

Blood Test for Breast Cancer Screening through the Detection of Tumor-Associated Circulating Transcripts

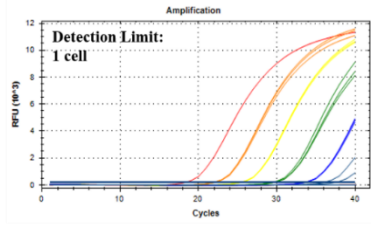
Sunyoung Park^{1,†}, Sungwoo Ahn^{1,†}, Jee Ye Kim², Jungho Kim³, Hyun Ju Han⁴, Dasom Hwang¹, Jungmin Park², Hyung Seok Park², Seho Park², Gun Min Kim⁵, Joohyuk Sohn⁵, Joon Jeong⁶, Yong Uk Song⁷, Hyeyoung Lee^{1,*}, Seung Il Kim^{2,*}

Table S1. The lists of profiled 28 genes for selection of TACTs

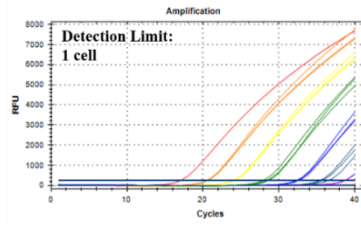
GENE ID	No of Pilot study (Clinical validation)	Characteristics of expression levels	<i>p</i> value
<i>EPCAM</i>	Breast cancer (282), healthy control (220)	Increase in BC	< 0.05
<i>KRT19</i>	Breast cancer (282), healthy control (220)	Increase in BC	< 0.05
<i>ERBB2</i>	Breast cancer (282), healthy control (220)	Increase in BC	< 0.05
<i>MKI67</i>	Breast cancer (282), healthy control (220)	Increase in BC	< 0.05
<i>TERT</i>	Breast cancer (282), healthy control (220)	Increase in BC	< 0.05
<i>VIM</i>	Breast cancer (282), healthy control (220)	Increase in BC	< 0.05
<i>SNAI2</i>	Breast cancer (282), healthy control (220)	Increase in BC	< 0.05
<i>FOXA2</i>	Breast cancer (282), healthy control (220)	Increase in BC	< 0.05
<i>MCAM</i>	Breast cancer (282), healthy control (220)	Increase in BC	< 0.05
<i>NPTN</i>	Breast cancer (282), healthy control (220)	Increase in BC	< 0.05
<i>TWIST1</i>	Breast cancer (43), healthy control (4)	Decrease in BCs	<0.05
<i>ZEB2</i>	Breast cancer (30), healthy control (40)	Decrease in BCs	0.001
<i>CXCR4</i>	Breast cancer (12), healthy control (4)	No significant difference	0.21
<i>SNAI1</i>	Breast cancer (43), healthy control (4)	Decrease in BCs	< 0.05
<i>CCND1</i>	Breast cancer (43), healthy control (4)	Decrease in BCs	< 0.05
<i>BMI1</i>	Breast cancer (15), healthy control (20)	No significant difference	0.14
<i>MAGEA3</i>	Breast cancer (4), healthy control (30)	Decrease in BCs	< 0.05
<i>PTEN</i>	Breast cancer (43), healthy control (4)	No significant difference	0.86
<i>PROM1</i>	Breast cancer (12), healthy control (4)	No significant difference	0.2
<i>FOXA1</i>	Breast cancer (12), healthy control (4)	No amplification of targets	-
<i>CDH1</i>	Breast cancer (12), healthy control (4)	No amplification of targets	-
<i>CDH2</i>	Breast cancer (43), healthy control (4)	No significant difference	0.71
<i>RRM1</i>	Breast cancer (43), healthy control (4)	Decrease in BCs	< 0.05
<i>CRYAB</i>	Breast cancer (12), healthy control (4)	No amplification of targets	-
<i>KL</i>	Breast cancer (12), healthy control (4)	No amplification of targets	-
<i>SCGB2A2</i>	Breast cancer (72), healthy control (19)	No significant difference	0.4
<i>TFF1</i>	Breast cancer (12), healthy control (4)	No amplification of targets	-
<i>MUC1</i>	Breast cancer (72), healthy control (19)	No significant difference	0.4

