCATACTTCTCTATGG-3′

GAPDH was used as the internal control, and relative expression was calculated using the $2^{-\Delta\Delta Ct}$ method.

Transient cell transfection

To overexpress *FAM126A*, MDA-MB-231 and T47D cells at 70% to 90% confluence were transiently transfected with a *FAM126A* plasmid (Miaoling Bioscience, China), while an empty vector was used as a

negative control. Transfection was performed with Lipo8000 transfection reagent (Beyotime, China) following the manufacturer's instructions, with a ratio of 3 μL of reagent per 2.5μg of plasmid DNA. Cells were harvested 24 hours after transfection for subsequent functional assays.

Cell viability assay

Cells were seeded at a density of 2,000 to 3,000 cells per well in 96-well plates, depending on the growth characteristics of the cell lines, and incubated for 0, 24, 48, 72, and 96 hours. Cell viability was assessed using the Cell Counting Kit-8 (Abbkine, USA), and absorbance was measured at 450 nm using a microplate reader.

EdU incorporation assay

Cell proliferation was evaluated using the EdU Imaging Kit (Cy3) (APExBio, USA). Cells were seeded into confocal plates at a density of 4,000 cells per well and incubated with 10 μM EdU solution at 37°C for 2-4 hours. After incubation, cells were fixed with Immunol Staining Fix Solution (Beyotime) and permeabilized with Immunostaining Permeabilization Solution with Saponin (Beyotime). EdU labeling was followed by nuclear staining with Hoechst, and the results were visualized under a fluorescence microscope.

Transwell Migration and Invasion Assays

Cell migration assays were performed using a Transwell apparatus (Corning, USA). 1×10^4 MDA-MB-231 cells were plated in DMEM without FBS, and 5×10^4 T47D cells in RPMI-1640 without FBS, both on fibronectin-coated polycarbonate membrane inserts. The lower chamber was filled with medium containing 20% FBS. After 24 to 48 hours of incubation, depending on the cell line's migration capacity, non-migrated cells on the upper surface of the insert were removed with a cotton swab. Cells that had migrated to the lower surface were fixed with methanol, stained with Giemsa, and counted under a microscope. For invasion assays, the membrane was pre-coated with Matrigel (BD Biosciences, USA) at 200 μg/mL. All experiments were independently repeated at least three times.

Method 4

Statistical analysis

Development of BC diagnostic models

To assess the diagnostic efficacy of methylation markers obtained from cfDNA mddPCR assays, we constructed diagnostic models using multiple logistic regression analysis. Predicted risk scores for each individual were computed based on methylation levels and estimated regression coefficients of markers (Table S4). The formula of the risk score was as follows: Risk score = $\sum_{i=1}^{n}(\boldsymbol{x_i} * \boldsymbol{Coef_i} + \boldsymbol{c})$

Here, n represents the number of methylation markers; $\boldsymbol{x_i}$ represents the methylation level of each marker; $\boldsymbol{Coef_i}$ represents the coefficients; $\boldsymbol{c}$ represents the intercept; Risk score represents a weighted sum of the methylation level of each marker.

Subsequently, we also constructed diagnostic models with three common machine learning algorithms, including Random Forest (RF), Support Vector Machine (SVM), and Elastic Net (EN). For these three algorithms, a supervised learning approach was used, then 355 plasma cfDNA were randomly divided into

a training set and a test set with a ratio of 7:3. The training set (141 BC, 59 healthy controls, and 50 benign tumors) was used to construct a diagnostic model, and the test set (60 BC, 24 healthy controls, and 21 benign tumors) was employed for validation. The algorithms were implemented using R version 4.3.1 with "randomForest", "e1071", and "glmnet" packages.

References

1 Stefansson, O. A., Moran, S., Gomez, A., Sayols, S., Arribas-Jorba, C., Sandoval, J. *et al.* A DNA methylation-based definition of biologically distinct breast cancer subtypes. *Mol Oncol* **9**, 555-568 (2015).

2 Gao, Y., Jones, A., Fasching, P. A., Ruebner, M., Beckmann, M. W., Widschwendter, M. *et al.* The integrative epigenomic-transcriptomic landscape of ER positive breast cancer. *Clin Epigenetics* **7**, 126 (2015).

3 Fleischer, T., Frigessi, A., Johnson, K. C., Edvardsen, H., Touleimat, N., Klajic, J. *et al.* Genome-wide DNA methylation profiles in progression to in situ and invasive carcinoma of the breast with impact on gene transcription and prognosis. *Genome Biol* **15**, 435 (2014).

4 Hannum, G., Guinney, J., Zhao, L., Zhang, L., Hughes, G., Sadda, S. *et al.* Genome-wide methylation profiles reveal quantitative views of human aging rates. *Mol Cell* **49**, 359-367 (2013).

5 Cerapio, J. P., Marchio, A., Cano, L., López, I., Fournié, J. J., Régnault, B. *et al.* Global DNA hypermethylation pattern and unique gene expression signature in liver cancer from patients with Indigenous American ancestry. *Oncotarget* **12**, 475-492 (2021).

6 Dietz, S., Lifshitz, A., Kazdal, D., Harms, A., Endris, V., Winter, H. *et al.* Global DNA methylation reflects spatial heterogeneity and molecular evolution of lung adenocarcinomas. *Int J Cancer* **144**, 1061-1072 (2019).

7 Li, M., Sun, X., Yao, H., Chen, W., Zhang, F., Gao, S. *et al.* Genomic methylation variations predict the susceptibility of six chemotherapy related adverse effects and cancer development for Chinese colorectal cancer patients. *Toxicol Appl Pharmacol* **427**, 115657 (2021).

8 Ishak, M., Baharudin, R., Rose, I. M., Sagap, I., Mazlan, L., Azman, Z. A. M. *et al.* Genome-Wide Open Chromatin Methylome Profiles in Colorectal Cancer. *Biomolecules* **10** (2020).

9 Su, X., Lu, X., Bazai, S. K., Compérat, E., Mouawad, R., Yao, H. *et al.* Comprehensive integrative profiling of upper tract urothelial carcinomas. *Genome Biol* **22**, 7 (2021).

10 Sanegre, S., Eritja, N., de Andrea, C., Diaz-Martin, J., Diaz-Lagares, Á., Jácome, M. A. *et al.* Characterizing the Invasive Tumor Front of Aggressive Uterine Adenocarcinoma and Leiomyosarcoma. *Front Cell Dev Biol* **9**, 670185 (2021).

11 Ravi, N., Yang, M., Mylona, N., Wennerberg, J. & Paulsson, K. Global RNA Expression and DNA Methylation Patterns in Primary Anaplastic Thyroid Cancer. *Cancers (Basel)* **12** (2020).

12 Sehl, M. E., Carroll, J. E., Horvath, S. & Bower, J. E. The acute effects of adjuvant radiation and chemotherapy on peripheral blood epigenetic age in early stage breast cancer patients. *NPJ Breast Cancer* **6**, 23 (2020).

13 Zhang, L., Li, D., Gao, L., Zhang, D., Fu, Q., Sun, H. *et al.* Pan-cancer analysis of genome-wide methylation profiling discover type-specific markers targeting circulating free DNA for the detection of colorectal cancer. *Clin Transl Med* **13**, e1370 (2023).

299 **Tables**

300 **Table S1. In-house and public datasets used in this study.**

Dataset source	Sample type	Cancer type annotation	Sample type	Platform	Number of samples
Discovery cohort (N = 24)					
In-house study	BC	Breast Cancer	Tissue	850K array	Tumor = 14, Tumor-adjacent = 10
Marker selection cohort (N = 5,748)					
TCGA	BC	Breast Cancer	Tissue	450K array	Tumor = 774, Tumor-adjacent = 97
TCGA	ACC	Adrenocortical Cancer	Tissue	450K array	Tumor = 49
TCGA	BLCA	Bladder Urothelial Carcinoma	Tissue	450K array	Tumor = 108, Tumor-adjacent = 10
TCGA	CESC	Cervical & Endocervical Cancer	Tissue	450K array	Tumor = 307, Tumor-adjacent = 3
TCGA	CHOL	Bile Duct Cancer	Tissue	450K array	Tumor = 20, Tumor-adjacent = 3
TCGA	CRC	Colorectal Cancer	Tissue	450K array	Tumor = 180, Tumor-adjacent = 21
TCGA	DLBC	Diffuse Large B-Cell Lymphoma	Tissue	450K array	Tumor = 26
TCGA	ESCA	Esophageal Carcinoma	Tissue	450K array	Tumor = 27, Tumor-adjacent = 6
TCGA	GBM	Glioblastoma Multiforme	Tissue	450K array	Tumor = 58
TCGA	HNSC	Head & Neck Squamous Cell Carcinoma	Tissue	450K array	Tumor = 142, Tumor-adjacent = 12
TCGA	KICH	Kidney Chromophobe	Tissue	450K array	Tumor = 27, Tumor-adjacent = 6
TCGA	KIRC	Kidney Clear Cell Carcinoma	Tissue	450K array	Tumor = 114, Tumor-adjacent = 54
TCGA	KIRP	Kidney Papillary Cell Carcinoma	Tissue	450K array	Tumor = 73, Tumor-adjacent = 14
TCGA	LAML	Acute Myeloid Leukemia	Blood	450K array	Tumor = 89
TCGA	LGG	Brain Lower Grade Glioma	Tissue	450K array	Tumor = 230
TCGA	LIHC	Liver Hepatocellular Carcinoma	Tissue	450K array	Tumor = 122, Tumor-adjacent = 20
TCGA	LUAD	Lung Adenocarcinoma	Tissue	450K array	Tumor = 244, Tumor-adjacent = 15
TCGA	LUSC	Lung Squamous Cell Carcinoma	Tissue	450K array	Tumor = 96, Tumor-adjacent = 13
TCGA	MESO	Mesothelioma	Tissue	450K array	Tumor = 16
TCGA	OV	Ovarian Cancer	Tissue	450K array	Tumor = 10
TCGA	PAAD	Pancreatic Adenocarcinoma	Tissue	450K array	Tumor = 82, Tumor-adjacent = 4
TCGA	PCPG	Pheochromocytoma & Paraganglioma	Tissue	450K array	Tumor = 101, Tumor-adjacent = 1
TCGA	SARC	Sarcoma	Tissue	450K array	Tumor = 142, Tumor-adjacent = 3
TCGA	SKCM	Skin Cutaneous Melanoma	Tissue	450K array	Tumor = 43
TCGA	STAD	Stomach Adenocarcinoma	Tissue	450K array	Tumor = 136, Tumor-adjacent = 2
TCGA	THCA	Thyroid Carcinoma	Tissue	450K array	Tumor = 371, Tumor-adjacent = 42
TCGA	THYM	Thymoma	Tissue	450K array	Tumor = 60, Tumor-adjacent = 2
TCGA	UCEC	Uterine Corpus Endometrioid	Tissue	450K array	Tumor = 431, Tumor-adjacent = 34

Dataset source	Sample type	Cancer type annotation	Sample type	Platform	Number of samples
		Carcinoma			
TCGA	UCS	Uterine Carcinosarcoma	Tissue	450K array	Tumor = 57
TCGA	UVM	Uveal Melanoma	Tissue	450K array	Tumor = 35
GSE40279	HC	Healthy Control	White blood cells	450K array	Normal = 338
GSE51032	BC & HC	Breast Cancer $ Healthy Control	White blood cells	450K array	Tumor = 233, Normal = 340
GSE159898	CRC	Colorectal Cancer	Tissue	850K array	Tumor = 9, Tumor-adjacent = 9
GSE149282	CRC	Colorectal Cancer	Tissue	850K array	Tumor = 6, Tumor-adjacent = 6
GSE114989	LUAD	Lung Adenocarcinoma	Tissue	850K array	Tumor = 7, Tumor-adjacent = 2
GSE136380	LIHC	Liver Hepatocellular Carcinoma	Tissue	850K array	Tumor = 7, Tumor-adjacent = 7
GSE136583	LIHC	Liver Hepatocellular Carcinoma	Tissue	850K array	Tumor = 11, Tumor-adjacent = 11
GSE192967	OV	Ovarian Cancer	Tissue	850K array	Tumor = 32
GSE161651	UTUC	Upper-Tract Urothelial Carcinoma	Tissue	850K array	Tumor = 5
GSE200484	STAD	Stomach Adenocarcinoma	Tissue	850K array	Tumor = 2, Tumor-adjacent = 2
GSE171142	UADC/LMS	Uterine Adenocarcinomas and Leiomyosarcomas	Tissue	850K array	Tumor = 30
GSE1460038	THCA	Thyroid Carcinoma	Tissue	850K array	Tumor = 6, Tumor-adjacent = 3
GSE140038	BC	Breast Cancer	White blood cells	850K array	Tumor = 144
In-house study	HC	Healthy control	White blood cells	850K array	Normal = 12
Tissue validation cohort (N = 1018)					
GSE52865	BC	Breast Cancer	Tissue	450K array	Tumor = 40, Tumor-adjacent = 17
GSE66695	BC	Breast Cancer	Tissue	450K array	Tumor = 80, Tumor-adjacent = 40
GSE69914	BC	Breast Cancer	Tissue	450K array	Tumor = 305, Tumor-adjacent = 92
GSE60185	BC	Breast Cancer	Tissue	450K array	Tumor = 239, Tumor-adjacent = 46
GSE72251	BC	Breast Cancer	Tissue	450K array	Tumor = 118*
In-house study	BC	Breast Cancer	Tissue	qMSP	Tumor = 22, Tumor-adjacent = 18
Plasma cfDNA validation cohort (N = 375)					
In-house study	BC	Breast Cancer	cfDNA	Droplet digital PCR	Tumor = 201
In-house study	BE	Breast Benign tumor	cfDNA	Droplet digital PCR	Benign = 71
In-house study	HC	Healthy control	cfDNA	Droplet digital PCR	Normal = 83
In-house study	CRC	Colorectal Cancer	cfDNA	Droplet digital PCR	Tumor = 20

Note: All samples were from female participants; cfDNA, Cell-free DNA; * presents the cohort validation for prognostic effect of markers, one patient with missing survival status has been removed from the original cohort.

Table S2. Primer and probe sequences of BC-specific methylation markers and reference gene.

Gene/Region	CpG sites	Primer/Probe	Primer Sequence (5'-3')	Tm ℃	GC%	Length (bp)	Product size (bp)
C1orf230	cg06100368	Forward primer	CGTTTAGCGTAGCGTAGTTGGA	58.3	50	22	104
	cg22790257	MGB probe	A**CG**CTAAATAAAAACG**CG**	69	39	18	
		Reverse primer	AAACCCCAATCCTAAAACAACTACA	58.1	36	25	
C2orf88	cg22003435	Forward primer	GCGCGTAGTTGTCGGGTAAC	59.2	60	20	118
		MGB probe	C**CG**ACAAACACCCGC	68	67	15	
		Reverse primer	CCATTAAAAAATCTAAAAAACACAACTAACA	58.6	23	31	
CCND2	cg12382902	Forward primer	TTTT**CG**CGATTTGTTTTTTGC	58.9	33	21	55
	cg16994506	MGB probe	CCG**CG**ACGACTTC	70	69	13	
		Reverse primer	TCCTAACCGCAACGCAAAC	58.4	53	19	
ETS1	cg18565473	Forward primer	TTGGTATTTTGCGGCGAAG	58.4	47	19	54
		MGB probe	CCCGAACG**CG**CTC	69	77	13	
		Reverse primer	GCGACGCGAACGAAACTAA	58.7	53	19	
ITPRIPL1	cg06341177	Forward primer	TCGCGTTGTTTTTTTAAGATCGT	58.9	35	23	115
		MGB probe	CCTCGAAATATCTCC**CG**	69	53	17	
		Reverse primer	CGCCTATCCGTAAAAAACCAAA	59	41	22	
ABO	cg11698867	Forward primer	ATTTTTGTTAGGAAGGGACGGC	59.4	45	22	67
		MGB probe	AATA**CG**CGCGAAAAA	70	40	15	
		Reverse primer	CACGAAAAACGCGCGAAT	59.4	50	18	
EBF1	cg20617821	Forward primer	GGGAGTCGGTTAAAGTTTTTCGT	58.6	43	23	68
		MGB probe	A**CG**CTCCTTCGCTCG	70	67	15	
		Reverse primer	CCCAACCCAAACGTCAATACTC	59.2	50	22	
NFYC	cg01573554	Forward primer	TTAAAATAGGAAGTGTATAAAAGAGGTAATTG	56.9	25	32	98
	cg23934465	MGB probe	CCA**CG**ATTAACTATTCA**CG**	69	42	19	

Gene/Region	CpG sites	Primer/Probe	Primer Sequence (5'-3')	Tm ℃	GC%	Length (bp)	Product size (bp)
ACTB		Reverse primer	TAATCCCATTAACTAAACTACTTTATCGAC	56.4	30	30	80
		Forward primer	GGTTAAGTGTGATTTTGTGGTGTG	57.2	42	24	
		MGB probe	ACCCTCTACTACCC	65	57	14	
		Reverse primer	CCTTTTACAAAATTCACCCTCCT	56.6	39	23	

Note: These are the sequences after the bisulfite conversion; The CGs in bold were the sites detected in the 850K array; *ACTB* is the reference gene.

