Breast Cancer: Targets and Therapy

Dovepress
Taylor & Francis Group

Open Access Full Text Article

ORIGINAL RESEARCH

# A Retrospective Analysis of a Serum Multi-Biomarker (CA153, TSGF, and CYFRA 21-1) Logistic Regression Model for Predicting Pathologic Complete Response to Neoadjuvant Chemotherapy in Breast Cancer

Limei He[1,*], Weihong Zhang[2,*], Libin Huang[1,*], Xi Lin[1], Weiming Huang[1]

[1]Department of Medical Oncology, The First Hospital of Putian, Teaching Hospital, Fujian Medical University, Putian, Fujian Province, 351100, People's Republic of China; [2]Department of Gastroenterological Surgery Unit 1, The First Hospital of Putian, Teaching Hospital, Fujian Medical University, Putian, Fujian Province, 351100, People's Republic of China

*These authors contributed equally to this work

Correspondence: Weiming Huang, Department of Medical Oncology, The First Hospital of Putian, Teaching Hospital, Fujian Medical University, 449 Nanmen West Road, Chengxiang District, Putian, Fujian Province, 351100, People's Republic of China, Email 13959590295@139.com



**Background:** Accurately predicting pathologic complete response (pCR) to neoadjuvant chemotherapy (NAC) in breast cancer remains challenging when relying on single biomarkers. We aimed to establish a composite serum model integrating CA153, tumor-supplied group of factors (TSGF), and CYFRA 21–1 to enhance predictive performance.

**Methods:** This retrospective study included 258 breast cancer patients who received NAC at The First Hospital of Putian between January 2021 and December 2022. Eligible patients had histologically confirmed invasive breast cancer and complete clinical, pathological, and biomarker data. Patients with distant metastasis or incomplete serum data were excluded. Serum biomarkers (CA153, TSGF, and CYFRA 21–1) were measured before NAC. Univariate and multivariate logistic regression analyses were conducted to identify independent predictors of pCR, and receiver operating characteristic (ROC) curves were used to evaluate model performance.

**Results:** Serum CA153, TSGF, and CYFRA 21–1 levels were significantly lower in patients achieving pCR than in non-pCR patients (CA153: 72.10 vs. 103.82 U/mL; TSGF: 129.77 vs. 188.12 U/mL; CYFRA 21–1: 10.16 vs. 17.05 ng/mL; all $P < 0.001$). Moderate positive correlations were observed among the three markers. Multivariate logistic regression confirmed CA153 (OR = 1.185), TSGF (OR = 1.062), and CYFRA 21–1 (OR = 1.395) as independent predictors of non-pCR. The composite serum model demonstrated excellent discrimination (AUC = 0.967, 95% CI: 0.949–0.985), with high sensitivity (0.980) and negative predictive value (0.968), outperforming each biomarker alone.

**Conclusion:** The CA153–TSGF–CYFRA 21–1 serum composite model provides a simple, accurate, and non-invasive approach for predicting NAC response in breast cancer, with potential to support individualized treatment strategies.

**Keywords:** breast cancer, pathological complete response, CA153, tumor supplied group of factors, CYFRA 21-1

## Introduction

Breast cancer remains the most common malignancy among women worldwide, and neoadjuvant chemotherapy (NAC) plays a critical role in reducing tumor burden, improving breast conservation rates, and evaluating chemosensitivity before surgery.[1–3] Achieving a pathologic complete response (pCR) after NAC is a key prognostic indicator associated with improved long-term survival. However, not all patients benefit equally from NAC, and accurate prediction of pCR remains a major clinical challenge.[3,4]

https://doi.org/10.2147/BCTT.S567301
Received: 13 September 2025
Accepted: 6 March 2026
Published: 25 March 2026

Current imaging and tissue-based predictors have limited accuracy, whereas serum biomarkers such as CA153, TSGF, and CYFRA 21–1 offer a noninvasive and cost-effective means to monitor tumor burden and chemotherapy response. While some studies have explored the relationship between clinical and molecular features and the response to neoadjuvant chemotherapy and prognosis in breast cancer patients, existing predictive models often lack consistency and reliability. The use of traditional clinical features, histological parameters, and molecular markers alone has been shown to have certain limitations in predicting outcomes.[5,6] Liscia et al[7] investigated the use of CYFRA 21–1 in predicting axillary lymph node metastasis in breast cancer, suggesting its potential to clearly separate positive samples from negative ones. Fisher et al[8] employed machine learning to predict the response to neoadjuvant therapy in triple-negative breast cancer patients, but due to the relatively poor interpretability of machine learning models, further research is needed for clinical application. Additionally, these methods are expensive, complex to operate, and have low repeatability, limiting their clinical utility. Therefore, rapidly, accurately, conveniently, and economically predicting pCR in NAC patients remains a key issue in clinical practice. Consequently, the search for interpretable and powerful biomarkers or predictive models to help identify breast cancer patients who may benefit from neoadjuvant chemotherapy has become urgent.

In this study, we explored the potential value of CA153 - Tumor Supplied Group of Factors - CYFRA 21–1 biomarkers in predicting pCR in breast cancer patients undergoing neoadjuvant chemotherapy. In contrast to previous studies, we combined multiple serum markers, which have previously shown potential predictive value in tumor diagnosis and prognostic assessment, to enhance the predictive ability of the model through their joint utilization.

In this retrospective study, we aimed to establish and validate a serum multi-biomarker logistic regression model incorporating CA153, TSGF, and CYFRA 21–1 to predict pCR after NAC in breast cancer patients. We hypothesized that the combined model would outperform single biomarkers and provide a practical, noninvasive tool to guide individualized treatment decisions.

