Identification of a subpopulation of metastatic breast cancer patients with very high HER2 expression levels and possible resistance to trastuzumab


1Division of Clinical Research; 2Division of Research and Development, Monogram Biosciences, South San Francisco, USA; 3Clinical Division of Oncology, Department of Medicine I, Medical University of Vienna, Vienna, Austria; 4Department of Medicine, Lebanon Veterans Affairs Medical Center, Lebanon; 5Department of Medicine, Division of Hematology/Medical Oncology, Penn State Hershey Medical Center, Hershey; 6Department of Translational Medicine and Biomarker Development, Division of Biostatics and Bioinformatics, Monogram Biosciences, South San Francisco, USA; Departments of 7Clinical Pathology; 8Obstetrics and Gynecology, Medical University of Vienna, Vienna, Austria

Received 21 July 2010; revised 3 November 2010; accepted 4 November 2010

Background: Patients with metastatic breast cancer (MBC) overexpressing HER2 (human epidermal growth factor receptor 2) are currently selected for treatment with trastuzumab, but not all patients respond.

Patients and methods: Using a novel assay, HER2 protein expression (H2T) was measured in formalin-fixed, paraffin-embedded primary breast tumors from 98 women treated with trastuzumab-based therapy for MBC. Using subpopulation treatment effect pattern plots, the population was divided into H2T low (H2T < 13.8), H2T high (H2T ≥ 68.5), and H2T intermediate (13.8 ≤ H2T < 68.5) subgroups. Kaplan–Meier (KM) analyses were carried out comparing the groups for time to progression (TTP) and overall survival (OS). Cox multivariate analyses were carried out to identify correlates of clinical outcome. Bootstrapping analyses were carried out to test the robustness of the results.

Results: TTP improved with increasing H2T until, at the highest levels of H2T, an abrupt decrease in the TTP was observed. KM analyses demonstrated that patients with H2T low tumors [median TTP 4.2 months, hazard ratio (HR) = 3.7, P = 0.0001] or H2T high tumors (median TTP 4.6 months, HR = 2.7, P = 0.008) had significantly shorter TTP than patients whose tumors were H2T intermediate (median TTP 12 months). OS analyses yielded similar results.

Conclusions: MBC patients with very high levels of H2T may represent a subgroup with de novo resistance to trastuzumab. These results are preliminary and require confirmation in larger controlled clinical cohorts.

Key words: HER2, metastatic breast cancer, trastuzumab

Introduction

HER2 (human epidermal growth factor receptor 2), a member of the HER family of receptor tyrosine kinases, has been shown to confer an adverse prognosis in breast cancer and has been identified as a predictor of clinical response to treatment with the humanized monoclonal antibody trastuzumab. Trastuzumab targets HER2 and is effective in the treatment of both metastatic breast cancer (MBC) and primary HER2-overexpressing breast cancer when used in combination with chemotherapy [1–6]. HER2 positivity is routinely assessed by immunohistochemistry (IHC), which measures HER2 protein expression semiquantitatively, or by FISH, which measures HER2 gene copy number. Whereas patients who overexpress HER2 protein (IHC) or display evidence of HER2 gene amplification (FISH) have been shown to experience better outcomes on trastuzumab than those scored negative by these assays, differences in the degree of expression or amplification by these methods have generally not been shown to discriminate between groups of patients with different outcomes. While several small studies have suggested possible correlations between differing levels of HER2 gene amplification by FISH and response to trastuzumab [7–9], examination of large well-controlled datasets has failed to find a significant association [10, 11].

The HERmark® assay uses a multiple antibody approach to make precise measurements of HER2 protein expression (H2T) with greater sensitivity than IHC [12, 13]. Assay validation data developed at Monogram comparing HERmark measurements with those of centrally carried out IHC in cell lines prepared as formalin-fixed, paraffin-embedded (FFPE) pellets, as well as with quantitative HER2 by enzyme-linked immunosorbent assay and quantitative HER2 expression by flow cytometry, demonstrate excellent concordance with all three measurements [12]. However, HERmark showed greater ability to accurately measure low-level HER2 expression than IHC and
also provided a more granular assessment of HER2 expression levels in a population viewed by IHC simply as ‘3+’. As a result, HERmark is able to measure a distribution of HER2 expression extending over a large dynamic range corresponding to \( \sim 2500 \) to 2 million receptors per cell.

