HER2 – a discussion of testing approaches in the USA

A. Thor
Department of Pathology, Northwestern University Medical School, Chicago, and Evanston Northwestern Healthcare, Evanston, Illinois, USA

Summary

HER2 (neu, erbB-2), a receptor related to the human epidermal growth factor receptor, has now become more important as a predictive marker of treatment response. While the value and direction of the treatment/HER2 interaction may vary, depending on the agents, dose, or schedule of drug administration, there is little disagreement that HER2 testing is an important part of breast cancer evaluation. In 1998, trastuzumab (Herceptin) was approved for the treatment of HER2-positive metastatic breast cancer patients by the Food and Drug Administration of the USA. Patients with abnormal HER2 in their breast cancer cells (generally 2 or 3+ with the HercepTest, overexpression by other immunohistochemical assays or amplification by fluorescence in situ hybridization [FISH] assay) have demonstrated the greatest response to trastuzumab treatment. It is unclear which test (method, reagent, cut-off points, etc.) is best to use to evaluate HER2 for this purpose because parallel testing of the same cancers from patients who received trastuzumab has only recently been initiated and the data are limited. It is widely believed that breast cancers without HER2 alterations will not be responsive to trastuzumab, although a clinical trial to test this specific hypothesis has not been initiated. There are also concerns that clonal heterogeneity for HER2 within a tumor, or between primary and metastatic cancer foci, may affect treatment response; yet we do not currently evaluate these parameters. Consensus regarding the best methods, reagents, or cut-off points to define HER2 status for determining trastuzumab responsiveness has not yet been reached. HER2 testing for other prognostic or predictive purposes, e.g. to determine whether patients are likely to respond to other agents, such as dose-intensive doxorubicin, may be less. Data from the Cancer and Leukemia Group B trial 8541 (companion 8869) suggest that, with proper controls in high-volume laboratories, many of the available methods produce comparable results.

Key words: doxorubicin, HER2, Herceptin, prediction, prognosis, tamoxifen, trastuzumab

Introduction

All normal cells and the majority of breast cancer cells contain two copies of the HER2 (HER2/neu, c-erbB2) gene and produce low levels of the corresponding encoded HER2 protein, also known as p185HER2. Mechanisms of HER2 gene amplification, HER2 protein overexpression, and receptor activation are complex but result in receptor phosphorylation, internalization and upregulation of a complex downstream signaling cascade. HER2 activation often involves heterodimer formation with related receptors (EGFR, erbB-3 and erbB-4) that are bound to receptor-specific ligands. The primary mechanism for HER2 alteration in breast cancer is an increase in gene copy number. This process (amplification) is associated with protein overexpression and occurs in 20%-30% of invasive and up to two-thirds of in situ ductal breast carcinomas. HER2 abnormalities are acquired during breast carcinogenesis; therefore, it is not an inherited genetic lesion.

HER2 abnormalities have been statistically associated with other markers of poor outcome (nodal metastases, high histologic grade, steroid-receptor negativity, younger patient age), although it independently associates with a poor outcome as well, i.e. a marker of poor prognosis, when specific treatment data is not considered. This prognostic value has been reproducibly demonstrated in a myriad of node-positive breast cancer studies, although its value in node-negative disease is less certain.

The ability of HER2 to predict outcome for patients treated with specific chemotherapies may be associated with either a worse or more favorable outcome, depending on the agents utilized. This predictive value of HER2 is the most commonly cited reason for HER2 testing. HER2 testing is now widely applied to samples of primary breast cancer to determine patient candidacy for several regimens, including dose-intensive doxorubicin and trastuzumab. Its predictive value in relation to other agents (paclitaxel, tamoxifen, methotrexate, etc.) is less certain.

Assays to evaluate HER2 generally measure either gene copy number (Southern blot, fluorescence in situ hybridization [FISH], polymerase chain reaction [PCR]) or protein expression (Western blot, immunohistochemistry [IHC]), although measurement of mRNA can also be performed. Using FISH methods, HER2 gene copy number greater than two (or four for cells under-
going division) is considered abnormal. Commercially available FISH kits include instructions for interpretation and scoring, which may involve data modification by establishing a ratio for the HER2 gene number/centromere copy number. This correction attempts to provide a more consistent measure of HER2 gene copy number by excluding the confounding factor of polyploidy, although in reality this correction changes a HER2 amplified to non-amplified case in fewer than 5% of breast cancers. Furthermore, there is limited treatment response data to be certain that the patients with polyploidy and HER2 amplification (viewed as normal when a ratio is established) will be unresponsive to trastuzumab. Genetic deletion, resulting in abnormally low protein expression, also occurs rarely in some cases of breast cancers. Although detectable by FISH, there is limited data regarding the prognostic or predictive value of HER2 deletion.