# Materials and Methods

## Patient Selection

This study included 258 patients with breast cancer who were treated at The First Hospital of Putian, Teaching Hospital, Fujian Medical University between January 2021 and December 2022 consecutively. All histologically confirmed breast cancer cases scheduled for neoadjuvant chemotherapy were eligible. In neoadjuvant chemotherapy, all patients received chemotherapy based on paclitaxel and/or anthracycline. Trastuzumab was added to the treatment regimen for HER2-positive patients. Tumor size was assessed by physical examination and ultrasound at each cycle, with MRI evaluation every two cycles. Pathological response was assessed on surgical specimens after completion of NAC according to institutional pathology practice and guideline-based criteria. Two experienced breast pathologists independently reviewed the specimens. Any discrepant interpretation was resolved by consensus (with senior pathologist adjudication when required). Pathologic complete response (pCR) was defined as ypT0/is ypN0 (no residual invasive carcinoma in the breast and axillary lymph nodes, with residual DCIS permitted). All patients were strictly selected based on inclusion and exclusion criteria: (1) unilateral primary breast cancer confirmed by biopsy; (2) age 18 years and above; (3) clinical stages I, II, and III; (4) meeting the requirements of the 2019 National Comprehensive Cancer Network guidelines. Exclusion criteria included: (1) previous receipt of other treatments; (2) incomplete pathological results or clinical data; (3) patients not receiving standardized chemotherapy or surgery; (4) incomplete follow-up information. Finally, 258 patients were included. All pathological and clinical data of the patients were obtained from electronic medical records by two clinical teams and verified by a third-party team. All clinical factors and laboratory indicators were tested within one week prior to the initiation of NAC. This retrospective study was approved by the Ethics Committee of The First Hospital of Putian, Teaching Hospital, Fujian Medical University, Helsinki Declaration. All patients provided informed consent. Comprehensive laboratory and imaging examinations were performed for all patients. Two experienced breast pathologists independently examined the specimens and determined pCR based on hematoxylin and eosin-stained slides. Pathologic complete response (pCR) was defined as the absence of residual invasive carcinoma in the breast and axillary lymph nodes (ypT0/is ypN0), allowing for residual ductal carcinoma in situ (DCIS). This study was reported in

accordance with the TRIPOD (Transparent Reporting of a multivariable prediction model for Individual Prognosis or Diagnosis) statement (Supplementary Table 1).

## Measurement of Plasma CA153, Tumor Supplied Group of Factors and CYFRA 21-1 Levels

CA153, also known as carbohydrate antigen 153 or breast cancer-associated antigen (BRCA), is a glycoprotein antigen. It is primarily produced by breast cancer cells but can also occur in other cancers such as ovarian cancer and colon cancer. Levels of CA153 are typically assessed through serum testing, as it can be detected in the blood of cancer patients. Tumor Supplied Group of Factors (TSGF) is a group of tumor-associated biomarkers that includes various molecules such as cytokines, growth factors, and cell adhesion molecules. These factors are mainly produced by tumor cells and are involved in tumor growth, spread, and metastasis through different pathways. Detection of TSGF can be performed using serum, urine, or tissue samples to assess tumor activity and clinical status. CYFRA 21–1 is a cytokeratin fragment produced mainly by epithelial cells, particularly as a result of cancer-induced epithelial cell damage. It is primarily detected in various cancers such as non-small cell lung cancer and can also be used for monitoring esophageal cancer, head and neck cancer, etc. Levels of CYFRA 21–1 are typically assessed through serum testing and can be used for assisting in cancer diagnosis, prognosis assessment, and treatment monitoring.

In this study, all three markers were assessed through automated blood testing conducted after morning blood draw.

## Clinical Data Collection

Clinical data on breast cancer patients meeting inclusion and exclusion criteria were collected from the First Hospital of Putian, Teaching Hospital, Fujian Medical University medical record system, including: (1) Basic hospitalization data: gender, age, clinical stage, tumor size, lymph node metastasis, NAC (neoadjuvant chemotherapy) regimen, Breast surgery, Therapeutic effect, etc. (2) Serological results within one week before NAC: CA153, Tumor Supplied Group of Factors, and CYFRA 21–1.Clinical and laboratory data were retrospectively extracted from the hospital's electronic medical record system using a standardized data abstraction form. Two investigators independently verified data entries, and missing or ambiguous values were cross-checked with the pathology and laboratory databases to ensure completeness and accuracy. Multicollinearity was assessed using Variance Inflation Factor (VIF), with all variables showing VIF < 2.0, indicating no significant multicollinearity.

Clinical covariates, including estrogen receptor (ER), progesterone receptor (PR), human epidermal growth factor receptor-2 (HER2), Ki-67 index, histologic grade, and chemotherapy regimen, were included based on established associations with neoadjuvant chemotherapy response and prognosis. ER, PR, and HER2 statuses were determined by immunohistochemistry according to ASCO/CAP guidelines, and Ki-67 ≥20% was defined as high proliferation.

Clinical stage (cTNM stage I–III) was assigned before initiation of NAC using the AJCC [8th edition] staging system based on pretreatment clinical and imaging assessment, including physical examination, breast and axillary ultrasound, and breast MRI (with additional imaging examinations when clinically indicated). Clinical staging was determined by the treating breast multidisciplinary team (or by two breast oncologists/radiologists, according to actual practice).

## Data Analysis

Continuous variables with normal distribution were assessed using t-tests, while those not conforming to normal distribution were analyzed using the Wilcoxon rank-sum test. Categorical variables were examined using either the chi-square test or Fisher's exact test. Initially, univariate logistic regression analysis was conducted to identify variables with a significance level of $P < 0.05$. Subsequently, these variables were included in a multivariate logistic regression analysis to identify risk factors and construct a composite inflammation index model. The optimal cutoff value for predicted probability was determined based on sensitivity and specificity, selecting the threshold corresponding to the maximum Youden index (sensitivity + specificity − 1) derived from the ROC curve within the same dataset. Receiver operating characteristic (ROC) curves were then utilized to evaluate the discriminative ability of individual serum indices and the composite inflammation model. A forced-entry multivariable logistic regression model was used, with all predictors

entered simultaneously based on prior clinical relevance rather than stepwise selection. Linearity of continuous predictors with the log-odds of pCR was evaluated using restricted cubic splines. The spline plots did not reveal meaningful departures from linearity; therefore, predictors were retained in linear form.

