



HHS Public Access

Author manuscript

Cancer Discov. Author manuscript; available in PMC 2022 April 01.

Published in final edited form as:

Cancer Discov. 2021 October ; 11(10): 2474–2487. doi:10.1158/2159-8290.CD-20-1557.

Impact of HER2 heterogeneity on treatment response of early-stage HER2-positive breast cancer: phase II neoadjuvant clinical trial of T-DM1 combined with pertuzumab

Otto Metzger Filho^{*,1,2,+}, Giuseppe Viale^{*,3,4}, Shayna Stein^{*,5,6}, Lorenzo Trippa⁵, Denise A. Yardley⁷, Ingrid A. Mayer⁸, Vandana G. Abramson⁸, Carlos L. Arteaga⁹, Laura M. Spring¹⁰, Adrienne G. Waks^{1,2}, Eileen Wrabel^{1,2}, Michelle K. DeMeo^{1,2}, Aditya Bardia¹⁰, Patrizia Dell'Orto³, Leila Russo³, Tari A. King^{2,11}, Kornelia Polyak^{1,13,14}, Franziska Michor^{5,6,12,13,14,15,+}, Eric P. Winer^{1,2}, Ian E. Krop^{1,2,+}

¹Medical Oncology, Dana-Farber Cancer Institute, Harvard Medical School, Boston, MA, USA

²Breast Oncology Program, Dana-Farber/Brigham and Women's Cancer Center, Boston, MA, USA

³Division of Pathology, European Institute of Oncology, IRCCS, Milan, Italy

⁴University of Milan, Milan, Italy

⁵Data Science, Dana-Farber Cancer Institute, Boston, MA, USA

⁶Department of Biostatistics, Harvard T.H. Chan School of Public Health, Boston, MA, USA

⁷Sarah Cannon Research Institute and Tennessee Oncology, Nashville, TN, USA

⁸Vanderbilt-Ingram Cancer Center, Nashville, TN, USA

⁹UT Southwestern Simmons Cancer Center, Dallas, TX, USA

¹⁰Massachusetts General Hospital, Boston, MA, USA

¹¹Division of Breast Surgery, Department of Surgery, Brigham and Women's Hospital, Boston, MA, USA

¹²Department of Stem Cell and Regenerative Biology, Harvard University, Cambridge, MA, USA

¹³Ludwig Center at Harvard, Boston, MA, USA

¹⁴Center for Cancer Evolution, Dana-Farber Cancer Institute, Boston, MA, USA

¹⁵The Broad Institute of MIT and Harvard, Cambridge, MA, USA

Abstract

⁺Authors for correspondence. **Corresponding authors:** Franziska Michor, PhD, Dana-Farber Cancer Institute, 450 Brookline Avenue, DFCI, Mailstop CLS-11007, Boston, MA 02215, Tel: 617-632-5045, michor@jimmy.harvard.edu; Otto Metzger Filho, MD, Dana-Farber/Brigham and Women's Cancer Center, 450 Brookline Avenue, Boston, MA 02215, Tel: 617-632-3800, Otto_Metzger@DFCI.HARVARD.EDU; Ian E. Krop, MD, PhD (lead contact), Dana-Farber/Brigham and Women's Cancer Center, 450 Brookline Avenue, Boston, MA 02215, Tel: 617-632-3800, Ian_Krop@dfci.harvard.edu.

^{*}These authors contributed equally

Intratumor heterogeneity is postulated to cause therapeutic resistance. To prospectively assess the impact of HER2 (*ERBB2*) heterogeneity on response to HER2-targeted therapy, we treated 164 patients with centrally confirmed HER2-positive early-stage breast cancer with neoadjuvant trastuzumab emtansine plus pertuzumab. HER2 heterogeneity was assessed on pretreatment biopsies from 2 locations of each tumor. HER2 heterogeneity, defined as an area with *ERBB2* amplification in > 5% but < 50% of tumor cells, or a HER2-negative area by FISH, was detected in 10% (16/157) of evaluable cases. The pathologic complete response rate was 55% in the non-heterogenous subgroup and 0% in the heterogenous group ($p < 0.0001$, adjusted for hormone receptor status). Single cell *ERBB2* FISH analysis of cellular heterogeneity identified the fraction of *ERBB2* non-amplified cells as a driver of therapeutic resistance. These data suggest HER2 heterogeneity is associated with resistance to HER2-targeted therapy and should be considered in efforts to optimize treatment strategies.

Keywords

breast cancer; heterogeneity; HER2

INTRODUCTION

Targeted therapies have now replaced or are used in conjunction with conventional cytotoxic chemotherapy in the treatment of the majority of cancers. This evolution in cancer care has resulted in improved outcomes and often less toxicity compared to previous chemotherapy regimens, which have led to interest in broadening the use of targeted agents and reducing the use of chemotherapy. However, resistance to targeted therapy almost always occurs. One potential mechanism of resistance is heterogeneous expression of the therapeutic target within the tumor (1), which is potentially of greatest concern in the curative setting in which a subclone lacking the target could escape the effects of the targeted therapy and lead to tumor recurrence. Thus, understanding how intratumor heterogeneity impacts response to targeted therapies is critical.

One example of the shift to targeted therapy is the treatment of HER2-positive breast cancer. Historically, amplification of *ERBB2* (encoding human epidermal growth factor receptor 2 [HER2]) defined a subset of breast cancers with aggressive clinical features and poor outcomes (2,3). However, the development of the HER2-specific monoclonal antibodies trastuzumab and pertuzumab and the antibody-drug conjugate trastuzumab emtansine (T-DM1), which consists of trastuzumab covalently linked to an anti-microtubule cytotoxic agent, have markedly decreased recurrence rates of patients with early-stage HER2-positive breast cancer (4–6).

The relevance of HER2 as a therapeutic target underscores the importance of accurate HER2 testing. The routine use of HER2 immunohistochemistry (IHC) and *in situ* hybridization (ISH) assays followed a seminal study demonstrating that the benefit of trastuzumab was restricted to patients diagnosed with HER2-positive tumors (7,8). Over the years, published guidelines from the American Society of Clinical Oncology (ASCO) and the College of American Pathologists (CAP) have optimized thresholds and recommendations to

define HER2 positivity (9–11). With the widespread use of HER2 testing, retrospective studies have reported different patterns of HER2 expression, coining the term “HER2 heterogeneity” (12). In parallel, diagnostic guidelines proposed definitions for HER2 heterogeneity (13–15), but its relevance in clinical practice has not been prospectively evaluated. Defining the impact of HER2 heterogeneity on responses to targeted anti-HER2 therapies is of particular importance as we endeavor to “de-escalate” standard therapeutic regimens and rely more on targeted anti-HER2 therapies for patients diagnosed with early-stage HER2-positive breast cancer (16,17).

In this study, we aimed to determine the effect of HER2 heterogeneity on response to therapy. We hypothesized that tumors with heterogeneity for HER2 amplification would have lower rates of pathologic complete response (pCR) when treated with a HER2-targeted regimen in the absence of conventional chemotherapy. To test this hypothesis, we conducted a prospective study in which patients diagnosed with HER2-positive breast cancer were treated with T-DM1 in combination with pertuzumab prior to surgery. The specificity and potency of T-DM1 and pertuzumab against HER2-amplified cells were critical components to the study design. Image-guided research biopsies performed prior to treatment initiation allowed a central pathology evaluation of HER2 heterogeneity. The study was powered to assess the impact of HER2 heterogeneity on the probability of achieving a pCR after a course of targeted anti-HER2 therapy.

RESULTS

Patients and Treatment

A total of 164 patients were enrolled in the study from January 2015 to January 2018 (Fig. 1A and (Supplementary Table S1). Patients received 6 cycles of T-DM1 and pertuzumab (Fig. 1B). The baseline demographic and clinical characteristics of the enrolled patients are listed in Table 1. Of all patients, 163 were treated with at least one dose of T-DM1 and pertuzumab. Central confirmation of HER2 status to define eligibility classified 74% (121/163) of cases as HER2 3+ by IHC and 25% (40/163) as HER2 2+. HER2 2+ cases were confirmed to be HER2-positive by fluorescence in situ hybridization (FISH) prior to study enrollment. HER2 positivity was defined by FISH without IHC information in two cases (1%, 2/163). All but one patient (99.4%) had either stage II or III cancer at presentation. Two-thirds (68.7%) of tumors were classified as hormone receptor (HR)-positive and the remaining tumors as HR-negative.

T-DM1 plus pertuzumab was associated with a favorable toxicity profile (Table 2, Supplementary Table S1). Mild fatigue was the most common adverse event and observed in 67% of patients. Grade 3 or higher toxicities were rare. Overall, 95% of the study population completed the planned 6 cycles of therapy. T-DM1 dose reductions occurred in 11% of patients. Two patients (1.2%) discontinued treatment due to adverse events.

HER2 Heterogeneity and Treatment Response

Central pathology evaluation of HER2 heterogeneity was successful in 97% (158/163) of cases. The 5 unsuccessful evaluations were due to a lack of invasive disease in the research

core biopsies. To assess HER2 heterogeneity, two spatially distinct biopsies performed at baseline were evaluated by FISH; ~50 cells were counted in three areas of each biopsy (Fig. 1C, D). The assessment of HER2 heterogeneity determined by central pathology review classified 10% (16/157) of tumors as HER2 heterogeneous (Supplementary Fig. S1). The study met its primary objective by demonstrating a significant association between HER2 heterogeneity and pCR adjusted by HR status; no cases of pCR, defined as a residual cancer burden (RCB) of 0, were observed in the heterogenous subgroup while the pCR rate was 55% (77/141) in the non-heterogenous subgroup ($p = 0.0001$, χ^2 test) (Fig. 2A). When using a less strict definition of pathological response defined as RCB 0 or I (i.e., including minimal residual disease), we found that the association between HER2 heterogeneity and response remained significant (odds ratio [OR] = 5.6; $p=0.002$, χ^2 test), (Fig. 2B).

