

Diagnostic Evaluation of HER-2 as a Molecular Target: An Assessment of Accuracy and Reproducibility of Laboratory Testing in Large, Prospective, Randomized Clinical Trials

Michael F. Press,^{1,2} Guido Sauter,⁵ Leslie Bernstein,^{1,3} Ivonne E. Villalobos,^{2,5} Martina Mirlacher,⁵ Jian-Yuan Zhou,² Rooba Wardeh,² Yong-Tian Li,² Roberta Guzman,² Yanling Ma,² Jane Sullivan-Halley,³ Angela Santiago,² Jinha M. Park,⁴ Alessandro Riva,⁶ and Dennis J. Slamon⁴

Abstract Purpose: To critically assess the accuracy and reproducibility of human epidermal growth factor receptor type 2 (HER-2) testing in outside/local community-based hospitals versus two centralized reference laboratories and its effect on selection of women for trastuzumab (Herceptin)-based clinical trials.

Experimental Design: Breast cancer specimens from 2,600 women were prospectively evaluated by fluorescence *in situ* hybridization (FISH) for entry into Breast Cancer International Research Group (BCIRG) clinical trials for HER-2-directed therapies.

Results: HER-2 gene amplification by FISH was observed in 657 of the 2,502 (26%) breast cancers successfully analyzed. Among 2,243 breast cancers with central laboratory immunohistochemistry (10H8-IHC) analysis, 504 (22.54%) showed overexpression (2+ or 3+). Outside/local laboratories assessed HER-2 status by immunohistochemistry in 1,536 of these cases and by FISH in 131 cases. Overall, the HER-2 alteration status determined by outside/local immunohistochemistry showed a 79% agreement rate [κ statistic, 0.56; 95% confidence interval (95% CI), 0.52-0.60], with FISH done by the central laboratories. The agreement rate comparing BCIRG central laboratory 10H8-IHC and outside/local laboratory immunohistochemistry was 77.5% (κ statistic, 0.51; 95% CI, 0.46-0.55). Finally, HER-2 status, determined by unspecified FISH assay methods at outside/local laboratories, showed a 92% agreement rate (κ statistic, 0.83; 95% CI, 0.73-0.93), with FISH done at the BCIRG central laboratories.

Conclusions: Compared with the HER-2 status determined at centralized BCIRG reference laboratories, these results indicate superiority of FISH to accurately and reproducibly assess tumors for the HER-2 alteration at outside/local laboratories for entry to clinical trials.

Authors' Affiliations: ¹Women's Cancers Program and Departments of ²Pathology and ³Preventive Medicine, Norris Comprehensive Cancer Center, University of Southern California Keck School of Medicine; ⁴Jonsson Comprehensive Cancer Center, University of California at Los Angeles School of Medicine, Los Angeles, California; ⁵Department of Pathology, University of Basel, Basel, Switzerland; and ⁶Breast Cancer International Research Group, Paris, France
Received 3/22/05; revised 6/10/05; accepted 6/21/05.

Grant support: Aventis, Inc., Genentech, and Hoffmann-La Roche Ltd. (BCIRG); National Cancer Institute grant CA48780, National Institute for Child Health and Human Development grant NO1-HD-3-3175, U.S. Army Medical Research and Materiel Command grant DAMD17-00-1-0186, and California Breast Cancer Research Program grants BCRP 5EB-3101 and BCRP 7PB-0051 (M.F. Press); and National Institute for Child Health and Human Development contract NO1-HD-3-3175 (L. Bernstein). Abbott-Vysis, formerly Vysis, Inc., provided the FISH assay reagents used in this study.

The costs of publication of this article were defrayed in part by the payment of page charges. This article must therefore be hereby marked *advertisement* in accordance with 18 U.S.C. Section 1734 solely to indicate this fact.

Note: The findings were presented in part at the 25th Annual San Antonio Breast Cancer Symposium, December 11 to 13, 2002, San Antonio, TX and at the 92nd Annual Meeting of the U.S.-Canadian Academy of Pathology, March 22 to 28, 2003, Washington, DC.

Requests for reprints: Michael F. Press, University of Southern California/Norris Comprehensive Cancer Center, NOR5409, 1441 Eastlake Avenue, Los Angeles, CA 98195-6043. Phone: 323-865-0563; Fax: 323-865-0122; E-mail: villalob@usc.edu.

©2005 American Association for Cancer Research.
doi:10.1158/1078-0432.CCR-05-0636

The human epidermal growth factor receptor type 2 (HER-2) gene encodes a membrane receptor protein in the epidermal growth factor receptor family (1, 2) and is amplified and overexpressed in ~25% of human breast cancers (3, 4). This alteration is associated with shorter disease-free and overall survival in women with either lymph node-negative (5-8) or lymph node-positive breast cancer (3, 4, 8-13). The HER-2 alteration, which is gene amplification and resulting overexpression, is also associated with lack of responsiveness to tamoxifen antiestrogen therapy (14, 15) and altered responsiveness to a variety of cytotoxic chemotherapy regimens (16-19). Trastuzumab (Herceptin), a recombinant humanized monoclonal anti-HER-2 antibody, was approved in September 1998 by the U.S. Food and Drug Administration (FDA) for the treatment of women with metastatic breast cancers whose tumors contain the HER-2 alteration. In addition, other therapeutic antibodies as well as small-molecule receptor tyrosine kinase inhibitors directed against this receptor are currently in development. These clinical associations and, more importantly, the availability of a targeted therapeutic make accurate assessment of the HER-2 alteration in clinical materials critical. There are two commercially available immunohistochemical assays and one fluorescence *in situ* hybridization

(FISH) assay that are FDA approved to determine HER-2 status for selection to Herceptin therapy; however, considerable controversy still exists regarding the relative accuracy of these assay methods.