Table S3. The assay mix preparation for one reaction.

	Multiplex ddPCR assay 1 (μL)		Multiplex ddPCR assay 2 (μL)			Multiplex ddPCR assay 3 (μL)		
Target region	C1orf230	C2orf88	CCND2	ETS1	ITPRIPL1	ABO	EBF1	NFYC
Forward (20 μM)	0.4	0.4	0.4	0.4	0.4	0.4	0.4	0.4
Reverse (20 μM)	0.4	0.4	0.4	0.4	0.4	0.4	0.4	0.4
Taqman probe 6FAM (10 μM)	0.9	0.9	0.9	0.45	0.75	0.8	0.8	0.15
Taqman probe VIC (10 μM)		0.9			0.75		0.8	
ACTB (reference gene)								
Forward (20 μM)	0.4			0.4			0.4	
Reverse (20 μM)	0.4			0.4			0.4	
Taqman probe VIC (10 μM)	0.3			0.3			0.4	
2× ddPCR Supermix for Probes (no dUTP)	10			10			10	
DNA template (sample)	variable			variable			variable	
Volume input for droplet generation for ddPCR	21			21			21	

Note: The working concentrations of primers and probes are 20 μM and 10 μM, and the probe solutions should be protected from light.

Table S4. The corresponding coefficients in each diagnostic model.

	BC and healthy control	BC and benign tumor	BC and non-cancer	BC and colorectal cancer	Combined
Intercept	-0.364	0.003	-0.883	2.485	-2.283
C1orf230	0.204	1.033	1.147	-0.141	0.062
C2orf88	2.490	0.465	0.858	0.970	0.317
CCND2	13.621	1.206	1.775	0.196	1.714
ETS1	0.059	-0.044	0.010	0.346	0.022
ITPRIPL1	15.744	0.887	1.364	0.414	1.140
ABO	1.454	0.609	0.950	0.098	1.026
EBF1	0.872	0.436	0.632	0.020	0.682
NYFC	0.008	0.021	0.014	-0.043	0.031
Mammography	-	-	-	-	1.926
Utrasound	-	-	-	-	1.386

Note: -, not available.

Table S5. The methylation differences of 21 BC-specific CpG sites in in-house 850K cohort and TCGA cohort.

Cpg sites	Gene	850K				TCGA			
		BC	Normal	Δ*β*	*p*-value	BC	Normal	Δ*β*	*p*-value
cg06100368	*C1orf230*	0.214	0.025	0.189	< 0.001	0.359	0.058	0.301	< 0.001
cg22790257	*C1orf230*	0.235	0.041	0.194	< 0.001	0.327	0.041	0.286	< 0.001
cg22003435	*C2orf88*	0.150	0.027	0.123	< 0.001	0.261	0.054	0.207	< 0.001
cg12382902	*CCND2*	0.252	0.027	0.225	< 0.001	0.278	0.029	0.250	< 0.001
cg16994506	*CCND2*	0.189	0.056	0.134	< 0.001	0.273	0.034	0.239	< 0.001
cg25454116	*CCND2*	0.224	0.038	0.186	< 0.001	0.233	0.021	0.212	< 0.001
cg18565473	*ETS1*	0.184	0.032	0.152	< 0.001	0.289	0.041	0.249	< 0.001
cg13002945	*FAM126A*	0.228	0.063	0.164	< 0.001	0.248	0.046	0.202	< 0.001
cg01934296	*FAM189A1*	0.815	0.940	-0.125	< 0.001	0.769	0.974	-0.204	< 0.001
cg01147665	*GPD1L*	0.770	0.958	-0.189	< 0.001	0.725	0.925	-0.200	< 0.001
cg06341177	*ITPRIPL1*	0.151	0.025	0.127	< 0.001	0.262	0.057	0.205	< 0.001
cg16508600	*MDM4*	0.802	0.937	-0.135	< 0.001	0.642	0.934	-0.292	< 0.001
cg00538599	*OGDHL*	0.577	0.823	-0.245	< 0.001	0.741	0.967	-0.226	< 0.001
cg13212963	*PACRG*	0.813	0.967	-0.154	< 0.001	0.717	0.954	-0.237	< 0.001
cg11519710	*SKI*	0.816	0.930	-0.114	< 0.001	0.663	0.910	-0.246	< 0.001
cg11698867	*ABO*	0.263	0.048	0.215	< 0.001	-	-	-	-
cg20617821	*EBF1*	0.293	0.044	0.248	< 0.001	-	-	-	-
cg09002832	*GLIS3*	0.265	0.039	0.226	< 0.001	-	-	-	-
cg01573554	*NFYC*	0.261	0.027	0.235	0.001	-	-	-	-
cg23934465	*NFYC*	0.255	0.035	0.221	0.001	-	-	-	-
cg25299666	*TTC28*	0.269	0.017	0.252	0.003	-	-	-	-

Note: $\Delta\beta$ is the average β values of BC group minus the average β values of normal group; The p-value has been adjusted by Benjamini-Hochberg method.

Table S6. The genomic characteristics of 21 BC-specific CpG sites.

CpG sites	Gene symbol	Gene description	Genomic coordinate	Strand	Relation to Island	UCSC refGene group
cg06100368	*C1orf230*	Chromosome 1 Open Reading Frame 230	chr1:151694123	+	Island	Body
cg22790257	*C1orf230*		chr1:151694108	-	Island	Body
cg22003435	*C2orf88*	Chromosome 2 Open Reading Frame 88	chr2:191045885	+	S_Shore	5'UTR
cg12382902	*CCND2*	Cyclin D2	chr12:4383699	+	Island	Body
cg16994506	*CCND2*		chr12:4383724	+	Island	Body
cg25454116	*CCND2*		chr12:4381471	+	Island	TSS1500
cg18565473	*ETS1*	ETS Proto-Oncogene 1, Transcription Factor	chr11:128392102	-	Island	Body
cg13002945	*FAM126A*	Family With Sequence Similarity 126, Member A	chr7:23053827	+	Island	TSS200
cg06341177	*ITPRIPL1*	Inositol 1,4,5-Triphosphate Receptor Interacting Protein-Like 1	chr2:96991327	-	S_Shore	TSS1500
cg01934296	*FAM189A1*	Family With Sequence Similarity 189 Member A1	chr15:29429225	-	OpenSea	Body
cg01147665	*GPD1L*	Glycerol-3-Phosphate Dehydrogenase 1 Like	chr3:32146075	-	N_Shore	TSS1500
cg16508600	*MDM4*	MDM4 Regulator of P53	chr1:204531383	-	OpenSea	3'UTR
cg00538599	*OGDHL*	Oxoglutarate Dehydrogenase L	chr10:50957749	+	OpenSea	Body
cg13212963	*PACRG*	Parkin Coregulated	chr6:163321435	-	OpenSea	Body
cg11519710	*SKI*	Sloan-Kettering Institute Proto-Oncogene	chr1:2239646	-	S_Shore	3'UTR
cg11698867	*ABO*	ABO, Alpha 1-3-N-Acetylgalactosaminyltransferase and Alpha 1-3-Galactosyltransferase	chr9:136150802	+	Island	TSS200
cg20617821	*EBF1*	EBF Transcription Factor 1	chr5:158527560	-	Island	TSS1500
cg09002832	*GLIS3*	GLIS Family Zinc Finger 3	chr9:4300181	-	Island	TSS200
cg01573554	*NFYC*	Nuclear Transcription Factor Y	chr1:41174907	+	OpenSea	TSS200
cg23934465	*NFYC*	Subunit Gamma	chr1:41174893	+	OpenSea	TSS200
cg25299666	*TTC28*	Tetratricopeptide Repeat Domain 28	chr22:28839093	+	OpenSea	Body

Table S7. Basic information of breast cancer tissues using for diagnostic validation in TCGA and GEO databases.

Characteristics	TCGA	GSE52865	GSE66695	GSE60185	GSE69914
Number of cases					
Tumor	774	40	80	239	305
Normal	97	17	40	46	92
Age (median, IQR), years	58.0 (48.0-67.0)	-	-	59.2 (50.0-71.0)	-
ER (%)					
Negative	169 (21.8%)	-	40 (50.0%)	-	48 (15.7%)
Positive	561 (72.5%)	-	40 (50.0%)	-	254 (83.3%)
Unknow	44 (5.7%)	40 (100.0%)	0 (0.0%)	239 (100.0%)	3 (1.0%)
PR (%)					
Negative	237 (30.6%)	-	-	-	81 (26.6%)
Positive	490 (63.3%)	-	-	-	217 (71.1%)
Unknow	47 (6.1%)	40 (100.0%)	80 (100.0%)	239 (100.0%)	7 (2.3%)
HER2					
Negative	392 (50.7%)	-	-	-	224 (73.4%)
Positive	90 (11.6%)	-	-	-	43 (14.1%)
Unknow	292 (37.7%)	40 (100.0%)	80 (100.0%)	239 (100.0%)	38 (12.5%)
Subtype					
Luminal	576 (74.4%)	19 (47.5%)	-	-	259 (84.9%)
TNBC/Basal_like	83 (10.7%)	3 (7.5%)	-	-	30 (9.8%)
HER2+	17 (2.2%)	14 (35.0%)	-	-	10 (3.3%)
Unknow	98 (12.7%)	4 (10.0%)	80 (100.0%)	239 (100.0%)	6 (2.0%)
Stage					
I	127 (16.4%)	-	-	-	-
II	431 (55.7%)	-	-	-	-
III	197 (25.4%)	-	-	-	-
IV	10 (1.3%)	-	-	-	-
Unknow	9 (1.2%)	40（100.0%）	80 (100.0%)	239 (100.0%)	305 (100.0%)

Note: -, not available; ER, estrogen receptor; PR, progesterone receptor; HER, human epidermal growth factor receptor; TNBC, triple-negative breast cancer.

Table S8. The methylation differences of candidate CpG sites between BC and tumor-adjacent tissues in GEO.

Cpg sites	Gene	GSE52865				GSE60185				GSE66695				GSE69914			
		BC	Normal	Δβ	*p*-value	BC	Normal	Δβ	*p*-value	BC	Normal	Δβ	*p*-value	BC	Normal	Δβ	*p*-value
cg06100368	*C1orf230*	0.201	0.066	0.136	< 0.001	0.174	0.025	0.149	< 0.001	0.243	0.056	0.187	< 0.001	0.269	0.046	0.223	< 0.001
cg22790257	*C1orf230*	0.176	0.029	0.147	< 0.001	0.255	0.044	0.211	< 0.001	0.172	0.023	0.149	< 0.001	0.239	0.048	0.191	< 0.001
cg22003435	*C2orf88*	0.179	0.060	0.119	< 0.001	0.171	0.041	0.130	< 0.001	0.163	0.058	0.105	< 0.001	0.194	0.040	0.155	< 0.001
cg12382902	*CCND2*	0.201	0.036	0.165	< 0.001	0.235	0.036	0.198	< 0.001	0.254	0.043	0.211	< 0.001	0.208	0.050	0.157	< 0.001
cg16994506	*CCND2*	0.232	0.063	0.170	< 0.001	0.245	0.055	0.190	< 0.001	0.260	0.068	0.192	< 0.001	0.202	0.056	0.146	< 0.001
cg25454116	*CCND2*	0.202	0.027	0.174	< 0.001	0.171	0.031	0.140	< 0.001	0.207	0.036	0.171	< 0.001	0.166	0.041	0.125	< 0.001
cg18565473	*ETS1*	0.189	0.037	0.153	< 0.001	0.232	0.065	0.168	< 0.001	0.172	0.031	0.142	< 0.001	0.241	0.055	0.187	< 0.001
cg13002945	*FAM126A*	0.157	0.019	0.138	< 0.001	0.202	0.043	0.159	< 0.001	0.189	0.028	0.161	< 0.001	0.242	0.071	0.172	< 0.001
cg01934296	*FAM189A1*	0.803	0.952	-0.149	< 0.001	0.815	0.956	-0.141	< 0.001	0.781	0.916	-0.135	< 0.001	0.783	0.942	-0.159	< 0.001
cg01147665	*GPD1L*	0.786	0.890	-0.104	< 0.001	0.706	0.950	-0.244	< 0.001	0.755	0.895	-0.141	< 0.001	0.826	0.952	-0.126	< 0.001
cg06341177	*ITPRIPL1*	0.217	0.065	0.152	< 0.001	0.166	0.043	0.123	< 0.001	0.202	0.064	0.138	< 0.001	0.198	0.042	0.155	< 0.001
cg16508600	*MDM4*	0.755	0.898	-0.143	< 0.001	0.734	0.967	-0.233	< 0.001	0.687	0.879	-0.192	< 0.001	0.787	0.954	-0.167	< 0.001
cg00538599	*OGDHL*	0.780	0.926	-0.146	< 0.001	0.790	0.942	-0.152	< 0.001	0.751	0.884	-0.134	< 0.001	0.781	0.924	-0.143	< 0.001
cg13212963	*PACRG*	0.766	0.925	-0.159	< 0.001	0.808	0.978	-0.170	< 0.001	0.755	0.905	-0.150	< 0.001	0.851	0.972	-0.120	< 0.001
cg11519710	*SKI*	0.789	0.857	-0.068	0.001	0.734	0.932	-0.197	< 0.001	0.723	0.881	-0.158	< 0.001	0.803	0.934	-0.131	< 0.001

Note: Δβ is the average β values of BC group minus the average β values of normal group; The *p*-value has been adjusted by Benjamini-Hochberg method.

Table S9. Diagnostic performance analysis of candidate CpG sites for distinguishing breast cancer from tumor-adjacent tissues in TCGA dataset.

CpG site	Gene	AUC	Cut-off	TCGA			
				TP	Sen	TN	Spe
cg06100368	*C1orf230*	0.863	0.131	578	0.747	91	0.938
cg22790257	*C1orf230*	0.807	0.134	564	0.729	92	0.948
cg22003435	*C2orf88*	0.820	0.046	608	0.787	74	0.763
cg12382902	*CCND2*	0.842	0.066	522	0.677	90	0.928
cg16994506	*CCND2*	0.837	0.057	536	0.695	89	0.918
cg25454116	*CCND2*	0.846	0.034	497	0.643	92	0.948
cg18565473	*ETS1*	0.783	0.156	468	0.606	94	0.969
cg13002945	*FAM126A*	0.728	0.073	367	0.474	93	0.959
cg01934296	*FAM189A1*	0.839	0.963	483	0.629	88	0.907
cg06341177	*ITPRIPL1*	0.885	0.121	544	0.704	91	0.938
cg01147665	*GPD1L*	0.821	0.887	538	0.695	91	0.938
cg16508600	*MDM4*	0.852	0.850	540	0.700	96	0.990
cg00538599	*OGDHL*	0.859	0.937	533	0.689	95	0.979
cg13212963	*PACRG*	0.915	0.945	645	0.836	87	0.897
cg11519710	*SKI*	0.879	0.842	579	0.748	93	0.959

Note: AUC, Area under the receiver operating characteristic curve; TP, Number of true positive; Sen, Sensitivity; TN, Number of true negative; Spe, Specificity. Sensitivity and specificity were determined based on cut_off values.

Table S10. Diagnostic performance analyses of 15 CpG sites for distinguishing breast cancer from tumor-adjacent tissues in four datasets from GEO.

CpG site	Gene	Cut-off	GSE52865				GSE60185				GSE66695				GSE69914			
			TP	Sen	TN	Spe	TP	Sen	TN	Spe	TP	Sen	TN	Spe	TP	Sen	TN	Spe
cg06100368	*C1orf230*	0.131	21	0.525	17	1.000	120	0.502	46	1.000	46	0.613	36	1.000	192	0.630	86	0.935
cg22790257	*C1orf230*	0.134	20	0.500	16	0.941	177	0.741	43	0.935	35	0.467	36	1.000	198	0.649	86	0.935
cg22003435	*C2orf88*	0.046	39	0.975	4	0.235	206	0.862	35	0.761	72	0.960	5	0.139	242	0.793	71	0.772
cg12382902	*CCND2*	0.066	26	0.650	16	0.941	172	0.720	43	0.935	55	0.733	32	0.889	196	0.643	84	0.913
cg16994506	*CCND2*	0.057	31	0.775	6	0.353	202	0.845	31	0.674	64	0.853	12	0.333	226	0.741	75	0.815
cg25454116	*CCND2*	0.034	28	0.700	14	0.824	189	0.791	32	0.696	65	0.867	19	0.528	225	0.738	64	0.696
cg18565473	*ETS1*	0.156	18	0.450	17	1.000	140	0.586	43	0.935	28	0.373	36	1.000	175	0.574	86	0.935
cg13002945	*FAM126A*	0.073	13	0.325	17	1.000	127	0.531	43	0.935	28	0.373	36	1.000	224	0.734	67	0.728
cg01934296	*FAM189A1*	0.963	38	0.950	4	0.235	213	0.891	17	0.370	67	0.893	4	0.111	299	0.980	11	0.120
cg06341177	*ITPRIPL1*	0.121	29	0.725	17	1.000	133	0.556	46	1.000	48	0.640	36	1.000	158	0.518	90	0.978
cg01147665	*GPD1L*	0.887	23	0.575	10	0.842	164	0.686	41	0.891	49	0.653	26	0.722	165	0.541	89	0.967
cg16508600	*MDM4*	0.85	23	0.575	17	1.000	150	0.628	46	1.000	43	0.573	31	0.861	157	0.515	90	0.978
cg00538599	*OGDHL*	0.937	36	0.900	5	0.294	193	0.808	35	0.761	68	0.907	8	0.222	281	0.921	24	0.261
cg13212963	*PACRG*	0.945	39	0.975	2	0.118	153	0.640	44	0.957	0	0.000	36	1.000	154	0.505	88	0.957
cg11519710	*SKI*	0.842	25	0.625	12	0.706	154	0.644	45	0.978	51	0.680	32	0.889	158	0.518	87	0.946

Note: TP, Number of true positive; Sen, Sensitivity; TN, Number of true negative; Spe, Specificity; Sensitivity and specificity were determined based on cut_off values.