Regarding the study's secondary objectives, the overall pCR rate was 49% ($n = 77/157$). Among cases with residual disease, the distribution of RCB scores was RCB-I 14% (22/157); RCB-II 26% (41/157); and RCB-III 11% (17/157). Higher pCR rates were observed in the subset of HR-negative cases compared to HR-positive cases (65% vs 42%, $p = 0.016$, χ^2 test) (Fig. 2C). The breakdown by HR status in the non-heterogenous subgroup revealed pCR rates of 69% (31/45) in the HR-negative and 48% (46/96) in the HR-positive subset (Fig. 2C).

The subset of cases classified as heterogeneous for HER2 had a similar distribution of HR status as the overall study population with 13 (81%) classified as HR-positive and 3 (19%) as HR-negative. HER2 IHC evaluation performed on diagnostic core biopsies classified 75% ($n=12$) of heterogeneous cases as HER2 2+, and 25% ($n=4$) as HER2 3+ disease. The majority (81%) of non-heterogeneous cases were HER2 3+.

In order to evaluate whether HER2 protein levels were confounding the association between heterogeneity and pCR, we performed an additional exploratory analysis and found that, when adjusted by HER2 protein IHC levels (2+ vs 3+) and HR status, the association between HER2 heterogeneity and pCR remained statistically significant ($p=0.002$, stratified Mantel-Haenszel χ^2 test). The association between HER2 heterogeneity and HER2 protein IHC levels (2+ vs 3+) was also significant ($p=2e-06$, χ^2 test).

HER2 Heterogeneity Assessed at the Single Cell Level

To better understand the relationship between spatial HER2 heterogeneity and clinical outcome at a more granular level, we assessed HER2 copy number at the single cell level using FISH (Supplementary Tables S2 and S3). A median of 6 tumor areas per patient were analyzed and each area contained a median of 75 cells (95% confidence interval [CI] 67 to 83). We first quantified HER2 cellular heterogeneity using the Shannon equitability index (18), a diversity measure that is commonly used to quantify heterogeneity in cancer biology (Materials and Methods). We found that cases with pCR had higher levels of heterogeneity than cases without pCR ($p=0.0032$, Wilcoxon test; Fig. 3A). This observation was unexpected because previous retrospective studies found that heterogeneous tumors are more likely to be resistant to treatment (19–21). Upon further investigation, we determined that the Shannon equitability index was negatively correlated with the fraction of *ERBB2* non-amplified cells in our cohort ($R^2=-0.64$, $p < 2.2e-16$, Pearson correlation; Fig. 3B),

suggesting that patients with a low extent of heterogeneity tend to have uniformly low HER2 levels and patients with higher heterogeneity tend to have populations of cells with high *ERBB2* copy numbers. We additionally found that patients meeting the clinical definition of heterogeneity have a lower Shannon index ($p = 3.2 \times 10^{-9}$, Wilcoxon test; Fig. 3C) and a higher fraction of *ERBB2* non-amplified cells ($p = 2.1 \times 10^{-10}$, Wilcoxon test; Fig. 3D).

We then quantified the extent of HER2 cellular heterogeneity using the Gini index (22), a metric that is frequently used to assess dispersion (Supplementary Fig. S2). We found that 95% of patients had a Gini coefficient between 0.18 and 0.45, suggesting that overall dispersion of *ERBB2* copy number is relatively low across the patient cohort. Moreover, we found that there was no significant difference in the Gini coefficient between cases with and without pCR ($p = 0.39$; Supplementary Fig. S2A). The Gini coefficient is weakly correlated with the Shannon equitability index ($R^2 = 0.25$, $p = 0.0015$, Pearson's product moment correlation, Supplementary Fig. S2B). Lastly, the Gini coefficient was not correlated with the percentage of *ERBB2* non-amplified cells ($R^2 = -0.082$, $p = 0.31$, Pearson's product moment correlation, Supplementary Fig. S2C). Altogether, these results suggest that the Gini index is not informative for predicting the response to T-DM1 plus pertuzumab.

Based on these observations, we quantified HER2 heterogeneity as a continuous variable according to the fraction of *ERBB2* non-amplified cells across the areas in each tumor. When analyzing the association of this metric with outcome, we found that patients without pCR had a higher overall fraction of *ERBB2* non-amplified cells compared to patients with pCR ($p = 0.0064$, Wilcoxon test; Fig. 4A). This relationship holds for HR-negative ($p = 0.0029$, Wilcoxon test) and HR-positive ($p = 0.061$, Wilcoxon test) patients, although the latter is not statistically significant (Fig. 4B). We additionally calculated the fraction of *ERBB2* non-amplified cells separately for each of the six area biopsies from each patient, and observed that pCR was significantly negatively associated with the fraction of *ERBB2* non-amplified cells per area ($p = 0.0045$ for HR-positive, $p = 0.017$ for HR-negative cases, F-test with Satterthwaite's method; Fig. 4C).

Next, we investigated whether the relationship between response and the fraction of *ERBB2* non-amplified cells was driven by a subpopulation of cells in one area. To this end, we determined the fraction of *ERBB2* non-amplified cells for each area, and used the maximum value across areas to investigate whether an area with a particularly high fraction of *ERBB2* non-amplified cells was driving resistance to treatment. We found that the relationship between pCR rate and the maximum area fraction of *ERBB2* non-amplified cells (Supplementary Fig. S3A) was similar to the relationship between pCR rate and the total tumor fraction of *ERBB2* non-amplified cells (Fig. 4A). This observation suggests that one area with a particularly high fraction of *ERBB2* non-amplified cells was not driving resistance to treatment. Moreover, we found that the minimum and maximum fraction of *ERBB2* non-amplified cells in an area were highly correlated with each other and with the total fraction of *ERBB2* non-amplified cells per tumor (Supplementary Fig. S3B–D). Taken together, these observations suggest that there was a low degree of spatial heterogeneity of HER2+ cells in this patient cohort and that nonresponders had a higher fraction of *ERBB2* non-amplified cells overall rather than in one specific area that is responsible for resistance.

We then investigated the relationship between *ERBB2* copy number and pCR in order to identify whether pCR was driven by the level of *ERBB2* copy number in addition to the fraction of *ERBB2* non-amplified cells. We found that patients with pCR have significantly higher median *ERBB2* copy number compared to patients without pCR ($p = 0.048$, Wilcoxon test; Supplementary Fig. S4A), and that this relationship held for HR-negative ($p = 0.028$, Wilcoxon test) patients (Supplementary Fig. S4B). Likewise, we calculated the median *ERBB2* copy number separately for each of the six core biopsy areas from each tumor, and found that pCR was associated with the median *ERBB2* copy number per area ($p = 0.081$ for HR-positive, $p = 0.09$ for HR-negative, F-test with Satterthwaite's method; Supplementary Fig. S4C). These results are consistent with results of the analysis of the fraction of *ERBB2* non-amplified cells, likely because biopsies with a higher fraction of *ERBB2* non-amplified cells have a lower median *ERBB2* copy number.

We then assessed whether we could predict pCR using the single cell HER2 FISH data. We first predicted pCR by classifying patients based on the fraction of *ERBB2* non-amplified cells and achieved an area under the curve (AUC) of 0.62 (0.59 HR+; 0.75 HR-; Table 3, Supplementary Fig. S5A). We then fit seven different logistic regression models: (1) including the fraction of *ERBB2* non-amplified cells and hormone receptor status; (2) including the median *ERBB2*/CEP17 ratio and hormone receptor status; (3) including all terms in model 1 plus median *ERBB2*/CEP17 ratio; (4) including all terms in model 1 plus an interaction term between hormone receptor status and the fraction of *ERBB2* non-amplified cells; (5) including all terms in model 1 plus clinical variables including stage (T and N) and HER2 IHC by central review; (6) including all terms in model 1 plus the Shannon equitability index; and (7) including all terms in model 1 plus heterogeneity (as per the College of American Pathologists definition). The resulting coefficient estimates and significance values are shown in Supplementary Table S4 and the resulting AUC values are shown in Table 3 (Supplementary Fig. S5B, Supplementary Fig. S6A–G). We observed that the fraction of *ERBB2* non-amplified cells (model 1) was slightly more predictive than the median *ERBB2*/CEP17 ratio (model 2), and that the median *ERBB2*/CEP17 ratio coefficient was not significant in model 3. Together, these results suggest that the fraction of *ERBB2* non-amplified cells was more predictive of pCR than the median *ERBB2*/CEP17 ratio, and that adding median *ERBB2*/CEP17 ratio to a model that includes the fraction of *ERBB2* non-amplified cells did not significantly improve prediction of pCR.