Table S2. Sensitivity and specificity of artificial neural network analysis in training and validation sets

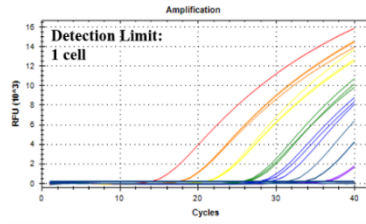
Model	ANN		DT		LR		SVM	
	Training Set	Validation Set	Training Set	Validation Set	Training Set	Validation Set	Training Set	Validation Set
Sensitivity	88.3% (95% CI 0.840–0.918)	90.2% (95% CI 0.835–0.948)	90.8% (95% CI 0.868–0.939)	79.5% (95% CI 0.713–0.863)	75.2% (95% CI 0.697–0.801)	76.2% (95% CI 0.677–0.835)	89.0% (95% CI 0.848–0.924)	86.1% (95% CI 0.786–0.917)
Specificity	87.7% (95% CI 0.827–0.918)	80.0% (95% CI 0.705–0.875)	84.6% (95% CI 0.791–0.891)	85.3% (95% CI 0.765–0.917)	63.6% (95% CI 0.569–0.700)	80.0% (95% CI 0.705–0.875)	76.8% (95% CI 0.707–0.822)	75.8% (95% CI 0.659–0.840)
Positive Predictive Value	90.2% (95% CI 0.861–0.935)	85.3% (95% CI 0.780–0.909)	88.3% (95% CI 0.840–0.917)	87.4% (95% CI 0.797–0.929)	72.6% (95% CI 0.671–0.776)	83.0% (95% CI 0.748–0.895)	83.1% (95% CI 0.784–0.872)	82.0% (95% CI 0.743–0.883)
Negative Predictive Value	85.4% (95% CI 0.801–0.897)	86.4% (95% CI 0.774–0.928)	87.7% (95% CI 0.826–0.918)	76.4% (95% CI 0.672–0.841)	66.7% (95% CI 0.599–0.730)	72.4% (95% CI 0.628–0.807)	84.5% (95% CI 0.787–0.892)	80.9% (95% CI 0.712–0.885)

A. EPCAM

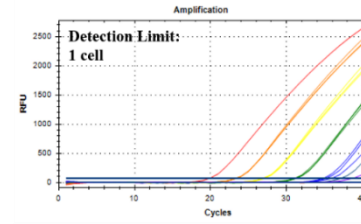
Blood + SK-BR-3 10^5
 Blood + SK-BR-3 10^4
 Blood + SK-BR-3 10^3
 Blood + SK-BR-3 10^2
 Blood + SK-BR-3 10^1
 Blood + SK-BR-3 10^0

B. KRT19

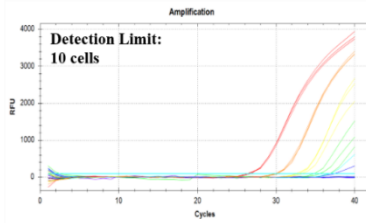
Blood + MCF-7 10^5
 Blood + MCF-7 10^4
 Blood + MCF-7 10^3
 Blood + MCF-7 10^2
 Blood + MCF-7 10^1
 Blood + MCF-7 10^0

C. ERBB2

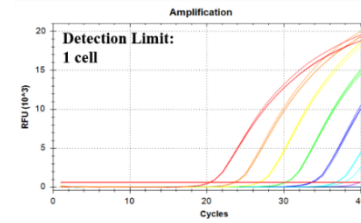
Blood + BT-474 10^5
 Blood + BT-474 10^4
 Blood + BT-474 10^3
 Blood + BT-474 10^2
 Blood + BT-474 10^1
 Blood + BT-474 10^0