Table S11. Basic information of breast cancer tissues using for prognostic validation in TCGA and GEO databases.

Characteristics	TCGA	GSE72251
Number of cases		
Tumor	754	118
Age (median, IQR, years)	58.0 (48.0-67.0)	58.5 (47.7-67.9)
Subtype		
Luminal	567 (75.2%)	57 (48.3%)
TNBC/Basal_like	79 (10.5%)	25 (21.2%)
HER2+	17 (2.2%)	34 (28.8%)
Unknow	91 (12.1%)	2 (1.7%)
Stage		
I	125 (16.6%)	-
II	422 (56.0%)	-
III	189 (25.0%)	-
IV	9 (1.2%)	-
Unknow	9 (1.2%)	118 (100.0%)
Vital status		
Alive	653 (86.6%)	106 (89.8%)
Dead	101 (13.4%)	12 (10.2%)

Note: -, not available; ER, estrogen receptor; PR, progesterone receptor; HER, human epidermal growth factor receptor; TNBC, triple-negative breast cancer.

Table S12. Univariate and multivariate Cox analyses for overall survival of CpG sites in the TCGA cohort.

CpG sites	Risk categories	Univariate analyses			Multivariate analyses		
		coefficient	HR (95% CI)	*p* value	coefficient	HR (95% CI)	*p* value
cg16994506	High (≥ 0.59) vs. Low (< 0.59)	0.903	2.470 (1.470-4.130)	0.001	0.990	2.692 (1.591-4.555)	**0.000**
cg16508600	High (≥ 0.42) vs. Low (< 0.42)	-0.702	0.495 (0.325-0.756)	0.001	-0.612	0.542 (0.348-0.844)	**0.007**
cg00538599	High (≥ 0.97) vs. Low (< 0.97)	0.668	1.950 (1.170-3.260)	0.011	0.454	1.574 (0.936-2.648)	0.087
cg22003435	High (≥ 0.55) vs. Low (< 0.55)	0.550	1.730 (1.060-2.830)	0.028	0.531	1.700 (1.024-2.823)	**0.040**
cg25454116	High (≥ 0.50) vs. Low (< 0.50)	0.491	1.630 (1.050-2.540)	0.030	0.384	1.468 (0.928-2.322)	0.101
cg13002945	High (≥ 0.04) vs. Low (< 0.04)	-0.489	0.614 (0.390-0.964)	0.034	-0.474	0.623 (0.391-0.990)	**0.045**
cg11519710	High (≥ 0.40) vs. Low (< 0.40)	-0.559	0.572 (0.334-0.980)	0.042	-0.697	0.498 (0.287-0.865)	**0.013**
cg01147665	High (≥ 0.55) vs. Low (< 0.55)	-0.470	0.625 (0.397-0.984)	0.042	-0.572	0.564 (0.353-0.901)	**0.017**
cg12382902	High (≥ 0.56) vs. Low (< 0.56)	0.541	1.720 (1.020-2.900)	0.043	0.532	1.702 (0.995-2.912)	0.052
cg06100368	High (≥ 0.63) vs. Low (< 0.63)	0.430	1.540 (0.932-2.540)	0.092	0.474	1.607 (0.972-2.655)	0.064
cg06341177	High (≥ 0.17) vs. Low (< 0.17)	0.327	1.390 (0.914-2.100)	0.124	0.316	1.372 (0.900-2.090)	0.141
cg01934296	High (≥ 0.98) vs. Low (< 0.98)	0.395	1.480 (0.891-2.470)	0.130	0.495	1.640 (0.978-2.752)	0.061
cg22790257	High (≥ 0.20) vs. Low (< 0.20)	-0.282	0.754 (0.500-1.140)	0.178	-0.385	0.680 (0.446-1.037)	0.073
cg18565473	High (≥ 0.53) vs. Low (< 0.53)	0.312	1.370 (0.857-2.180)	0.190	0.265	1.304 (0.805-2.111)	0.281
cg13212963	High (≥ 0.95) vs. Low (< 0.95)	0.286	1.330 (0.807-2.190)	0.263	0.243	1.274 (0.752-2.160)	0.368

Notes:HR, hazard ratio; CI, confidence interval; All analyses were adjusted for age, subtype, and stage. Bold font represents *p*-value < 0.05. The methylation risk score (MRS) = cg16994506 x 0.990 + cg16508600 x (-0.612) + cg22003435 x 0.531 + cg13002945 x (-0.474) + cg11519710 x (-0.697) + cg01147665 x (-0.572).

Table S13. Univariate and multivariate Cox analyses of baseline characteristics and methylation risk score on overall survival in TCGA and GSE72251 cohorts.

Variates	TGCA				GSE72251			
	Univariate analyses		Multivariate analyses		Univariate analyses		Multivariate analyses	
	HR (95% CI)	***p* value**	**HR (95% for HR)**	***p* value**	**HR (95% for HR)**	***p* value**	**HR (95% CI)**	***p* value**
Age (Years)								
≤ 60	1.000		1.000		1.000		1.000	
> 60	2.179 (1.470-3.231)	0.000	2.172 (1.450-3.254)	0.000	1.108 (0.337-3.642)	0.865	1.062 (0.322-3.505)	0.921
Subtype								
non-TNBC	1.000		1.000		1.000		1.000	
TNBC	1.459 (0.813-2.620)	0.206	2.887 (1.537-5.423)	0.001	3.965 (1.187-13.24)	0.025	6.660 (1.6867-26.303)	0.007
Stage								
I/II	1.000		1.000					
III/IV	2.392 (1.595-3.587)	0.000	2.790 (1.847-4.215)	0.000				
Unknow	2.217 (0.867-5.669)	0.097	1.965 (0.766-5.042)	0.160				
MRS								
Low-risk			1.000				1.000	
High-risk	2.694 (1.809-4.011)	0.000	2.826 (1.841-4.338)	0.000	2.034 (0.59-7.015)	0.261	4.286 (1.044-17.592)	0.043

Notes: HR, hazard ratio; CI, confidence interval; All analyses were adjusted for age, subtype, and stage.

Table S14. Diagnostic effect of the single or combined markers for distinguishing BC from healthy control, benign tumor, non-cancer, and colorectal cancer groups.

Model	Cut_off	Sensitivity	Specificity	AUC (95% CI)	*p*-value
BC and Healthy control					
All markers	1.154	0.662	0.952	0.856 (0.814-0.898)	-
C1orf230	1.610	0.164	0.988	0.577 (0.549-0.605)	< 0.001
C2orf88	1.651	0.313	0.988	0.652 (0.619-0.686)	< 0.001
CCND2	7.886	0.124	1.000	0.562 (0.539-0.585)	< 0.001
ETS1	1.008	0.169	0.976	0.572 (0.541-0.603)	< 0.001
ITPRIPL1	7.689	0.124	1.000	0.562 (0.539-0.585)	< 0.001
ABO	1.606	0.224	0.988	0.607 (0.576-0.638)	< 0.001
EBF1	0.672	0.547	0.855	0.719 (0.672-0.767)	< 0.001
NYFC	0.713	0.736	0.530	0.626 (0.553-0.699)	< 0.001
BC and Benign tumor					
All markers	1.378	0.512	0.958	0.742 (0.684-0.801)	-
C1orf230	1.627	0.164	0.986	0.576 (0.547-0.604)	< 0.001
C2orf88	1.060	0.313	0.887	0.602(0.554-0.651)	< 0.001
CCND2	2.393	0.114	1.000	0.556 (0.5230-0.582)	< 0.001
ETS1	1.525	0.144	0.986	0.536 (0.493-0.579)	< 0.001
ITPRIPL1	2.182	0.109	1.000	0.550 (0.520-0.579)	< 0.001
ABO	1.960	0.189	1.000	0.578 (0.538-0.619)	< 0.001
EBF1	1.268	0.433	0.887	0.661 (0.601-0.720)	0.003
NYFC	1.179	0.338	0.915	0.641 (0.569-0.713)	0.009
BC and Non-cancer					
All markers	0.326	0.637	0.844	0.804 (0.760-0.848)	-
C1orf230	0.924	0.164	0.987	0.576 (0.549-0.603)	< 0.001
C2orf88	0.415	0.313	0.942	0.629 (0.592-0.666)	< 0.001
CCND2	1.034	0.124	0.994	0.559 (0.536-0.583)	< 0.001
ETS1	0.682	0.154	0.974	0.556 (0.523-0.588)	< 0.001
ITPRIPL1	0.810	0.124	0.987	0.556 (0.532-0.581)	< 0.001
ABO	1.216	0.224	0.974	0.594 (0.561-0.626)	< 0.001
EBF1	0.711	0.428	0.916	0.692 (0.646-0.739)	< 0.001
NYFC	0.091	0.731	0.519	0.633 (0.575-0.691)	< 0.001
BC and Colorectal cancer					
All markers	2.039	0.806	0.900	0.851 (0.761-0.940)	-
C1orf230	2.299	0.164	0.950	0.555 (0.496-0.614)	< 0.001
C2orf88	2.846	0.294	0.950	0.614 (0.546-0.681)	< 0.001
CCND2	2.883	0.104	1.000	0.539 (0.489-0.590)	< 0.001
ETS1	2.619	0.134	1.000	0.519 (0.443-0.595)	< 0.001
ITPRIPL1	3.024	0.100	1.000	0.539 (0.488-0.590)	< 0.001
ABO	2.530	0.194	0.950	0.563 (0.490-0.636)	< 0.001
EBF1	2.299	0.458	0.650	0.533 (0.409-0.658)	< 0.001
NYFC	2.458	0.647	0.950	0.789 (0.720-0.859)	0.102

Note: Sensitivity and specificity were determined based on cut_off values in the logistic regression analysis; The *p*-value was obtained from DeLong test to assess the significance of diagnostic performance differences between single and combined markers.

Table S15. Diagnostic performance of the mammography, ultrasound, CEA, CA153, and combined markers for distinguishing breast cancer from benign tumors.

Marker	Cut_off	Sensitivity	Specificity	AUC (95% CI)	*p*-value
Mammography	0.500	0.726	0.873	0.787 (0.729-0.845)	< 0.001
Ultrasound	0.500	0.940	0.754	0.754 (0.667-0.842)	< 0.001
CEA	0.500	0.174	0.901	0.537 (0.493-0.581)	< 0.001
CA153	0.500	0.154	0.915	0.534 (0.492-0.575)	< 0.001
Combined model	2.429	0.547	0.958	0.898 (0.858-0.938)	-

Note: CEA, carcinoembryonic antigen; CA153, cancer antigen 15-3; Combined model: combining methylation markers with mammography and ultrasound; Sensitivity and specificity were determined based on cut_off values; The *p*-value was obtained from DeLong test to assess the significance of diagnostic performance differences between single and combined model.

Table S16. Sensitivity and specificity of the combined model.

	Total	False predict	Correct predict	Sensitivity	Specificity
Stage I	65	42	23	0.354	-
Stage II	93	40	53	0.570	-
Stage III	37	9	28	0.757	-
Stage IV	6	0	6	1.000	-
All cancer	201	91	110	0.547	-
Benign tumor	71	3	68	-	0.958

Note: Sensitivity and specificity were determined based on cut_off value of 2.429.

Table S17. Demographic and clinical characteristics of training and test cohorts.

	Training set (N=250)			Test set (N = 105)		
Characteristic	**BC**	**Benign tumor**	**Healthy control**	**BC**	**Benign tumor**	**Healthy control**
Total (n)	**141**	**50**	**59**	**60**	**21**	**24**
Age (years)						
Mean ± SD	57.6 ± 9.6	45.3 ± 9.1	54.5 ± 11.0	57.5 ± 10.6	44.5 ± 10.6	56.5 ± 12.6
Range	36-82	27-68	37-79	34-81	27-60	39-78
BI-RADS category in mammography [a], n (%)						
1-4a	41 (29.1%)	43 (86.0%)	-	14 (23.3%)	19 (90.5%)	-
4b-6	83 (58.9%)	3 (6.0%)	-	39 (65.0%)	2 (9.5%)	-
Unknown	17 (12.0%)	4 (8.0%)	-	7 (11.7%)	0 (0.0%)	-
BI-RADS category in ultrasound [a], n (%)						
1-4a	7 (5.0%)	36 (72.0%)	-	5(8.3%)	16 (76.2%)	-
4b-6	108 (76.6%)	2 (4.0%)	-	45 (75.0%)	1 (4.8%)	-
Unknown	26 (18.4%)	12 (24.0%)	-	10 (16.7%)	4 (19.0%)	-
Tumor size, n (%)						
Tumor size ≤ 2 cm	70 (49.6%)	-	-	29 (48.3%)	-	-
Tumor size > 2 cm	71 (50.4%)	-	-	31 (51.7%)	-	-
Lymph node, n (%)						
Negative	81 (57.4%)	-	-	34 (56.7%)	-	-
Positive	60 (42.6%)	-	-	25 (41.7%)	-	-
Unknown	0 (0.0%)	-	-	1 (1.6%)	-	-
Stage [b], n (%)						
Stage I	48 (34.1%)	-	-	17 (28.3%)	-	-
Stage II	61 (43.3%)	-	-	32 (53.3%)	-	-
Stage III	27 (19.1%)	-	-	10 (16.7%)	-	-
Stage IV	5 (3.5%)	-	-	1 (1.7%)	-	-
Molecular subtype, n (%)						
Luminal	98 (69.5%)	-	-	51 (85.0%)	-	-
HER2+	10 (7.1%)	-	-	3 (5.0%)	-	-
TNBC	29 (20.6%)	-	-	4 (6.7%)	-	-
Unknown	4 (2.8%)	-	-	2 (3.3%)	-	-
CEA, n (%)						
≤ 5 ng/mL	120 (85.1%)	46 (92.0%)	-	46 (76.7%)	18 (85.7%)	-
> 5 ng/mL	3 (2.1%)	1 (2.0%)	-	6 (10.0%)	0 (0.0%)	-

Characteristic	Training set (N=250)			Test set (N = 105)		
	BC	Benign tumor	Healthy control	BC	Benign tumor	Healthy control
Unknown	18 (12.8%)	3 (6.0%)	-	8 (13.3%)	3 (14.3%)	-
CA153, n (%)						
≤ 25 U/mL	122 (86.5%)	47 (94.0%)	-	48 (80.0%)	18 (85.7%)	-
> 25 U/mL	1 (0.7%)	0 (0.0%)	-	4 (6.7%)	0 (0.0%)	-
Unknown	18 (12.8%)	3 (6.0%)	-	8 (13.3%)	3 (14.3%)	-

Note: [a], Categories of the Breast Imaging Reporting and Data System; [b], Clinical staging was determined according to the eighth edition of the classification for breast cancer of the American Joint Commission of Cancer; BC, Breast cancer; SD, Standard deviation; TNBC, Triple-negative breast cancer; CEA, Carcinoembryonic antigen; CA153, Cancer antigen 15-3; -, not available.

Table S18. Diagnostic performance of RF-, SVM-, and EN- assisted model for discriminating breast cancer from healthy control.