We also found that the interaction terms in model 4 were not significant, suggesting that the interaction between hormone receptor status and the fraction of *ERBB2* non-amplified cells was not predictive of response in this patient cohort. Moreover, we found that the additional clinical variables in model 5 were not significant, suggesting that the single cell *ERBB2* FISH data are more predictive of treatment response than clinical stage or HER2 IHC score. Finally, we found that neither Shannon equitability index in model 6 nor College of American Pathologists-defined heterogeneity in model 7 were significant, suggesting that these metrics did not add additional information to model 1. Together, the results of this analysis suggest that the model using the fraction of *ERBB2* non-amplified cells and hormone receptor status was optimal among the models investigated for predicting pCR. We then examined whether assessment of HER2 heterogeneity was affected by sampling. Only a section of each cell's nucleus was assessed when measuring gene copy number, which

could potentially lead to an artefactual appearance of heterogeneity in cases with low levels of *ERBB2* amplification. To investigate whether this potential issue affected our predictions, we generated a simulated dataset by adding a random Poisson($\lambda=ERBB2/2$) value to the *ERBB2* copy number of each cell and repeating the simulation 1,000 times. We chose this λ under the assumption that the number of uncounted *ERBB2* copies would, on average, be half the number of observed copies in any given cell. Note that a Poisson($\lambda=ERBB2/2$) random variable has variance $\lambda=ERBB2/2$, which allows for variability in the number of uncounted *ERBB2* copies. The resulting AUC based on logistic regression model 1 applied to this simulated dataset was 0.69 (95% CI: 0.65–0.72, Table 3), which is very close to the AUC achieved by model 1 using the observed data (0.70). This finding suggests that our results were unlikely caused by artefactual heterogeneity of the samples.

Finally, we assessed whether the addition of data from the second biopsy improves our prediction accuracy by randomly sampling one biopsy from each patient to use in the prediction model. The resulting AUC values from this sampling approach are shown in Table 3. The 95% confidence intervals of the AUC using this sampling approach contain the AUC obtained when using all observed data for every model. These results suggest that the addition of the second biopsy does not improve prediction accuracy based on our patient cohort (Table 3). We also assessed whether analyzing three areas rather than one area per core biopsy improved our prediction accuracy by randomly sampling one of the three areas from the selected biopsy in the analysis above. Similar to our analysis of using one vs. two biopsies, we found that the 95% confidence intervals of the AUC using one vs three areas sampling contain the AUC obtained when using all three areas from a sampled biopsy for every model (Table 3). As expected, we also found that the confidence intervals were wider when sampling one area from each sampled biopsy compared to using information from all three areas from each sample biopsy for every model tested. This observation suggests that using less information increases the variability of the prediction accuracy. Furthermore, we found that our logistic regression model improves the sensitivity of predicting pCR without sacrificing specificity, compared to using the current metric of heterogeneity suggested by the College of American Pathologists. For reference, we achieved a specificity of 1 and a sensitivity of 0.162 when using the current guidelines for assessing heterogeneity. Depending on the threshold chosen, the logistic regression model can achieve a sensitivity above 0.3 with specificity above 0.9, or a sensitivity above 0.5 with specificity above 0.75. Thus, our approach allows the identification of a larger number of patients who are likely resistant to neoadjuvant T-DM1 plus pertuzumab as compared to the currently clinical definition of heterogeneity.

DISCUSSION

Cancer, almost by definition, implies the presence of intratumor heterogeneity, as tumors arise from accumulating genetic and epigenetic alterations (23–25). As treatments have evolved from older drugs able to disrupt the DNA of cancer cells (i.e., chemotherapy) to targeted therapies, studying how heterogeneity of the drug target impacts responses to therapies becomes increasingly important. In this study, we prospectively evaluated how intratumor heterogeneity of HER2 affects response to T-DM1 plus pertuzumab. Our study met its primary objective by demonstrating a significant association between HER2

heterogeneity and pCR. In the subset of heterogeneous tumors, all cases had residual disease at the time of surgery, while the pCR rate was 55% in the subset without heterogeneity. The effect of heterogeneity remained significant when controlling for HR status, HER2 protein levels, and less strict definitions of pathologic response including RCB 0 or I. Our analysis of heterogeneity at the single cell level confirmed the presence of a direct relationship between HER2 heterogeneity and response to targeted anti-HER2 therapy. Our findings suggest that HER2 heterogeneity defined by FISH does not reflect methodological limitations of the test, but rather represents a true biological readout of the impact of HER2 heterogeneity on response to targeted anti-HER2 therapy.

To the best of our knowledge, this is the first clinical trial designed to prospectively evaluate the impact of intratumor heterogeneity on response to targeted therapy in breast oncology. The results of this study suggest that efforts to reduce the amount of conventional chemotherapy and rely more on HER2-targeted therapies for patients diagnosed with HER2-positive early-stage disease may not be successful if intratumor heterogeneity is not taken into consideration.

It is important to note that the T-DM1 and pertuzumab regimen used in this study represents a hybrid model of targeted anti-HER2 therapy linked to chemotherapy and an anti-HER2 therapy, respectively. One significant limitation of our study was that we did not have a control group treated with conventional chemotherapy and HER2-directed therapy. Such a control arm would have allowed us to investigate the benefits of chemotherapy in the subset of HER2 heterogeneous cancers. Moreover, having an investigational arm containing targeted anti-HER2 agents only (i.e., no cytotoxic component) could also be of interest. The latter was considered as a possibility while designing the study but deemed infeasible as it would require a larger sample size due to expected lower pCR rates with targeted anti-HER2 therapies only. It is possible that cancers with HER2 heterogeneity may be less sensitive not only to a T-DM1-based regimen but also to HER2 therapy with or without chemotherapy. Indeed, supporting this possibility, a retrospective analysis of the Kristine phase III study (26) comparing neoadjuvant T-DM1 plus pertuzumab versus TCHP (taxotere, carboplatin, trastuzumab, and pertuzumab) observed numerically lower pCR rates in both arms of the study in the subset of tumors with focal or variable HER2 IHC staining compared to those with homogenous HER2 staining. These results, together with the data from the current study, suggest that the ability to identify HER2 heterogeneous tumors may be useful in the development of novel targeted anti-HER2 agents with the capacity to target tumors with less uniform levels of HER2 amplification. For one such agent, trastuzumab deruxtecan, preclinical and clinical evidence already exists suggesting effectiveness against such tumors (27).

The favorable toxicity profile observed with T-DM1 and pertuzumab in this study is similar to previous studies (28–30). In the Kristine study, T-DM1 was associated with a lower incidence of grade 3 adverse events (13% vs. 64%) and lower incidence of treatment discontinuation (3% vs. 8%) when compared with TCHP (26,31).

The neoadjuvant setting in this study provided us with an ideal platform for evaluating HER2 heterogeneity. Image-guided research biopsies were obtained from all participants

and HER2 heterogeneity was evaluable in 97% of cases. In HER2-positive disease, pCR after neoadjuvant therapy is a surrogate for improved survival outcomes (32). It is important to note, however, that pCR has not been validated as a trial-level surrogate for survival outcomes and the current study was not designed to establish a direct link between the lack of pCR associated with heterogeneity and long-term outcomes. A larger sample size and long-term follow-up would be needed to evaluate the association between HER2 heterogeneity and survival outcomes. Nonetheless, data from the Kristine study indicated a favorable 3-year IDFS outcome of 97% among patients treated with T-DM1 and pertuzumab who achieved a pCR (31). The integration of HER2 heterogeneity evaluation in the selection of patients for treatment with targeted HER2 therapies in the early-stage setting could prove to be a successful strategy in future clinical trials.

The definition of HER2 heterogeneity in this study follows recommendations from treatment guidelines (13,15). Retrospective series evaluating the prevalence of HER2 heterogeneity have reported variable prevalence of heterogeneity varying from 5% to 40%, likely reflecting differences in sample size or variations in the definition of heterogeneity (12). A retrospective series found a significant association between HER2 heterogeneity and inferior survival outcomes (33). The potential clinical impact of intratumor heterogeneity observed in that retrospective study and our current prospective study highlight the need to develop a uniform, reproducible definition of HER2 heterogeneity for future studies.

To elucidate the relationship between spatial HER2 heterogeneity and response to targeted therapy at a more granular level, we leveraged the two spatially distinct tumor biopsy samples obtained from each patient and analyzed HER2 copy number data at the single cell level. Using these data, we found that the main driver of resistance to therapy (as defined by lack of pCR) was the fraction of HER2 non-amplified cells across the tumor. Interestingly, it did not appear that one tumor area with a high fraction of *ERBB2* non-amplified cells accounted for the lack of pCR; instead, the overall tumor fraction of *ERBB2* non-amplified cells was the key factor. Consistent with this observation, pCR could be predicted as well with the single cell data from one biopsy per patient as with two biopsies. Furthermore, we found that models using the overall fraction of *ERBB2* non-amplified cells to predict patient pCR achieve higher prediction sensitivity without sacrificing specificity compared to using the current recommendations from treatment guidelines. Together, the results from these single cell analyses highlight the potential for defining novel thresholds for HER2 heterogeneity and response to HER2-targeted therapies. In addition, it will be important to determine if similar aspects of heterogeneity drive resistance to therapies against other targets and in other cancer types.

The decision to use FISH to measure HER2 in our study facilitates reproducibility but is not meant to account for the full spectrum of HER2 heterogeneity. Certainly, our selection of a single primary modality to evaluate HER2 heterogeneity in this study represents a limitation. Previous studies in both the early disease and metastatic setting have demonstrated a significant association between lower HER2 levels quantified at the mRNA level and decreased T-DM1 benefit. A multimodality approach to evaluate HER2 heterogeneity including HER2 mRNA and additional features associated with resistance (e.g., *PIK3CA* mutation status) could be informative. With the development of modern

techniques able to evaluate the genomic (and perhaps proteomic) landscape at the single cell level, we should expect to understand the biological significance of intratumor heterogeneity in an unprecedented manner.