Using molecularly characterized breast cancer cohorts, two independent studies have shown previously that FISH is the most accurate commercially available assay method (20, 21). Based on these published data, the Breast Cancer International Research Group (BCIRG) decided to determine the HER-2 status of breast cancer patients being screened prospectively for adjuvant and metastatic Herceptin-based therapy using FISH. The assay was done in two centralized laboratories, one in the United States and one in Europe. This approach was undertaken to circumvent the potentially variable results of either immunohistochemistry or FISH results obtained from a variety of local testing facilities or commercial service laboratories. This approach now provides an opportunity to critically assess the accuracy and reproducibility of HER-2 testing in a large, prospectively accrued randomized study and to compare the two assay methods (FISH and immunohistochemistry) as well as to evaluate the consistency of centralized versus local (or outside) testing.

Materials and Methods

Tissue specimens. Breast cancer specimens from the first 2,600 women prospectively submitted for screening by FISH for entry into BCIRG clinical trials were considered eligible for inclusion in this study. Three BCIRG clinical trials (BCIRG005, BCIRG006, and BCIRG007) required central laboratory evaluation of HER-2 gene amplification by FISH before patients were entered into the research protocol. BCIRG005, a multicenter, phase III, randomized trial, is designed to compare disease-free survival and overall survival among node-positive patients with operable breast cancer who are HER-2 negative and treated either with docetaxel (Taxotere) in combination with doxorubicin and cyclophosphamide (TAC) or with doxorubicin and cyclophosphamide followed by docetaxel (AC→T). BCIRG006 is a multicenter phase III randomized trial to evaluate the adjuvant use of trastuzumab and compare doxorubicin and cyclophosphamide followed by docetaxel (AC→T) as the control arm with doxorubicin and cyclophosphamide followed by docetaxel and trastuzumab (AC→TH) or with docetaxel, platinum salt, and trastuzumab (TCH). The primary end points are disease-free survival and overall survival in HER-2-positive, node-positive, and high-risk node-negative patients with operable breast cancer. The BCIRG007 study is a multicenter, phase III, randomized trial designed to evaluate docetaxel and trastuzumab versus docetaxel, platinum salt (cisplatin or carboplatin), and trastuzumab as first-line chemotherapy for HER-2-positive patients with metastatic breast cancer, which is previously untreated. These three clinical trials required HER-2 screening of a total of 10,948 breast cancer specimens with accrual of 3,301 patients to BCIRG005, 3,222 to BCIRG006, and 263 to BCIRG007.

Two BCIRG central laboratories received tumor specimens: the University of Southern California (USC; Los Angeles, CA) and the University of Basel (Basel, Switzerland). The specimens included in this report were those that met the following criteria: (a) patient eligibility for one of three BCIRG clinical trials (BCIRG005, BCIRG006, or BCIRG007), (b) availability of a paraffin-embedded tissue specimen suitable for FISH analysis (either a paraffin-embedded tissue block or unstained tissue sections), and (c) histologic confirmation of invasive breast cancer in the tissue specimen. Written informed patient consent for investigation of breast cancer tissue and local institutional review board approvals were obtained before analysis.

Among the first 1,400 cases received at USC, 1,368 were considered suitable for clinical trials entry and tissue processing by FISH. Of the 32

cases excluded, 26 showed either no carcinoma or no invasive carcinoma, 2 were excluded from clinical trials before FISH testing, 1 was submitted in error, 1 could not be processed (only H&E-stained slides were received), and 2 were submitted in duplicate. Among the first 1,200 cases received at the University of Basel, 1,175 were suitable for clinical trial entry processed by FISH. The remaining 25 showed no invasive carcinoma. Overall, 2,543 of the first 2,600 (97.8%) breast cancer cases were suitable for FISH analysis and form the basis of this report.

Fluorescence in situ hybridization assay. Both central laboratories did FISH assays using the PathVysion assay (Abbott-Vysis, Inc., Downers Grove, IL) as described elsewhere (20). A slide was stained with H&E to confirm the presence of invasive tumor. The corresponding area of invasive carcinoma was enumerated on the FISH slides after hybridization was complete.

Slides were evaluated for HER-2 gene amplification by determining the HER-2/CEP17 signal ratio in 60 tumor nuclei. If the ratio was <2.0, the specimen was considered to lack gene amplification; if the ratio was ≥ 2.0 , the specimen was considered to show HER-2 gene amplification. For ratios near the cutoff value (i.e., 1.8-2.2), an additional 60 nuclei were evaluated and the ratio was recalculated. All results were verified by independent slide review.

FISH processing at the USC and University of Basel central laboratories was standardized before receipt of any specimens and was confirmed as identical in independent site visits to each laboratory.