Missing data represented less than 3% of all variables and were handled using complete-case analysis, as the pattern of missingness was random and unlikely to bias results. Data analysis was performed using SPSS 25.0 (IBM, Armonk, New York, USA), with statistical significance defined as a two-tailed P-value < 0.05. Graphs were generated using R language (version 4.0.5) and GraphPad Prism (version 8.0). Sample size adequacy was evaluated using events per variable (EPV) analysis, ensuring ≥10 EPV in the multivariable logistic regression. Internal validation was performed through bootstrapping with 1000 resamples to test model stability and minimize overfitting.

# Results

## Baseline Information of Patients with Breast Cancer After Neoadjuvant Chemotherapy Stratified into pCR and Non-pCR Groups

The study included a total of 258 patients after exclusion criteria were applied (see Figure 1), with 150 patients in the pCR group and 108 patients in the Non-pCR group. Regarding basic demographic characteristics, there were no statistically significant differences in age and BMI between the two groups (P=0.707 and 0.893, respectively), with a similar proportion of patients over and under 50 years old in both groups. In terms of tumor characteristics, the pCR group had smaller tumor diameters, with 70.7% classified as T1-T2, compared to 50.0% in the Non-pCR group, showing a statistically significant difference. Additionally, there was a statistical difference in cTNM stage between the two groups, with more patients in the Non-pCR group classified as stage III (67.3%). Although the prevalence of documented lymph node metastasis in Table 1 was relatively low, a substantial proportion of patients were classified as stage III because stage III classification may also be driven by primary tumor extent (eg., T3/T4 disease), not nodal status alone. This pattern can be seen in NAC cohorts enriched for locally advanced tumors. The positivity rates of certain tumor molecular markers were comparable between the two groups, with no statistically significant differences in ER, PR, and HER2 positivity rates. Among pCR patients, the proportions of positive cases for these three markers were 76.7%, 64.7%, and 81.3%, respectively. Regarding preoperative serum markers, the serum levels of CA153, TSGF, and CYFRA 21–1 in Non-pCR patients were 103.82±22.49 U/mL, 188.12±36.63 U/mL, and 17.05±5.43, respectively, while in pCR patients, they were72.10±15.95 U/mL, 129.77±28.90 U/mL, and 10.16±3.88 ng/mL, respectively. Visual representation using bar graphs showed statistically significant differences between the two groups for all three markers (P<0.001) (see Figure 2). For detailed data, please refer to Table 1. The variable "lymph node metastasis" in Table 1 refers to

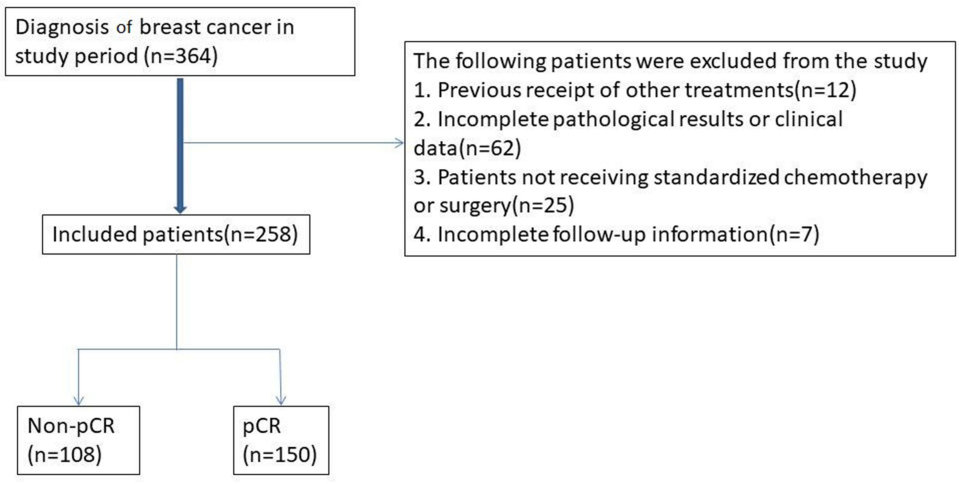


**Figure 1** Study flowchart illustrating patient inclusion and exclusion criteria. A total of 312 patients receiving neoadjuvant chemotherapy for breast cancer were screened. After applying predefined inclusion and exclusion criteria, 258 eligible patients were included in the final analysis and categorized into pCR (n = 150) and non-pCR (n = 108) groups.