From a clinical perspective, the integration of tumor heterogeneity measurements in the selection of patients for treatment with targeted therapies could prove to be a successful strategy to reduce the use of conventional cytotoxic chemotherapy. As an example, patients identified as having non-heterogeneous HER2+ tumors could be offered participation in a trial comparing T-DM1 plus pertuzumab versus conventional chemotherapy and HER2-directed therapy. Those achieving a pCR with T-DM1 plus pertuzumab would continue therapy with anti-HER2 therapies, limiting the use of standard cytotoxic agents only to those patients with residual disease at the time of surgery. In the control arm of the study patients would be treated with standard anti-HER2 therapy and chemotherapy. Such a treatment strategy could translate into a significant reduction in the use of standard chemotherapy for patients diagnosed with HER2+ breast cancer and improve quality of life outcomes. If this type of approach is successful in HER2+ breast cancer, it could then be explored in other cancer types in which targeted therapies are utilized.

In conclusion, our study illustrates the critical importance of understanding the impact of intratumor heterogeneity on response to targeted therapies in the early disease (i.e., curative) setting. This issue is particularly important in the modern era where the number of targets and targeted therapies has increased dramatically and there is a strong interest in using these targeted agents in place of conventional chemotherapy. In the subset of HER2-positive breast cancer, HER2 heterogeneity strongly impacts response to T-DM1 given in combination with pertuzumab. If these results are validated in additional studies, assessment of HER2 heterogeneity may facilitate optimal selection of therapy for patients with early stage HER2-positive cancers.

METHODS

Study Design, Eligibility and Enrollment

We conducted an investigator-initiated single arm Phase 2 study. The study was open at the Dana-Farber Cancer Institute, Massachusetts General Hospital, Sarah Cannon Research Institute, Tennessee Oncology, and Vanderbilt-Ingram Cancer Center. Enrollment required a pathological diagnosis of carcinoma of the breast with immunohistochemical staining for the HER2 protein of 3+ intensity or amplification of the *ERBB2* gene based on FISH (ratio of *ERBB2* to chromosome 17 centromere [CEP17] ≥ 2 , or *ERBB2* copy number ≥ 6). HER2 status was centrally confirmed before study enrollment. The invasive tumor had to measure at least 2 cm in the greatest dimension assessed by physical examination or imaging; there was no upper limit on tumor size or axillary nodal status. Other requirements included willingness to undergo a research biopsy prior to treatment initiation, adequate hematopoietic and liver function and a left ventricular ejection fraction of 55% or greater. The institutional review board at each participating institution approved the study. Written informed consent was provided by all participants. The protocol is available as a supplemental file.

Assessment of HER2 Heterogeneity

Patients underwent image-guided research biopsies performed in two different geographic areas of the same tumor prior to treatment initiation. Central pathology evaluation of HER2 heterogeneity was performed at the European Institute of Oncology in Milan and blinded to treatment outcome. Formalin-fixed research biopsies were embedded into paraffin blocks. HER2 status was assessed by IHC and FISH. HER2 heterogeneity was evaluated following recommendations from the College of American Pathologists (13) and the United Kingdom HER2 testing guidelines (34). An entire tissue section of each core biopsy site was scanned before selecting 3 tumor areas for HER2 counting. *ERBB2* and CEP17 signals were counted in approximately 50 cells per area (i.e., approximately 300 cells counted per tumor). HER2 heterogeneity was defined as the existence of an area of tumor cells with *ERBB2* amplification (i.e., *ERBB2*/CEP 17 ratio ≥ 2.0 or a gene copy number ≥ 6) representing more than 5% but less than 50% of infiltrating tumor cells, or an area of tumor that tested HER2-negative by FISH. HER2 IHC results did not influence the definition of HER2 heterogeneity for the evaluation of primary endpoint.

Treatment Regimen

Treatment consisted of six cycles of T-DM1 given in combination with pertuzumab. Participants received T-DM1 at a dose of 3.6 mg per kilogram of body weight, and pertuzumab at a loading dose of 840 mg followed by 420 mg. T-DM1 and pertuzumab were given every three weeks intravenously for a total duration of six cycles. Patients underwent breast surgery within 42 days of the last cycle of therapy. The type of breast surgery was at the discretion of the patient and surgeon. Decisions regarding choices of adjuvant radiotherapy and adjuvant systemic therapy were made by the treatment team and not mandated per protocol.

Endpoints and Study Assessments

The primary objective was to evaluate the relationship between pCR and intratumor heterogeneity of HER2 amplification. Key secondary objectives included assessment of safety, the relationship between pCR and intratumor heterogeneity assessed as a continuous variable.

Pathologic response was reported using the RCB calculator. RCB 0, no residual invasive cancer in the breast or axillary nodes, defined pCR for the primary endpoint of the study. Left ventricular ejection fraction was assessed at baseline, end of cycle 2, and at the pre-surgery visit. Laboratory monitoring was performed prior to each treatment cycle and adverse events were assessed with each treatment cycle according to Common Terminology Criteria for Adverse Events v4.0. Single cell HER2 counting by FISH was used to evaluate HER2 heterogeneity as a continuous variable. The fraction of HER2 non-amplified cells was evaluated in three areas per core biopsy site at the central pathology laboratory in Milan.

The protocol was reviewed by the Dana-Farber/Harvard Cancer Center Data and Safety Monitoring Committee throughout the study to monitor toxicity and review accrual.

Statistical Analysis

The study sample size was determined by the expected percentage of the study population classified as HER2 heterogeneous, the overall pCR rate and the ability to demonstrate differences in pCR rates between heterogeneous and non-heterogeneous subsets. Considering that the study sample size would not be adjusted according to the prevalence of HER2 heterogeneity, we included a sensitivity analysis with different effect sizes (pCR heterogeneous versus non-heterogeneous) for which there would be 80% and 90% power to detect under varying prevalences of HER2 heterogeneity (e.g., 10%, 20%, 30%). By estimating that the overall pCR with T-DM1 plus pertuzumab would be approximately 40%, and 20% of the population classified as heterogeneous, the study would have 80% power with 136 evaluable patients to detect a difference in pCR of 44.9% in the homogenous versus 20.3% in the heterogeneous subgroup. The final sample size was defined as 165 participants, assuming a 15% failure rate in the assessment of heterogeneity. As detailed in the protocol (see supplement), the study had 90% power to detect differences in pCR of 43.4% in the homogenous versus 8.8% in the heterogeneous subgroup if the observed prevalence of HER2 heterogeneity was 10%. To evaluate the association between pCR and HER2 heterogeneity, we used a stratified Mantel-Haenszel chi-square test across HR-positive and HR-negative patients. The use of a stratified model was deemed critical to prevent confounding due to the known relationship between HR status and pCR.

Assessing Tumor Heterogeneity using HER2 FISH

We used Shannon equitability index to assess tumor heterogeneity based on single cell HER2 FISH. For each patient i , where $i = 1, 2, \dots, 157$, denote the number of tumor images as n_i . For each tumor image j , where $j = 1, 2, \dots, n_i$, we obtained the HER2, x_{ijk} , and CEP17, y_{ijk} , levels for each cell $k = 1, \dots, K$ in the image. Each cell was then denoted by the tuple $(x_{ijk} = h, y_{ijk} = c)$, where $h = 1, 2, \dots, H$ is the HER2 level and $c = 1, 2, \dots, C$ is the CEP17 copy number. We then calculated the frequency of each tuple in each image, where Z_{ijhc} is the frequency of (h, c) cells in image j from patient i and $|Z_{ij}|$ is the total number of species in image j from patient i . Each species then has frequency $p_{ijhc} = Z_{ijhc}/|Z_{ij}|$. Finally, we calculated the Shannon index for each image: $H_{ij} = -\sum_{(h,c)} p_{ijhc} \ln(p_{ijhc})$ and the Shannon equitability index: $E_{ij} = H_{ij}/\ln(|Z_{ij}|)$. We also calculated the overall Shannon index for each patient i : $H_i = -\sum_{(h,c)} p_{ihc} \ln(p_{ihc})$, where p_{ihc} is the overall frequency of (h, c) cells among all of the images from patient i . Likewise, the overall Shannon equitability index for each patient i is given by $E_i = H_i/\ln(|Z_i|)$, where $|Z_i|$ is the total number of species among all of the images from patient i .

We used the Gini coefficient to assess dispersion of the HER2/CEP17 ratio within each patient. We used the `gini` function from the `reldist` package in R to calculate the gini index for each patient.

We used the fraction of HER2 non-amplified cells to assess the level of HER2 amplification within each patient. A cell was considered non-amplified if the HER2/CEP17 ratio < 2 and the gene copy number ≤ 6 . The fraction of HER2 non-amplified cells was calculated to be the percentage of non-amplified cells out of the entire population.

Predicting Patient Response using *ERBB2* FISH

To predict patient response using the fraction of *ERBB2* non-amplified cells, we tested every threshold between 0 and 1 at 0.01 intervals. Patients below the threshold were predicted to be responders; patients at or above the threshold were considered to be non-responders. We then calculated the sensitivity and specificity of the prediction for every threshold. These values were used to create a receiver operating characteristic (ROC) curve and calculate the AUC. This analysis was done both on the entire patient cohort and subcohorts stratified by HR status.

We additionally used logistic regression to model patient response as a function of the fraction of *ERBB2* non-amplified cells, median *ERBB2* copy number, and HR status. We used the ``glm`` function in R. We tested the model both with and without interaction terms between HR status and fraction of *ERBB2* non-amplified cells, and HR status and median *ERBB2* copy number. We then used the ``predict`` function to compute the predicted probability of response for each patient, given the estimated logistic regression model. The ``roc`` function in R was used to create ROC curves and estimate the AUC.

