Supplement Materials

Appendix A**:** Treatment Prediction Models using Supervised Machine Leaning algorithm-

Logistic Regression

Using a supervised machine learning algorithm based on logistic regression (Eq. 1 in main manuscript), we developed treatment prediction models to identify the best predictors and the best early treatment windows for predicting response. Our approach estimates the model parameters $\boldsymbol{\beta}_{n\,,}$ n=0, 1, …k, from training data selected from a set of predictor variables and response outcomes and then test each model using the corresponding test set of predictor variables to evaluate its AUC under the corresponding ROC.

The 38 patient data obtained from this study and an early data obtained from 22 patients using a similar prototype system were combined to increase the robustness of the models (Table 1S). The combined group has a total of 60 patients (mean age=50, range 24-74). Twenty-six patients were HER2+, 16 were triple-negative breast cancer (TNBC), and 18 were ER+/HER2- (n=16) or PR+/HER2- (n=2). Thirteen had stage 3 disease, 39– stage 2, and 8– stage 1. Based upon MP grade, 6 patients had no response (pNR) (MP1), 19 patients had a partial response (pPR), including 4 with a minor response (MP2) and 15 with an intermediate response (MP3), while 6 had a near-complete response (MP4) and 29 had a complete response (pCR) (MP5). Table 1s summaries patient and tumor characteristics, treatment response, and treatment regimens.

The predictors were evaluated based on the Spearman's rho correlation with the Miller-Payne grade. For tumor Nottingham score and tumor biomarkers, HER2 is a significant predictor of treatment response (P<0.001) and ER is marginally significant (P=0.075). For hemoglobin parameters, pre-treatment HbT, and fraction of HbT normalized to pre-treatment, %HbT, measured at the end of cycles 1-3 are significant predictors (P=0.003, P<0.001, P<0.001, P<0.001). For US measurements, pre-treatment maximum diameter US, and fraction of maximum diameter, %US, measured at the end of cycles 1-3 are significant predictors (P=0.008, P<0.001, P<0.001, P<0.001).

Two-thirds of the patients were used for training the logistic regression models and rest for testing. Each training set had 40 patients of 20 responders and 20 non-responders while the test set had the remaining 20 cases. Hyperparameter tuning was performed by 5-fold cross-validation on the training set. 50 random train/test split were used to train and test each prediction model and an average **testing** ROC curve was computed. The area under the ROC (AUC) was used as a performance measure to assess the model.

We also used the 50 AUC values to construct the 95% confidence interval for the mean AUC for each model, using binomial formula. These confidence intervals can provide summary information on comparisons of the different models in terms of their AUC values. For example, if model I has a higher mean AUC than model II, and if their corresponding confidence intervals do not overlap, then this is an indication that model I may have a higher prediction power than the model II in terms of the AUC criterion. However, this interpretation should be understood with the caution that the 50 values are not true random samples.

As shown in Table 2S, when the HER2 and ER status are included as predictors, the accuracy of

AUC is 0.799 (95% CI: 0.688-0.910). With the addition of pre-treatment HbT, the AUC is increased slightly to 0.814 (95% CI: 0.706-0.922). With an earlier predictor of %US or %HbT measured at the end of cycle 1 (EOC1) included, which is 2-3 weeks into the neoadjuvant therapy, AUCs are substantially increased to 0.878 (95% CI: 0.788-0.969) and 0.887 (95% CI: 0.799-0.975), respectively. The AUC can be further improved to 0.929 (95% CI:0.858-1.0) when both %US and %HbT measured at EOC1 are included as predictors. The best AUC = 0.958 (95% CI: 0.903-1.014) is achieved when %US EOC1 and %HbT EOC3 are selected as predictors, which is about 9 to 12 weeks into the neoadjuvant therapy. The AUC improvement of "HER2 and ER status, %US EOC1 and %HbT EOC3" is statistically significant than that of "HER2 and ER status, %US EOC1 and %HbT EOC1" (P=0.002) and both are statistically significant when compared with "HER2 and ER status and %US at EOC1" (P<0.001).

Also shown in the Table 2S, within the TNBC subtype, the accuracy of AUC in prediction is 0.487 (95% CI:0.349-0.626) which is about the level of chance. Pre-treatment US is not predictive with AUC of 0.464 (95% CI: 0.327-0.603), however, HbT prediction accuracy is AUC=0.704 (95% CI: 0.577-0.830). With the addition of the best earlier predictors, %HbT and %US measured at EOC1, AUC is substantially increased to 0.920 (95% CI: 0.845-0.995). The AUC can be further improved to 0.966 (95% CI: 0.916-1.016) when %HbT EOC3 replaces %HbT EOC1 regardless of tumor subtypes. The AUC improvement of "%US EOC1 and %HbT EOC3" is statistically significant than that of "%US EOC1 and %HbT EOC1" (P<0.001) and both are statistically significant when compared with %US at EOC1 (P<0.001). These results agree with data reported in the main body of this manuscript and suggest that our models are robust for predicting breast cancer neoadjuvant treatment.