Within-laboratory observer variation of fluorescence in situ hybridization assay results. At the outset of the study, 59 breast cancers were evaluated by independent observers to evaluate interobserver variation for FISH. Thirty-six were enumerated by two observers at USC and 23 were enumerated by two observers at the University of Basel. All observers were masked to all other data related to these cases. These test specimens were not from patients screened for clinical trial eligibility. These cases represented a wide range of FISH ratios (0.97-18.08), average HER-2 copy numbers (1.37-52.67 copies per nucleus), and chromosome 17 centromere averages (1.15-5.18 centromeres per nucleus; Fig. 1A). Fifty-seven of the 59 (96.6%) cases were concordant for HER-2 amplification status. One discordant case had a ratio of 1.99 by one observer and 2.02 by another observer. Disagreement also occurred for one other case (not near the cutoff ratio of 2.0). The resulting κ statistic was 0.93 [95% confidence interval (95% CI), 0.84-1.00].

Assessment of interlaboratory variability of fluorescence in situ hybridization assays. Fifty breast cancers with known gene amplification levels (20) previously analyzed by FISH in the USC central laboratory were sent to the Swiss central laboratory in a multitumor tissue block format to assess interlaboratory variability (22). A second group of 30 quality-control breast cancer cases was exchanged after 1 year of testing was completed at each site. The cases were analyzed independently by FISH in each laboratory and the results were compared. Both laboratories obtained FISH results for 76 of the 80 (95%) test-set cases. The 76 breast cancer specimens used for the interlaboratory comparison represented a wide range of FISH ratios (0.91-19.48), HER-2 copy numbers (1.50-46.27 copies per tumor nucleus), and chromosome 17 centromere averages (1.13-4.40 centromeres). Both laboratories classified 75 of 76 (98.7%) breast cancers identically with regard to HER-2 status (amplified versus not amplified). The single exception was the case, described above, near the 2.0 cutoff that had been assessed differently by two observers in the same laboratory. The resulting κ statistic was 0.97 (95% CI, 0.91-1.00).

Assessment of HER-2 immunohistochemical staining at the University of Southern California central laboratory. Because some investigators suggest that disagreements between HER-2 status determined by FISH and immunohistochemistry may be due to "single-copy overexpression" (high expression of HER-2 without amplification of the gene; ref. 23), we analyzed HER-2 protein by immunohistochemistry at the USC central laboratory. Previously published data have shown that two nonproprietary immunohistochemistry assays were more accurate than two proprietary, FDA-approved immunohistochemistry assays in a set

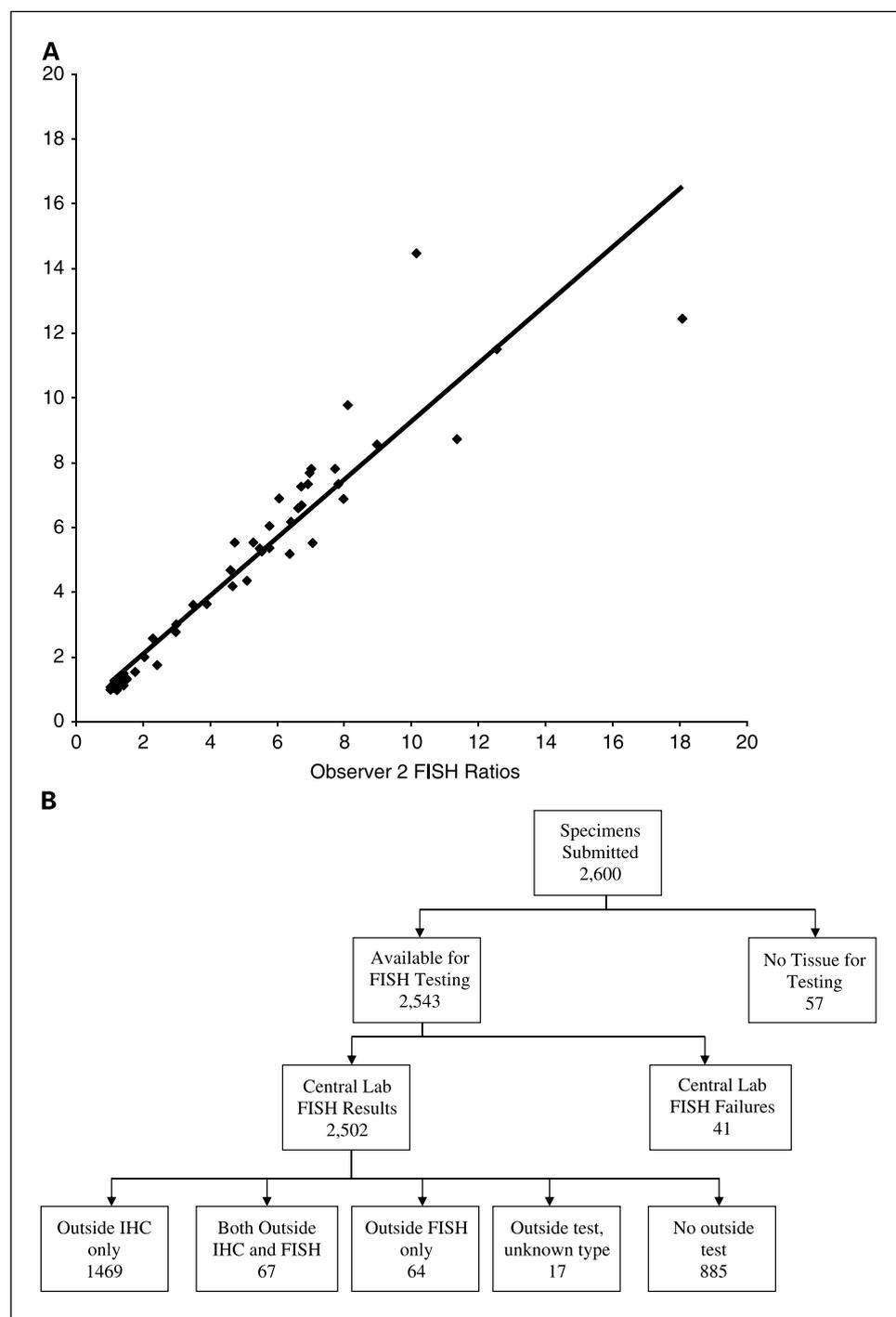


Fig. 1. BCIRG central laboratories, FISH assays. *A*, comparison of FISH ratios determined by two different observers in the same laboratory. *B*, specimen accountability. This schematic diagram summarizes the numbers of breast cancer specimens received, analyzed by FISH in the BCIRG central laboratories and by various methods at outside/local laboratories.