In a cohort of patients with MBC, we have previously shown that HER2 expression levels (H2T) correlated with response to trastuzumab-based treatment such that patients with higher H2T experienced better outcomes than those with lower H2T, regardless of HER2 gene amplification status [14]. Those analyses used categorical values to define high and low H2T and the method of Kaplan–Meier (KM) to examine outcomes in those subgroups. We had also considered H2T as a continuous variable in Cox proportional hazards analyses. In both sets of analyses, we made the general observation that patients with high H2T values experienced better outcomes on trastuzumab than those with low H2T values. However, we subsequently realized that neither of those analytical methods would allow us to detect subpopulations of outliers if the H2T–outcome relationship was nonlinear. Given the ability of the HERmark assay to measure the continuum of HER2 expression quantitatively, and in light of the fact that many patients who overexpress HER2 nonetheless fail to respond optimally to trastuzumab, we used the same patient population as previously described [14] but employed different analytical methods [subpopulation treatment effect pattern plots (STEPP)] to investigate whether particular subgroups of patients with varying levels of HER2 expression responded differently to trastuzumab. We hypothesized that clinical outcomes on trastuzumab would continue to improve as H2T levels increased and that this would be true even within the subgroup of patients characterized as central FISH positive for HER2 gene amplification and also H2T high (concordant positives). Thus, the important difference between the current study and the prior description of this MBC cohort lies in the identification of another subgroup of patients with very high H2T levels who experience clinical outcomes that are distinct from those of patients who also overexpress HER2 but to a lesser degree.

**methods**

**study population**

This cohort comprised patients with HER2 overexpressing (>10% of tumor cells IHC 3+ as determined by the HercepTest; DAKO Diagnostics, Vienna, Austria) and/or HER2-amplified MBC who were prospectively observed during trastuzumab-based therapy at the Medical University of Vienna between 1999 and 2006 [14]. The clinical characteristics of the cohort are described in supplemental Table S1 (available at *Annals of Oncology* online). Patients were required to be trastuzumab naïve at study entry and have bidimensionally measurable disease progressing within 4 weeks before initiation of treatment (excluding previously irradiated lesions). Pregnant or breast-feeding patients and those with a history of congestive heart failure, ischemic heart disease, second malignancy (except *in situ* cervical cancer, adequately treated basal cell or squamous cell carcinoma of the skin), severe hepatic or renal dysfunction, or altered mental status were not eligible. Written informed consent was obtained for documentation of disease and treatment-related data as well as subsequent analysis of tumor tissue before initiating trastuzumab; however, not all consented patients had tissue available for testing (Figure 2). All treatment decisions were made by the treating physicians, not the investigators. All consented patients received planned therapy. Response to treatment was documented by review of all imaging studies according to the Southwest Oncology Group criteria as previously described [15], and data on HER2 and hormone receptor status were retrieved from pathology reports. Exposure to chemotherapy, either before trastuzumab or during trastuzumab therapy, was heterogeneous. Tissue for analysis was derived from the primary tumor. Patients were initially selected for trastuzumab-based treatment using local IHC with confirmatory FISH in those patients assessed as IHC 2+. Central FISH was carried out, in blinded fashion, on all patients after HERmark testing for H2T. The research protocol was approved by the Penn State Hershey Medical Center and the Medical University of Vienna institutional review boards.

**the HERmark assay**

The HERmark H2T assay is an application of the VeraTag™ technology platform designed specifically for breast cancer. VeraTag is a proximity-based method designed to accurately and reproducibly quantify protein expression and protein–protein complexes, including cell surface dimers, in FFPE specimens [13]. The assay has been technically validated according to the specifications prescribed by the Clinical Laboratory Improvement Amendments and is carried out in a College of American Pathologists-certified clinical reference laboratory at Monogram Biosciences Inc. in South San Francisco, CA.

**central FISH testing**

Detection of HER2 gene amplification in HERmark-tested specimens was carried out in one batch by a single pathologist at the Medical University of Vienna using the Vysis assay (Abbott-Vysis Inc., Downers Grove, IL). The central pathologist was blinded to the local HER2 results and to the HERmark results. FISH assays were carried out using contiguous sections from the same tissue block as the HERmark measurements. For each specimen, the HER2 and CEP17 signals of >50 nuclei of invasive tumor areas were recorded. Tumors were classified as FISH positive if the HER2/CEP17 ratio exceeded 2.2.