We used a sampling method to assess whether the use of data from two biopsies provides significantly more information for predicting response compared to using one biopsy. For each patient, we randomly sampled one core biopsy site and used the FISH data to calculate the fraction of *ERBB2* non-amplified cells and median *ERBB2* copy number. We then used this sampled data to predict patient response using the two methods described above and estimate the AUC. We repeated this process 1,000 times, resulting in 1,000 sampled AUC values for each method where each sample is a permutation of randomly selecting one biopsy per patient. We then used the 2.5% and 97.5% quantiles to estimate 95% CIs of the AUC for each method where only one biopsy per patient is used.

Finally, we used a simulation method to assess the accuracy of copy number assessment using FISH, given that only a section of the nucleus is assessed in each individual cell. For each simulation, we randomly added additional *ERBB2* copies to each cell according to a poisson(λ) distribution, where $\lambda = \text{HER2}/2$ for each cell. We then used this randomly simulated data to predict patient response using logistic regression and estimate the AUC. We repeated this process 1,000 times, resulting in 1,000 randomly simulated AUC values. We then used the 2.5% and 97.5% quantiles to estimate the 95% CIs of the AUC.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

ACKNOWLEDGEMENTS

We thank Drs. Paul Catalano, Michalina Janiszewska, Simona Cristea, and Jamie Dean for their critical reading of our manuscript and useful suggestions.

Disclosure of Potential Conflicts of Interest:

O.M.F. reports consulting/advisory roles for Abbvie, G1 Therapeutics, and Grupo Oncoclinicas (Brazil); honoraria from Roche (Brazil); research funding from Abbvie (Inst), Cascadian Therapeutics (Inst), Eisai (Inst), Pfizer (Inst), Roche/Genentech (Inst), and Susan G. Komen for the Cure (Inst); and travel/accommodations/

expenses from Grupo Oncoclinicas. **G.V.** reports consulting/advisory roles for Astellas Pharma, Dako, Novartis and Roche/Genentech; honoraria from MSD Oncology and Pfizer; research funding from Dako/Agilent Technologies (Inst), Roche/Genentech, and Ventana Medical Systems (Inst); and travel/accommodations/expenses from Celgene, and Roche. **D.A.Y.** reports consulting/advisory roles for Biotheranostics (Inst), Bristol-Myers Squibb (Inst), Celgene (Inst), Daiichi Sankyo/Lilly (Inst), Eisai (Inst), Genentech/Roche (Inst), NanoString Technologies (Inst), and Novartis (Inst); speakers' bureau for Genentech/Roche and Novartis; research funding from Abbvie (Inst), AstraZeneca (Inst), Clovis Oncology (Inst), Daiichi Sankyo (Inst), Eisai (Inst), Genentech/Roche (Inst), Immunomedics (Inst), InventisBio (Inst), Lilly (Inst), MacroGenics (Inst), MedImmune (Inst), Medivation (Inst), Merck (Inst), Novartis (Inst), Oncothyreon (Inst), Pfizer (Inst), Syndax (Inst), and Tesaro (Inst); and travel/accommodations/expenses from Genentech/Roche and Novartis. **I.A.M.R.** reports consulting/advisory roles for AstraZeneca, Genentech, GlaxoSmithKline, Immunomedics, Lilly, MacroGenics, and Novartis; and research funding from Genentech (Inst), Novartis (Inst), and Pfizer (Inst). **V.G.A.** reports consulting/advisory roles for Eisai and Novartis; and research funding from Astellas Pharma, Genentech/Roche, and Lilly. **C.L.A.** reports consulting/advisory roles with Novartis, Lilly, Sanofi, TAIHO Oncology, PUMA Biotechnology, Merck, H3Biomedicine, RADIUS, OrigiMed, Immunomedics, Petra Pharma, G1 Therapeutics, Athenex, Daiichi Sankyo, Clovis, and Abbvie; research grants from Pfizer, Lilly, Takeda, and Bayer; stock/other ownership interests in Provista and Y-TRAP. He serves in the Scientific Advisory Board of the Susan G. Komen for the Cure Foundation. **L.S.** reports consulting/advisory role for Novartis; honoraria from DAVA Oncology; research funding from Tesaro (Inst); and travel/accommodations/expenses from Novartis. **A.B.** reports consulting/advisory roles for bioTheranostics, Genentech, Genentech/Roche (Inst), Immunomedics, Immunomedics (Inst), Innocrin Pharma (Inst), Merck, Merck (Inst), Novartis, Novartis (Inst), Pfizer, Pfizer (Inst), Radius Health (Inst), Radius Pharma, Sanofi, and Spectrum Pharmaceuticals; and research funding from bioTheranostics. **T.A.K.** reports honoraria from Genomic Health. **K. P.** reports consulting/advisory role for Acrivon Therapeutics, Scorpion Therapeutics, and twoXAR; equity shares from Scorpion Therapeutics; honoraria from Genentech and Blueprint Medicine; and patents on S100A7 antibody and BET inhibitor resistance. **F.M.** is the co-founder of an oncology company. **E.P.W.** reports consulting/advisory roles for InfiniteMD and Leap Therapeutics; honoraria from Carrick Therapeutics, Genentech/Roche, GlaxoSmithKline, InfiniteMD, Jounce Therapeutics, Lilly, and Tesaro; research funding from Genentech (Inst); and stock/other ownership interests in Verastem. **I.E.K.** reports grants from Genentech/Roche (Inst) during the conduct of the study; personal fees from Context Therapeutics, personal fees from Daiichi Sankyo, personal fees from Genentech/Roche, personal fees from MacroGenics, personal fees from Seagen, personal fees from Taiho Pharmaceutical, personal fees from Astra Zeneca, grants from Pfizer (Inst), personal fees from Merck, personal fees from Novartis, and personal fees from Celltrion outside the submitted work. All remaining authors report no conflicts.

Grant Support:

This work was supported by Roche, Susan G. Komen (I.E.K), the Dana-Farber Cancer Institute's Physical Sciences Oncology Center (NCI U54CA209988 to F.M. and K.P.), and the National Cancer Institute grants F31CA239565 (S.S.) and R35CA197623 (K.P.).

Data availability:

All data generated or analyzed during this study are included in this published article (and its supplementary information files).

REFERENCES

1. Marusyk A, Janiszewska M, Polyak K. Intratumor Heterogeneity: The Rosetta Stone of Therapy Resistance. *Cancer Cell* 2020;37(4):471–84 doi 10.1016/j.ccell.2020.03.007. [PubMed: 32289271]
2. King CR, Kraus MH, Aaronson SA. Amplification of a novel v-erbB-related gene in a human mammary carcinoma. *Science* 1985;229(4717):974–6. [PubMed: 2992089]
3. Slamon DJ, Clark GM, Wong SG, Levin WJ, Ullrich A, McGuire WL. Human breast cancer: correlation of relapse and survival with amplification of the HER-2/neu oncogene. *Science* 1987;235(4785):177–82. [PubMed: 3798106]
4. Hayes DF. HER2 and Breast Cancer - A Phenomenal Success Story. *N Engl J Med* 2019;381(13):1284–6 doi 10.1056/NEJMcibr1909386. [PubMed: 31502769]
5. von Minckwitz G, Huang CS, Mano MS, Loibl S, Mamounas EP, Untch M, et al. Trastuzumab Emtansine for Residual Invasive HER2-Positive Breast Cancer. *N Engl J Med* 2019;380(7):617–28 doi 10.1056/NEJMoa1814017. [PubMed: 30516102]

