HER-2 Testing in Breast Cancer Using Parallel Tissue-Based Methods

Hadi Yaziji, MD
Lynn C. Goldstein, MD
Todd S. Barry, MD, PhD
Robert Werling, MD
Harry Hwang, MD
Georgiana K. Ellis, MD
Julie R. Gralow, MD
Robert B. Livingston, MD
Allen M. Gown, MD

Context  Testing for HER-2 oncogene in breast cancer has increased because of its role as a prognostic and predictive factor. Some advocate gene testing by fluorescence in situ hybridization (FISH) vs protein testing by immunohistochemistry as the method which most accurately evaluates and predicts response to the anti–HER-2 antibody, trastuzumab. However, critical examination of FISH on a screening basis has yet to be performed.

Objectives  To determine the correlation between FISH and immunohistochemistry results by determining HER-2/neu gene status on tumor sections with indeterminate immunohistochemistry results (2+ score), confirm gene amplification on tumor sections with positive results (3+ score), and verify gene status on tumor sections with negative results (0 or 1+ score).

Design, Setting, and Patients  A quality control and quality assurance program for HER-2 testing by FISH, which used tumor specimens from 2963 patients (median age, 56 years) with breast cancer received from 135 hospitals and cancer centers in 29 states, was performed at a reference laboratory from January 1, 1999, to May 15, 2003. Every specimen evaluated by FISH was parallel tested with immunohistochemistry tests.

Main Outcome Measures  With FISH as the presumed standard testing method, the positive and negative predictive values and sensitivity and specificity of immunohistochemistry were calculated.

Results  A total of 3260 clinical HER-2 tests by FISH were performed on 2963 serially referred breast cancer specimens. Of these, 2933 tests were successful and 2913 breast cancer specimens had both FISH and immunohistochemistry results available. With FISH as the standard testing method, the positive predictive value of positive immunohistochemistry score (3+) was 91.6%, and the negative predictive value of negative immunohistochemistry score (0 or 1+) was 97.2%. The sensitivity of immunohistochemistry tests, including tumor sections with scores of 2+ or 3+, was 92.6% and the specificity of immunohistochemistry tests with scores of 3+ was 98.8%. The FISH test had a significantly higher failure rate (5% vs 0.08%) and reagent cost ($140 vs $10), and longer testing (36 hours vs 4 hours) and interpretation times (7 minutes vs 45 seconds) vs immunohistochemistry tests.

Conclusions  A testing algorithm for HER-2 determination is most efficient by using immunohistochemistry as the method of choice, with FISH performed for cancers with indeterminate results (2+ score). Successful quality control and quality assurance programs are a prerequisite for such approaches.


©2004 American Medical Association. All rights reserved.
methods of limited utility. This is further compounded because immunohistochemistry is more widely used than FISH in community laboratories, leading to inaccurate results.20,21 Thus, based on such studies, it may be incorrect to conclude that FISH testing should replace immunohistochemistry as the screening tool of choice in breast cancer patients.

The use of FISH as a widely used screening assay has not been critically evaluated. Some of its documented disadvantages include high reagent cost and longer procedure and interpretation time.13 We examined the relationship between FISH and immunohistochemistry testing for HER-2, using tests approved by the Food and Drug Administration (FDA), in a large number of breast tumor samples and recommend a practical testing algorithm to screen for HER-2 in breast cancer.

METHODS
Patients and Study Design
From January 1, 1999, to May 15, 2003, a quality control and quality assurance program for HER-2 testing was commenced at PhenoPath Laboratories, Seattle, Wash. For every FISH assay requested by a referring physician, an accompanying immunohistochemistry test was set up in parallel. A total of 2963 patients with breast cancer were accrued during this interval. Patients’ ages ranged from 28 to 87 years, with a median age of 56 years. Tumor sections were received from 135 hospitals and cancer centers in 29 states. The reason for requesting FISH testing by the referring physicians was to determine HER-2/neu gene status on tumor sections with indeterminate (2+ score) immunohistochemistry results (52%), to confirm gene amplification on tumor sections with positive (3+ score) immunohistochemistry results (24%), and to verify gene status on tumor sections with negative (0 or 1+ score) immunohistochemistry results (23%). The University of Washington Human Subjects Division has determined that this research activity qualifies as exempt from US federal regulations.31