**data blinding and statistical methods**

All HERmark assays were carried out retrospectively and in blinded fashion, and the treating physicians were blinded to the results of central FISH and the HERmark results. After testing, H2T results were transmitted to Penn State Hershey Medical Center before receipt of clinical data at Monogram. Statistical analyses were carried out independently at Monogram and Penn State and then compared. Determination of the original clinical cut-off for H2T (discriminating ‘low’ from ‘high’ H2T [14]) occurred before carrying out central FISH assays at Vienna, by positional scanning analysis and selection of the cut-off associated with the lowest \( P \) value for time to progression (TTP). The second clinical cut-off, suggested by STEPP analyses and discriminating between patients with ‘intermediate’ and ‘high’ H2T, was also defined using positional scanning and selection of the lowest \( P \) value for TTP, but this cut-off was identified after central FISH testing.

Correlation of H2T with TTP and overall survival (OS) by KM analysis used the log-rank test, and multivariate analyses used Cox proportional hazards regression in the R statistical package (version 2.10.1). Graphpad Prism (version 5) was used to generate all graphs. All \( P \) values are two-tailed. TTP was defined as the time from the initiation of trastuzumab-based treatment to progression or censor, and OS was defined as the time from initiation of trastuzumab-based treatment to death or censor. Patients were classified as low, intermediate, or high for H2T based on the clinical cut-off originally identified for TTP (H2T \( \geq 13.8 \)) [14] and the second cut-off suggested by STEPP analyses and identified by positional scanning.
(H2T = 68.5). High, low, and intermediate groups were compared for time-to-event end points using KM analysis. Cox proportional hazards models were fitted to determine the most significant correlates with outcome, and the assumption of proportional hazards was tested for each variable. Potential prognostic factors included H2T greater than or equal to versus lower than cut-off (using the cut-offs 13.8 and 68.5), continuous values of H2T, HER2 status by FISH (amplified versus not amplified where amplified was defined as having ≥2 HER2 copies per chromosome 17 centromere), HER2 total copy number by FISH, HER2 copy number per chromosome 17 centromere copy number by FISH (HER2/CEP17), IHC score, age, treatment group (trastuzumab only or trastuzumab plus chemotherapy), estrogen and progesterone receptor status, time from metastasis to treatment, and time from surgery to treatment. Total number of sites of metastasis was considered both as a continuous and a categorical (<3 versus ≥3 sites) variable. All possible Cox models were tested and those with the lowest P values were considered the best models.

Bootstrapping analyses were carried out to test the robustness of the H2T cut-off of 68.5. Two different approaches were used. First, we iteratively sampled, with replacement, a new population of the same size and calculated the hazard ratio (HR) comparing the very high H2T group and the intermediate H2T group for TTP and OS. A P value was calculated from the percentage of point estimate HRs of the 10 000 sample populations that were ≥1. The calculation was initially carried out without noise and then repeated with the addition of noise, modeling the H2T variable as log-normal distribution with σ = 0.075. Secondly, three datasets \[N = 98, N = 980 (each patient entered 10 times), and N = 9800 (each patient entered 100 times)] were created with the addition of noise given as log-normal distribution with σ = 0.075. One thousand iterations were carried out at each sample size, and HRs and P values were calculated for TTP.

**results**

The particular question we were interested in asking was whether patients that were ‘HER2 positive’ and predicted to have the best outcomes following trastuzumab therapy experienced different clinical outcomes on trastuzumab depending on the degree of HER2 overexpression that their tumors exhibited. To examine this, we carried out a STEPP analysis [16] and interrogated the outcome of patients along the entire distribution of H2T from low to high, anticipating that higher levels of H2T would always correlate with better outcomes on trastuzumab when compared with patients with lower levels of H2T. However, we unexpectedly observed that the H2T–trastuzumab response relationship appeared to be nonlinear. Using progression-free survival at 12 months as the outcome measure, we noted that higher H2T levels correlated with better outcomes throughout most of the H2T dynamic range but that there was a marked dip in the curve near the top of the H2T distribution (Figure 1A). We also carried out a similar STEPP analysis using the HER2 gene copy number ratio (HER2/CEP17). After an initial improvement in outcome for patients with copy numbers above ∼2–2.5, the likelihood of being progression free at 12 months following initiation of trastuzumab appeared to plateau for copy numbers >3 (Figure 1B).