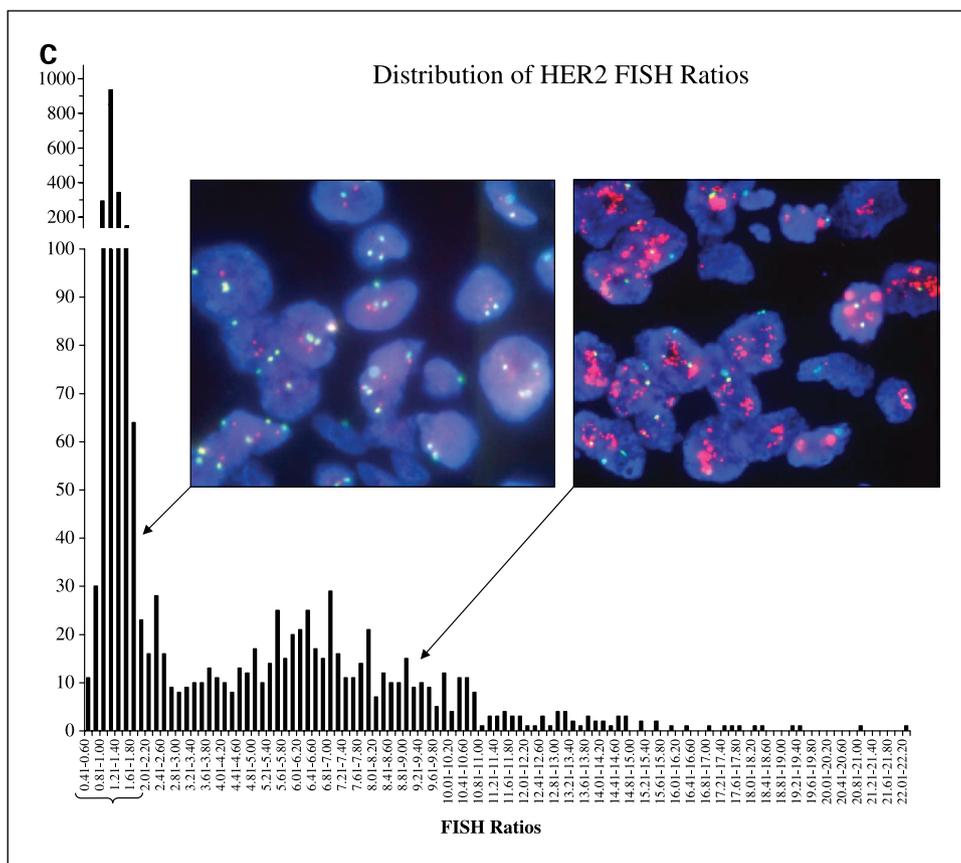
Downloaded from <http://aacrjournals.org/clinccancerres/article-pdf/11/18/6598/1960023/6598-6607.pdf> by guest on 04 October 2022

of samples with HER-2 status characterized by Southern hybridization, Northern hybridization, Western immunoblot, and frozen section immunohistochemistry (20). Because of this higher accuracy, the 10H8 monoclonal antibody immunohistochemical method (10H8-IHC) was used to characterize HER-2 immunostaining (20, 24). These assays were done masked to all other clinical and laboratory information. Tissue sections were incubated with two different antibody reagents: the primary mouse antibody specific for HER-2 receptor protein (10H8 monoclonal antibody, 5 µg/mL, 30 minutes, room temperature) followed by a goat anti-mouse immunoglobulin conjugated to a horseradish peroxidase-labeled dextran polymer (Envision+, 30 minutes, room temperature, DAKO, Corp., Carpinteria, CA). Prediluted

goat anti-mouse IgG immunoglobulin conjugated to horseradish peroxidase-labeled dextran polymer (Envision+) was used as specified by the manufacturer (DAKO). The site of immunoprecipitates was identified using a chromogen, diaminobenzidine, visualized microscopically. Immunostaining was scored subjectively with an Olympus bright-field microscope as 0, 1+, 2+, and 3+ as described for the Clinical Trials Assay system (25, 26) and the DAKO HercepTest (26); 0 and 1+ were considered as low expression and 2+ and 3+ as overexpression.

Although 10H8-IHC immunostaining was done in an identical fashion for all cases, the tissue section preparation varied between the central laboratories at USC and the University of Basel. At USC, a sufficient number of unstained tissue sections or those cut from

Fig. 1 Continued. C, distribution of HER-2 FISH ratios in 2,502 breast cancers successfully analyzed by FISH. FISH assays were either successful on the first attempt or repeated at least two additional times within 5 days of specimen receipt.



individual blocks were available from 1,364 breast cancer patients. HER-2 10H8-IHC was done within 2 weeks of sectioning the tissue block or receiving unstained tissue sections. At the University of Basel, tissue microarrays were prepared from 957 breast cancer cases as described elsewhere (27, 28). 10H8-IHC was done on tissue sections prepared from these tissue microarrays. Two hundred eighteen breast cancer cases were not represented in the tissue microarrays either because available tissue in the tissue block was insufficient or because only unstained tissue sections had been submitted. Forty specimens did not show invasive carcinoma in TMA sections.