The most widely applied testing method for HER2 abnormality in the USA is IHC. In part, this reflects the widespread utilization of this method in hospital-based laboratories and the general comfort of pathologists with this technology. For breast cancer, IHC subtyping for hormone receptors in breast cancer has been practiced for many years; therefore, adding a new and similar test for HER2 has been generally well received.

Unfortunately, the specific reagent used to detect HER2 expression in tumors from the initial trastuzumab trials is no longer available. A surrogate was commercially developed and approved by the FDA (the DAKO HercepTest) for selection of candidates for trastuzumab therapy. Many laboratories have used this reagent kit, although not all have embraced this method or reagent kit for a variety of reasons. Many laboratories developed HER2 testing in-house for prognostic purposes before the HercepTest, and for them switching to new reagents at potentially greater cost has not been justifiable. Others use automated IHC systems, and not all commercially developed kits can be adapted as marketed. Laboratory costs, test duration, reimbursement, and reporting issues often significantly influence testing patterns. The impact of these issues may vary from state to state and institution to institution within the USA.

Above and beyond these practical issues, however, others suggest biological reasons for why one test or method is superior to another. Some have argued that analysis of the HER2 protein (by IHC) is more valid because trastuzumab targets the protein receptor. Others contend that IHC assays are less reliable and less quantitative than gene-based assays (even if post-translational events can modify receptor levels). I would argue that the limited comparative data from some studies (such as CALGB 8541) suggest that IHC and FISH can be reliable, specific, and equally prognostic and predictive if appropriately performed. In general, when using any method for HER2 testing of invasive breast cancers, a bimodal distribution of data should be observed. Most invasive cancers will be negative (normal or low). A smaller group of cancers will demonstrate high levels of gene amplification or protein expression in the vast majority of cells (approximately one-third of cases). A limited number of cases will be heterogeneous for HER2, with some cells exhibiting either gene amplification or protein overexpression while other invasive cancer cells will have normal HER2.

HER2 as a prognostic factor in breast cancer

HER2 is a prognostic marker that was first reported by Slamon and colleagues in 1987 [1]. HER2 is an independent prognostic variable for invasive ductal carcinomas of the breast with level II evidence in approximately 5000 patients [2, 3]. The prognostic value of HER2 appears to be strongest in node-positive breast cancer patients. Its prognostic value in the node-negative invasive ductal cancer patient subgroup is less certain and requires further study.

Despite the independent and confirmed prognostic value, HER2 clinical applications based on prognosis alone have significantly declined. This has resulted, in large part, from the reported predictive value of HER2 and the fact that, with some therapeutic strategies, HER2 appears to portend a better rather than worse outcome (trastuzumab, dose-intensive doxorubicin). Utilization of HER2 as a prognostic variable for node-positive patients who generally receive treatment, therefore, may provide misleading and erroneous data. As long as outcome may be modified by treatment strategies, prognostic data in isolation no longer appear relevant.

HER2 predicts for response to therapy

HER2 and response to doxorubicin

The Cancer and Leukemia Group B (CALGB) became interested in looking at HER2 and p53, and added a companion trial to a prospective clinical trial of CAF (CALGB 8541) in 1987. The initial study included approximately 400 patients and was published in 1994 by Muss and colleagues [4]. These early data suggested an interaction between HER2 IHC data and response to doxorubicin. A follow-up study that included both molecular and IHC data, longer follow-up, as well as an additional 400 cases from the same trial was published in 1998 [5]. The most recent data [6] demonstrate that the predictive value of HER2 was observed in this trial regardless of whether IHC, molecular, or FISH data were used. That trial further demonstrated a lack of interobserver variability and stability of HER2 antigen in cut tissue sections that were stored for several years [5]. These three reports also emphasize the length of time and intricacies of companion studies of this nature.

The evidence suggesting an interaction between HER2 and the level of response to doxorubicin is clearly believed in the USA, as much of it has been derived from
While the US FDA has approved three HER2 test kits for different uses, it is likely that the data derived from these tests are roughly comparable. Most laboratories, therefore, do not utilize and separately report each of the three FDA-approved tests for HER2. In fact, many hospital-based laboratories in the USA are not using any of the three FDA-approved assays for HER2 testing. For those that do, many do not use the reagent kit exactly according to the FDA-approved methodology. Because randomized trial design for trastuzumab requires HER2 testing, these issues are a major concern for the oncology, pathology, and patient communities.