	RF	SVM	EN
AUC (95% CI)	0.907 (0.843-0.972)	0.771 (0.666-0.876)	0.862 (0.784-0.940)
Accuracy (95% CI)	0.798 (0.696-0.878)	0.833 (0.736-0.906)	0.833 (0.736-0.906)
No Information Rate	0.714	0.714	0.714
P-Value [Acc > NIR]	0.055	0.008	0.008
Kappa	0.510	0.570	0.559
Sensitivity	0.850	0.917	0.933
Specificity	0.667	0.625	0.583
Positive Predictive Value	0.864	0.859	0.849
Negative Predictive Value	0.640	0.750	0.778
Prevalence	0.714	0.714	0.714
Detection Rate	0.607	0.655	0.667
Detection Prevalence	0.702	0.762	0.786
Balanced Accuracy	0.758	0.771	0.758
Positive Class	BC	BC	BC

Note: RF, random forest algorithm; SVM, support vector machine algorithm; EN, elastic net algorithm; AUC, area under the curve; No Information Rate is the accuracy of a prediction only based on the frequency of the most common class (usually a negative class) without any model; Detection Rate is the percentage of all samples that are correctly predicted to be positive; Detection prevalence is the proportion of samples that the model predicts to be positive, whether or not those predictions are correct.

Table S19. Diagnostic performance of RF-, SVM-, and EN- assisted model for discriminating breast cancer from benign tumors.

	RF	SVM	EN
AUC (95% CI)	0.683 (0.557-0.808)	0.500 (0.500-0.500)	0.706 (0.583-0.830)
Accuracy (95% CI)	0.753 (0.645-0.842)	0.741 (0.631-0.832)	0.617 (0.503-0.723)
No Information Rate	0.741	0.741	0.741
P-Value [Acc > NIR]	0.458	0.558	0.995
Kappa	0.109	0.000	0.248
Sensitivity	0.983	1.000	0.567
Specificity	0.095	0.000	0.762
Pos Pred Value	0.756	0.741	0.872
Neg Pred Value	0.667	Na	0.381
Prevalence	0.741	0.741	0.741
Detection Rate	0.728	0.741	0.420
Detection Prevalence	0.963	1.000	0.482
Balanced Accuracy	0.539	0.500	0.664
Positive Class	BC	BC	BC

Note: RF, random forest algorithm; SVM, support vector machine algorithm; EN, elastic net algorithm; AUC, area under the curve; No Information Rate is the accuracy of a prediction only based on the frequency of the most common class (usually a negative class) without any model; Detection Rate is the percentage of all samples that are correctly predicted to be positive; Detection prevalence is the proportion of samples that the model predicts to be positive, whether or not those predictions are correct.

Table S20. Diagnostic performance of RF-, SVM-, and EN- assisted model for discriminating breast cancer from non-cancer.

	RF	SVM	EN
AUC (95% CI)	0.819 (0.742-0.897)	0.686 (0.601-0.771)	0.797 (0.712-0.882)
Accuracy (95% CI)	0.695 (0.598-0.781)	0.667 (0.568-0.756)	0.695 (0.598-0.781)
No Information Rate	0.571	0.571	0.571
P-Value [Acc > NIR]	0.006	0.029	0.006
Kappa	0.407	0.354	0.398
Sensitivity	0.583	0.550	0.633
Specificity	0.844	0.822	0.778
Pos Pred Value	0.833	0.805	0.792
Neg Pred Value	0.603	0.578	0.614
Prevalence	0.571	0.571	0.571
Detection Rate	0.333	0.314	0.362
Detection Prevalence	0.400	0.391	0.457
Balanced Accuracy	0.714	0.686	0.706
Positive Class	BC	BC	BC

Note: RF, random forest algorithm; SVM, support vector machine algorithm; EN, elastic net algorithm; AUC, area under the curve; No Information Rate is the accuracy of a prediction only based on the

frequency of the most common class (usually a negative class) without any model; Detection Rate is the percentage of all samples that are correctly predicted to be positive; Detection prevalence is the proportion of samples that the model predicts to be positive, whether or not those predictions are correct.

Table S21. cfDNA detection rates for breast cancer and benign tumors in mammography and ultrasound diagnostic groups.

Group	BC	BE
Mammography-negative	50.01% (28/55)	95.16 (59/62)
Mammography-positive	48.36% (59/122)	100.00% (5/5)
Ultrasound-negative	66.67% (8/12)	98.08% (51/52)
Ultrasound-positive	52.29% (80/153)	100.00% (3/3)

Figures

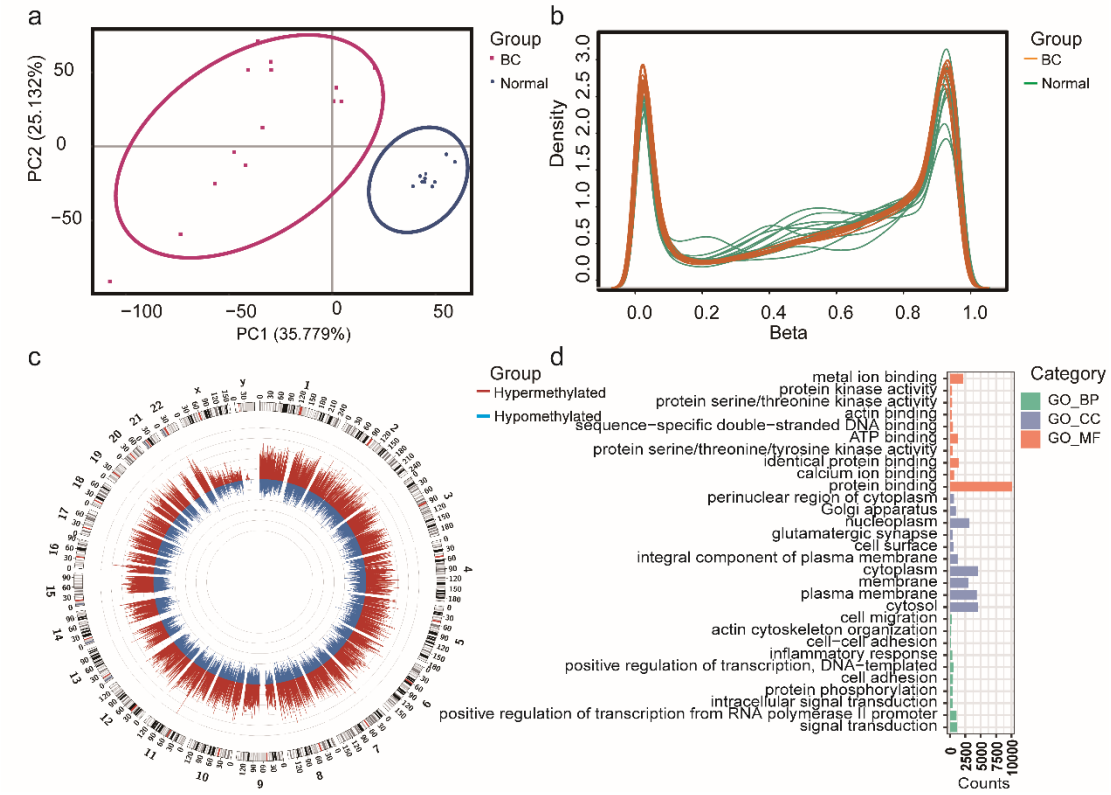


Figure S1. Overview of DNA methylation changes in BC compared to tumor-adjacent tissues. a, Principal component analysis of 850K array data from in-house subjects (n = 24). b, Density distribution plot of probes for BC and tumor-adjacent tissues. c, The distribution of CpG sites among different chromosomes. d, The Gene Ontology (GO) enrichment results of differentially methylated genes.

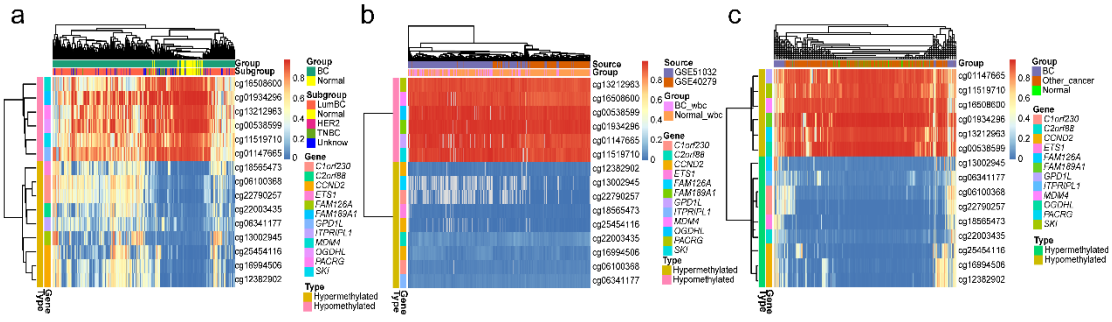


Figure S2. Methylation levels of breast cancer specific CpG sites in the DNA of breast cancer tissues, tumor-adjacent tissues, white blood cells, and other cancer-type tissues in 450K array. a, Unsupervised hierarchical clustering of 15 breast cancer (BC)-specific DNA methylation markers that are differentially methylated between BC (n = 774) and tumor-adjacent tissues (n = 97) from the TCGA database. b, Unsupervised hierarchical clustering of 15 BC-specific DNA methylation markers in white blood cells (n = 911) from the GEO database. c, Unsupervised hierarchical clustering of 15 BC-specific DNA methylation markers in 29 other cancer type tissues (n = 3655) from the TCGA database.

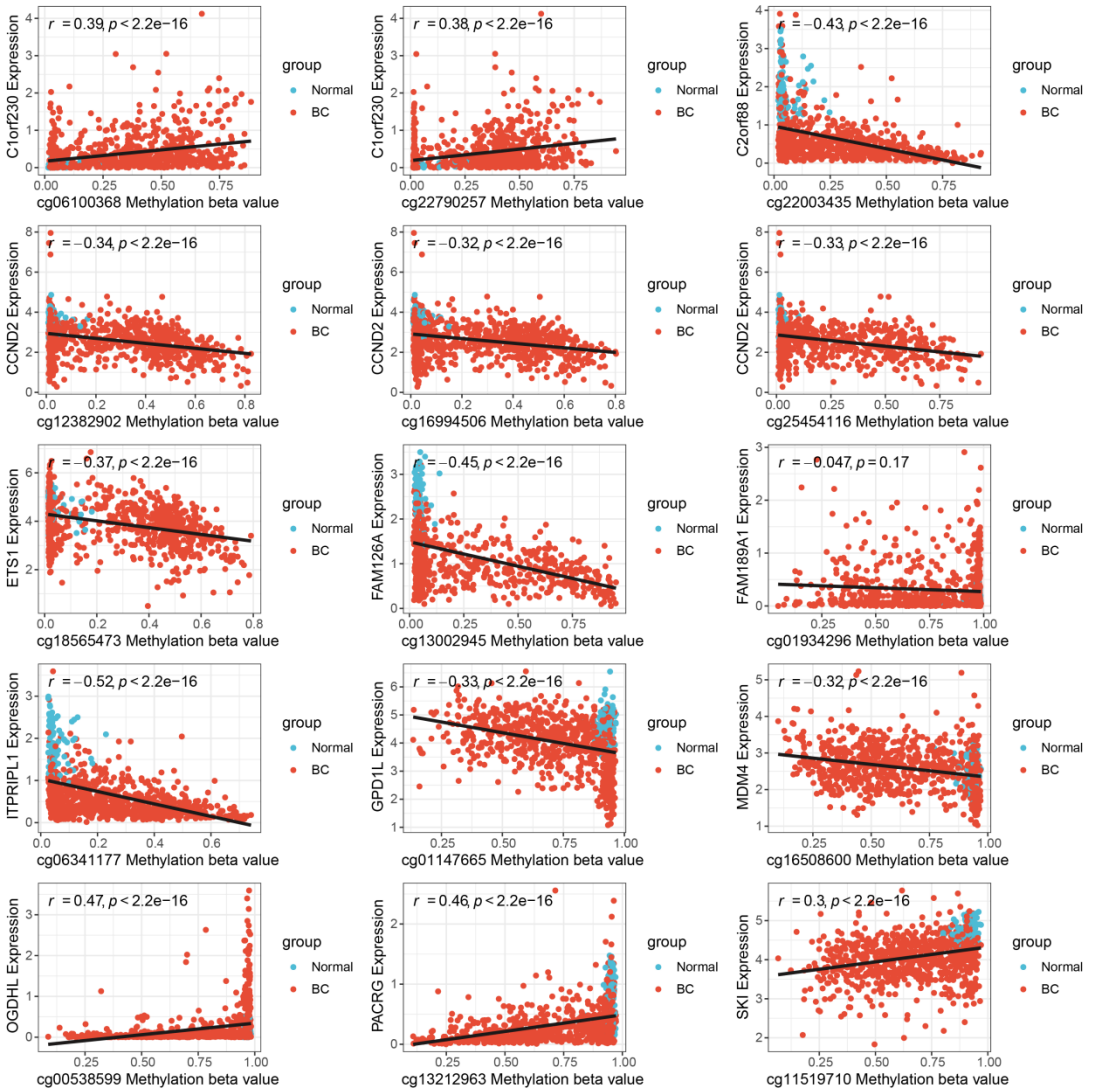


Figure S3. Spearman correlation analysis between methylation levels of specific CpG sites and expression levels of corresponding genes in the TCGA BC dataset. The x-axis represents the methylation level and the y-axis represents the expression level. r, correlation coefficient; BC, breast cancer.

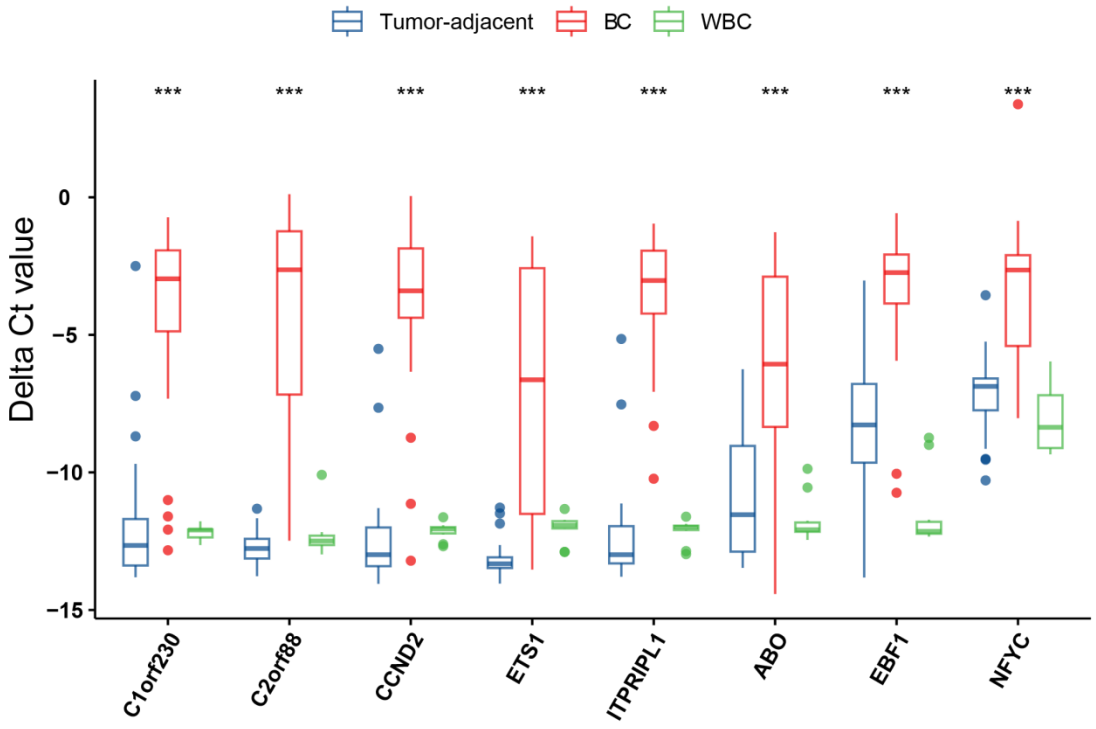


Figure S4. The methylation levels of the eight candidate markers in the qMSP cohort. Boxplot of methylation levels of eight candidate markers of *C1orf230*, *C2orf88*, *CCND2*, *ETS1*, *ITPRIPL1*, *ABO*, *EBF1*, and *NFYC* in BC tissues (n = 22), tumor-adjacent tissues (n = 18), and WBC (n = 10). The y-axis represents methylation levels (ΔCt = CTreference (VIC) – CTtarget (FAM)), in which a higher value represents a higher methylation level. BC, breast cancer; ***, *p*-value ≤ 0.001; *p*-values were determined by ANOVA test.