While STEPP analyses are useful in detecting the presence of subpopulations with different outcomes within a larger cohort, they are limited in their ability to define the boundaries of such populations because the variables, as well as the outcomes, are average values derived from all the patients included in each

**Figure 1.** (A) Subpopulation treatment effect pattern plots (STEPPs) examining the relationship between the continuous variable H2T (HER2 protein expression) and the probability of being progression free at 12 months following the initiation of trastuzumab-based treatment for metastatic breast cancer. The bin size used for this analysis was 30. The second-order curve fit is shown in red. (B) STEPP analyses examining the probability of being progression free at 12 months according to the HER2/CEP17 copy number ratio.

bin. Therefore, we used positional scanning to determine the cut-off that discriminated between the subgroup at the high end of the H2T distribution that appeared to be driving this discontinuity in the STEPP curve and the rest of the population (data not shown). The cut-off was determined to be 68.5. Using this cut-off in addition to the previously described cut-off of 13.8 [14], we divided the population into three subgroups: (i) H2T low (H2T < 13.8), (ii) H2T intermediate (13.8 ≤ H2T < 68.5), and (iii) H2T high (H2T ≥ 68.5). Essentially, we had now taken the group previously characterized as ‘H2T high’ in our previous study [14] and subdivided that group into two groups—‘H2T intermediate’ and ‘H2T high’ (Figure 2).

Assessments of HER2 positivity by local IHC and central FISH in these three subgroups of patients were examined. In the low H2T group, 97% were IHC 3+ but only 34% were central FISH positive. In the H2T intermediate group, 94% were IHC 3+ and 94% were central FISH positive. In the H2T high group, 87% were IHC 3+ and 100% were central FISH positive (Table 1). When compared using total or CEP17-corrected HER2 copy numbers by FISH, the three groups showed overlapping distributions, with statistical differences observed for comparisons of the H2T low with the intermediate and high groups (Figure 3A and B). Importantly, on the basis of total or CEP17-corrected HER2 copy numbers, one would not have been able to discriminate whether an individual patient was in the H2T intermediate or H2T high group.

We used KM analysis to compare the outcomes of the three H2T groups following trastuzumab treatment using TTP and OS as end points. As shown in Figure 4A and B, the high H2T subgroup experienced TTP and OS that were indistinguishable compared with patients who had low H2T levels. The subgroup that had intermediate H2T levels experienced the longest TTP and OS, at 12 and 38.6 months, respectively. Compared with the intermediate group, the low H2T group (median TTP = 4.2 months, HR = 3.7, P < 0.0001; median OS = 28.7 months, HR 2.0, P = 0.04) as well as the high H2T group (median TTP = 4.6
months, HR = 2.7, P = 0.008; median OS = 28.6 months, HR = 2.4, P = 0.04) experienced significantly worse outcomes (Figure 4C).

Finally, we carried out multivariate Cox proportional hazards regression analyses to examine the impact of multiple variables, including this new subgroup with H2T values ≥68.5, on TTP and OS in this cohort. The best model for TTP included the high H2T subgroup, as defined by the cut-off 68.5, as a significant correlate of outcome (HR = 4.03, P < 0.001) (Table 2). The high H2T subgroup was not included in the best model for OS, but some OS models did include it as a significant variable (data not shown). The assumption of proportional hazards was tested and there was no evidence of time dependence for any of the significant variables.