 https://doi.org/10.2147/BCTT.S567301 

**Table 1** Demographic and Clinical Characteristics of Breast Cancer Patients with pCR Versus Non-pCR (N = 258)

| | | pCR Group (n=150) | Non-pCR Group (n=108) | P-value‡ |
|---|---|---|---|---|
| Age | | | | 0.707 |
| | <50 y | 76(50.7) | 52(48.1) | |
| | ≥50 y | 74(49.3) | 56(51.9) | |
| BMI | | | | 0.893 |
| | < 24 | 96(64.0) | 70(64.8) | |
| | ≥ 24 | 54(36.0) | 38(35.2) | |
| Tumor size | | | | 0.001 |
| | T1-T2 | 106(70.7) | 54(50.0) | |
| | T3-T4 | 44(29.3) | 54(50.0) | |
| Lymph node metastases | | | | 0.901 |
| | Negative | 134(89.3) | 97(89.8) | |
| | Positive | 16(10.7) | 11(10.2) | |
| cTNM stage | | | | 0.002 |
| | I–II | 80(53.3) | 32(32.7) | |
| | III | 70(46.7) | 66(67.3) | |
| NAC regimen | | | | 0.915 |
| | Non-anthracycline based | 123(82.0) | 88(81.5) | |
| | Anthracycline based | 27(18.0) | 20(18.5) | |
| Chemotherapy cycle | | | | 0.805 |
| | < 6 | 48(32.0) | 33(30.6) | |
| | ≥ 6 | 102(68.0) | 75(69.4) | |
| Breast surgery | | | | 0.790 |
| | Breast-conserving surgery | 58(38.7) | 40(37.0) | |
| | Mastectomy | 92(61.3) | 68(63.0) | |
| Family history | | | | 0.680 |
| | No | 105(70.0) | 73(67.6) | |
| | Yes | 45(30.0) | 35(32.4) | |
| Estrogen receptor | | | | 0.076 |
| | Negative | 35(23.3) | 36(33.3) | |
| | Positive | 115(76.7) | 72(66.7) | |
| Progesterone receptor | | | | 0.246 |
| | Negative | 53(35.3) | 46(42.6) | |
| | Positive | 97(64.7) | 62(57.4) | |
| HER2 status | | | | 0.220 |
| | Negative | 38(25.3) | 14(13.0) | |
| | Positive | 112(74.7) | 94(87.0) | |
| Ki-67 proliferation index | | | | 0.015 |
| | ≤ 20 | 38(25.3) | 25(23.1) | |
| | > 20 | 112(74.7) | 83(76.9) | |
| Indicators | | | | |
| CA153 | U/mL | 72.10±15.95 | 103.82±22.49 | <0.001 |
| TSGF | U/mL | 129.77±28.90 | 188.12±36.63 | <0.001 |
| CYFRA 21–1 | ng/mL | 10.16±3.88 | 17.05±5.43 | <0.001 |

**Notes**: The values in parentheses are percentages unless indicated otherwise. ‡ The $\chi 2$ test with Yates' correction was used for categorical data, and the *t*-test was used for measures conforming to a normal distribution; Lymph node metastasis denotes pretreatment clinical axillary nodal involvement assessed by imaging and/or needle biopsy when available.

**Abbreviations**: pCR, pathological complete response; BMI, Body Mass Index; TNM, Tumor, Node, Metastasis; NAC regimen, Neoadjuvant Chemotherapy regimen; HER2 status, Human Epidermal Growth Factor Receptor 2 status; TSGF, Tumor-Specific Growth Factor; CYFRA 21–1, Cytokeratin-19 Fragment 21–1.


https://doi.org/10.2147/BCTT.S567301

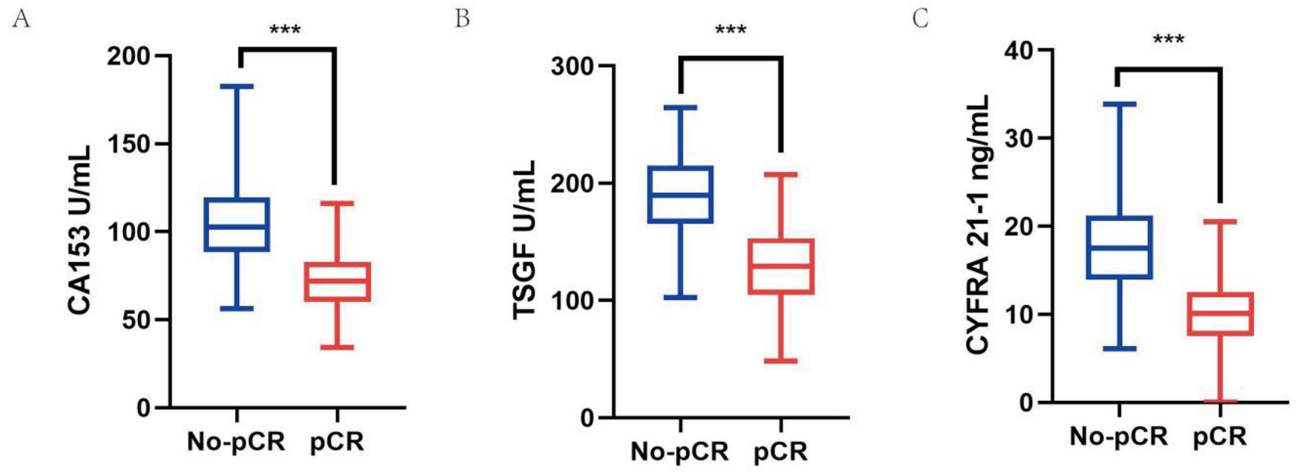


**Figure 2 (A-C)** Comparison of serum biomarker levels (CA153, TSGF, and CYFRA 21–1) between pCR and non-pCR groups. Box-and-whisker plots display the distribution of each biomarker. Patients achieving pCR showed significantly lower median levels of all three markers compared with non-pCR patients (all P < 0.001), suggesting their potential value in predicting chemotherapy response.*** represents P< 0.001.

[pretreatment clinically suspected/biopsy-confirmed axillary nodal metastasis OR postoperative pathological nodal metastasis]; this has been clarified in the table footnote. Detailed neoadjuvant treatment patterns are summarized in Supplementary Table 2.

## Correlation Analysis Between the Scores of the Three Indicators

By analyzing the two-by-two Pearson coefficients between the values of the three preoperative serum markers, the results showed that the Pearson coefficient was 0.512 (P<0.001) between CA153 and TSGF, 0.449 (P<0.001) between CA153 and CYFRA 21–1, and 0.428 (P<0.001) between TSGF and CYFRA 21–1. There was a positive correlation between the three markers (see Table 2).