HER-2 status determined at other laboratories. Physicians referring patients to the BCIRG clinical trials were queried as to whether the patient's breast cancer specimen had been previously tested to determine HER-2 status, what methods were used, and the results of the outside/local assay recorded. The most frequently used outside immunohistochemistry assay methods were the DAKO HercepTest and

Ventana Pathway Assay, both of which use a similar subjective scoring system as approved by the U.S. FDA with low expression defined as 0 and 1+ immunostaining and overexpression defined as 2+ and 3+ immunostaining. We use this definition for the current report and document any deviation from this definition. We have addressed previously the correspondence of HER-2 immunostaining levels in frozen tissue to HER-2 gene amplification levels (4, 9, 29).

A total of 1,062 of 1,368 (77.6%) cases tested at USC had HER-2 testing at outside laboratories, whereas 564 of 1,175 (48.0%) cases tested at the University of Basel had outside/local HER-2 testing. Sixty-seven cases were tested by both immunohistochemistry and FISH, 64 by FISH alone, and 1,469 by immunohistochemistry alone (Fig. 1B).

Statistical analyses. To measure the degree of concordance (agreement corrected for chance), the κ statistic and its 95% CI were calculated (30). In general, $\kappa > 0.91$ is considered to represent almost perfect concordance (31, 32).

Table 1. Immunohistochemical assay results by any method with 0, 1+, 2+, and 3+ scoring from outside laboratories compared with FISH assay results from the BCIRG central laboratories

	Outside laboratory immunohistochemistry scores (0-3+)				Total
	0	1+	2+	3+	
FISH negative (%)	538 (96.4)	230 (93.9)	165 (83.3)	89 (21.9)	1,022 (72.6)
FISH positive (%)	20 (3.6)	15 (6.1)	33 (16.7)	317 (78.1)	385 (27.4)
Total (%)	558 (100)	245 (100)	198 (100)	406 (100)	1,407 (100)

NOTE: There were also 24 FISH failures specimens in the BCIRG central laboratory that had outside IHC results with 0 to 3+ scoring: 14 were scored as 0, 4 as 1+, 2 as 2+, and 5 as 3+.

Table 2. DAKO HercepTest immunohistochemical assay results from outside laboratories compared with FISH assay results from the BCIRG central laboratories

	Outside laboratory immunohistochemistry scores (0-3+)				Total
	0	1+	2+	3+	
FISH negative (%)	296 (96.4)	142 (94.7)	103 (83.1)	57 (21.8)	598 (71.0)
FISH positive (%)	11 (3.6)	8 (5.3)	21 (16.9)	204 (78.2)	244 (29.0)
Total (%)	307 (100)	150 (100)	124 (100)	261 (100)	842 (100)

Results

Analysis of Breast Cancer International Research Group clinical trial breast cancer specimens by fluorescence in situ hybridization.

Among the first 2,600 breast cancer specimens submitted to the BCIRG for HER-2 evaluation, 2,543 met the entry criteria for inclusion in this report (Fig. 1B). A total of 2,502 (98.4%) were successfully analyzed by FISH; 655 (26.2%) had *HER-2* gene amplification and 1,845 did not. Two specimens had a FISH ratio of <2.0 but high average *HER-2* gene copy number (27.50 and 15.35 copies per tumor cell nucleus) as well as *HER-2* gene signals arranged in aggregates, a characteristic of amplified genes located in homogeneous staining regions of a chromosome. We reclassified these breast cancers as *HER-2* amplified, bringing the total number of breast cancers with *HER-2* gene amplification to 657 (26.3%). The FISH ratios ranged from 0.51 to 48.87 (Fig. 1C). Only 40 of the 2,502 (1.6%) breast cancer cases had FISH ratios within 10% of the 2.0 cutoff (between 1.80 and 2.20) for gene amplification.

Comparison of fluorescence in situ hybridization results for HER-2 status in the Breast Cancer International Research Group central laboratories with results from outside/local laboratories.

A total of 1,600 of 2,502 (63.9%) breast cancers with a central laboratory FISH result also had at least one *HER-2* assay previously done at an outside laboratory (1,536 using immunohistochemistry, 131 using FISH, and 17 using unspecified methods; Fig. 1B). Among those tested in outside/local laboratories by immunohistochemistry, we were able to classify the results for 1,457: 637 (43.7%) showed overexpression and 820 (56.3%) showed low expression. Four hundred five (27.8%) of these were amplified by FISH in a BCIRG central laboratory, whereas 1,052 (72.2%) were not.

Central FISH assay failures occurred in 25 specimens for which we had outside/local immunohistochemistry results (18 with low expression and 7 with overexpression). The overall agreement rate between immunohistochemistry results in outside laboratories and FISH in a BCIRG central laboratory for the 1,457 specimens with both results was 79%. Thirty-six of the 404 (8.9%) FISH amplified cases had low immunohistochemistry expression. These represented only 4.4% (36 of 820) of the breast cancers with low expression and are considered to be false-negative results by immunohistochemistry as explained below. Conversely, using FISH as a "gold standard" for determining the presence or absence of the *HER-2* alteration, 268 of 637 (42.1%) cases classified as showing overexpression by immunohistochemistry were not amplified by FISH and are considered to be false-positive results by immunohistochemistry as described below.