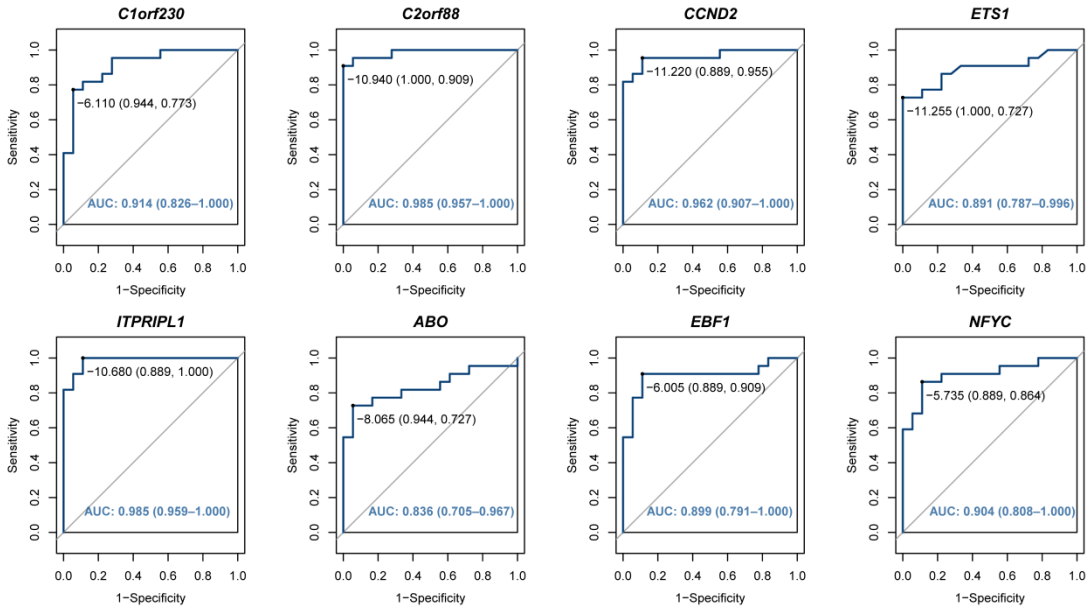


Figure S5. The receiver operating characteristic (ROC) curves of the eight candidate markers in the quantitative methylation-specific PCR (qMSP) cohort.

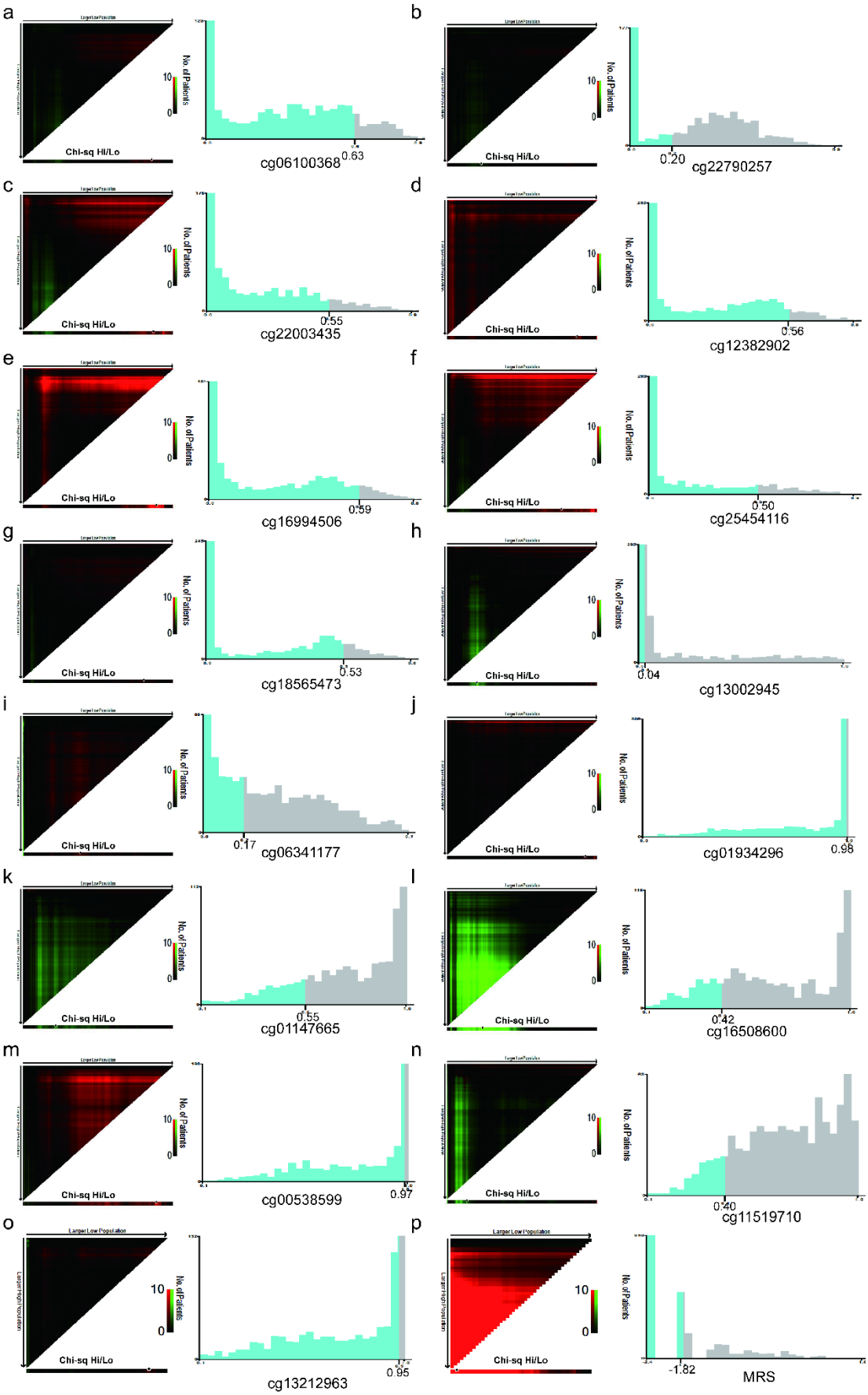

a
Chi-sq Hi/Lo
No. of Patients
10
0
0.63
cg06100368
b
Chi-sq Hi/Lo
No. of Patients
10
0
0.20
cg22790257
c
Chi-sq Hi/Lo
No. of Patients
10
0
0.55
cg22003435
d
Chi-sq Hi/Lo
No. of Patients
10
0
0.56
cg12382902
e
Chi-sq Hi/Lo
No. of Patients
10
0
0.59
cg16994506
f
Chi-sq Hi/Lo
No. of Patients
10
0
0.50
cg25454116
g
Chi-sq Hi/Lo
No. of Patients
10
0
0.53
cg18565473
h
Chi-sq Hi/Lo
No. of Patients
10
0
0.04
cg13002945
i
Chi-sq Hi/Lo
No. of Patients
10
0
0.17
cg06341177
j
Chi-sq Hi/Lo
No. of Patients
10
0
0.98
cg01934296
k
Chi-sq Hi/Lo
No. of Patients
10
0
0.55
cg01147665
l
Chi-sq Hi/Lo
No. of Patients
10
0
0.42
cg16508600
m
Chi-sq Hi/Lo
No. of Patients
10
0
0.97
cg00538599
n
Chi-sq Hi/Lo
No. of Patients
10
0
0.40
cg11519710
o
Larger Low Population
Larger High Population
Chi-sq Hi/Lo
No. of Patients
10
0
0.95
cg13212963
p
Larger Low Population
Larger High Population
Chi-sq Hi/Lo
No. of Patients
10
0
-1.82
MRS

Figure S6. X-tile analysis of overall survival was performed to determine the cutoffs for markers.

X-tile plots are created by dividing marker data into two populations: low and high (left panels), red coloration of cut points indicates an inverse correlation with survival, whereas green coloration represents direct associations. The cut-point in the left panel is shown on a histogram (right panels).

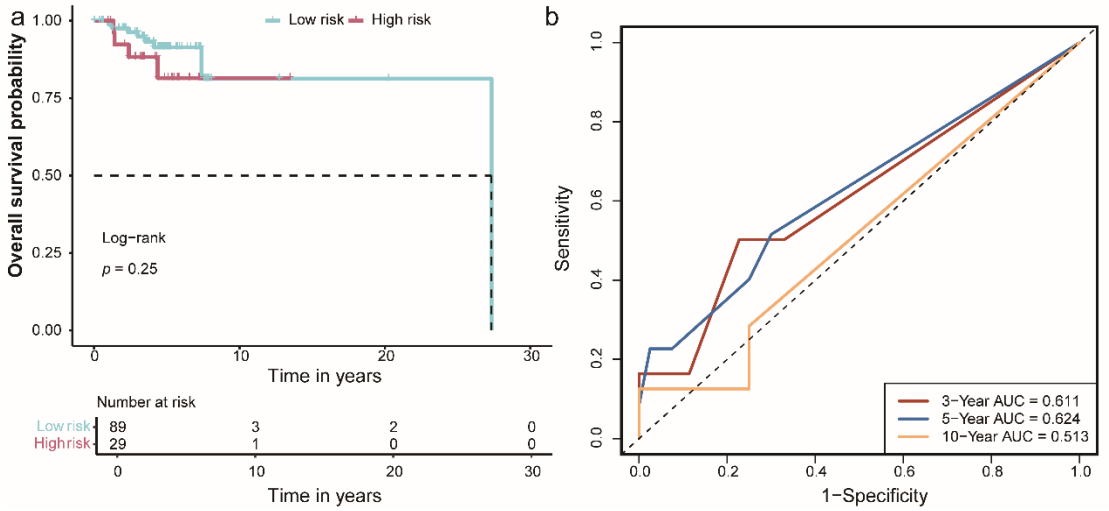


Figure S7. Predictive overall survival performance of methylation risk score using Kaplan–Meier survival curve (a) and time-dependent ROC analysis (b) in GSE72251 cohort.

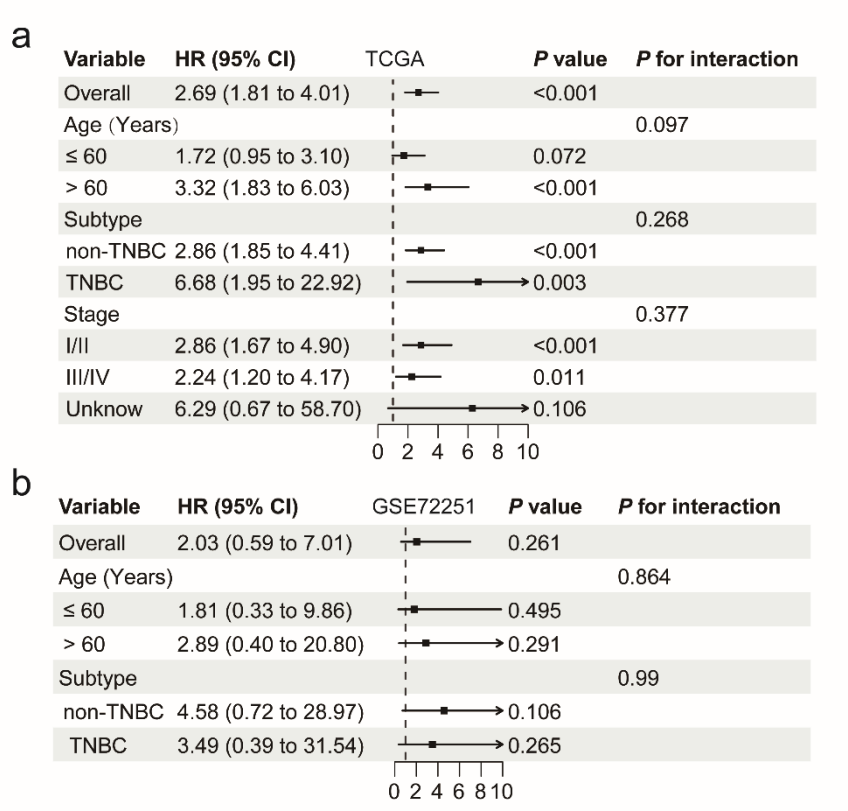
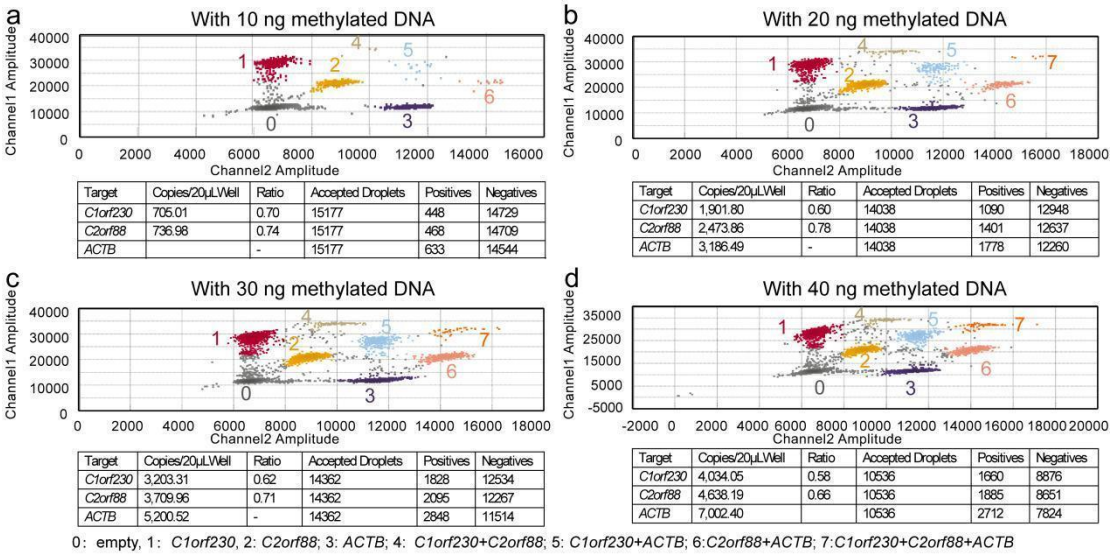


Figure S8. Association between methylation risk score and overall survival by subgroup analysis in the TCGA (a) and GSE72251 (b) cohorts.



Target	Copies/20µLWell	Ratio	Accepted Droplets	Positives	Negatives
C1orf230	705.01	0.70	15177	448	14729
C2orf88	736.98	0.74	15177	468	14709
ACTB		-	15177	633	14544

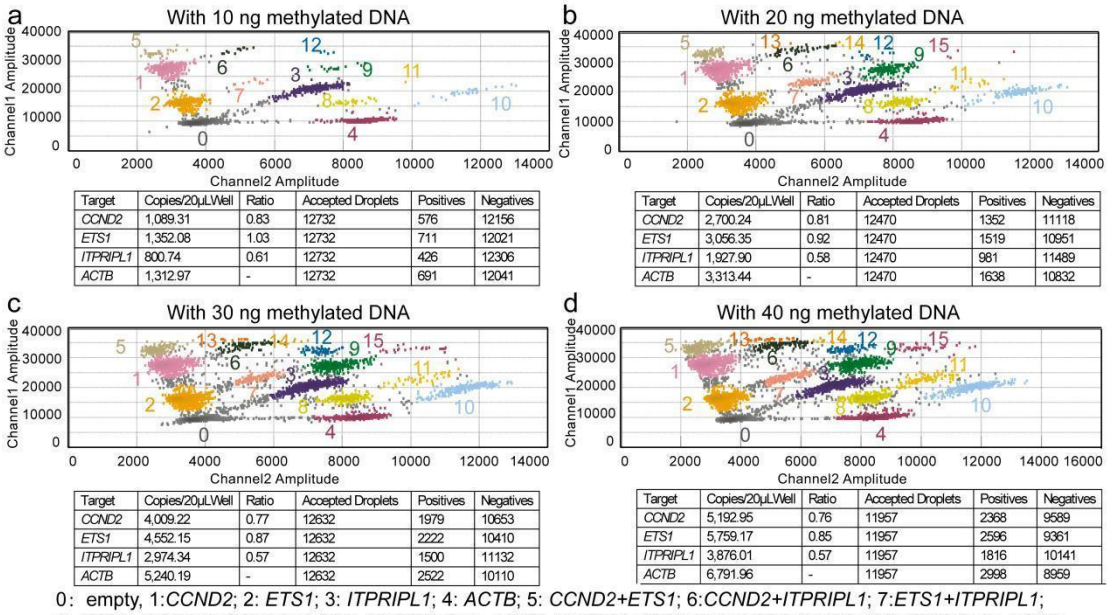
Target	Copies/20µLWell	Ratio	Accepted Droplets	Positives	Negatives
C1orf230	1,901.80	0.60	14038	1090	12948
C2orf88	2,473.86	0.78	14038	1401	12637
ACTB	3,186.49	-	14038	1778	12260

Target	Copies/20µLWell	Ratio	Accepted Droplets	Positives	Negatives
C1orf230	3,203.31	0.62	14362	1828	12534
C2orf88	3,709.96	0.71	14362	2095	12267
ACTB	5,200.52	-	14362	2848	11514

Target	Copies/20µLWell	Ratio	Accepted Droplets	Positives	Negatives
C1orf230	4,034.05	0.58	10536	1660	8876
C2orf88	4,638.19	0.66	10536	1885	8651
ACTB	7,002.40		10536	2712	7824

0: empty, 1: *C1orf230*, 2: *C2orf88*; 3: *ACTB*; 4: *C1orf230+C2orf88*; 5: *C1orf230+ACTB*; 6:*C2orf88+ACTB*; 7:*C1orf230+C2orf88+ACTB*

Figure S9. Analysis of a 3-plex ddPCR assay (mddPCR assay1) for cluster definitions. For this experiment, 10 ng (a), 20 ng (b), 30 ng (c), and 40 ng (d) of methylated DNA control were used. The objective of this experiment was to test cluster discrimination based on fluorescence amplitude shift and cluster position.