## Analysis of Risk Factors Affecting Pathologic Complete Response Using Univariate and Multivariate Logistic Regression and Construction of an CTC (CA153 - Tumor Supplied Group of Factors- CYFRA 21-1) Serum Composite Model

Variables with P<0.05 in univariate logistic regression were included in the multivariate logistic regression model. Significant variables in univariate analysis included Tumor size, cTNM stage, Lymph node metastases, CA153, TSGF, and CYFRA 21–1. The final multivariate logistic regression results showed that Tumor size as T3-T4 (OR= 1.158, 95% CI: 1.076–1.339, P<0.001 vs T1-T2), Lymph node metastases (OR = 1.295, 95% CI: 1.077–1.582, P = 0.014 vs. Negative), cTNM stage III (OR= 1.304, 95% CI: 1.123–1.591, P = 0.008 vs I–II), CA153 (OR= 1.185, 95% CI: 1.091–1.310, P <0.001 per U), TSGF (OR= 1.062, 95% CI: 1.028–1.117, P <0.001 per U), and CYFRA 21–1 (OR= 1.395, 95% CI: 1.137–1.604, P <0.001 per 1 ng) were significantly associated with Non-pCR (see Table 3). Based on logistic regression results, an CTC model was constructed with the formula = −19.509+0.067×CA153 (U/mL)

**Table 2** Pearson Correlation Analysis

| | CA153 | TSGF | CYFRA 21–1 |
|---|---|---|---|
| CA153 | 1 | | |
| TSGF | 0.512(P<0.001) | 1 | |
| CYFRA 21-1 | 0.449(P<0.001) | 0.428(P<0.001) | 1 |

**Abbreviations**: TSGF, Tumor-Specific Growth Factor; CYFRA 21–1, Cytokeratin-19 Fragment 21–1.


https://doi.org/10.2147/BCTT.S567301

**Table 3** Univariate and Multivariate Analysis of Risk Factors for Non-pCR Patients

| Variables | | Univariate Analysis | | | Multivariate Analysis | | |
|---|---|---|---|---|---|---|---|
| | | P | OR | 95% CI | P | OR | 95% CI |
| Age | >50 y/≤50 y | 0.554 | 1.025 | 0.912–1.169 | | | |
| BMI | >25/≤25 | 0.134 | 1.199 | 0.813–1.455 | | | |
| Tumor size | T3-T4/T1-T2 | <0.001 | 1.266 | 1.091–1.428 | <0.001 | 1.158 | 1.076–1.339 |
| Lymph node metastases | P/N | 0.033 | 1.305 | 1.083–1.644 | 0.014 | 1.295 | 1.077–1.582 |
| cTNM stage | III/I–II | 0.012 | 1.246 | 1.046–1.500 | 0.008 | 1.304 | 1.123–1.591 |
| NAC regimen | Anthracycline based/Non-anthracycline based | 0.246 | 1.210 | 0.791–1.399 | | | |
| Chemotherapy cycle | ≥ 6/<6 | 0.822 | 1.052 | 0.762–1.448 | | | |
| Breast surgery | Mastectomy/Breast-conserving surgery | 0.612 | 0.961 | 0.728–1.414 | | | |
| Family history | Yes/No | 0.228 | 1.043 | 0.792–1.511 | | | |
| Estrogen receptor | P/N | 0.366 | 1.310 | 0.754–1.720 | | | |
| Progesterone receptor | P/N | 0.431 | 1.255 | 0.688–1.489 | | | |
| HER2 status | P/N | 0.297 | 1.442 | 0.699–1.802 | | | |
| Ki-67 proliferation index | > 20/≤20 | 0.196 | 1.192 | 0.801–1.561 | | | |
| CA153 | Per 10 U | <0.001 | 1.159 | 1.087–1.236 | <0.001 | 1.185 | 1.091–1.310 |
| TSGF | Per 10 U | <0.001 | 1.054 | 1.024–1.083 | <0.001 | 1.062 | 1.028–1.117 |
| CYFRA 21-1 | Per 10 ng | <0.001 | 1.422 | 1.192–1.695 | <0.001 | 1.395 | 1.137–1.604 |

**Abbreviations**: pCR, pathological complete response; BMI, Body Mass Index; TNM, Tumor, Node, Metastasis; NAC regimen, Neoadjuvant Chemotherapy regimen; HER2 status, Human Epidermal Growth Factor Receptor 2 status; TSGF, Tumor-Specific Growth Factor; CYFRA 21–1, Cytokeratin-19 Fragment 21–1.

+0.056×TSGF (U/mL)+0.347×CYFRA 21–1 (ng/mL). The full model specification with standardized predictor scaling, including regression coefficients (β), standard errors, and odds ratios, is provided in Supplementary Table 3.

## Plotting Receiver Operating Curves (ROC) and Decision Curve Analysis for Individual Metrics as Well as Combined Models and the Effect of Each Metric

ROC curves were plotted for CA153, TSGF, CYFRA 21–1, and the combined CTC model, with 1-specificity on the horizontal axis and sensitivity on the vertical axis. The area under the curve was 0.860 (0.812–0.909) for CA153, 0.906 (0.870–0.942) for TSGF, 0.878 (0.835–0.920) for CYFRA 21–1, and 0.967 (0.949–0.985) for the combined serum CTC model (see Figure 3). The optimal cutoff values for each marker were determined using the maximum Youden index. The optimal cutoff values for CA153, TSGF, and CYFRA 21–1 were 84.4 U/mL, 159.54 U/mL, and 12.75 ng/mL, respectively. The sensitivity, specificity, positive predictive value, and negative predictive value were as follows: for CA153, 0.787, 0.824, 0.861, and 0.736; for TSGF, 0.840, 0.833, 0.875, and 0.789; for CYFRA 21–1, 0.787, 0.852, 0.881, and 0.742; and for the combined CTC, 0.980, 0.852, 0.902, and 0.968 (see Table 4). Internal validation was performed using bootstrapping with 1,000 resamples. Model calibration was assessed using the Hosmer–Lemeshow test and calibration plots (Supplementary Figure 1; Supplementary Table 4). Decision curve analysis demonstrated that the CTC composite model provided a higher net benefit than the treat-all and treat-none strategies across a wide range of clinically relevant threshold probabilities, supporting its potential clinical utility for pCR prediction (Supplementary Figure 2).