Among the 1,457 breast cancers first analyzed at outside laboratories by immunohistochemistry assay were 1,407 breast cancers classified using an immunohistochemistry standard 0, 1+, 2+, or 3+ immunohistochemistry scoring system (Table 1). A total of 20 of 558 (3.6%) of the specimens with immunohistochemistry classified as 0 and 15 of 245 (6.1%) specimens with immunohistochemistry classified as 1+ showed gene amplification by FISH. Conversely, only 33 of 198 (16.7%) specimens scored in the 2+ immunohistochemistry category and 317 of 406 (77.1%) specimens in the 3+ immunohistochemistry category were amplified by FISH.

The 1,536 outside immunohistochemistry assays were done by a variety of techniques, including 885 by the DAKO HercepTest, 157 by the CB11 Ventana Pathway Assay, 67 by CB11 assays other than the Ventana kit, 42 by TAB250 (Zymed Laboratories, Inc., San Francisco, CA), 24 by CB11 plus

Table 3. HER-2 status determined by immunohistochemistry or FISH done in outside laboratories compared with HER-2 status determined in the BCIRG central laboratory by FISH

Assay	No. positive/negative	Sensitivity (95% CI)	Specificity (95% CI)	κ (95% CI)	Accuracy (95% CI)
All immunohistochemistry (n = 1,457)	637/820	369/405 = 0.91 (0.88-0.94)	784/1,052 = 0.75 (0.72-0.77)	0.56 (0.52-0.60)	0.79 (0.77-0.81)
DAKO (n = 862)	396/466	232/251 = 0.92 (0.88-0.95)	447/611 = 0.73 (0.69-0.77)	0.56 (0.51-0.61)	0.79 (0.76-0.81)
Ventana (n = 153)	65/88	34/36 = 0.94 (0.81-0.99)	86/117 = 0.74 (0.65-0.81)	0.53 (0.40-0.66)	0.78 (0.71-0.85)
T250 and CB11 (n = 23)	8/15	5/5 = 1.0 (0.48-1.0)	15/18 = 0.83 (0.59-0.96)	0.68 (0.37-1.00)	0.87 (0.66-0.97)
T250 (n = 37)	9/28	5/7 = 0.71 (0.29-0.96)	26/30 = 0.87 (0.69-0.96)	0.52 (0.19-0.85)	0.84 (0.68-0.94)
CB11 (n = 63)	25/38	18/19 = 0.95 (0.74-1.0)	37/44 = 0.84 (0.70-0.93)	0.72 (0.55-0.90)	0.87 (0.77-0.94)
FISH (n = 131)	50/81	41/42 = 0.98 (0.87-1.0)	80/89 = 0.90 (0.82-0.95)	0.83 (0.73-0.93)	0.92 (0.86-0.96)

TAB250 together, 24 by the ChromaVision ACIS computerized imaging system, 108 by a variety of other available antibodies, and 229 by unspecified immunohistochemical methods. Immunohistochemistry results using the standard scoring system (0, 1+, 2+, and 3+) were available for 842 specimens evaluated using the DAKO HercepTest (Table 2). Of these, 385 (45.7%) showed overexpression (2+ or 3+ immunohistochemistry staining) and 457 (54.3%) showed low expression (0 or 1+ immunohistochemistry staining). When compared with HER-2 status determined by FISH in the two BCIRG central laboratories, overall agreement was 78.6% (662 of 842) with a κ statistic of 0.56 (95% CI, 0.51-0.61). Eleven cases (3.6%) with gene amplification determined by FISH lacked HER-2 immunostaining (scores of 0) by immunohistochemistry and an additional 8 FISH amplified cases (5.3%) had only weak HER-2 immunostaining (scores of 1+) by immunohistochemistry. These cases were considered as immunohistochemistry false-negative results. Only 58.4% (225 of 385) of cases that scored 2+ or 3+ by immunohistochemistry had HER-2 gene amplification by FISH. Nearly two-thirds of the disagreements (103 of 160 or 64%) were immunohistochemistry 2+, FISH-negative breast cancers, but a high proportion of disagreements (57 of 160 or 35.6%) was also found in the immunohistochemistry 3+ category. Only 21 of 124 (16.9%) cases with 2+ immunohistochemistry staining had gene amplification, whereas 204 of 261 cases (78.2%) with 3+ immunohistochemistry staining had gene amplification.

The standard 0 to 3+ scoring system was used for 151 of 157 breast cancer cases tested using the Ventana Pathway HER-2 CB11 immunohistochemistry assay method; 88 (58.3%) showed low expression (0 and 1+ immunostaining) and 63 (41.7%) showed overexpression (2+ and 3+ immunostaining). The overall agreement rate was 78% (33 specimens with both amplification and overexpression and 86 with neither amplification nor overexpression) and the κ statistic was 0.53 (95% CI, 0.40-0.66; Table 3). The 3+ immunohistochemistry staining category contained 39 of 151 (25.8%) breast cancers, which was similar to the FISH gene amplification rate (23.2%). If classification of overexpression were limited to the 3+ immunohistochemistry category, the agreement rate with FISH would have been substantially higher (93.4%) than the 78% agreement rate (Table 3) observed using the Ventana package insert recommendations.