Target	Copies/20µLWell	Ratio	Accepted Droplets	Positives	Negatives
CCND2	1,089.31	0.83	12732	576	12156
ETS1	1,352.08	1.03	12732	711	12021
ITPRIPL1	800.74	0.61	12732	426	12306
ACTB	1,312.97	-	12732	691	12041

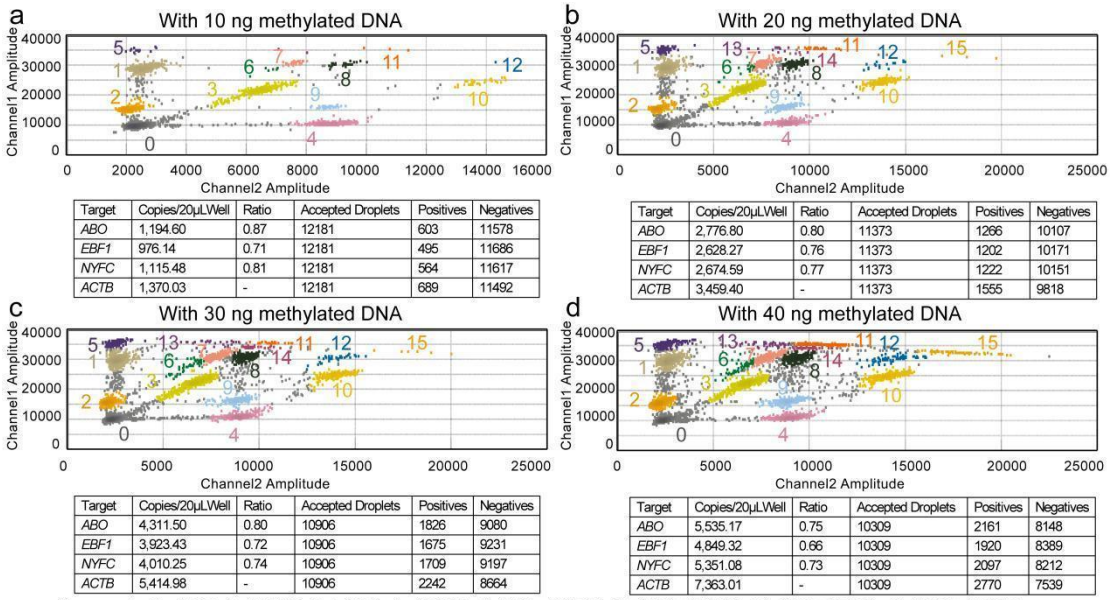
Target	Copies/20µLWell	Ratio	Accepted Droplets	Positives	Negatives
CCND2	2,700.24	0.81	12470	1352	11118
ETS1	3,056.35	0.92	12470	1519	10951
ITPRIPL1	1,927.90	0.58	12470	981	11489
ACTB	3,313.44	-	12470	1638	10832

Target	Copies/20µLWell	Ratio	Accepted Droplets	Positives	Negatives
CCND2	4,009.22	0.77	12632	1979	10653
ETS1	4,552.15	0.87	12632	2222	10410
ITPRIPL1	2,974.34	0.57	12632	1500	11132
ACTB	5,240.19	-	12632	2522	10110

Target	Copies/20µLWell	Ratio	Accepted Droplets	Positives	Negatives
CCND2	5,192.95	0.76	11957	2368	9589
ETS1	5,759.17	0.85	11957	2596	9361
ITPRIPL1	3,876.01	0.57	11957	1816	10141
ACTB	6,791.96	-	11957	2998	8959

0: empty, 1:*CCND2*; 2: *ETS1*; 3: *ITPRIPL1*; 4: *ACTB*; 5: *CCND2+ETS1*; 6:*CCND2+ITPRIPL1*; 7:*ETS1+ITPRIPL1*; 8: *ETS1+ ACTB*; 9:*CCND2+ACTB*; 10:*ITPRIPL1+ ACTB*; 11:*ETS1+ITPRIPL1+ACTB*; 12:*CCND2+ITPRIPL1+ ACTB*; 13: *CCND2+ETS1+ITPRIPL1*; 14: *CCND2+ETS1+ACTB*; 15:*CCND2+ETS1+ITPRIPL1+ ACTB*

Figure S10. Analysis of a 4-plex ddPCR assay (mddPCR assay2) for cluster definitions. For this experiment, 10 ng (a), 20 ng (b), 30 ng (c), and 40 ng (d) of methylated DNA control were used.



Target	Copies/20µLWell	Ratio	Accepted Droplets	Positives	Negatives
ABO	1,194.60	0.87	12181	603	11578
EBF1	976.14	0.71	12181	495	11686
NYFC	1,115.48	0.81	12181	564	11617
ACTB	1,370.03	-	12181	689	11492

Target	Copies/20µLWell	Ratio	Accepted Droplets	Positives	Negatives
ABO	2,776.80	0.80	11373	1266	10107
EBF1	2,628.27	0.76	11373	1202	10171
NYFC	2,674.59	0.77	11373	1222	10151
ACTB	3,459.40	-	11373	1555	9818

Target	Copies/20µLWell	Ratio	Accepted Droplets	Positives	Negatives
ABO	4,311.50	0.80	10906	1826	9080
EBF1	3,923.43	0.72	10906	1675	9231
NYFC	4,010.25	0.74	10906	1709	9197
ACTB	5,414.98	-	10906	2242	8664

Target	Copies/20µLWell	Ratio	Accepted Droplets	Positives	Negatives
ABO	5,535.17	0.75	10309	2161	8148
EBF1	4,849.32	0.66	10309	1920	8389
NYFC	5,351.08	0.73	10309	2097	8212
ACTB	7,363.01	-	10309	2770	7539

0: empty, 1: *ABO*; 2: *NYFC*; 3: *EBF1*; 4: *ACTB*; 5:*ABO+NYFC*; 6:*EBF1+NYFC* ; 7:*ABO+EBF1* ; 8:*ABO+ ACTB*; 9: *NYFC+ACTB*; 10: *EBF1+ACTB*; 11:*ABO+EBF1+ACTB* ; 12: *EBF1+NYFC+ACTB*; 13:*ABO+NYFC+ EBF1*; 14: *ABO+NYFC+ACTB*; 15: *ABO+NYFC+ EBF1+ACTB*

Figure S11. Analysis of a 4-plex ddPCR assay (mddPCR assay3) for cluster definitions. For this experiment, 10 ng (a), 20 ng (b), 30 ng (c), and 40 ng (d) of methylated DNA control were used.

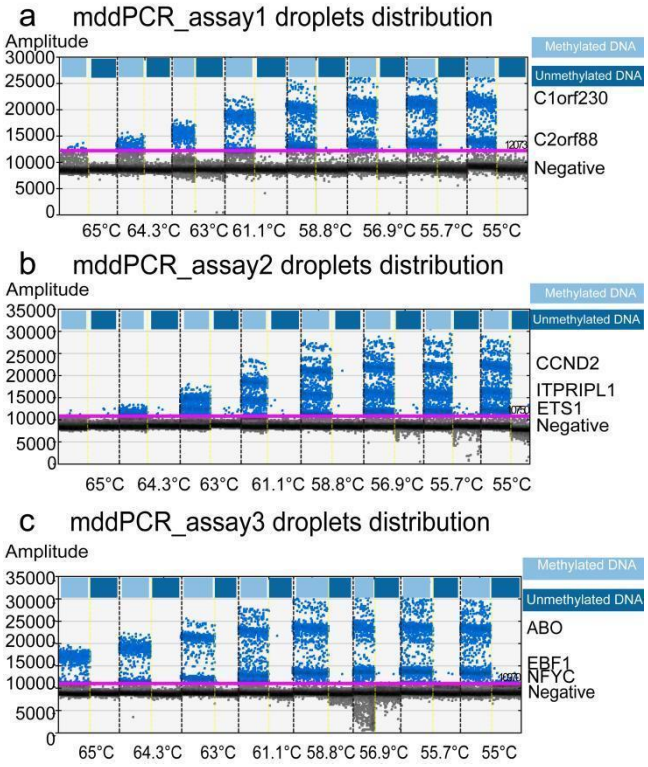


Figure S12. Optimization of annealing temperature for three mddPCR assays. a, For mddPCR assay 1, 56.9°C was considered as the optimal annealing temperature yielded the best performance; For mddPCR assay 2 (b) and mddPCR assay 3 (c), 58.8°C was considered as the optimal annealing temperature. Each column (x-axis) indicating one PCR plate well, filling with methylated DNA and

unmethylated DNA. Y-axis: fluorescence amplitude of droplet detection in the FAM channel. For each target, visually evaluating the fluorescence shift from empty droplets (grey) to amplified clusters (blue).

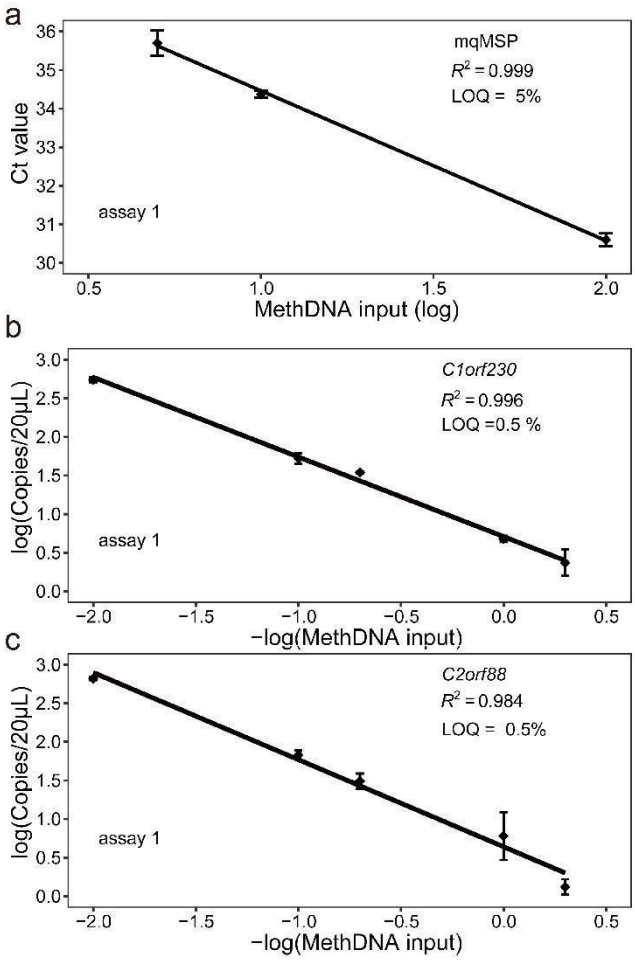


Figure S13. Comparative analysis of limit of quantification (LOQ) for multiplex qMSP (mqMSP assay 1) and 3-plex ddPCR assay 1. Calibration curves showing the linear dynamic range of mqPCR assay 1 (a), *C1orf230* (b), and *C2orf88* (c) using the mixture of 10 ng of the methylated DNA control with the unmethylated DNA control and diluting it at 100%, 10%, 5%, 1%, 0.5%, 0.1%, 0.01%, and 0%.

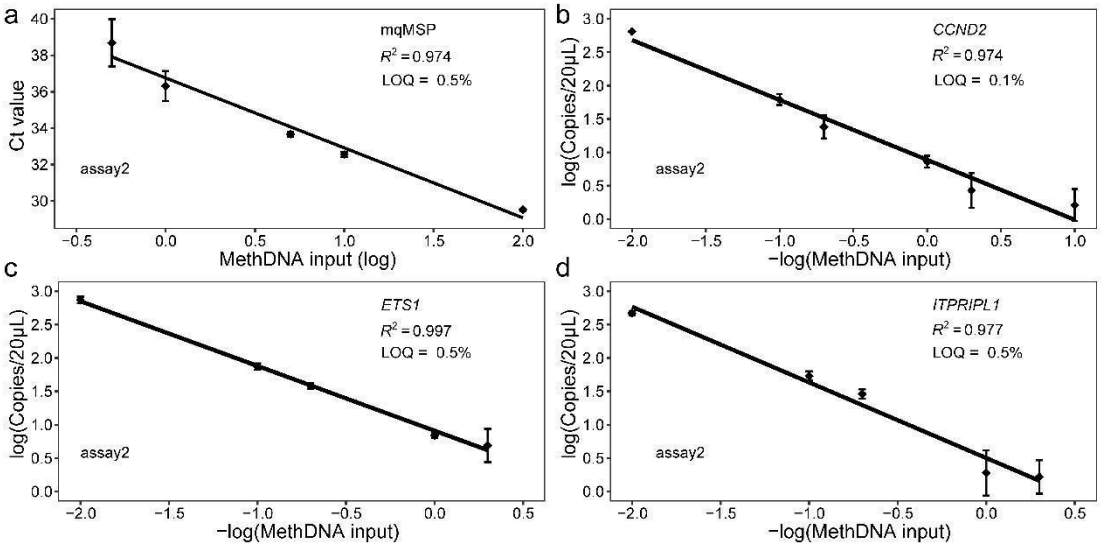


Figure S14. Comparative analysis of limit of quantification (LOQ) for multiplex qMSP (mqMSP assay 2) and 4-plex ddPCR assay 2. Calibration curves showing the linear dynamic range of mqPCR assay 2 (a), *CCND2* (b), *ETS1* (c), and *ITPRIPL1* (d) using the mixture of 10 ng of the methylated DNA control with the unmethylated DNA control and diluting it at 100%, 10%, 5%, 1%, 0.5%, 0.1%, 0.01%, and 0%.

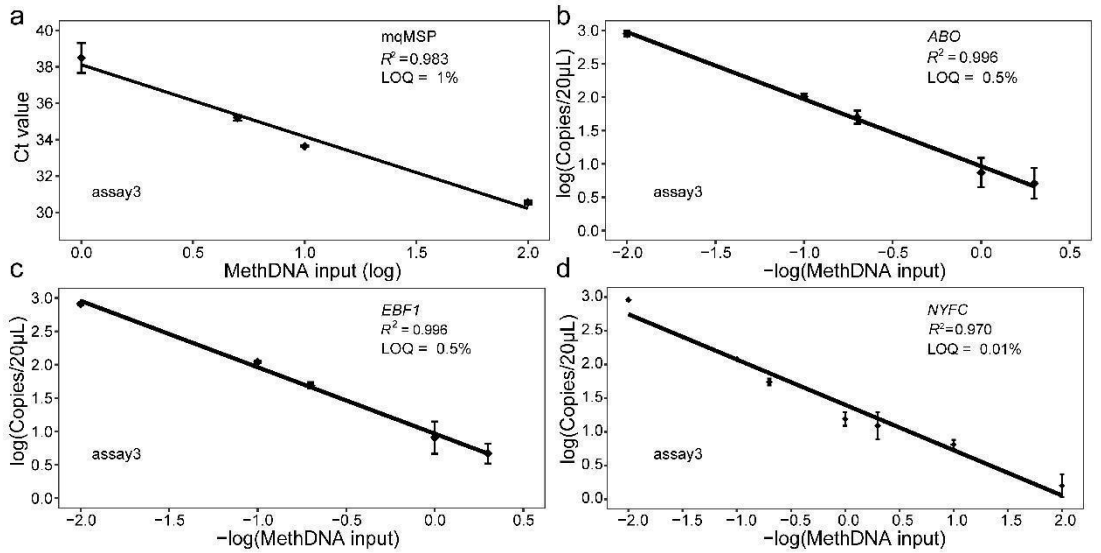


Figure S15. Comparative analysis of limit of quantification (LOQ) for multiplex qMSP (mqMSP assay 3) and 4-plex ddPCR assay 3. Calibration curves showing the linear dynamic range of mqPCR assay 2 (a), *ABO* (b), *EBF1* (c), and *NYFC* (d) using the mixture of 10 ng of the methylated DNA control with the unmethylated DNA control and diluting it at 100%, 10%, 5%, 1%, 0.5%, 0.1%, 0.01%, and 0%.