6. von Minckwitz G, Procter M, de Azambuja E, Zardavas D, Benyunes M, Viale G, et al. Adjuvant Pertuzumab and Trastuzumab in Early HER2-Positive Breast Cancer. *N Engl J Med* 2017;377(2):122–31 doi 10.1056/NEJMoa1703643. [PubMed: 28581356]
7. Cobleigh MA, Vogel CL, Tripathy D, Robert NJ, Scholl S, Fehrenbacher L, et al. Multinational study of the efficacy and safety of humanized anti-HER2 monoclonal antibody in women who have HER2-overexpressing metastatic breast cancer that has progressed after chemotherapy for metastatic disease. *J Clin Oncol* 1999;17(9):2639–48 doi 10.1200/JCO.1999.17.9.2639. [PubMed: 10561337]
8. Slamon DJ, Leyland-Jones B, Shak S, Fuchs H, Paton V, Bajamonde A, et al. Use of chemotherapy plus a monoclonal antibody against HER2 for metastatic breast cancer that overexpresses HER2. *N Engl J Med* 2001;344(11):783–92 doi 10.1056/NEJM200103153441101. [PubMed: 11248153]
9. Wolff AC, Hammond ME, Schwartz JN, Hagerty KL, Allred DC, Cote RJ, et al. American Society of Clinical Oncology/College of American Pathologists guideline recommendations for human epidermal growth factor receptor 2 testing in breast cancer. *J Clin Oncol* 2007;25(1):118–45. [PubMed: 17159189]
10. Wolff AC, Hammond ME, Hicks DG, Dowsett M, McShane LM, Allison KH, et al. Recommendations for human epidermal growth factor receptor 2 testing in breast cancer: American Society of Clinical Oncology/College of American Pathologists clinical practice guideline update. *J Clin Oncol* 2013;31(31):3997–4013 doi 10.1200/JCO.2013.50.9984. [PubMed: 24101045]
11. Wolff AC, Hammond MEH, Allison KH, Harvey BE, Mangu PB, Bartlett JMS, et al. Human Epidermal Growth Factor Receptor 2 Testing in Breast Cancer: American Society of Clinical Oncology/College of American Pathologists Clinical Practice Guideline Focused Update. *J Clin Oncol* 2018;36(20):2105–22 doi 10.1200/JCO.2018.77.8738. [PubMed: 29846122]
12. Hanna WM, Ruschoff J, Bilous M, Coudry RA, Dowsett M, Osamura RY, et al. HER2 in situ hybridization in breast cancer: clinical implications of polysomy 17 and genetic heterogeneity. *Mod Pathol* 2014;27(1):4–18 doi 10.1038/modpathol.2013.103.
13. Vance GH, Barry TS, Bloom KJ, Fitzgibbons PL, Hicks DG, Jenkins RB, et al. Genetic heterogeneity in HER2 testing in breast cancer: panel summary and guidelines. *Arch Pathol Lab Med* 2009;133(4):611–2 doi 10.1043/1543-2165-133.4.611. [PubMed: 19391661]
14. Bernasconi B, Chiaravalli AM, Finzi G, Milani K, Tibiletti MG. Genetic heterogeneity in HER2 testing may influence therapy eligibility. *Breast Cancer Res Treat* 2012;133(1):161–8 doi 10.1007/s10549-011-1744-3. [PubMed: 21901388]
15. Bartlett JM, Starczynski J, Atkey N, Kay E, O’Grady A, Gandy M, et al. HER2 testing in the UK: recommendations for breast and gastric in-situ hybridisation methods. *J Clin Pathol* 2011;64(8):649–53 doi 10.1136/jcp.2011.089847. [PubMed: 21690244]
16. Curigliano G, Burstein HJ, Winer EP, Gnant M, Dubsy P, Loibl S, et al. De-escalating and escalating treatments for early-stage breast cancer: the St. Gallen International Expert Consensus Conference on the Primary Therapy of Early Breast Cancer 2017. *Ann Oncol* 2017;28(8):1700–12 doi 10.1093/annonc/mdx308. [PubMed: 28838210]
17. Burstein HJ, Curigliano G, Loibl S, Dubsy P, Gnant M, Poortmans P, et al. Estimating the benefits of therapy for early-stage breast cancer: the St. Gallen International Consensus Guidelines for the primary therapy of early breast cancer 2019. *Ann Oncol* 2019;30(10):1541–57 doi 10.1093/annonc/mdz235. [PubMed: 31373601]
18. Magurran AE. *Measuring biological diversity*. Malden: Blackwell; 2004. 256 p.
19. Janiszewska M, Liu L, Almendro V, Kuang Y, Paweletz C, Sakr RA, et al. In situ single-cell analysis identifies heterogeneity for PIK3CA mutation and HER2 amplification in HER2-positive breast cancer. *Nat Genet* 2015;47(10):1212–9 doi 10.1038/ng.3391. [PubMed: 26301495]
20. Rye IH, Trinh A, Saetersdal AB, Nebdal D, Lingjaerde OC, Almendro V, et al. Intratumor heterogeneity defines treatment-resistant HER2+ breast tumors. *Molecular oncology* 2018;12(11):1838–55 doi 10.1002/1878-0261.12375. [PubMed: 30133130]
21. Zhao J, Krishnamurti U, Zhang C, Meisel J, Wei Z, Suo A, et al. HER2 immunohistochemistry staining positivity is strongly predictive of tumor response to neoadjuvant chemotherapy in HER2 positive breast cancer. *Pathol Res Pract* 2020;216(11):153155 doi 10.1016/j.prp.2020.153155.

22. Gini B On the Measure of Concentration with Special Reference to Income and Statistics. Colorado College Publication, General Series 1936;208:73–9.
23. Reiter JG, Makohon-Moore AP, Gerold JM, Heyde A, Attiyeh MA, Kohutek ZA, et al. Minimal functional driver gene heterogeneity among untreated metastases. *Science* 2018;361(6406):1033–7 doi 10.1126/science.aat7171. [PubMed: 30190408]
24. Hinohara K, Polyak K. Intratumoral Heterogeneity: More Than Just Mutations. *Trends Cell Biol* 2019;29(7):569–79 doi 10.1016/j.tcb.2019.03.003. [PubMed: 30987806]
25. Koren S, Bentires-Alj M. Breast Tumor Heterogeneity: Source of Fitness, Hurdle for Therapy. *Mol Cell* 2015;60(4):537–46 doi 10.1016/j.molcel.2015.10.031. [PubMed: 26590713]
26. Hurvitz SA, Martin M, Symmans WF, Jung KH, Huang CS, Thompson AM, et al. Neoadjuvant trastuzumab, pertuzumab, and chemotherapy versus trastuzumab emtansine plus pertuzumab in patients with HER2-positive breast cancer (KRISTINE): a randomised, open-label, multicentre, phase 3 trial. *Lancet Oncol* 2018;19(1):115–26 doi 10.1016/S1470-2045(17)30716-7. [PubMed: 29175149]
27. Modi S, Park H, Murthy RK, Iwata H, Tamura K, Tsurutani J, et al. Antitumor Activity and Safety of Trastuzumab Deruxtecan in Patients With HER2-Low-Expressing Advanced Breast Cancer: Results From a Phase Ib Study. *J Clin Oncol* 2020;38(17):1887–96 doi 10.1200/JCO.19.02318. [PubMed: 32058843]
28. Metzger-Filho O, Winer EP. Breast cancer: T-DM1 - an important agent in the history of breast cancer management. *Nat Rev Clin Oncol* 2017;14(11):651–2 doi 10.1038/nrclinonc.2017.123. [PubMed: 28786416]
29. Harbeck N, Gluz O, Christgen M, Kates RE, Braun M, Kuemmel S, et al. De-Escalation Strategies in Human Epidermal Growth Factor Receptor 2 (HER2)-Positive Early Breast Cancer (BC): Final Analysis of the West German Study Group Adjuvant Dynamic Marker-Adjusted Personalized Therapy Trial Optimizing Risk Assessment and Therapy Response Prediction in Early BC HER2- and Hormone Receptor-Positive Phase II Randomized Trial-Efficacy, Safety, and Predictive Markers for 12 Weeks of Neoadjuvant Trastuzumab Emtansine With or Without Endocrine Therapy (ET) Versus Trastuzumab Plus ET. *J Clin Oncol* 2017;35(26):3046–54 doi 10.1200/JCO.2016.71.9815. [PubMed: 28682681]
30. Bergh JCS, Andersson A, Bjohle J, Bosch A, Carlsson L, Dreifaldt AC, et al. Docetaxel, trastuzumab, pertuzumab versus trastuzumab emtansine as neoadjuvant treatment of HER2-positive breast cancer: Results from the Swedish PREDIX HER2 trial identifying a new potential de-escalation standard? *Journal of Clinical Oncology* 2019;37(15_suppl):501- doi 10.1200/JCO.2019.37.15_suppl.501.
31. Hurvitz SA, Martin M, Jung KH, Huang CS, Harbeck N, Valero V, et al. Neoadjuvant Trastuzumab Emtansine and Pertuzumab in Human Epidermal Growth Factor Receptor 2-Positive Breast Cancer: Three-Year Outcomes From the Phase III KRISTINE Study. *J Clin Oncol* 2019;37(25):2206–16 doi 10.1200/JCO.19.00882. [PubMed: 31157583]
32. Cortazar P, Zhang L, Untch M, Mehta K, Costantino JP, Wolmark N, et al. Pathological complete response and long-term clinical benefit in breast cancer: the CTNeoBC pooled analysis. *Lancet* 2014;384(9938):164–72 doi 10.1016/S0140-6736(13)62422-8. [PubMed: 24529560]
33. Lee HJ, Kim JY, Park SY, Park IA, Song IH, Yu JH, et al. Clinicopathologic Significance of the Intratumoral Heterogeneity of HER2 Gene Amplification in HER2-Positive Breast Cancer Patients Treated With Adjuvant Trastuzumab. *Am J Clin Pathol* 2015;144(4):570–8 doi 10.1309/AJCP51HCGPOPWSCY. [PubMed: 26386078]
34. Walker RA, Bartlett JM, Dowsett M, Ellis IO, Hanby AM, Jasani B, et al. HER2 testing in the UK: further update to recommendations. *J Clin Pathol* 2008;61(7):818–24. [PubMed: 18381380]

SIGNIFICANCE

HER2-targeted therapies improve cure rates in HER2-positive breast cancer, suggesting chemotherapy can be avoided in a subset of patients. We show that HER2 heterogeneity, particularly the fraction of *ERBB2* non-amplified cancer cells, is a strong predictor of resistance to HER2 therapies and could potentially be used to optimize treatment selection.