D. MKI67

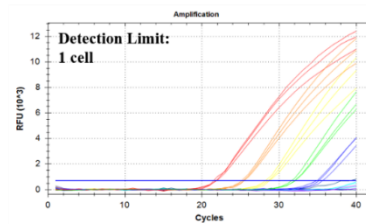
Blood + SK-BR-3 10^5
 Blood + SK-BR-3 10^4
 Blood + SK-BR-3 10^3
 Blood + SK-BR-3 10^2
 Blood + SK-BR-3 10^1
 Blood + SK-BR-3 10^0

E. TERT

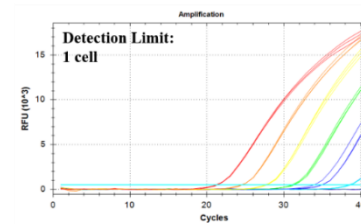
Blood + MDA-MB-231 10^5
 Blood + MDA-MB-231 10^4
 Blood + MDA-MB-231 10^3
 Blood + MDA-MB-231 10^2
 Blood + MDA-MB-231 10^1
 Blood + MDA-MB-231 10^0

F. VIM

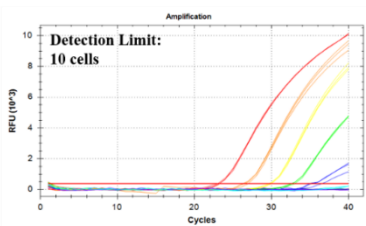
Blood + MDA-MB-231 10^5
 Blood + MDA-MB-231 10^4
 Blood + MDA-MB-231 10^3
 Blood + MDA-MB-231 10^2
 Blood + MDA-MB-231 10^1
 Blood + MDA-MB-231 10^0

G. NPTN

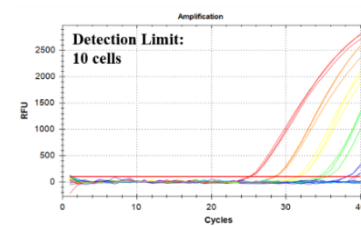
Blood + MDA-MB-231 10^5
 Blood + MDA-MB-231 10^4
 Blood + MDA-MB-231 10^3
 Blood + MDA-MB-231 10^2
 Blood + MDA-MB-231 10^1
 Blood + MDA-MB-231 10^0

H. MCAM

Blood + MDA-MB-231 10^5
 Blood + MDA-MB-231 10^4
 Blood + MDA-MB-231 10^3
 Blood + MDA-MB-231 10^2
 Blood + MDA-MB-231 10^1
 Blood + MDA-MB-231 10^0

I. SNAI2

Blood + MDA-MB-231 10^5
 Blood + MDA-MB-231 10^4
 Blood + MDA-MB-231 10^3
 Blood + MDA-MB-231 10^2
 Blood + MDA-MB-231 10^1

J. FOXA2

Blood + MDA-MB-231 10^5
 Blood + MDA-MB-231 10^4
 Blood + MDA-MB-231 10^3
 Blood + MDA-MB-231 10^2
 Blood + MDA-MB-231 10^1

Figure S1. Analytical sensitivity of ten tumor-related transcripts in breast cancer cells. Ten tumor-related transcripts (*EPCAM*, *KRT19*, *ERBB2*, *MKI67*, *TERT*, *VIM*, *NPTN*, *MCAM*, *SNAI2*, and *FOXA2*) were designed for highly sensitive detection, and evaluated in four representative breast cancer cell lines SKBR-3, MDA-MB-231, MCF-7, and BT-474. For the analytical sensitivity of RT-qPCR assay, breast cancer cell lines were serially diluted 10-fold from 1×10^6 to 1×10^0 cells/mL, and the last serial linear concentration that yielded 3 positive replicates was identified. The overall detection limit of this RT-qPCR assay for the ten tumor-related transcripts ranged from 10^1 to 10^0 cells/mL, which indicated high sensitivity. The 10 tumor-related transcripts could be detected in all four representative breast subtypes (luminal A, luminal B, HER2, and triple-negative breast cancer).