Tissue sections were deparaffinized and rehydrated before incubating them in 0.01-M citrate buffer at pH 6.0 in a steamer for 40 minutes at more than 95°C. A polyclonal antibody to HER-2/neu (A0485, DAKOCytomation, Carpinteria, Calif) was applied to sections and incubated for 40 minutes at room temperature. With intervening wash steps in phosphate-buffered saline, slides were incubated for 30 minutes at room temperature in a rabbit-specific labeled polymer (DAKOCytomation, Carpinteria, Calif), followed up by 10 minutes at 37°C in a solution containing 3% hydrogen peroxide and 3′,3′-diaminobenzidine. Slides were counterstained with hematoxylin. Immunostained slides were scored according to a modification of the scoring system approved by the FDA, as described previously.32 In brief, because of the multitude of sources of specimens with a wide range of fixatives and processing techniques, we developed a “subtraction scoring” method, which counts any visible signal of the nonneoplastic breast epithelium as negative and subtracts the score of the tumor cells from that of the benign cells.32 When the subtraction score is 2+, we used the term indeterminate, given the uncertainty about predictability of FISH results among tumor sections with a score of 2+. Such results are not be confused with a failed test, which was defined in this study as failure to produce interpretable results after an additional testing attempt.

The FISH studies to assess HER-2 gene status were performed by using reagents (Vysis Inc, Downers Grove, Ill), according to the manufacturer’s guidelines. Pathologists (H.Y., A.M.G., L.C.G., T.S.B., R.W., H.H.) scored the FISH studies, counting at least 60 tumor cells. Results were reported as a ratio of HER-2 to chromosome 17 signals.

Quality Control and Quality Assurance Programs
Digital images of FISH and immunohistochemistry studies were stored and subsequently used for quality control conferences. For every batch of tumor sections, appropriate quality control sections were used for FISH testing. Sections were declared not interpretable when it was impossible to accurately determine the number of gene or chromosome copies on the tumor cells, even with adequate quality control sections. All FISH tests on these sections were repeated once before results were considered inconclusive. Quality control sections for immunohistochemistry tests included individual 3+ positive tumor sections, multiple cell pellet sections with known ranges of protein expression, and gene status. Each FISH test was simultaneously accompanied by a hematoxylin and eosin-stained section and immunohistochemistry section. In the same order with each case, the pathologist examined the hematoxylin and eosin-stained section and scored the immunohistochemistry results before scoring the FISH test. Following the determination of FISH status, pathologists reviewed the immunohistochemistry-stained section on the same case without modifying the score, regardless of any immunohistochemistry-FISH discrepancy. Results of the immunohistochemistry and FISH scores were entered in a computerized database (Fourth Dimension Inc, San Jose, Calif), transferred to Microsoft Excel spreadsheet (Microsoft Corp, Redmond, Wash) after each test batch, and then exported to JMP version 5.0 statistical software (SAS Institute Inc, Cary, NC).

A number of retrospective analyses were conducted on a regular basis. In addition to a monthly review of small random batches of tumor sections (10-40), the retrospective analysis also included targeted reviews of all immunohistochemistry false-negative and false-positive tumor sections and possible sources of discrepancy were identified and arranged for future reviews. The quality control and quality assurance program also included periodic blinded review of the immunohistochemistry-stained sections to the FISH results, with intraobserver and interobserver comparisons.
RESULTS
A total of 3260 clinical HER-2 tests by FISH were performed on 2963 specimens from patients with breast cancer tumors. Of these, 2933 tests were successful and 2913 breast cancer specimens had both FISH and immunohistochemistry test results available. The FISH failure rate was 5% (n=163), and the immunohistochemistry failure rate was 0.08% (n=23) after a second and final test attempt. Table 1 shows the relationship between FISH and immunohistochemistry test results.

When FISH was presumed to be the standard testing method, the positive predictive value of positive immunohistochemistry scores (3+) was 91.6% and the negative predictive value of a negative immunohistochemistry score (0 or 1+) was 97.2%. We also calculated the sensitivity and specificity of immunohistochemistry. These analyses were performed with and without the tumor group that had indeterminate (2+) results (Table 2).

The mean direct reagent cost by vendors of each FISH and immunohistochemistry test was $140 and $10, respectively. The mean (SD) procedure time was 36 hours (30 minutes) for FISH and 4 hours (12 minutes) for immunohistochemistry. Mean (SD) time for the pathologist to interpret the test was 7 minutes (2.5 minutes) for FISH and 45 seconds (13 seconds) for immunohistochemistry (Table 3).

The concordance rate between immunohistochemistry and FISH tests was 64.9% when scores of 2+ and 3+ positive tumor sections were grouped together. However, when the 2+ score group (1151 tumor sections) was excluded from the analysis, the concordance rate improved to 96.1%. This number is very high because a large subset of tumor sections (52.0%) was initially submitted for FISH testing as a result of an indeterminate score (2+) by immunohistochemistry by the referring pathologists. Among the FISH-negative tumor sections (n=2371), only 1.2% (n=28) had an immunohistochemistry score of 3+, and the remaining tumor sections were either negative (58.5%, n=1388) or indeterminate (40.3%, n=955).