## Exploratory Subtype-Stratified Model Performance

As recommended by the reviewer, we further assessed model performance across major molecular subtypes. Subtype-stratified results are provided in Supplementary Table 5. The model demonstrated high discriminative ability across subgroups, with the highest AUC observed in HER2-positive disease, followed by HR+/HER2− and TNBC. Because subgroup sample sizes—particularly in TNBC—were limited, these analyses are considered exploratory and hypothesis-generating.

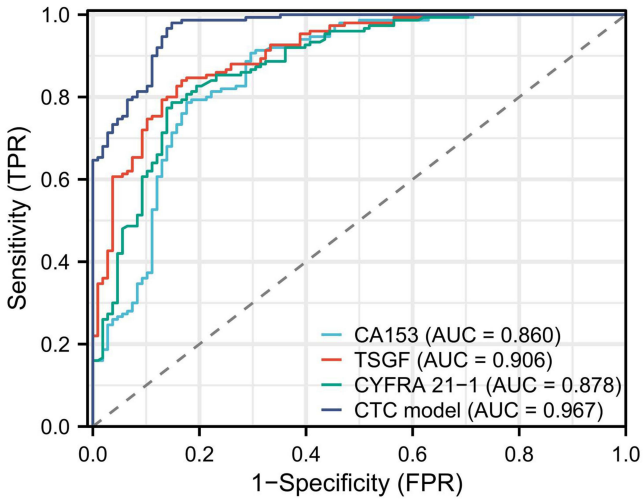


**Figure 3** Receiver operating characteristic (ROC) curves comparing the predictive performance of individual biomarkers and the combined logistic regression model. The combined CA153–TSGF–CYFRA 21–1 model demonstrated superior discriminative ability (AUC = 0.97, 95% CI: 0.95–0.99) compared with individual markers. The figure also illustrates the optimal cutoff point determined by the Youden index for maximizing sensitivity and specificity.

## Discussion

Breast cancer comprises various molecular subtypes and is a highly heterogeneous solid tumor, where neoadjuvant therapy yields different treatment responses.[9,10] Therefore, predicting the pathological complete response (PCR) of breast cancer patients to neoadjuvant chemotherapy (NAC) is crucial for optimizing treatment strategies and improving clinical outcomes.[3] In this study, we investigated the predictive value of a serum biomarker model composed of CA153, tumor-specific growth factor (TSGF), and CYFRA 21–1 for breast cancer patients undergoing NAC. Our findings suggest that this composite biomarker model is an effective tool for predicting treatment response and guiding personalized treatment decisions.

The combination of CA153, TSGF, and CYFRA 21–1 was selected based on their complementary biological relevance to tumor activity and chemotherapy response. CA153, a mucin-1–related glycoprotein, reflects epithelial tumor burden and proliferative activity. TSGF is associated with tumor-induced angiogenesis and metabolic reprogramming, while CYFRA 21–1 represents cytokeratin-19 fragments released during apoptosis or necrosis of epithelial tumor cells. Together, these markers provide an integrative assessment of tumor proliferation, metabolic status, and cellular turnover, which may correlate with chemosensitivity and residual disease after neoadjuvant therapy. The primary objective of our research was to address the existing challenges of predicting PCR using a single biomarker by exploring the synergistic effects of multiple serum markers. Previous studies relying solely on clinical characteristics, histological parameters, or single biomarkers have shown limitations in accurately predicting treatment response. Xu et al[11] proposed a scoring system based on breast tumor and local lymph nodes, suggested in 2007, to evaluate residual lesions after NAC, which they found to be indicative of post-neoadjuvant chemotherapy outcomes in breast cancer. However, this index is cumbersome to use and outdated. Ilie et al[12] constructed a model using pathological and immunohistochemical data to collectively predict prognosis after neoadjuvant chemotherapy, achieving good results. In our study, by combining

**Table 4** Predictive Value of Independent Predictors and Combination Predictors for Non-pCR and pCR

| Predictors | Optimal Cut-off Value | AUC | Accuracy | Sensitivity | Specificity | PPV | NPV |
|---|---|---|---|---|---|---|---|
| CA153 | 84.4 | 0.860(0.812–0.909) | 0.802 | 0.787 | 0.824 | 0.861 | 0.736 |
| TSGF | 159.54 | 0.906(0.870–0.942) | 0.837 | 0.840 | 0.833 | 0.875 | 0.789 |
| CYFRA 21-1 | 12.75 | 0.878(0.835–0.921) | 0.813 | 0.787 | 0.852 | 0.881 | 0.742 |
| CTC model | | 0.967(0.949–0.985) | 0.926 | 0.980 | 0.852 | 0.902 | 0.968 |

**Abbreviations**: TSGF, Tumor-Specific Growth Factor; CYFRA 21–1, Cytokeratin-19 Fragment 21–1; AUC, areas under the receiver operating characteristic curve; PPV, positive predictive value; NPV, negative predictive value.


https://doi.org/10.2147/BCTT.S567301

CA153, TSGF, and CYFRA 21–1, which hold promise as prognostic indicators in various cancer types, we aimed to enhance the predictive accuracy of the model. This multifaceted approach enabled us to capture the complexity of tumor biology and better reflect the heterogeneity of breast cancer. Previous studies have demonstrated that serum biomarkers, such as CA153, CYFRA 21–1, and TSGF, are associated with tumor burden, treatment response, and prognosis in breast cancer.[13–15] Our findings are consistent with these observations and further suggest that combining multiple serum markers may enhance predictive accuracy for pathological complete response (pCR) following neoadjuvant chemotherapy.[16,17] This integrated approach aligns with recent efforts to establish multi-marker models for individualized treatment prediction.