In general, the consensus meetings reiterated that for a marker to be useful it should have a significant and independent predictive value that is validated by clinical testing (level II evidence). The assay should also be feasible, reproducible, and widely available with quality controls. Data that are derived should be clearly interpretable. Testing must not consume the tissue sample that is needed for routine pathologic evaluation; hence, it must be adaptable to routinely fixed, processed, and embedded tissues as well as stored, archived materials.

**CAP guidelines**

The CAP Breast Cancer Working Group guidelines were issued following a meeting in June 1999 [19]. A multidisciplinary panel, including statisticians, oncologists, surgeons, and pathologists, generated the guidelines that pertain to tumor markers. They examined HER2 as a prognostic, predictive and marker test, evaluating methods, controls and standardization, reagents, reagent suppliers, and the issue of FDA approval. Although the group considered general usage, it was widely agreed that clinical and statistical colleagues rather than pathologists should determine when and if laboratory testing should be applied.

The CAP panel recommended that HER2 testing be classified as category 2, meaning that it should not necessarily be applied to all newly diagnosed breast cancer patients (as of June 1999). The panel fully acknowledged, however, that guidelines from the NCI or ASCO should be used to define standard of care and thus determine when HER2 should be elevated to category 1 (always applied). This category 2 classification reflected the concerns of the group regarding laboratory testing issues in large part, rather than a failure of the group to recognize independent prognostic and predictive value.

However, the CAP Working Group recommended several steps that should be implemented by pathologists to improve HER2 testing and reporting of results. Furthermore, they recommended that consideration of HER2 testing should be a routine part of the diagnostic work-up of newly diagnosed invasive breast cancers (Table 1). The guidelines state that both the method and primary reagent, including the name and the supplier, should be reported in writing with the assay result, regardless of the type of test used. Because controls are a very important part of the assay, these should be

### HER2 and resistance to tamoxifen

Many reports have suggested some form of resistance to tamoxifen in estrogen-receptor (ER)-positive, HER2-positive, or epidermal growth factor receptor (EGFR)-positive tumors [7-11]. However, recent trials have not supported this interaction in HER2-positive patients. These include the GUN-1 trial [12], the Uppsala Swedish trial [13], the CALGB 8541 trial [14], the NSABP B-14 trial [15] and the South Western Oncology Group (SWOG) 8828 trial [16]. Despite the lack of clinical trial consensus, several molecular studies of cell lines or model systems suggest an interaction between these critical pathways [17, 18]. Further exploration of these signal transduction pathways and possibly new clinical trials designed to address this question may be required for final resolution of this issue.

### HER2 – laboratory issues

In the USA, governmental or professional regulation of tumor markers, their development, and application has been sparse. Laboratories often look to the Food and Drug Administration (FDA) for guidance. Guidelines from professional groups like the College of American Pathology (CAP), the National Cancer Institute (NCI), and the American Society of Clinical Oncology (ASCO) also often impact on clinical practice by helping to define what is considered 'standard of care'. Specific guidelines related to HER2 testing have just recently been published [19]. These include recommendations for testing, controls, interpretation, and reporting. The goal of these recommendations (which include several different breast cancer markers) was to begin the process of standardization and to ensure that the manner of testing and reporting becomes more clearly documented in the test report.

Both the recent CAP and NCI consensus meetings revealed that there is not yet consensus regarding a superior method, reagent, or supplier based on extensive discussion of published and unpublished data. While the US FDA has approved three HER2 test kits...
described as well in the report. Fixed embedded cell lines with amplified as well as non-amplified HER2 are recommended as controls, in addition to human breast samples with both benign and malignant breast epithelium. In this way, over-detection of normal levels of HER2 should be easily identifiable. Hence, if any normal tissues demonstrate membranous reactivity, the assay should be considered indeterminate. They further prescribe that only the invasive component of the carcinoma is to be scored and only a membranous pattern of reactivity should be considered positive.

In terms of reporting IHC HER2 data, the CAP committee recommended that in every case an estimate of the percentage of positive invasive cells should be reported, e.g. 30% immunopositivity, rather than positive or negative. Separate scoring systems (such as the 0/1+/2+/3+ DAKO reporting system) may also be used, but they should be reported in addition to an estimation of the percentage of positive invasive cells. If a dichotomous scoring system is also used, it should be clearly defined in writing in the laboratory report. The committee also recommended that indeterminate cases should be confirmed by another method, such as FISH. Furthermore, each laboratory should document and review their HER2 data on a regular and ongoing basis to determine if the anticipated one-third of cases are positive, and that most of the positive cases are high-grade invasive ductal carcinomas (rather than low-grade invasive ductal or invasive lobular carcinomas, which are more often HER2 negative).