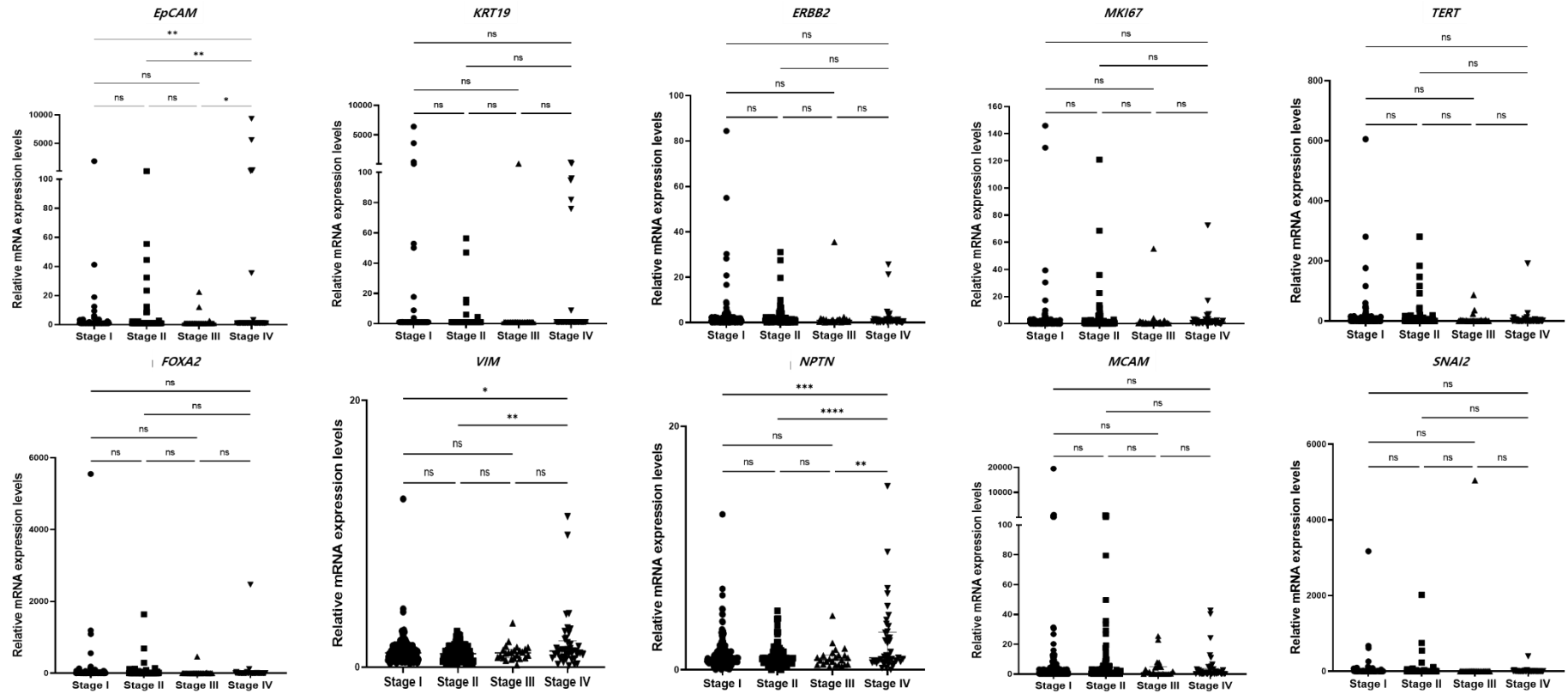


Figure S2. Clinical relevance between each TACT marker and TNM stage. There was statistical difference between stage IV and stage I, stage II in *EpCAM*, *VIM*, and *NPTN*. In addition, *EpCAM* showed significant difference between stage III compared with stage IV.). * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$. **** $p < 0.0001$.