Our results indicate that the combined serum biomarker model significantly predicts the likelihood of PCR in breast cancer patients receiving NAC. CA153 is a major glycoprotein antigen produced by breast cancer cells, associated with tumor progression and metastasis. Elevated serum levels of CA153 are associated with advanced disease and poor prognosis in breast cancer patients. Ren et al[18] found that CA153 can predict recurrence in breast cancer patients, and a multicenter study by Luo et al[19] also suggested good efficacy of CA153 in routine physical examinations and early breast cancer screening. TSGF represents a group of tumor-related biomarkers involving various aspects of tumor biology, including growth, angiogenesis, and immune evasion. By combining markers from this group, we aimed to capture the dynamic interactions between tumor cells and the tumor microenvironment. Wang et al[20] found that TSGF in nipple discharge could serve as a novel biomarker for breast cancer diagnosis and prognosis. CYFRA 21–1, a cytokeratin fragment indicating epithelial cell damage, provides further insights into tumor cell turnover and response to treatment. Its elevation is associated with tumor invasiveness and resistance to treatment in various cancer types. Yooh et al[21] found that CYFRA 21–1 plays an important role in preoperative diagnosis in breast cancer patients' axillary lymph fluid. However, in our study, we found that combining these markers further enhances the effectiveness of predicting patients' response to neoadjuvant therapy.

The strength of our predictive model lies in its simplicity, reliability, and clinical utility. Unlike complex machine learning algorithms that may lack interpretability and practicality in clinical settings,[22,23] our model offers a straightforward approach to predicting treatment response. By leveraging routinely available serum markers and established cutoff values, our model can be easily implemented in clinical practice without the need for specialized equipment or expertise. This accessibility is particularly advantageous in resource-limited settings where complex diagnostic tools may be impractical to use.

Furthermore, our research underscores the importance of personalized medicine in breast cancer management. By stratifying patients based on their likelihood, our model enables clinicians to tailor treatment strategies for individual patients. Patients identified as high responders to NAC may benefit from intensified chemotherapy regimens or targeted therapies, while those with a lower predicted PCR likelihood may require alternative treatment approaches, such as surgical intervention or participation in clinical trials. This personalized approach maximizes the likelihood of treatment success while minimizing unnecessary toxicity and treatment-related morbidity.The study population was a treatment-selected neoadjuvant chemotherapy (NAC) cohort from a tertiary referral center rather than an unselected breast cancer population; therefore, biological subtypes associated with NAC indication (including HER2-positive disease) were overrepresented. Accordingly, the subtype distribution in this cohort does not reflect the epidemiologic distribution of breast cancer in the general population, and external generalizability should be interpreted with caution.

Although our study yields promising results, several limitations should be acknowledged. Firstly, the retrospective nature of our analysis may introduce inherent biases and confounding variables, and validation in independent cohorts is necessary to confirm the robustness and generalizability of our findings. Secondly, the relatively limited sample size may reduce statistical power to detect small but clinically meaningful effects. Moreover, subgroup analyses stratified by molecular subtype (HR+/HER2–, HER2+, TNBC) were not performed due to the small number of cases within each category, which may restrict the applicability of the model across heterogeneous breast cancer populations. Thirdly, we acknowledge an important limitation regarding endpoint definition. In this study, pCR was defined using the commonly adopted neoadjuvant breast cancer endpoint ypT0/is ypN0, which permits residual DCIS. However, there is long-standing controversy regarding whether residual DCIS should be considered equivalent to "true" pCR and whether DCIS response to NAC parallels invasive tumor response. Therefore, our binary pCR/non-pCR endpoint may combine

biologically heterogeneous residual disease patterns. In our cohort, among pCR-defined cases (ypT0/is ypN0), a subset of patients had no residual invasive disease with residual DCIS present, whereas others had no residual invasive or in situ disease; however, the sample size was insufficient for separate endpoint modeling. Because of the retrospective design and sample size, we did not perform a separate analysis of residual DCIS-only cases. Future studies with larger cohorts should evaluate whether model performance differs across stricter endpoint definitions (eg., ypT0 ypN0 vs ypT0/is ypN0). In addition, the serum biomarkers included in our model—CA153, TSGF, and CYFRA 21–1—are relatively nonspecific, and their mechanistic linkage to chemosensitivity remains indirect. Future multicenter studies with larger, molecularly and mechanistically stratified cohorts are warranted to validate these findings and explore biological pathways underlying biomarker responsiveness. Finally, integrating additional biomarkers, clinical parameters, and novel modalities such as imaging or genomic profiling may further enhance predictive accuracy and provide deeper insights into tumor biology and treatment response.

## Conclusion

In summary, this study developed a serum biomarker–based predictive model integrating CA153, TSGF, and CYFRA 21–1 for assessing pathological complete response (pCR) following neoadjuvant chemotherapy in breast cancer. The model demonstrated good internal discrimination and calibration, suggesting potential utility as a non-invasive tool for individualized response prediction. However, given the retrospective single-center design and limited sample size, these findings should be interpreted cautiously. Prospective, multicenter studies with external validation are warranted to confirm the robustness and generalizability of this biomarker-based model before its application in clinical decision-making.

## Data Sharing Statement

The datasets used and analyzed during the current study are available from the corresponding author on reasonable request.

## Ethics and Consent to Participate

This study was conducted in accordance with the principles of the Declaration of Helsinki. Ethical approval was obtained from the Ethics Committee of The First Hospital of Putian, Teaching Hospital, Fujian Medical University. Informed consent was obtained from all participants prior to inclusion in the study.

## Acknowledgments

Thanks to the nurses in the department for their help with the project.

## Author Contributions

All authors made a significant contribution to the work reported, whether that is in the conception, study design, execution, acquisition of data, analysis and interpretation, or in all these areas; took part in drafting, revising or critically reviewing the article; gave final approval of the version to be published; have agreed on the journal to which the article has been submitted; and agree to be accountable for all aspects of the work.

## Funding

This study did not receive any funding in any form.

## Disclosure

The authors have no conflicts of interest to declare.