A FISH assay method was used to evaluate HER-2 status for 132 breast cancer cases in outside laboratories. Only 1 sample (0.8%) was not analyzed successfully at BCIRG central laboratories. Among the 131 cases with FISH results from both the BCIRG and the outside/local laboratories, 50 (38.2%) showed gene amplification, whereas 81 (61.8%) did not (Table 3). Forty-one cases were found to be amplified in both outside and BCIRG laboratories, whereas 80 were not amplified in either the outside or the BCIRG laboratory (agreement rate, 92.4%; κ statistic, 0.83; 95% CI, 0.73-0.93). The specific FISH assays used by the outside laboratories were not available.

Fluorescence in situ hybridization versus immunohistochemistry, both done in the Breast Cancer International Research Group central laboratories. Because some investigators have suggested that discrepancies in HER-2 status determined by FISH and immunohistochemistry are the result of HER-2 gene single-copy overexpression, we did immunohistochemistry assays for HER-2 expression on 2,243 specimens using a single, accurate

immunohistochemistry assay method with the 10H8 monoclonal antibody (20, 24). This antibody does not involve antigen retrieval of specimens and therefore avoids documented problems with false-positive assay results (33). Only 129 of 1,739 (7.4%) of the breast cancer cases scored as 0 or 1+ by 10H8-IHC showed HER-2 gene amplification, whereas 49 of 504 (9.7%) scored as 2+ or 3+ by 10H8-IHC showed no HER-2 gene amplification (Table 4). The agreement between immunohistochemistry and FISH results using the 10H8 antibody was 92% (2,065 of 2,243; κ statistic, 0.78; 95% CI, 0.75-0.81), which was a substantial improvement over the results obtained with any of the immunohistochemical methods used by outside laboratories, including the FDA-approved antibody methods (Table 3).

Comparison of immunohistochemistry done in the Breast Cancer International Research Group central laboratories with immunohistochemistry done in outside laboratories. We tested 1,291 of the 1,407 breast cancers analyzed in outside laboratories by immunohistochemistry with 0 to 3+ scoring by 10H8-IHC in the BCIRG central laboratories. Of the 748 breast cancers characterized as having low HER-2 expression (0 or 1+ immunostaining) in outside laboratories, 719 (96.1%) were also characterized as having low expression (0 or 1+ immunostaining) in the BCIRG central laboratories (Table 5), whereas only 281 of 543 (51.7%) breast cancers having HER-2 overexpression (2+ or 3+ immunostaining) at outside laboratories also had overexpression in the central laboratories. In addition, only 256 of 360 (71.1%) breast cancers determined to have 3+ immunohistochemistry immunostaining in outside laboratories had 2+ or 3+ immunohistochemistry immunostaining in the BCIRG central laboratories. Overall agreement between these results was 77.5% and the κ statistic was 0.51 (95% CI, 0.46-0.55). These data showed substantial disagreement between outside/local laboratories and BCIRG central laboratories, with only 281 of 543 (51.7%) breast cancers scored as 2+ or 3+ in outside laboratories showing 2+ or 3+ immunostaining in the central laboratories. Likewise, only 256 of the 360 (71.1%) breast cancers scored as immunohistochemistry 3+ in outside/local laboratories also showed either 2+ or 3+ immunostaining in the central laboratories.

Potential for selection bias among breast cancer specimens submitted to Breast Cancer International Research Group central laboratories. Because the breast cancer cases reported here were from women who were candidates for entry to BCIRG clinical trials, we were concerned about potential selection bias in the samples submitted to our laboratories. Because two of the three clinical trials (BCIRG006 and BCIRG007) were designed for women with HER-2 gene amplification and only one clinical trial (BCIRG005) was designed for women without HER-2 gene amplification, we considered the possibility that the previous outside HER-2 test results might have been used as a preliminary screen to identify suitable patients for specific clinical trials. To address this issue, we compared the HER-2 gene amplification rate among women who had previous HER-2 testing with that of women with no previous HER-2 testing. Overall, 1,018 of 1,368 (74.4%) cases tested at USC had outside test result information available as did 529 of 1,175 (45.0%) cases tested at the University of Basel. Among the 996 cases lacking outside test results, the HER-2 gene amplification rate was 23.2%; among those with outside test results, the amplification rate was 28.2%. These positivity rates, which were

Table 4. Comparison of HER-2 status determined by FISH and by 10H8-IHC in the BCIRG central laboratories

	10H8-IHC scores (0-3+)				Total
	0	1+	2+	3+	
FISH negative (%)	1,281 (94.6)	329 (85.5)	46 (19.4)	3 (1.1)	1,659 (74.0)
FISH positive (%)	73 (5.4)	56 (14.5)	191 (80.6)	264 (98.9)	584 (26.0)
Total (%)	1,354 (100)	385 (100)	237 (100)	267 (100)	2,243 (100)

different from one another ($P = 0.005$), did not vary meaningfully by central laboratory.