There is a general rule of thumb proposed by Harrell that the training size should be one order larger than the number of predictors to reduce the chance of overfitting. Thus, we have kept the maximum predictors to three or four based on our training size of 40. We did not observe higher training AUCs and lower testing AUCs in the reported prediction models.

S1. Harrell F. The PHGLM procedure. In: SUGI Supple- mental Library User's Guide. SAS Institute, Cary, North Carolina, 1983, pp. 267-294.

Figure Captions

Figure 1S. Average ROCs obtained from different set of predictor variables. (a) ROCs including HER2 and ER biomarkers with 4 sets of predictor variables, and (b) ROCs based on HbT, %HbT and %US regardless of biomarkers.

Table 1S. Patient and tumor characteristics, treatment response and regimen. 38 patient data from this study and 22 patient data from Ref. 19

Biomarkers	Tumor stage	Miller-Payne	Treatment Regimen
•HER2+ : N=26 •TNBC: N=16 •ER+ or PR+/HER2-: N=18	•T3: N=13 •T2: N=39 •T1: N=8	•MP 5: N=29 •MP 4: N=6 •MP3: N=15 •MP2: N=4 •MP1: N=6	•ACT: N=22 •TCHP: N=22 •CarboT: N=8 •Others: N=8

TCHP: Docetaxel, carboplatin, trastuzumab, and pertuzumab
ACT: AC (Doxorubicin hydrochloride and cyclophosphamide) every two weeks followed by weekly paclitaxel (Taxol) for 12 weeks
CarboT: Carboplatin (Paraplatin) and Docetaxel (Taxotere)
ClinicalTrials.gov Identifier: NCT02124902

Table 2S**.** Logistic regression models based on tumor subtypes (HER2, ER) and TNBC, US measurements, and hemoglobin parameters, AUC of testing data. Data were from this study of total 38 patients and an early study of 22 patients (19).

Prediction Models including HER2 and ER Biomarkers			
Biomarkers, HbT and US measured before NAT	Biomarkers, %US measured at EOCs 1-3	Biomarkers, %HbT measured at EOCs 1-3	Biomakers, %HbT and %US measured at EOCs 1-3
•HER2, ER AUC=0.799 95% CI: 0.688-0.910 •HER2, ER, HbT AUC=0.814 95% CI: 0.706-0.922	•HER2, ER, %US_EOC1 •AUC=0.878 95% CI: 0.788-0.969 •HER2, ER, %US_EOC2 AUC=0.790 95% CI: 0.677-0.902 •HER2, ER, %US_EOC3 AUC=0.838 95% CI: 0.735-0.940	•HER2, ER, HbT, %HbT_EOC1 AUC=0.887 95% CI: 0.799-0.975 •HER2, ER, HbT, %HbT_EOC2 AUC=0.854 95% CI: 0.755-0.952 •HER2, ER, HbT,%HbT_EOC3 •AUC=0.907 95% CI: 0.826-0.987	**•HER2, ER, %HbT_EOC1, %US_EOC1** **AUC=0.929** 95% CI: 0.858-1.0 •HER2, ER, %HbT_EOC2, %US_EOC2 AUC=0.898 95% CI:0.814-0.982 •HER2, ER, %HbT_EOC3, %US_EOC3 AUC=0.913 95% CI:0.835-0.991 **•HER2, ER, %HbT_EOC3, %US_EOC1** **AUC=0.958** 95% CI: 0.903-1.014
Prediction Models based on Imaging Parameters			
Biomarkers, HbT and US measured before NAT	%US measured at EOCs 1-3	%HbT measured at EOCs 1-3	%HbT and %US measured at EOCs 1-3
•TNBC AUC=0.487 95% CI: 0.349-0.626 •US AUC=0.465 95% CI:0.327-0.603 •HbT AUC=0.704 95% CI:0.577-0.830	• %US_EOC1 AUC=0.855 95% CI; 0.757-0.952 •%US_EOC2 AUC=0.777 95% CI:0.661-0.892 •%US_EOC3 AUC=0.787 95% CI: 0.673-0.900	•HbT, %HbT_EOC1 •AUC=0.833 95% CI:0.729-0.936 •HbT, %HbT_EOC2 •AUC=0.833 95% CI: 0.730-0.936 •HbT, %HbT_EOC3 AUC=0.892 95% CI: 0.806-0.978	**•%HbT_EOC1, %US_EOC1** **AUC=0.920** 95% CI: 0.845-0.995 %HbT_EOC2, %US_EOC2 AUC=0.901 95% CI:0.818-0.984 •%HbT_EOC3, %US_EOC3 AUC=0.908 95% CI:0.828-0.988 **• %HbT_EOC3, %US_EOC1** **AUC=0.966** 95% CI:0.916-1.016