## References

1. Barzaman K, Moradi-Kalbolandi S, Hosseinzadeh A, et al. Breast cancer immunotherapy: current and novel approaches. *Int Immunopharmacol*. 2021;98:107886. doi:10.1016/j.intimp.2021.107886

2. Montemurro F, Nuzzolese I, Ponzone R. Neoadjuvant or adjuvant chemotherapy in early breast cancer? *Expert Opin Pharmacother*. 2020;21(9):1071–1082. doi:10.1080/14656566.2020.1746273
3. Wang H, Mao X. Evaluation of the efficacy of neoadjuvant chemotherapy for breast cancer. *Drug Des Devel Ther*. 2020;14:2423–2433. doi:10.2147/DDDT.S253961
4. Zhang J, Wu Q, Yin W, et al. Development and validation of a radiopathomic model for predicting pathologic complete response to neoadjuvant chemotherapy in breast cancer patients. *BMC Cancer*. 2023;23(1):431. doi:10.1186/s12885-023-10817-2
5. Keskin S, Muslumanoglu M, Saip P, et al. Clinical and pathological features of breast cancer associated with the pathological complete response to anthracycline-based neoadjuvant chemotherapy. *Oncology*. 2011;81(1):30–38. doi:10.1159/000330766
6. Wang KN, Meng YJ, Yu Y, et al. Predicting pathological complete response after neoadjuvant chemotherapy: a nomogram combining clinical features and ultrasound semantics in patients with invasive breast cancer. *Front Oncol*. 2023;13:1117538. doi:10.3389/fonc.2023.1117538
7. Liscia DS, Detoma P, Zanchetta M, et al. The use of CYFRA 21-1 for the detection of breast cancer axillary lymph node metastases in needle washouts of fine-needle aspiration biopsies. *Appl Immunohistochem Mol Morphol*. 2017;25(3):190–195. doi:10.1097/PAI.0000000000000287
8. Fisher TB, Saini G, Rekha TS, et al. Digital image analysis and machine learning-assisted prediction of neoadjuvant chemotherapy response in triple-negative breast cancer. *Breast Cancer Res*. 2024;26(1):12. doi:10.1186/s13058-023-01752-y
9. Gao JJ, Swain SM. Luminal A breast cancer and molecular assays: a review. *Oncologist*. 2018;23(5):556–565. doi:10.1634/theoncologist.2017-0535
10. Yin L, Duan JJ, Bian XW, Yu SC. Triple-negative breast cancer molecular subtyping and treatment progress. *Breast Cancer Res*. 2020;22(1):61. doi:10.1186/s13058-020-01296-5
11. Xu X, Zhao W, Liu C, et al. *et al*: the residual cancer burden index as a valid prognostic indicator in breast cancer after neoadjuvant chemotherapy. *BMC Cancer*. 2024;24(1):13. doi:10.1186/s12885-023-11719-z
12. Ilie SM, Briot N, Constatin G, et al. *et al*: pathologic and immunohistochemical prognostic markers in residual triple-negative breast cancer after neoadjuvant chemotherapy. *Front Oncol*. 2023;13:1309890. doi:10.3389/fonc.2023.1309890
13. Fang Z, Ren G, Ke S, et al. Serum metabolomic profiling for predicting therapeutic response and toxicity in breast cancer neoadjuvant chemotherapy: a retrospective longitudinal study. *Breast Cancer Res*. 2025;27(1):2. doi:10.1186/s13058-024-01956-w
14. Giovanella L, Ceriani L, Giardina G, Bardelli D, Tanzi F, Garancini S. Serum cytokeratin fragment 21.1 (CYFRA 21.1) as tumour marker for breast cancer: comparison with carbohydrate antigen 15.3 (CA 15.3) and carcinoembryonic antigen (CEA). *Clin Chem Lab Med*. 2002;40(3):298–303. doi:10.1515/CCLM.2002.047
15. Song X, Liang B, Wang C, Shi S. Clinical value of color Doppler ultrasound combined with serum CA153, CEA and TSGF detection in the diagnosis of breast cancer. *Exp Ther Med*. 2020;20(2):1822–1828. doi:10.3892/etm.2020.8868
16. Elsayed B, Alksas A, Shehata M, et al. Exploring neoadjuvant chemotherapy, predictive models, radiomic, and pathological markers in breast cancer: a comprehensive review. *Cancers*. 2023;15(21):5288. doi:10.3390/cancers15215288
17. van den Ende NS, Nguyen AH, Jager A, Kok M, Debets R, van Deurzen CHM. Triple-negative breast cancer and predictive markers of response to neoadjuvant chemotherapy: a systematic review. *Int J Mol Sci*. 2023;24(3):2969. doi:10.3390/ijms24032969
18. Ren Z, Yang J, Liang J, et al. Monitoring of postoperative neutrophil-to-lymphocyte ratio, D-dimer, and CA153 in: diagnostic value for recurrent and metastatic breast cancer. *Front Surg*. 2022;9:927491.
19. Luo J, Xiao J, Yang Y, Chen G, Hu D, Zeng J. Strategies for five tumour markers in the screening and diagnosis of female breast cancer. *Front Oncol*. 2022;12:1055855.
20. Wang G, Qin Y, Zhang J, et al. Nipple discharge of CA15-3, CA125, CEA and TSGF as a new biomarker panel for breast cancer. *Int J Mol Sci*. 2014;15(6):9546–9565. doi:10.3390/ijms15069546
21. Yoon JH, Han KH, Kim EK, et al. Fine-needle aspirates CYFRA 21-1 is a useful tumor marker for detecting axillary lymph node metastasis in breast cancer patients. *PLoS One*. 2013;8(2):e57248. doi:10.1371/journal.pone.0057248
22. Jiang T, Gradus JL, Rosellini AJ. Supervised machine learning: a brief primer. *Behav Ther*. 2020;51(5):675–687. doi:10.1016/j.beth.2020.05.002
23. Uddin S, Khan A, Hossain ME, Moni MA. Comparing different supervised machine learning algorithms for disease prediction. *BMC Med Inf Decis Making*. 2019;19(1):281. doi:10.1186/s12911-019-1004-8