When FISH-positive results were divided into low and high levels of gene amplification (using a gene/chromosome ratio cutoff of 5), only 2.8% of the high-level group had negative immunohistochemistry results and the majority of these tumor sections had immunohistochemistry scores of either 3+ (77.9%) or 2+ (19.3%). In the low-level group (gene/chromosome ratio of 2 or more but less than 5), 32.7% had an immunohistochemistry score of 3+, and the remaining tumor sections showed predominantly (54.9%) a score of 2+, with 12.5% negative immunohistochemistry results.

Using a set of 117 tumor sections to test the variability among pathologists, complete agreement by 5 pathologists (H.Y., L.C.G., A.M.G., T.S.B., H.H.) was reached on 50.4% of the sections, with 4 of 5 pathologists reaching an agreement on an additional 24.7% of the sections. No 2-step discrepancy in immunohistochemistry scoring (ie, 3+ score vs 1+ or 0 score) was observed in any of the sections.

Table 1. Comparison of Fluorescence In Situ Hybridization and Immunohistochemistry Test Results

<table>
<thead>
<tr>
<th>Immunohistochemistry Score</th>
<th>Fluorescence In Situ Hybridization</th>
<th>Negative for Amplification</th>
<th>Positive for High-Level Gene Amplification*</th>
<th>Positive for Low-Level Gene Amplification†</th>
<th>Total</th>
</tr>
</thead>
<tbody>
<tr>
<td>2+ and 3+</td>
<td>983</td>
<td>502</td>
<td>277</td>
<td>225</td>
<td>1485</td>
</tr>
<tr>
<td>Positive (3+)</td>
<td>28</td>
<td>306</td>
<td>222</td>
<td>84</td>
<td>334</td>
</tr>
<tr>
<td>Indeterminate (2+)</td>
<td>955</td>
<td>196</td>
<td>55</td>
<td>141</td>
<td>1151</td>
</tr>
<tr>
<td>Negative (0 or 1+)</td>
<td>1388</td>
<td>40 (29, 11)</td>
<td>8 (6, 2)</td>
<td>32 (18, 14)</td>
<td>1428</td>
</tr>
<tr>
<td>Total</td>
<td>2371</td>
<td>542</td>
<td>285</td>
<td>257</td>
<td>2013</td>
</tr>
</tbody>
</table>

*Defined as gene/chromosome ratio of 5 or more.
†Defined as gene/chromosome ratio of 2 or more and less than 5.

Table 2. Sensitivity, Specificity, and Predictive Values of Immunohistochemistry With Fluorescence In Situ Hybridization Presumed to be the Standard Testing Method

<table>
<thead>
<tr>
<th>Numerator/Denominator %</th>
<th>Sensitivity (including 2+ and 3+)</th>
<th>Specificity Including 2+ and 3+</th>
<th>3+ only</th>
<th>Predictive value Positive (2+ or 3+)</th>
<th>Positive (3+ only)</th>
<th>Negative (0 or 1+)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>502/542</td>
<td>1388/2371</td>
<td>2343/2371</td>
<td>502/1485</td>
<td>306/334</td>
<td>1388/1428</td>
</tr>
</tbody>
</table>

Numerator/Denominator %

Table 3. Comparison of Immunohistochemistry and Fluorescence In Situ Hybridization as Screening Tools for HER-2 in Breast Cancer

<table>
<thead>
<tr>
<th>Fluorescence In Situ Hybridization</th>
<th>Immunohistochemistry</th>
</tr>
</thead>
<tbody>
<tr>
<td>Failure rate, %</td>
<td>5.0</td>
</tr>
<tr>
<td>Procedure time, mean (SD)</td>
<td>36 h (30 min)</td>
</tr>
<tr>
<td>Interpreting time by the pathologist, mean (SD)</td>
<td>7 min (2.5 min)</td>
</tr>
<tr>
<td>Mean direct reagent cost for laboratory, US $*</td>
<td>140</td>
</tr>
<tr>
<td>Projected annual cost of screening, US $ million†</td>
<td>55-75</td>
</tr>
</tbody>
</table>

*Costs based on prices charged by vendors to PhenoPath Laboratories.
†Based on estimated 211,300 invasive breast cancers in 2003, according to the American Cancer Society data.33
COMMENT

Testing methods for HER-2 are currently heatedly debated among pathologists and oncologists involved in breast cancer, with no agreed-upon standard practices. The Center for Biologics Evaluation and Research at the FDA has recognized the limitations of all current HER-2 testing methods, stating that it “views both immunohistochemistry and FISH as semi-quantitative if performed under ideal circumstances. Both methods require subjective interpretation.”