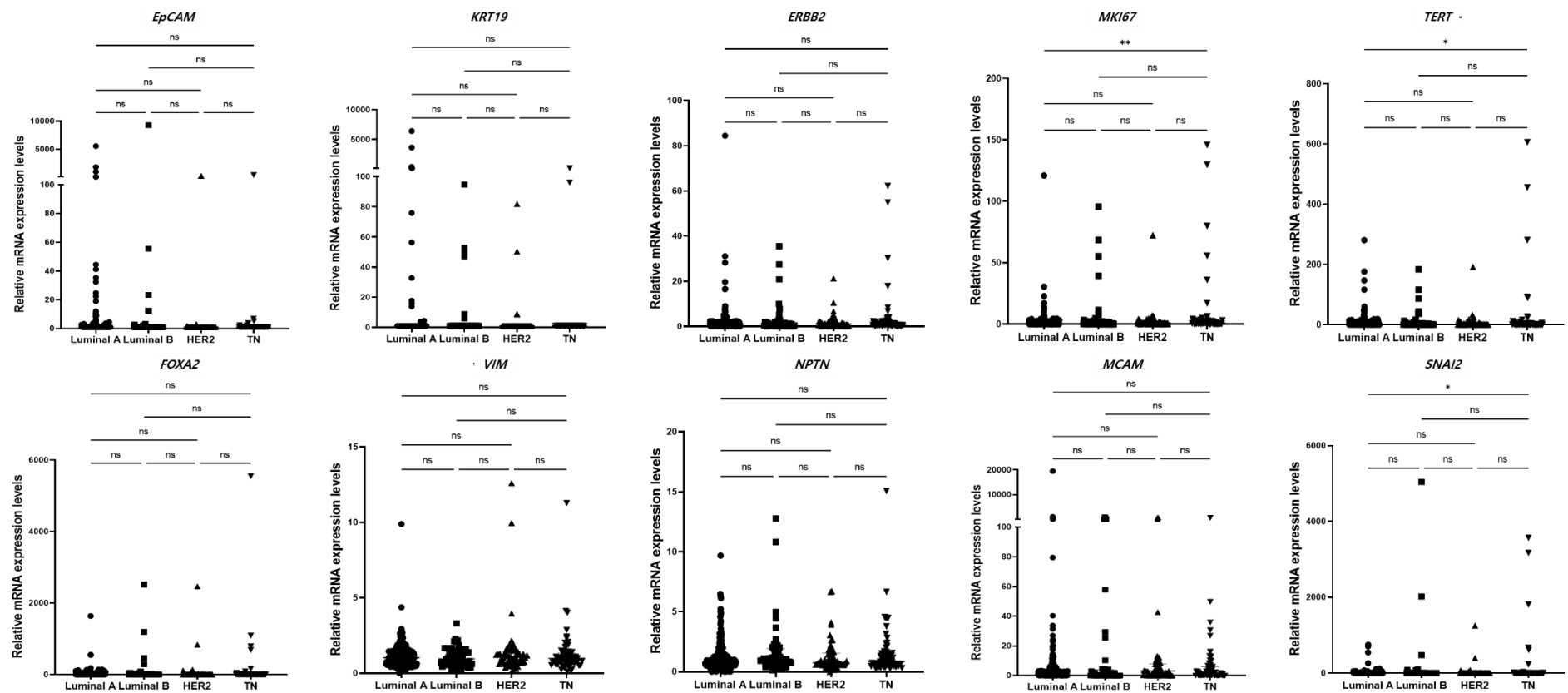


Figure S3. Clinical relevance between each TACT marker and subtypes. *MKI67*, *TERT* and *SNAI2* had significant difference between luminal A and basal-like subtype. * $p < 0.05$, ** $p < 0.01$.