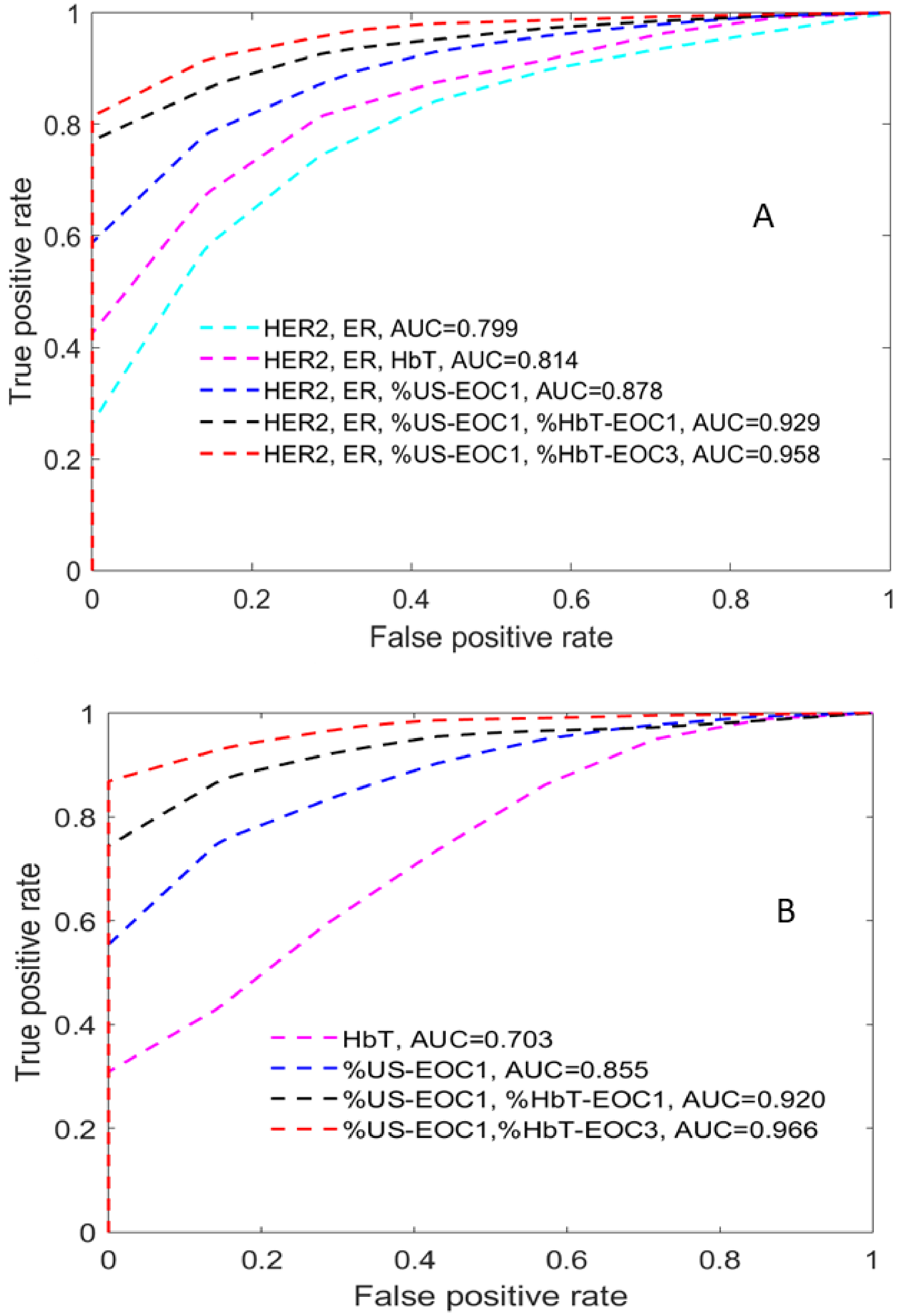

Figure 1S. Average ROCs obtained from different set of predictor variables. (a) ROCs including HER2 and ER biomarkers with 4 sets of predictor variables, and (b) ROCs based on HbT, %HbT and %US regardless of biomarkers.