Although pathologists commonly perform first-line screening by immunohistochemistry with confirmatory FISH testing for indeterminate immunohistochemistry test results, trastuzumab clinical trials have reported better use of FISH in predicting treatment response, suggesting that FISH should be used as a primary screening test. However, when positive immunohistochemistry test results (3+ score) were analyzed separately from indeterminate test results (2+ score) in the same trials, immunohistochemistry was predictive of response.

Moreover, the critical appraisal of FISH as a screening test for HER-2 has not been accomplished. A screening test to be routinely performed on a large number of clinical specimens should be easy to perform and interpret, show high rates of sensitivity, exhibit high levels of interlaboratory and intralaboratory reproducibility, and be relatively inexpensive. The FISH method for HER-2 testing should be critically assessed against these criteria before implementing FISH as the first-line screening tool for HER-2 in breast cancer.

Trastuzumab acts at the protein and not at the gene level. Although protein levels directly reflect the gene status and thus can be indirectly measured via FISH method, emerging data show response to trastuzumab in patients with positive immunohistochemistry score whether or not gene amplification is present, arguing for the existence of an alternative mechanism of protein overexpression other than gene amplification. Protein expression as a predictor for clinical response independent of the gene status supports a true biological basis for the discrepancy between the 2 methods. Additionally, patients whose test results show chromosome 17 polysomy without gene amplification (increased gene copy number secondary to aneuploidy of chromosome 17) are currently not judged to be eligible for trastuzumab by the FISH method because their test results are scored as negative. In our study, this subset of tumor sections is quite large, approaching one third (29.6%, n=870; data not shown) of the entire cohort comparable with what was previously reported (24.8%) in another large study.

Determination of response rate to trastuzumab for this group of patients is crucial given the large size of this cohort. Widely performed testing with HER-2 FISH assays that do not contain a probe for chromosome 17 report positive HER-2/neu gene status for tumor sections with polysomy of chromosome 17 and single copy HER-2/neu gene.

Another area of concern is the uncertain clinical significance of low-level gene amplification. To our knowledge, no clinical trials have reported the relationship between low-level amplification and response to trastuzumab. It is possible that this category of patients may well be clinically classified as indeterminate for amplification. Until data evaluating low-level amplification become available, all otherwise eligible patients whose test results are within this range will continue to receive trastuzumab without a critical assessment of their likely response to treatment. In our study, the fraction of patients with low-level gene amplification constitutes 47.4% of the positive tumor sections (Table 1).

Preliminary findings from 2 cooperative group trials testing the adjuvant use of trastuzumab, National Surgical Adjuvant Breast and Bowel Project B31 and North Central Cancer Treatment Group N9831, showed high rates of discordance in HER-2 immunohistochemistry testing between community practices and a central laboratory. When comparing FISH testing between laboratories, B31 and N9831 preliminary assessment showed 5% and 23% disagreement, respectively, on FISH results (positive vs negative) between 2 central laboratories. Additionally, repeat FISH testing on a large subset (n=198) of breast cancers by a central laboratory showed 16% disagreement with previously reported results by another central laboratory.

The current study demonstrates the very tight correlation between immunohistochemistry and FISH as part of a quality control and quality assurance program. Negative immunohistochemistry test results in our study accurately predicted FISH results in 97.2% of tumor sections, with a false-negative rate of 2.8%. The reasons for the very low rate of false-negative immunohistochemistry results are multifactorial. First, technologists and pathologists at the same institution are performing the technical and interpretation aspects of the FISH and immunohistochemistry tests. Also, FISH testing has improved the quality control and quality assurance of HER-2 testing by immunohistochemistry at this institution. In the past, our main immunohistochemistry monitoring method and scoring approach was periodic verification of the percentage of HER-2 positive breast cancers. The FISH test now provides another “fine-tuning” tool for evaluating HER-2 testing by immunohistochemistry. Finally, performing side-by-side parallel immunohistochemistry testing for all tumor sections tested by FISH has been a powerful quality control mechanism providing instant and continuous feedback on the varied technical and scoring parameters associated with both of these technologies. Similarly, some of the reported high rates of false-positive and false-negative immunohistochemistry results are secondary to multiple factors. Technically, lack of adherence to strict quality control measures in accepting or rejecting a test result, absence of internal quality control programs that identify overscoring or underscoring of immunohistochemistry...
try results are contributing factors. At times, the type of specimen fixation (particularly alcohol or alcoholic formalin) could significantly increase the sensitivity of the method so tumor sections with negative (0 or 1+ signal) or indeterminate (2+) scores may show stronger signal intensity. If the subtraction score was not performed on such tumor sections, they may inadvertently be scored as positive (3+) for overexpression.