**Special problems related to IHC**

The most commonly applied HER2 testing method is IHC, which detects p185 expression at levels above 'normal'. Although the methods are relatively straightforward, utilization of different primary, secondary, or visualization reagents or conditions can significantly alter the results and testing sensitivity or specificity. For example, increasing the concentration or time of incubation for the primary anti-HER2 antibody or utilization of ultra-sensitive detection reagents may result in increased HER2 detection/visualization. Microscopically, the slide will have many cell types, including normal epithelium and fibroblasts, which appear positive. While this may be 'true' positive (rather than an artifact) in a biologic sense, breast cancers tested with this method will almost always appear positive because the assay detects normal levels of protein expression. Likewise, inclusion of the **in situ** component in the microscopic interpretation may result in up to two-thirds of ductal cancers appearing positive, grossly overestimating the HER2 'status' of the invasive component. The goal of HER2 immunoassays is to detect overexpression rather than normal expression; hence, increases or decreases in sensitivity pose significant problems. Exacting utilization of cell line controls with each assay (using cells that have normal, slightly increased, and greatly increased protein expression) will help to ensure reproducibility and proper assay conditions at the local laboratory level and from day to day. Fixed embedded blocks of cell line pellets can be generated by laboratories themselves, or cut sections can be obtained commercially for this purpose. A commonly used control, an intra-laboratory positive case, is less effective in detecting inter-assay variance.

In summary, HER2 testing using IHC and its interpretation should be performed by those with some knowledge of HER2 and breast cancer biology. Modification of carefully scripted methodology in commercial kits (such as the DAKO HercepTest) should be attempted with caution. In general, controls and conditions prescribed by the company and approved by the FDA should be rigorously followed. In the USA, commercial suppliers of both IHC kits and FISH kits provide training programs to introduce and train laboratories before they initiate testing, which has been very effective. While CAP guidelines do not solve methodological and interpretive issues, they provide a starting point and common format for implementation of the HER2 laboratory test.

**FISH techniques**

The CAP group did not determine that FISH was superior to IHC, nor did it state whether it should be used as an adjunct, surrogate, or replacement. As with IHC, the panel recommended that, for the FISH method, the reagent and supplier should be documented with the results. Other issues were similar as well; controls were described as critical; any variance in methodology beyond the FDA-approved tests should be documented; indeterminate cases should be re-tested or evaluated using another testing method; and invasive cancer only should be interpreted. The following issues were identified as unresolved with regard to FISH testing: heterogeneity; the importance of probe to centromeric ratio correction; and the optimal (clinically relevant) cut-off for amplification (> 2, 4 or 10 gene copies). What level of gene amplification is most likely to predict treatment response, particularly with trastuzumab, also is yet to be determined.

---

**Table 1. CAP recommendations – general excerpts.**

<table>
<thead>
<tr>
<th>Recommendations</th>
</tr>
</thead>
<tbody>
<tr>
<td>The prognostic and predictive value is compelling and may warrant HER2 testing as a routine part of the diagnostic work-up</td>
</tr>
<tr>
<td>The predictive value of HER2 in some groups will require further validation</td>
</tr>
<tr>
<td>Many significant issues relating to testing are unresolved</td>
</tr>
<tr>
<td>HER2 testing work is in progress</td>
</tr>
</tbody>
</table>
A NCI Consensus Conference was held in Washington in November 1999 to bring multispecialty clinicians together with industrial, governmental, patient and health care representatives. A key point for discussion was the determination of patient eligibility for trastuzumab treatment. There was general agreement that the testing paradigm for entry should consider patient as well as laboratory and clinical variables, and the viewpoint for the best cut-off point might well vary, depending on who was considering the problem.

The primary concerns of the FDA were to assess clinical outcomes on trastuzumab trials in relation to method of patient selection. More specifically, what was the outcome for patients selected using the HercepTest as compared with other testing reagents or methods? What would be the response or lack thereof to trastuzumab treatment for the weakly positive patients (IHC 1+ or 2+ by the HercepTest or weak gene amplification by FISH)?

Clinician perspectives expressed at this meeting varied but included generalized concerns regarding comparability of testing methods and reagents, inter-institutional HER2 data variability, intra-patient cancer heterogeneity for HER2, and the effects of artificial cut-off points on entry to trastuzumab trials. More specifically, there were concerns regarding patient eligibility and how that might affect efficacy or impact morbidity/mortality from the trastuzumab trials. Clinicians also expressed distress over an apparent lack of standards and reproducibility.