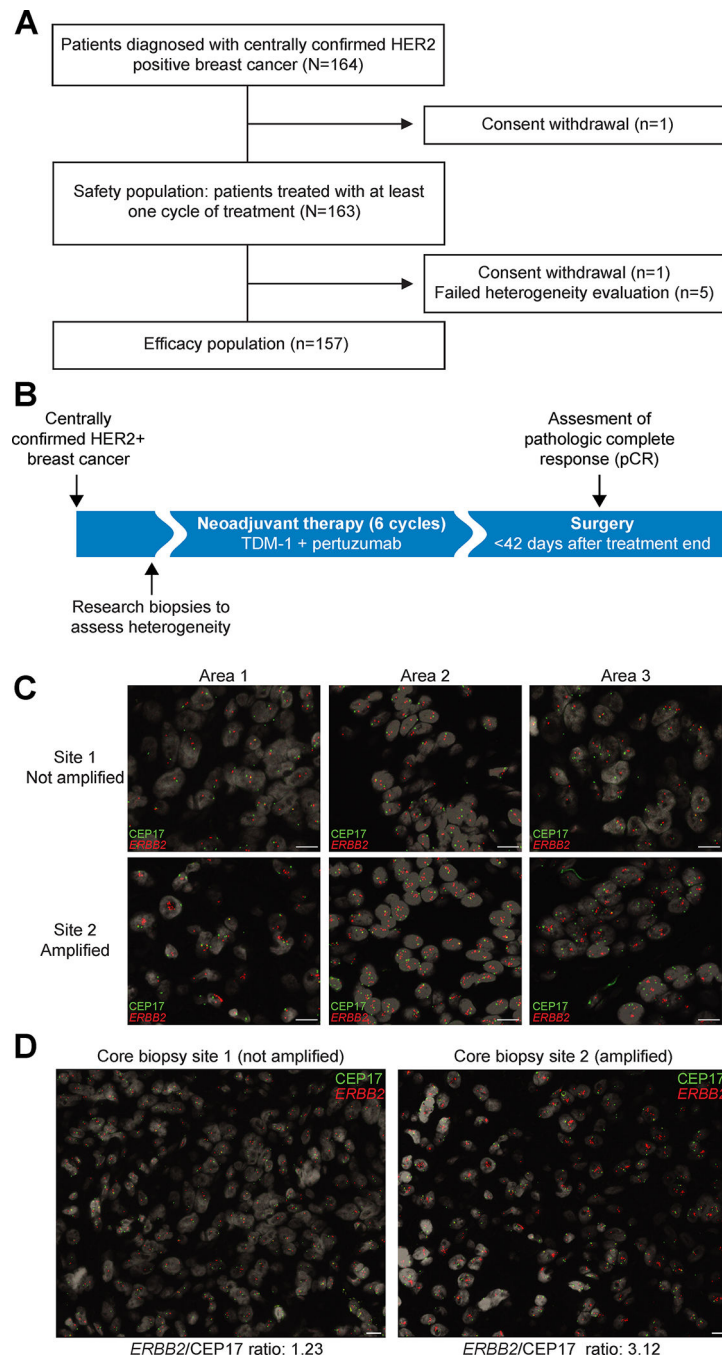


Figure 1. Study Design.

A, CONSORT diagram. **B**, Study design. Centrally confirmed HER2-positive breast cancer patients were treated in a single arm study. Treatment consisted of six infusions of T-DM1 given in combination with pertuzumab. **C**, Example of central pathology evaluation of HER2 heterogeneity assessed by FISH, with CEP17 probe in green and *ERBB2* in red. *ERBB2* copy number counting was performed in three different areas per core biopsy site counting approximately 50 cells in each area. Scale bar corresponds to 10 μ m. **D**, A representative

case of HER2 heterogeneity with core biopsy site 1 showing HER2-negative cancer and core biopsy site 2 HER2-amplified cancer. Scale bar corresponds to 10 μm .

Author Manuscript

Author Manuscript

Author Manuscript

Author Manuscript

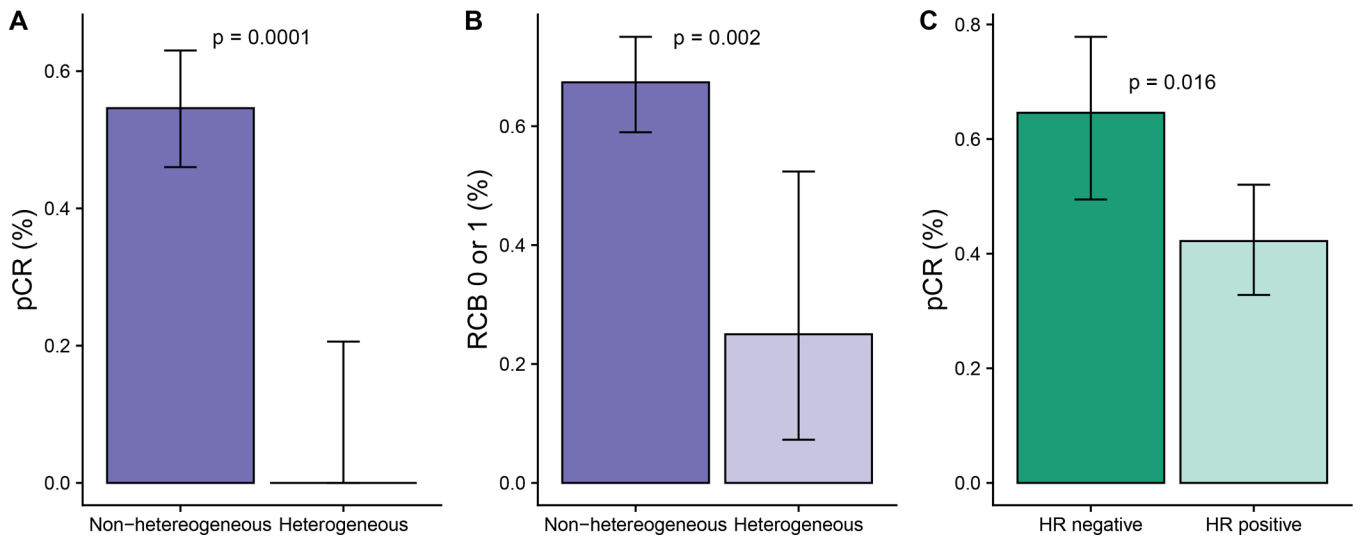


Figure 2. Pathological response according to HER2 heterogeneity and hormone receptor status. The mean and 95% confidence interval of the mean is shown in each graph. **A**, pCR by HER2 heterogeneity. **B**, Pathological response defined as RCB 0 or I by HER2 heterogeneity. **C**, pCR by hormone receptor (HR) status for the overall study population. **A-C**, P-values are from a χ^2 test.

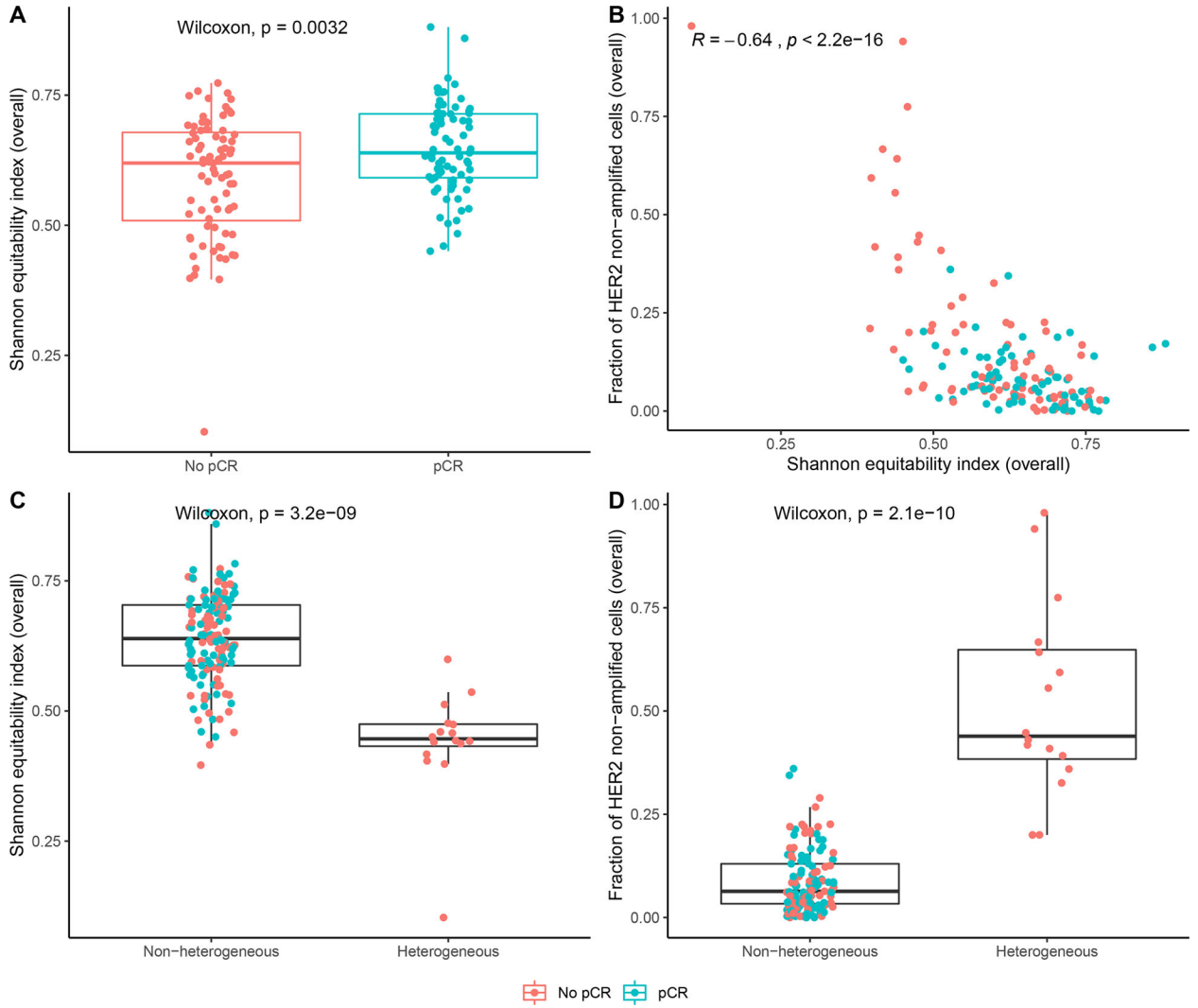


Figure 3. Assessment of HER2 heterogeneity at the single cell level.