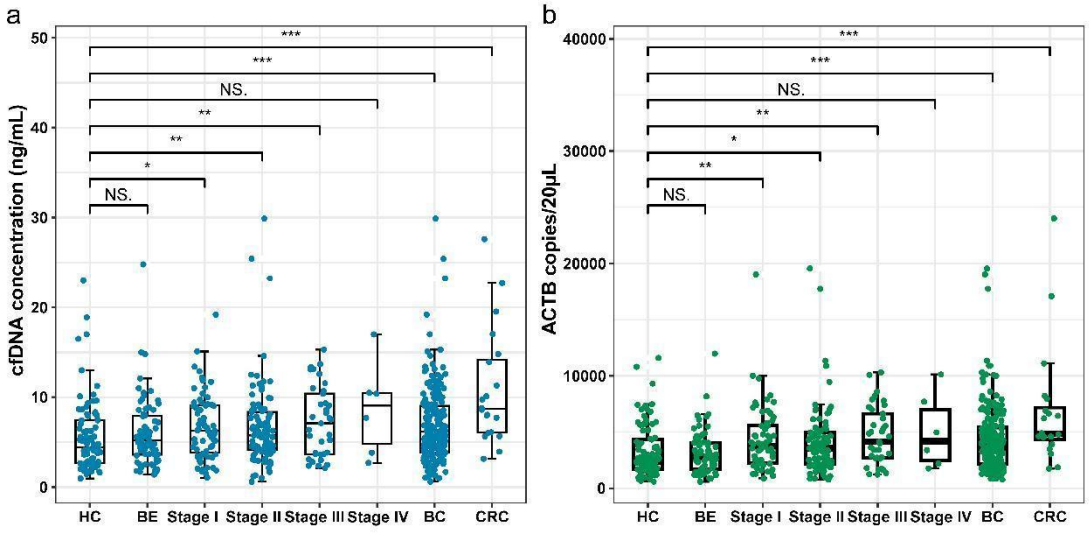


Figure S16. Circulating cell free DNA concentrations and number of ACTB copies analyzed for each sample. a, Circulating cell free DNA concentration for breast cancer patients, benign diseases, healthy controls, and colorectal cancer samples. b, Number of *ACTB* copies/20 µL analyzed in each ddPCR well. Comparison of healthy controls and other groups using *t* test. NS, no significance; *, *p*-value ≤ 0.05; **, *p*-value ≤ 0.01; ***, *p*-value ≤ 0.001. BC, breast cancer; HC, healthy controls; BE, benign diseases; NC, non-cancers; CRC, colorectal cancer.

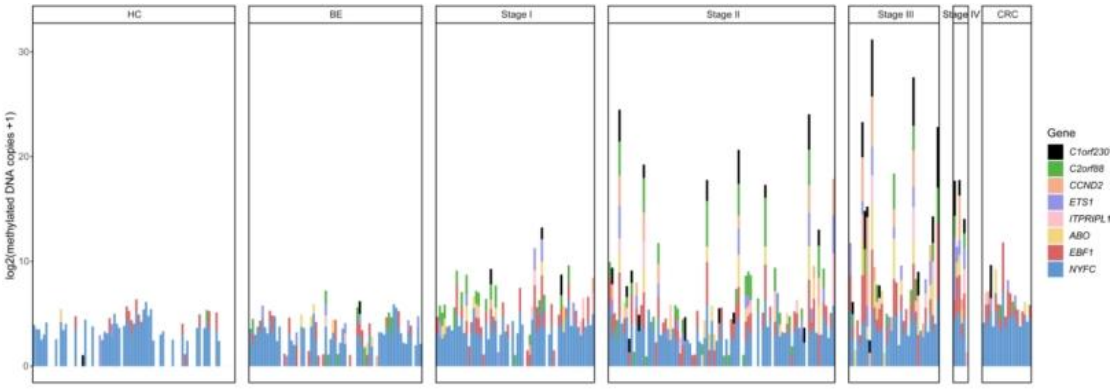


Figure S17. Histogram plots displaying data from the cfDNA cohort. For each sample (X-axis), the height of the histogram bar indicating the number of methylated DNA copies per mL plasma in the mddPCR assays, and the sum (Y-axis), and each colored segment representing an individual gene.

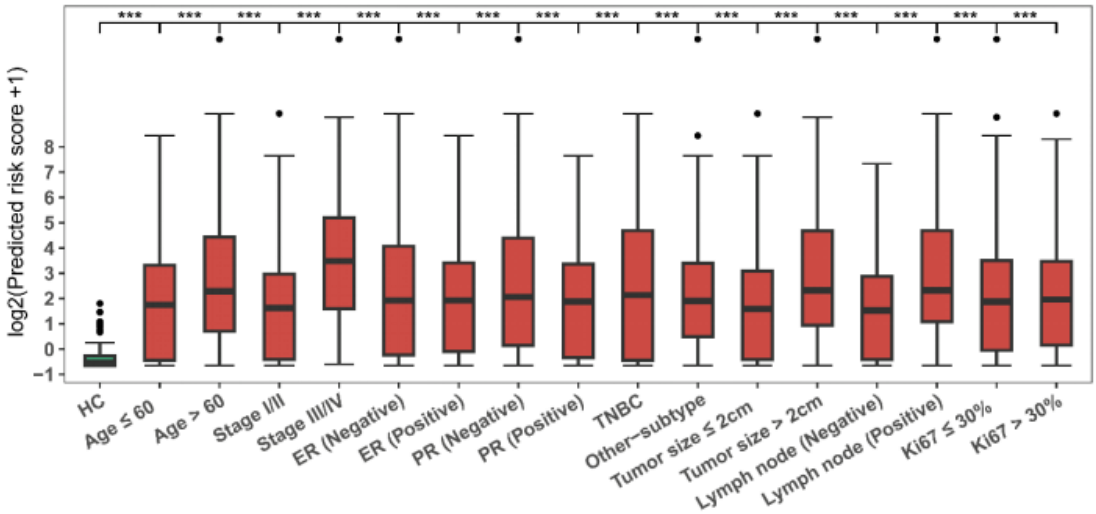


Figure S18. Differences of risk scores in BC patients of each subgroup clinical characteristics with healthy controls.

***: *p*-value ≤ 0.001; HC, healthy controls

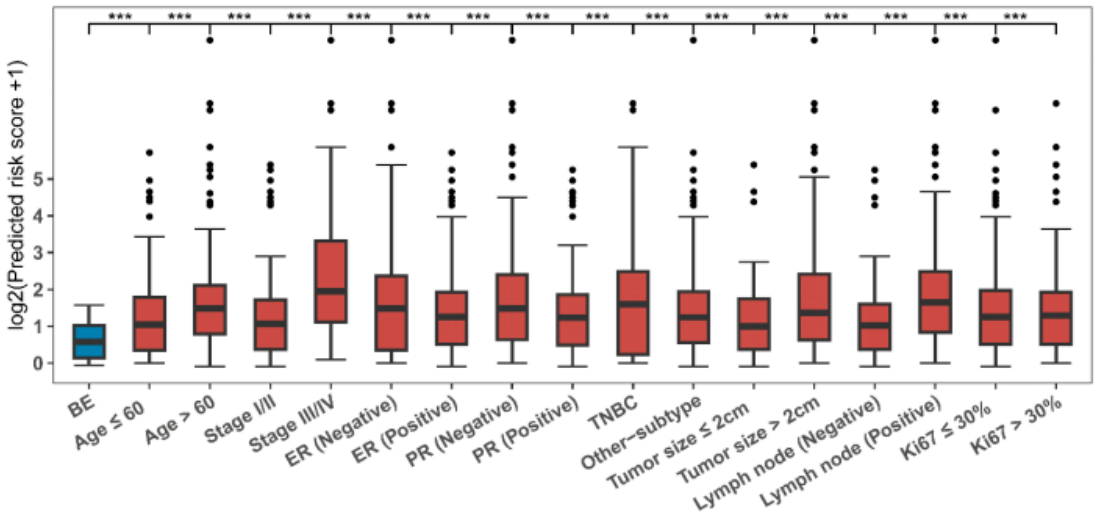


Figure S19. Differences of risk scores in BC patients of each clinical characteristics with benign tumors.

***: *p*-value ≤ 0.001; BE, benign-tumors.

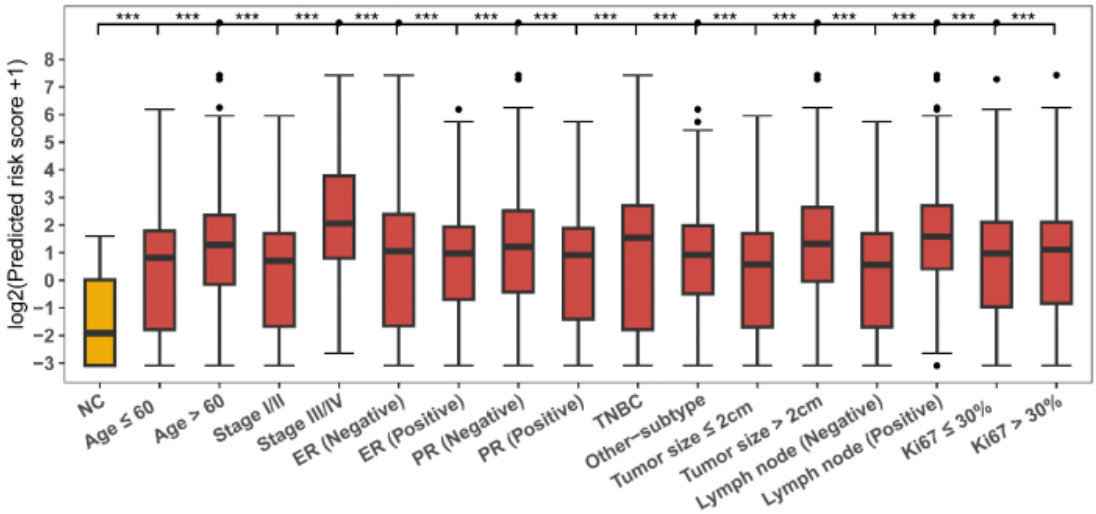


Figure S20. Differences of risk scores in BC patients of each subgroup clinical characteristics with non-cancers.

***: *p*-value ≤ 0.001; NC, non-cancers.

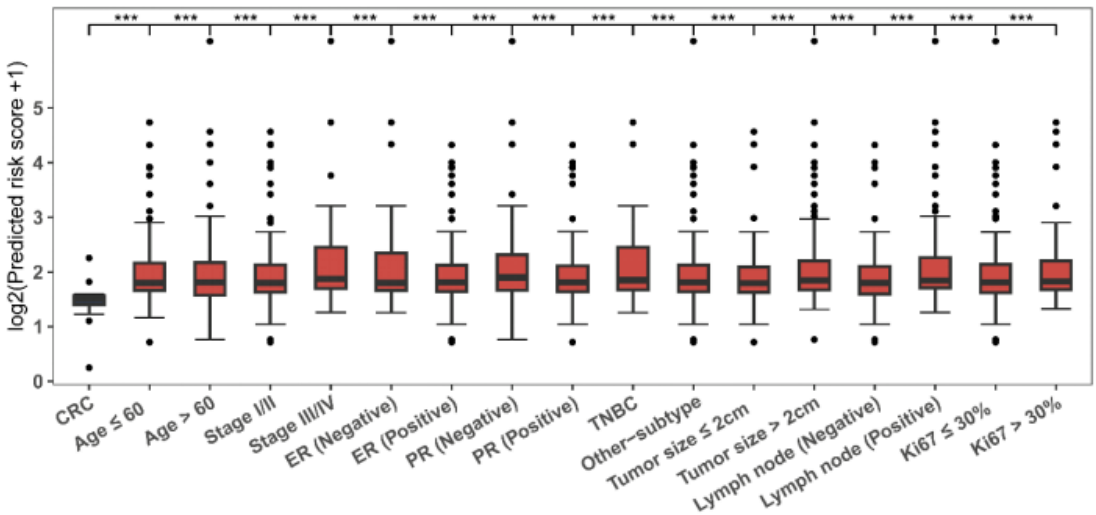


Figure S21. Differences of risk scores in BC patients of each subgroup clinical characteristics with colorectal cancer patients. ***: *p*-value ≤ 0.001; CRC, colorectal cancer.

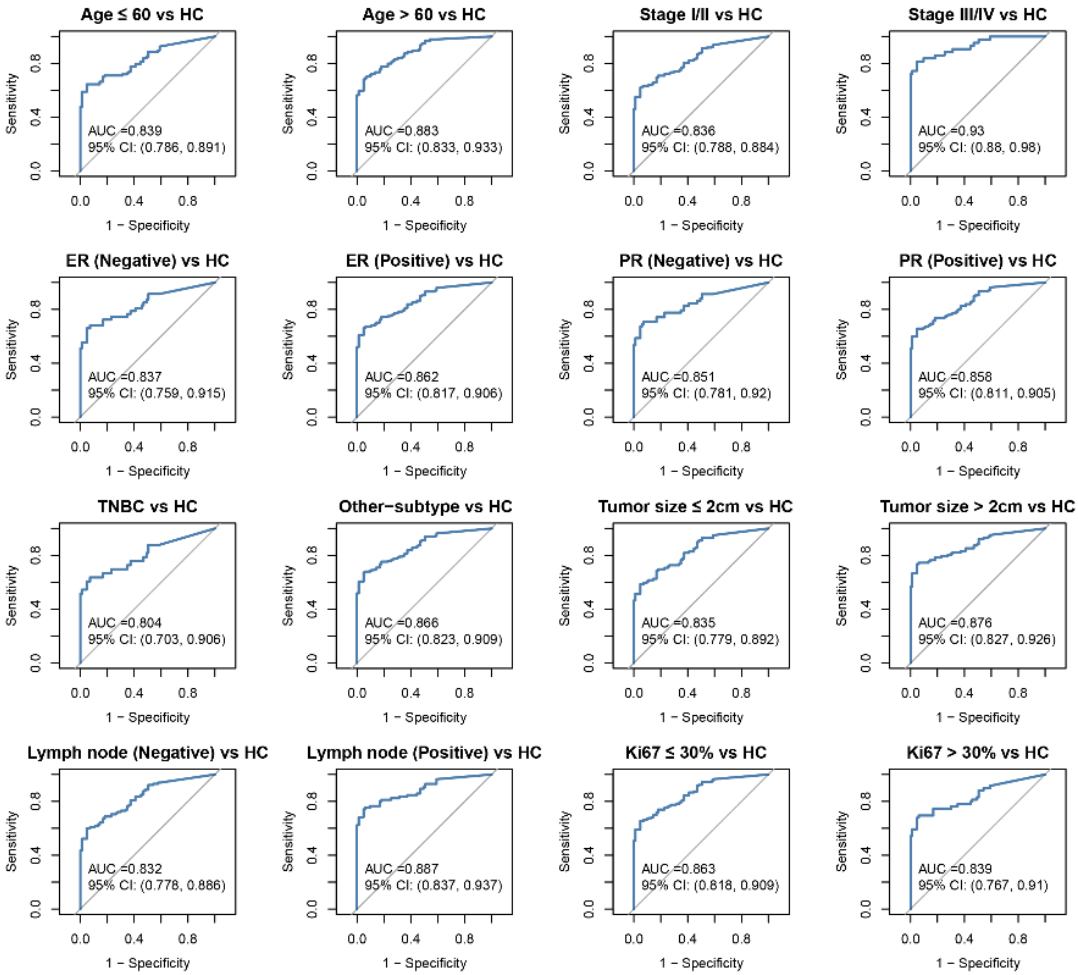


Figure S22. Receiver operating characteristic (ROC) curves between BC patients of each subgroup clinical characteristics and healthy controls.

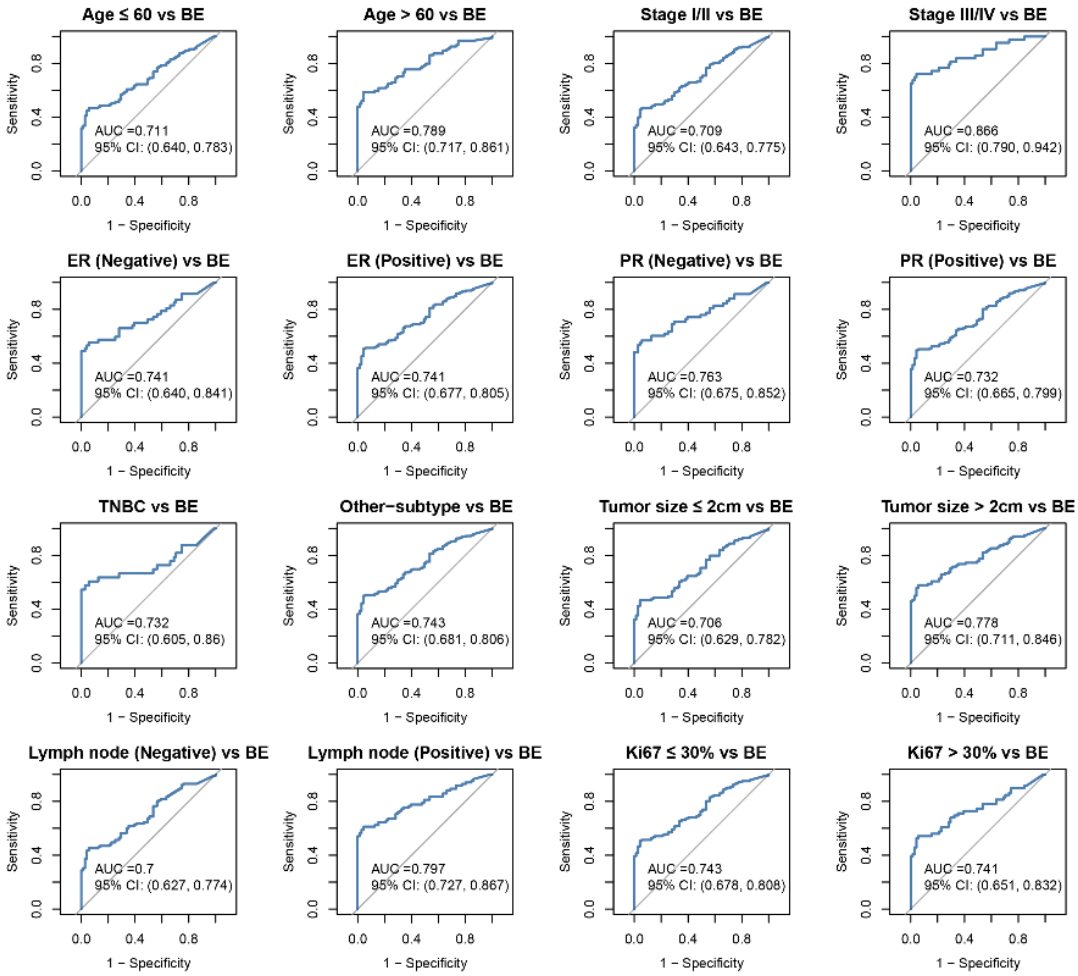


Figure S23. Receiver operating characteristic (ROC) curves between BC patients of each subgroup clinical characteristics and benign tumors.