Because this dataset is small and the very high H2T subgroup includes only 15 patients, there is a concern that the significant relationship we observe using a cut-off of 68.5 is the result of random variation, i.e. a statistical artifact. To address this issue, we carried out bootstrapping analyses using two different approaches. The first method, using iterative sampling with replacement (described in ‘Methods’ section), demonstrated P values associated with the probability of finding an HR ≤1 (i.e. different from our observation that the HR comparing outcomes of patients with very high H2T values with those with intermediate H2T values is >1) of 0.012 for TTP and 0.032 for OS without noise, and 0.086 for TTP and 0.17 for OS with the addition of noise. However, these results are complicated by the fact that the cut-off may have changed if positional scanning analyses were carried out on each dataset after the introduction of noise, and using such small sample sizes, small changes in the cut-off could have a significant effect. Thus, we carried out a second bootstrapping analysis where sample sizes of 98, 980, and 9800 were used with the introduction of noise. One thousand iterations were carried out for each sample size, and the HRs and associated P values were calculated for TTP. Using the dataset of 98 with noise, the TTP HR was 1.56 ± 0.26, P = 0.013. Using the dataset of 980, the TTP HR was 1.55 ± 0.084, P = 0.001. Taken together, these bootstrapping analyses suggest that the existence of this very high H2T subpopulation is unlikely to be attributable to statistical artifact.

discussion

While trastuzumab-based systemic therapy is effective in the treatment of patients with HER2-positive breast cancer in both the primary and the metastatic settings, not every HER2-overexpressing tumor responds to trastuzumab-based treatment. Understanding why patients fail this therapy is critically important if we are to maximize the potential of trastuzumab to benefit patients with breast cancer, while at the same time enhancing cost-efficiency for the health care...
system. A number of hypotheses have been proposed to explain why some patients fail to achieve optimal outcomes on trastuzumab, including activation of the PI3 kinase signaling pathway as a result of HER2–HER3 heterodimerization, as well as constitutive activation of downstream signaling mediated by a truncated form of the HER2 receptor, p95/HER2 [15, 17–29]. Thus far however, quantitative measures of HER2 expression by IHC or HER2 gene copy number by FISH have not been reproducibly shown to correlate with degree of benefit from trastuzumab-based treatment [10].

Here, we report the discovery of a subpopulation of patients with very high H2T and evidence of HER2 gene amplification (100%) by central FISH who appear to respond suboptimally to trastuzumab. Their clinical outcomes following trastuzumab treatment are similar to patients with low levels of HER2 expression, of whom approximately two-thirds (66%) lack evidence of HER2 gene amplification. However, the very high HER2 patients fall significantly short of the outcomes observed for patients with intermediate levels of HER2 expression, of whom nearly all (94%) possess evidence of HER2 gene amplification. Indeed, these data suggest that the relationship between quantitative levels of H2T and response to trastuzumab in this cohort of patients with MBC is nonlinear. These data support prior observations of nonlinearity in the HER2–outcome relationship, although that analysis did not involve trastuzumab-treated patients [30].

Interpretations based on this cohort alone need to be tempered, as these data are the result of an exploratory retrospective analysis of a nonrandomized cohort with heterogeneous exposure to concomitant chemotherapy. Thus, even though the bootstrapping analyses offer support, confirmation of these findings in independent, larger well-controlled cohorts of patients with HER2-positive breast cancer is required before the clinical utility of such data can be properly determined. That said, recent analyses of the FinHer study, a prospectively designed clinical trial that incorporated randomization to trastuzumab or not for patients who were confirmed HER2 gene amplified by centrally carried out chromogenic in situ hybridization, using the same quantitative measure of H2T (by HERmark), also identified a subgroup of patients with the highest H2T that did not appear to benefit from trastuzumab, while patients with intermediate levels of HER2 expression did benefit [31]. The very high HER2 subgroup represented 13% of the HER2 positives in FinHer, similar to the current study, where it represented 16%. The optimal cut-offs for the high HER2 subgroup were slightly different in these two studies, possibly reflecting differences in the study populations and study designs, particularly the different clinical end points used in each study.

Recent analyses of the HERA Trial found that there was no statistically significant relationship between HER2 gene copy number by FISH and benefit from trastuzumab [10]. However, subgroup analyses conducted by the HERA investigators suggested that, consistent with our results from the bootstrapping analyses, patients with the highest levels of HER2 gene amplification in that trial may have failed to benefit from trastuzumab (see Figure 3 in Dowsett et al. [10]). Of note, the subset of patients with very high HER2 gene copy number identified in HERA contained 494 patients, making it unlikely that the observation was purely an artifact of small sample size. Similar results, also examining HER2 gene copy number, were recently reported in an analysis of the N9831 trial [11]. Thus, there are several lines of evidence from different studies, including large well-designed clinical trials of trastuzumab in the adjuvant setting, which support the notion that this subpopulation exists.