A, Shannon equitability index for each patient, separated by response. P-value is from a Wilcoxon test. **B**, Shannon equitability index (x-axis) vs. frequency of *ERBB2* non-amplified cells. Each datapoint represents one patient. R and p-value are from a Pearson correlation test of association. Red represents non-responders and blue represents responders. **C**, Shannon equitability index for each patient, separated by central pathology evaluation of HER2 heterogeneity. P-value is from a Wilcoxon test. **D**, Frequency of *ERBB2* non-amplified cells in each patient, separated by central pathology evaluation of HER2 heterogeneity. P-value is from a Wilcoxon test. **A, C-D**, The lower and upper hinges correspond to the 1st and 3rd quartiles, respectively. The lower and upper whiskers extend to the smallest and largest values, respectively, no further than 1.5*IQR from the hinge, where the IQR is the distance between the 1st and 3rd quartiles. All data points are plotted.

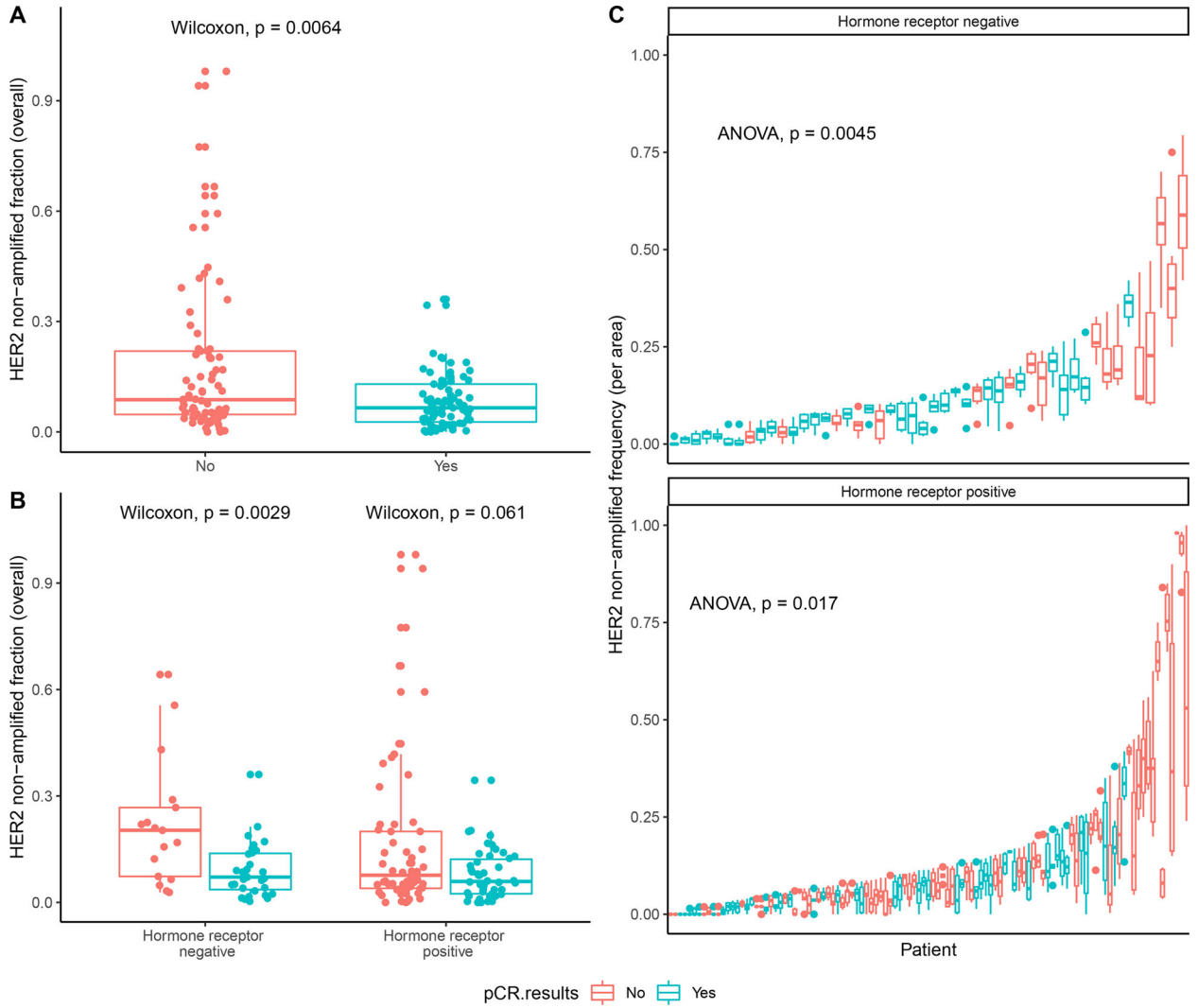


Figure 4. Association between HER2 negativity and pathologic complete response (pCR). **A**, Fraction of *ERBB2* non-amplified cells in each patient. P-value is from a Wilcoxon test. **B**, Fraction of *ERBB2* non-amplified cells in each patient separated by HR status: HR-negative cases and HR-positive cases. P-values are from a Wilcoxon test. **C**, Frequency of *ERBB2* non-amplified cells in each patient separated by HR status: HR-negative patients (top) and HR-positive cases (bottom). Each boxplot shows the fraction of HER2 non-amplified cells in each area for one patient, where patients are ordered by the maximum fraction of *ERBB2* non-amplified cells found in any area and stratified by hormone receptor status. Responders (i.e., patients achieving pCR) are enriched on the left side of the plot with a lower fraction of *ERBB2* non-amplified cells, and non-responders are enriched on the right side of the plot with a higher fraction of *ERBB2* non-amplified cells. P-values are from an F-test with Satterthwaite’s method. **A-C**, Red represents patients without pCR and blue represents patients with pCR. The lower and upper hinges correspond to the 1st and 3rd quartiles, respectively. The lower and upper whiskers extend to the smallest and largest values, respectively, no further than $1.5 \times \text{IQR}$ from the hinge, where the IQR is the distance

between the 1st and 3rd quartiles. All data points are plotted in **A-B**, and data points outside of the whiskers are plotted separately in **C**.

Table 1:

Patient and Disease Characteristics

Characteristic N (%)	T-DM1 plus pertuzumab N = 163
Age	
< 35	16 (9.8)
35–50	50 (30.7)
51–65	65 (39.9)
> 65	32 (19.6)
Race	
White	140 (85.8)
Black	6 (3.6)
Asian	10 (6.1)
Missing	7 (4.3)
Median tumor size (IQR)	2.8 cm (2.1–3.8 cm)
Hormone receptor status, n (%)	
ER+ and/or PR+	112 (68.7)
ER– and PR–	51 (31.3)
Clinical stage	
I	1 (0.6)
II	138 (84.7)
III	24 (14.7)
HER2 IHC (central evaluation)	
2+	40 (24.5)
3+	121 (74.2)
Missing	2 (1.2)
Histologic Grade	
1	3 (1.8)
2	56 (34.4)
3	103 (63.2)
Missing	1 (0.6)
Surgery *	
Wide local excision	83 (51.2)
Mastectomy	79 (48.8)

* One patient withdrew consent prior to surgery.

Abbreviations: ER, estrogen receptor; IHC, immunohistochemistry; IQR, interquartile range; PR, progesterone receptor

Table 2:

Summary of Adverse Events in the Safety Population *

Adverse Events	Pertuzumab + TDM1 (n=163)
Number of patients with at least one, n (%)	
Grade 1	155 (95.1)
Grade 2	85 (52.1)
Grade 3 AEs	12 (7.4)
AE leads to Tx discontinuation	2 (1.2)
AE with fatal outcome	0 (0)
Grade 3 AEs that occurred in 1% of patients	
Diarrhea	3 (1.8)
Hypokalemia	2 (1.2)
Platelet count decreased	2 (1.2)

* Listed are all AEs regardless of attribution to study drugs with an onset that occurred from the first dose of treatment through the first study visit after breast surgery.

Abbreviations: AE, adverse event; Tx, treatment

Author Manuscript

Author Manuscript

Author Manuscript

Author Manuscript

Table 3:

Model AUC values

Model		AUC value			
		All observed data	Randomly sampling one biopsy	Randomly sampling one area per core biopsy	Randomly adding HER2 copies
Classification based on the fraction of HER2 non-amplified cells	All patients	0.62	0.63 (0.60–0.66)	0.61 (0.56–0.65)	--
	HR+	0.59	0.61 (0.57–0.65)	0.60 (0.54–0.66)	--
	HR–	0.75	0.72 (0.67–0.77)	0.69 (0.61–0.77)	--
Logistic regression	Model 1	0.70	0.69 (0.66–0.72)	0.68 (0.65–0.72)	0.69 (0.65–0.72)
	Model 2	0.65	0.65 (0.64–0.67)	0.65 (0.64–0.68)	--
	Model 3	0.70	0.69 (0.67–0.72)	0.68 (0.65–0.72)	--
	Model 4	0.71	0.69 (0.66–0.72)	0.68 (0.65–0.72)	--
	Model 5	0.72	0.72 (0.69–0.75)	0.71 (0.68–0.75)	--
	Model 6	0.71	0.70 (0.68–0.73)	0.70 (0.66–0.73)	--
	Model 7	0.69	0.69 (0.67–0.71)	0.69 (0.66–0.72)	--

Author Manuscript

Author Manuscript

Author Manuscript

Author Manuscript