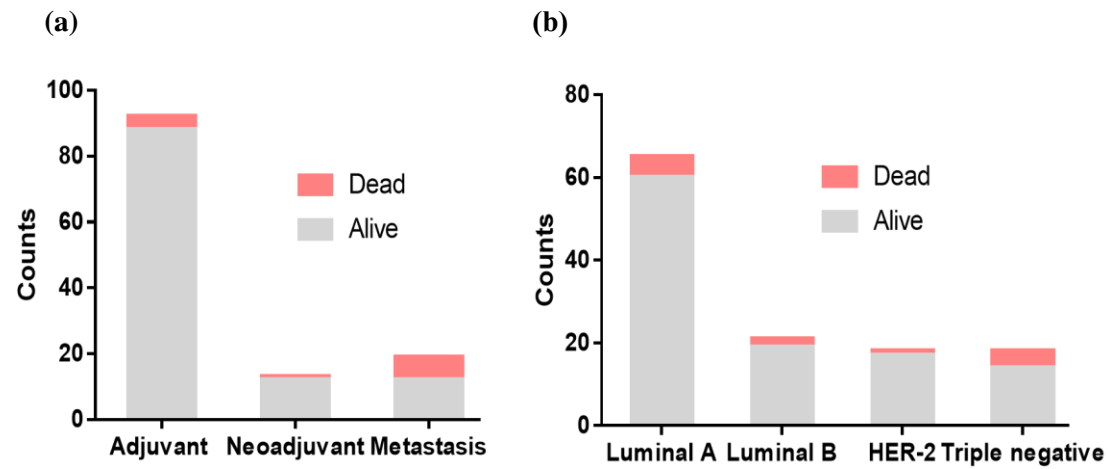


Figure S4. Vital status of stages and subtypes. a. Vital status of adjuvant, neoadjuvant, and metastasis chemotherapy types. b. Vital status of breast cancer subtypes.

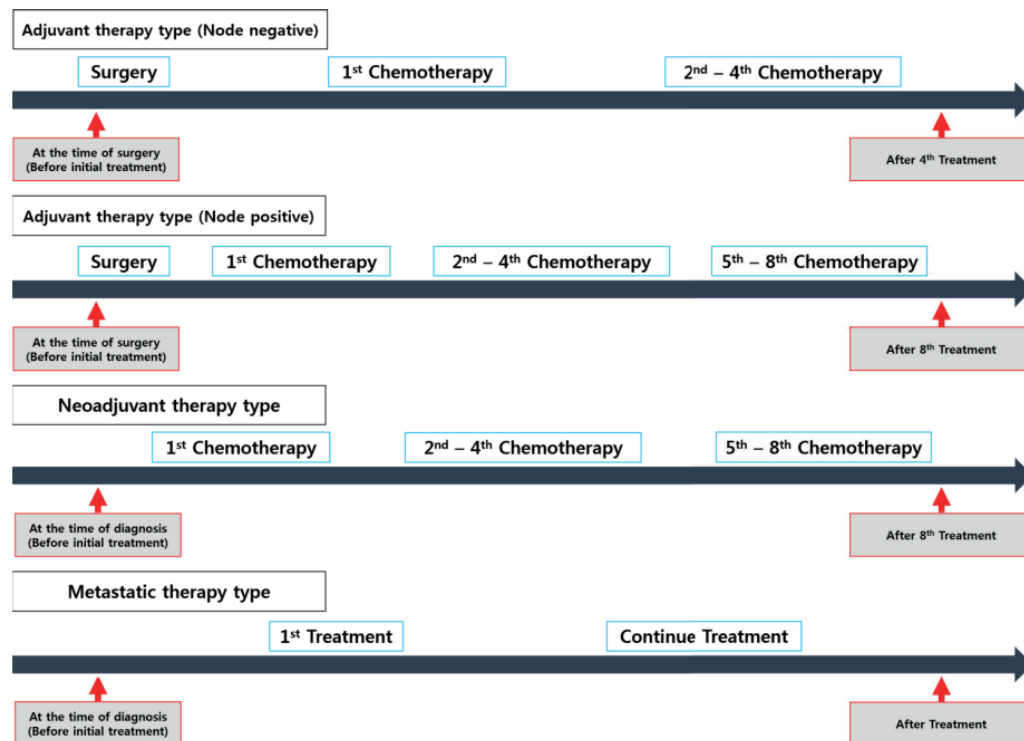


Figure S5. Scheme of adjuvant, neoadjuvant, and metastasis chemotherapy types and blood collection time from the breast cancer patients. Breast cancer patients with the adjuvant setting were treated with up to 4 to 8 chemotherapy, and blood collection was performed after the completion of last chemotherapy. Breast cancer belongs to the neoadjuvant therapy type were treated with 8 chemotherapy, and first blood collection was performed at the time of diagnosis of breast cancer. Breast cancer belongs to metastatic therapy type, the first blood collection was performed at the time of diagnosis, and the last blood collection was performed at the completion of treatment.