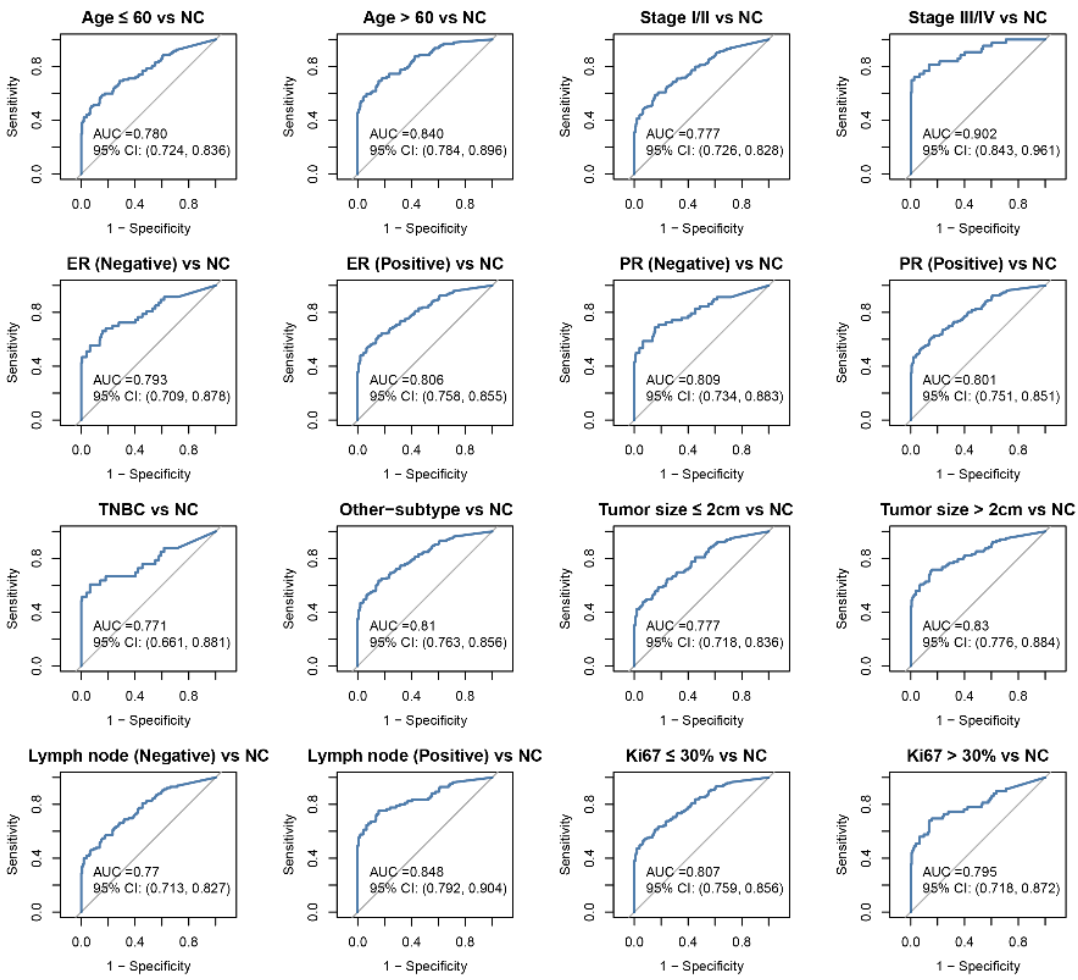


Figure S24. Receiver operating characteristic (ROC) curves between BC patients of each subgroup clinical characteristics and non-cancers.

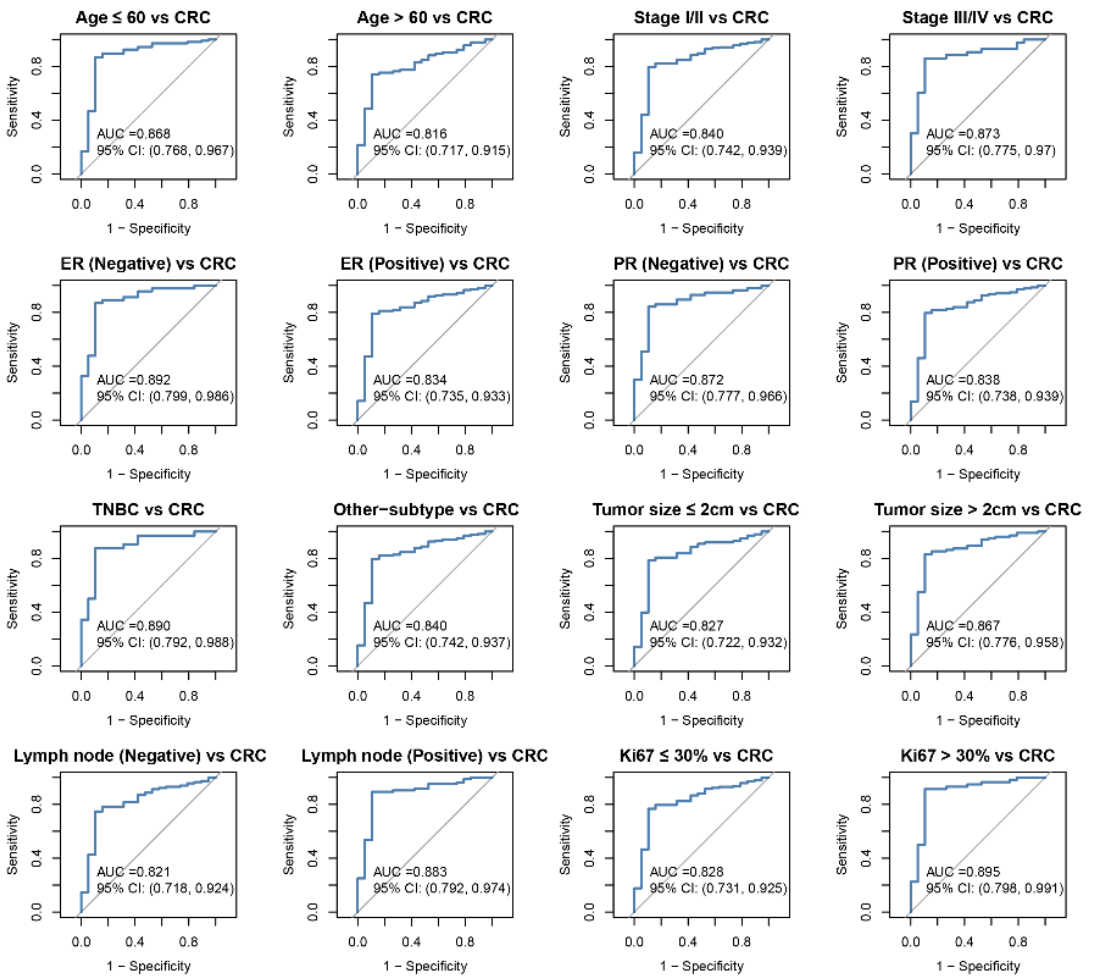


Figure S25. Receiver operating characteristic (ROC) curves between BC patients of each subgroup clinical characteristics and colorectal cancers.

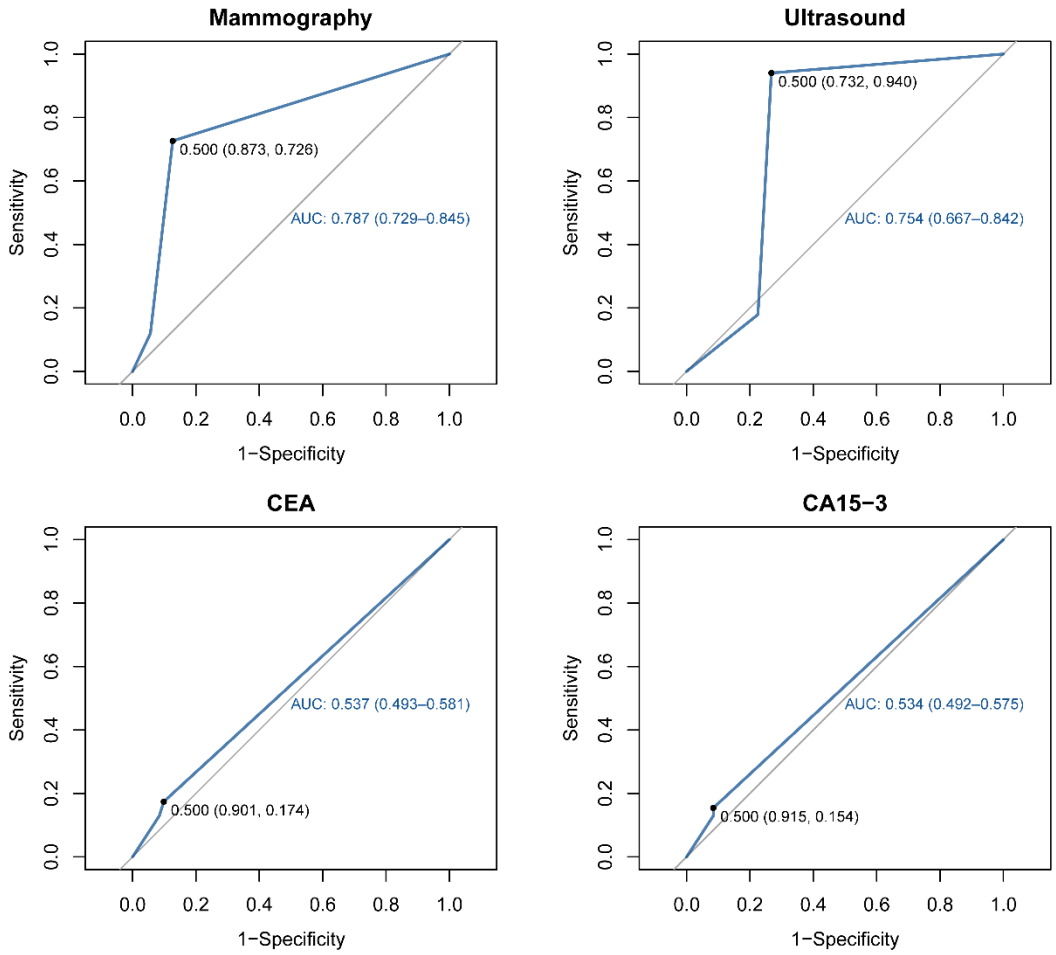


Figure S26. Receiver operating characteristic (ROC) curves of the diagnostic prediction model for distinguishing breast cancer patients from benign group, using mammography, ultrasound, CEA, and CA15-3.

The area under the curve (AUC) of the imaging (AUC = 0.754-0.787) for ultrasound and mammography or protein biomarkers (AUC = 0.534-0.537) for CA15-3 and CEA.

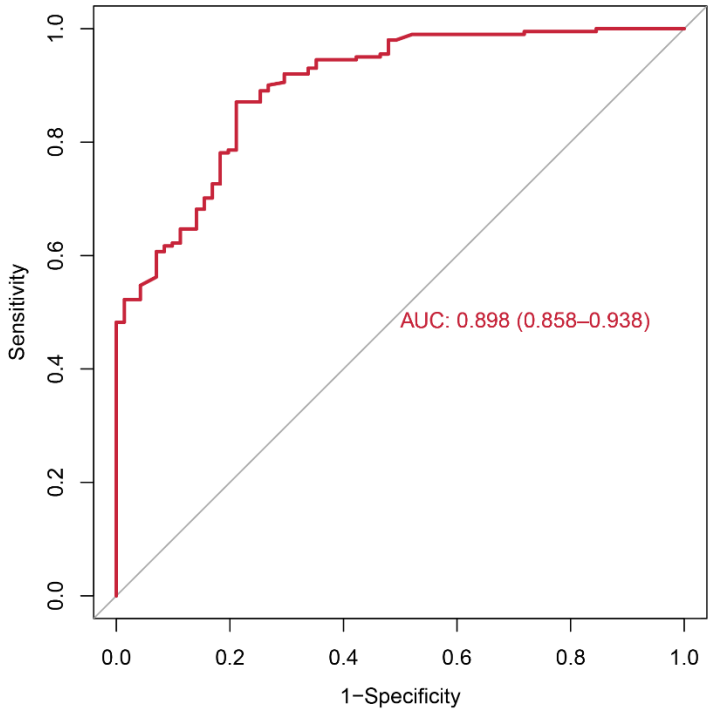


Figure S27. Receiver operating characteristic (ROC) curves of the combined diagnostic model for distinguishing breast cancer patients from benign group.

The combined model could differentiate BC patients from benign disease group with outstanding performance (AUC = 0.898, 95% CI = 0.858-0.938), which performed better than either one of the imaging.

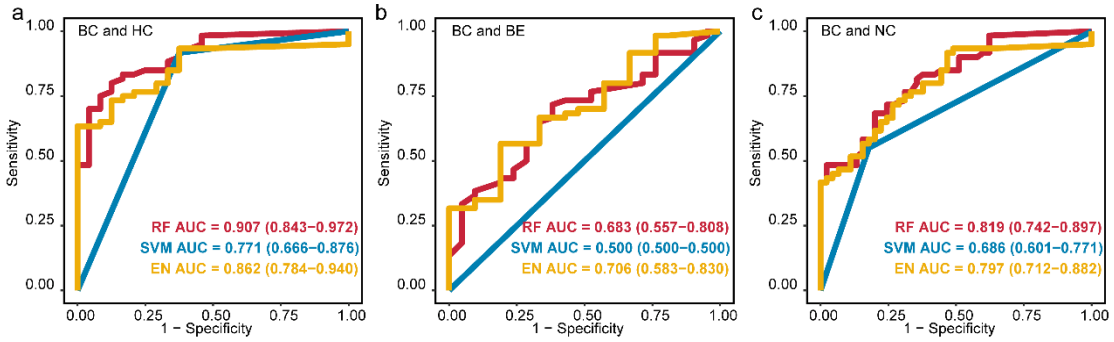


Figure S28. Diagnostic efficiency of DNA methylation markers using machine learning algorithms for breast cancer.

Receiver operating characteristic (ROC) for differentiating BC from healthy control (a), benign tumor (b), and non-cancer (c) groups based on supervised machine learning algorithms. RF, Random Forest; SVM, Support Vector Machine; EN, Elastic

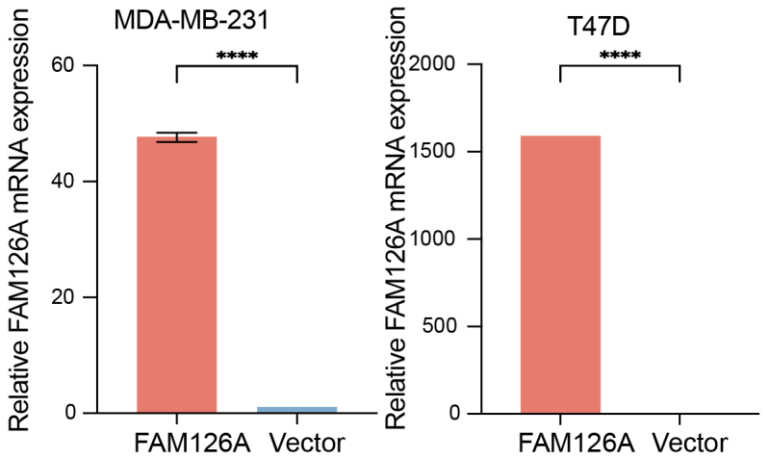


Figure S29. Overexpression of *FAM126A* was confirmed by qRT-PCR in MDA-MB-231 and T47D cells.

Data represent mean ± SEM from 3 independent experimentsand each had three replications. **** $p <$ 0.0001.