How can we explain this phenomenon? It is possible that extreme overexpression of HER2 on the cell surface may trigger biological changes that impact the activity of trastuzumab. Receptor crowding on the cell surface might restrict access of trastuzumab to its epitope target in the juxtamembrane domain of HER2. Pronounced overexpression of HER2 could lead to the generation of truncated p95/HER2 C-terminal fragments, hypothesized to be resistant to trastuzumab due to loss of the extracellular domain of HER2, and with it, the epitope target of the drug. Perhaps HER2 is more prone to form heterodimers...
with other signaling molecules at very high HER2 concentrations. Or, perhaps some patients simply need higher levels of trastuzumab to completely inhibit HER2 signaling when HER2 is overexpressed at very high levels. Whatever the mechanism, should the existence of this subpopulation be confirmed in yet additional cohorts, this observation offers the opportunity to understand why some patients fail to respond optimally to trastuzumab. In that case, experiments designed to explore the molecular mechanisms underlying this phenomenon are critically important, as they could lead to new therapeutic strategies and better outcomes for patients with HER2-positive breast cancer.

Figure 4. Kaplan–Meier analyses were carried out to compare the respective outcomes of patients in the low, intermediate, and high H2T subgroups following treatment with trastuzumab-based therapy. Results for (A) time to progression (TTP) and (B) overall survival (OS) are shown. (C) Using the H2T intermediate group as a comparator, hazard ratios and \( P \) values for the three groups are shown for the end points TTP and OS. H2T, HER2 protein expression.

Table 2. Cox proportional hazards multivariate analyses for time to progression (TTP) and overall survival (OS)

<table>
<thead>
<tr>
<th>Variable</th>
<th>Hazard ratio</th>
<th>( P ) value</th>
</tr>
</thead>
<tbody>
<tr>
<td>TTP</td>
<td></td>
<td></td>
</tr>
<tr>
<td>H2T (continuous)</td>
<td>0.43</td>
<td>0.027</td>
</tr>
<tr>
<td>H2T ( \geq ) 68.5</td>
<td>4.03</td>
<td>&lt;0.001</td>
</tr>
<tr>
<td>FISH ( \text{HER2/CEP17} ) copy number (continuous)</td>
<td>0.77</td>
<td>0.018</td>
</tr>
<tr>
<td>Trastuzumab only</td>
<td>1.99</td>
<td>0.043</td>
</tr>
<tr>
<td>Progesterone positive versus negative</td>
<td>0.44</td>
<td>0.009</td>
</tr>
<tr>
<td>No. of metastases (&lt;3 vs. ( \geq ) 3)</td>
<td>2.5</td>
<td>&lt;0.001</td>
</tr>
<tr>
<td>OS</td>
<td></td>
<td></td>
</tr>
<tr>
<td>H2T ( &lt;13.8 ) versus ( \geq 13.8 )</td>
<td>0.47</td>
<td>0.019</td>
</tr>
<tr>
<td>Age</td>
<td>0.97</td>
<td>0.01</td>
</tr>
<tr>
<td>No. of metastatic sites (continuous)</td>
<td>1.57</td>
<td>&lt;0.001</td>
</tr>
<tr>
<td>Estrogen positive versus negative</td>
<td>0.46</td>
<td>0.018</td>
</tr>
</tbody>
</table>

\( ^* \)Two patients excluded for missing data. H2T, HER2 protein expression.

acknowledgements

Data presented in part at the 45th ASCO annual meeting, 29 May to 2 June 2009 Orlando, FL.

funding

Komen for the Cure (BCTR0707722 to AL); Mayor of the City of Vienna (MWFB 47-07 to WJK); Monogram Biosciences.

disclosure

AL has received research support from Monogram Biosciences and has been an advisor to GlaxoSmithKline. CFS is an advisor to Roche, has received research support from Roche and Novartis. WJK has received research support from Monogram Biosciences. MB, JS, AP, YL, TS, WH, JMW, and JW are employees of Monogram Biosciences. SMA, KL, RH, and EMF have indicated no conflicts of interest.

references