The pathologist's perspective reflected a high level of frustration and significant pressure to choose a particular test and generate results in the absence of clear-cut guidelines. Pathologists felt they were being forced into choosing cut-off points, methods and interpretative systems without sufficient comparative data related to clinical trials. In fact, some laboratories have chosen the standardized HercepTest kit in order to avoid having to determine if it will be necessary to examine the HER2 laboratory methodology, test or assay criteria will be most predictive of trastuzumab response.

Data from my laboratory comparing FISH to IHC using a variety of IHC reagents (HercepTest, CB11 and TAB250) has revealed that these methods and reagents are fairly similar in terms of their false-positive and false-negative rates (Thor, unpublished data). We have also had greater success using IHC, rather than FISH, assays with significantly older, stored blocks. However, the question remains of when it is appropriate to test or re-test by FISH. Should these assays be performed in parallel? What is the best initial screening methodology? Answers to these questions will hopefully be revealed in studies that are ongoing.

In the USA, the approach to HER2 testing is highly variable. As part of the Eastern Cooperative Oncology Group (ECOG) Pathology Coordinating Office, we have implemented centralized testing for HER2 status using the DAKO HercepTest in three trials currently underway in breast and lung cancer. Entry criteria vary between the trials, with 2+ and 3+ patients being included in the breast cancer trials and 1+, 2+, and 3+ patients being included in the lung cancer trials. This approach is expensive and time-consuming because samples are sent from throughout the USA and data to determine patient entry are returned within 72 hours of receipt of the tissue sample. In each case, comparative studies using other IHC reagents and FISH methods will occur at a later time. Newer trials may include a combination of IHC and FISH, using the latter only in 2+ cases.

An alternative approach adopted by other cooperative oncology groups, namely the CALGB and NSABP, is entry testing for trastuzumab trials based on any test methodology at the local institution. Subsequent centralized testing using a variety of methods and reagents will then be performed. This kind of approach will allow an assessment of variance between methods, laboratories, and pathology groups. These types of painstaking analyses are vital to determine what effect different testing techniques have on patient outcomes.

**Discussion and conclusions**

In order to ensure that the remaining issues relating to HER2 testing are answered, it is necessary to use the
current data to define areas of uncertainty, and to stringently design new trials. There is a particular need for cancer tissue submission for every patient entered onto a randomized trial, because it is only through these studies that predictive markers can be adequately assessed. Individual markers generally do not act alone and, therefore, rigorous statistical testing should be applied to determine the relative value and independent effect of each marker being assessed. In particular, there is a range of other biologic factors that may also affect the response of patients in both trastuzumab and doxorubicin trials. These include topoisomerase 2α, levels of HER2 phosphorylation/activation, and activation of downstream effectors. In particular, heterodimerization partners with HER2 are crucial because HER2 in isolation does not bind to a ligand [22]. Ligand binding mainly occurs when HER2 is linked to another member of the family, particularly HER1 or HER3. Both heterodimerization and ligand binding are pivotal events for receptor signaling to occur. While testing for HER1 and HER3 is not yet commonplace, examination of their association with HER2 and patient response to therapy may be important. In addition, it may also be necessary to examine receptor (or receptor-antibody, receptor-ligand) internalization, recycling and degradation pathways.

A consensus recommendation from the CAP was recently published [19]. It is hoped that the NCI will produce a document summarizing its consensus meeting as well. ASCO is expected to publish its recommendations in the spring of 2000. Despite these eminent recommendations, however, it will be vital to examine the data from ongoing or recently completed trials since treatment combinations and regimens are always changing. Until treatment response and test data comparisons are available, it will be very difficult to identify the best method or reagents.

Finally, it is necessary to think beyond the prognostic and predictive power of HER2, perhaps to consider preventive strategies. There is interesting and provocative animal and cell line data examining HER2 and its role in the initiation or progression of breast carcinogenesis. Rare studies suggest that the identification of HER2 abnormalities prior to frank cancer may identify patients at risk. As preventative agents and strategies are available, the application of molecular testing pre-cancer is likely to emerge as an important new growth area.

Note
The author has not reported any financial relationships with companies whose products are mentioned in the text.

References


Correspondence to:
A. Thor, MD
Department of Pathology
Evanston Northwestern Healthcare
Evanston
Illinois IL 60201
USA
E-mail: a-thor@northwestern.edu