Economic and practical considerations would tend to favor immunohistochemistry over FISH as a screening tool, because these are very important given the high incidence of breast cancer. The American Cancer Society predicted 211,300 new cases of invasive breast cancer in the United States during 2003.33 Given the known incidence of HER-2 alterations in breast cancer, 70% to 75% of tumor sections tested by immunohistochemistry will be negative (0 or 1+ score), and the remainder divided almost equally between indeterminate results (2+ score) and positive (3+ score) overexpression. Therefore, only about 15% of patient samples that are screened primarily by immunohistochemistry would be indeterminate and require confirmatory FISH testing. In the United States, if FISH testing were adopted as the standard for all cases, we conservatively estimate the additional health care cost would range from $50 to $70 million in 2003, for supplies and reagents alone. The additional time allocated for FISH testing and scoring by pathologists is also disproportionately high compared with immunohistochemistry. At our institution, pathologists dedicate a substantial amount of time for scoring and reporting of FISH tests compared with immunohistochemistry (Table 3).

Screening by immunohistochemistry and confirming indeterminate results with FISH testing has been advocated by many investigators,14,15,18,20,36,37 with some suggesting confirmation of both indeterminate and positive immunohistochemistry results with FISH.26 Other studies favor FISH testing as the primary screening test.22,27,28,38 An international panel from 10 western countries has recently recommended immunohistochemistry to be the screening method of choice, with repeat FISH testing on the indeterminate subset. This panel also emphasized strict quality control and quality assurance measures for any given technology.37 The results in our study support these measures and recommend a standard procedure for any laboratory that provides clinicians with either test. Lack of such measures may explain the suboptimal reproducibility of the results of FISH tests performed in community settings and repeated by the central laboratories of the cooperative group trials.29,30 In fact, after amending the B31 protocol to require entry HER-2 testing from laboratories approved by the National Surgical Adjuvant Breast and Bowel Project (based on volume of testing and documented immunohistochemistry/FISH concordance), the false-negative rate has decreased to 3%.39 In the case of HER-2 testing, once the results of such programs are translated into successfully high rates of immunohistochemistry/FISH concordance, immunohistochemistry can be reliably offered as a first-line screening test requiring confirmation by FISH only in a small subset of cases that show indeterminate results (FIGURE).

In conclusion, accurate testing for HER-2 in breast cancer can be achieved using immunohistochemistry with complementary FISH testing, taking advantage of the strength of each of these technologies. Such an algorithmic approach is especially important given the increasing number of prognostic and predictive biomarkers in oncology.30

**Author Contributions:** Drs Gralow and Yaziji had full access to all the data in the study and take responsibility for the integrity of the data and the accuracy of the data analysis.

**Study concept and design:** Yaziji, Goldstein, Hwang, Ellis, Livingston, Gown.

**Acquisition of data:** Yaziji, Goldstein, Barry, Werling, Gown.

**Analysis and interpretation of data:** Yaziji, Goldstein, Barry, Werling, Hwang, Ellis, Gralow, Livingston, Gown.

**Drafting of the manuscript:** Yaziji, Goldstein, Hwang, Gown.

**Critical revision of the manuscript for important intellectual content:** Yaziji, Goldstein, Barry, Werling, Hwang, Ellis, Gralow, Livingston, Gown.

**Administrative, technical, or material support:** Yaziji, Barry, Werling, Gown.

**Supervision:** Gown.

**Funding/Support:** This study was supported by an internal preexisting budget from PhenoPath Laboratories that is allocated for quality assurance and quality control studies.

**Role of the Sponsor:** Authors from PhenoPath Laboratories participated in the design and conduct of the study, in the collection, analysis, and interpretation of the data, and in the preparation, review, and approval of the manuscript.

**Acknowledgment:** We thank the following individuals for their outstanding efforts in the technical part, data entry, and data analysis of the study: Phyllis Davie, MT, Jill Wirtala, Erik Hughes, Jana Hague, MT, Theresa Edwards, MT, Arian Giesholt, Patti Loykasek, Stephanie Scholl, Lynn Hopf, and Paul Kong. We also thank David Tauben, MD, for reviewing the manuscript.

**REFERENCES**