Discussion

The critical issue of clinical laboratory testing methods used for evaluation of the HER-2 alteration in breast cancer has remained surprisingly controversial. Some have argued that *HER-2* gene amplification, determined by FISH, does not correspond directly with *HER-2* gene overexpression, determined by immunohistochemistry (23); however, we disagree. Previous DNA testing that uses frozen breast cancer specimens eliminates fixation or processing artifacts that clearly affect protein detection. Studies with such tissues show that *HER-2* gene amplification and pathologic *HER-2* overexpression results agree in essentially all cases (4, 20, 21, 34). Furthermore, the FDA has approved selected immunohistochemistry and FISH assays as acceptable methods for identifying *HER-2* alterations (amplification/overexpression) when selecting patients for trastuzumab-based therapy. Both methods are being used independently to identify the *HER-2* alteration (amplification/overexpression) for selection of *HER-2*-directed therapy in clinical trials evaluating the efficacy of various experimental drug regimens as well as in the clinical setting (35, 36). It is, therefore, important to evaluate the performance of these assay methods in a large, prospectively acquired set of samples.

Our initial studies comparing gene amplification determined by Southern hybridization with gene expression determined by Northern hybridization, Western immunoblot, and frozen section immunohistochemistry identified 18 of 187 (10%) that were considered single-copy overexpressors. These cases lacked gene amplification by Southern blot analysis but clearly had overexpression (4). Subsequent analysis of these same specimens using FISH as opposed to Southern blot showed that

almost all of these single-copy overexpression breast cancers had gene amplification by FISH (21). The majority of these Southern hybridization false-negative results were attributed to a dilutional artifact in predominantly stroma-rich breast cancers containing relatively few tumor cells. Overall, 98% of frozen breast cancer specimens showed a direct association between *HER-2* gene amplification and gene expression (21). We consider the remaining 2% to be within the range expected for experimental error. These results show that *HER-2* gene amplification is directly associated with pathologic or altered *HER-2* overexpression, whereas nonamplified specimens contain expression levels directly associated with the low expression range observed in normal tissues (29). Others have described similar findings (34, 37).

Because of questions about the reproducibility of *HER-2* assays done in different laboratories, the BCIRG decided to centralize testing for *HER-2* for its Herceptin-based therapeutic clinical trials. FISH was selected as the method of choice based primarily on two considerations. First, compelling data exist demonstrating that FISH is the most accurate FDA-approved method for evaluating paraffin-embedded clinical specimens (20, 38). Secondly, and perhaps more relevant to the central question of selecting patients for Herceptin-based therapeutic trials, retrospective analysis of *HER-2* in the breast cancers of women entered in Herceptin clinical trials show that patients with amplification by FISH had greater responsiveness to Herceptin than patients with overexpression by immunohistochemistry (25).

In this report, we used FISH assay results as our gold standard for purposes of assessing sensitivity, specificity, and overall agreement because *HER-2* status as determined by FISH analysis is similar to the known "true" molecular status as determined by a comprehensive molecular analysis in frozen tissue (3, 20, 21). We have shown previously that FISH assays done in paraffin-embedded tissues provide an assessment of the *HER-2* gene status that is 97% accurate (20, 21). In addition, FISH is considerably less affected by tissue processing artifacts than is immunohistochemistry (20, 33, 34). In those cases where FISH is affected by tissue processing, one obtains a nonresult (i.e., FISH hybridization fails, resulting in no information). This was observed in only ~1.6% of the breast cancers in the current study. Conversely, an "immunohistochemistry-negative" result could be due to either minimal levels of *HER-2* expression or a true assay failure due to antigen alteration resulting from tissue processing artifacts, most notably fixation. The inability to distinguish between these two possibilities can lead to incorrectly identifying a patient as *HER-2* negative when she is not. This can lead to false-negative immunohistochemistry rates as high as 70% (39). In an attempt to overcome this potential pitfall of

Table 5. Immunohistochemical assay results from outside/local laboratories compared with 10H8-IHC assay results from the BCIRG central laboratories

	All outside/local laboratory immunohistochemistry				
	0	1+	2+	3+	Total (%)
0	409	158	113	65	745 (57.7)
1+	96	56	45	39	236 (18.3)
2+	13	8	15	99	135 (10.5)
3+	4	4	10	157	175 (13.6)
Total (%)	522 (40.4)	226 (17.5)	183 (14.2)	360 (27.9)	1,291 (100)

immunohistochemistry, many have used antigen retrieval methods to increase the immunostaining intensity and minimize the number of immunohistochemistry false-negatives. The FDA-approved immunohistochemistry assays routinely use "antigen retrieval" techniques for all specimens to reverse the effects of tissue fixation/processing and as a result enhance the amount of immunostaining. Antigen retrieval, such as heat-induced antigen retrieval, is used to boost the amount of immunostaining obtained with antibodies that have reduced sensitivity for paraffin-embedded HER-2 protein. Although the amount of immunostaining in these cases is increased by antigen retrieval, the increased immunostaining may not be proportional to amount of HER-2 protein in the tumor specimen. Therefore, antigen retrieval can be viewed as causing an increase in all immunostaining with a "shift to the right" (0 → 1+ → 2+ → 3+) in the amount of immunostaining observed compared with the amount of immunostaining observed with the same antibody in the absence of antigen retrieval. This results in an increased number of false-positive cases by immunohistochemistry [i.e., cases that show increased immunohistochemistry staining (2+ or 3+)] and is considered immunohistochemistry positive but have neither *HER-2* gene amplification nor *HER-2* overexpression.

We do not consider immunohistochemistry screening for entry to clinical trials or for selection to Herceptin immunotherapy to be an acceptable strategy. Approximately 8% (26, 40) to 22% of women whose breast cancers have immunohistochemistry 3+ *HER-2* staining do not have *HER-2* gene amplification (or overexpression); therefore, inclusion of these women in the clinical trial will lead to a substantial proportion of study subjects who have breast cancers that do not contain the molecular target for Herceptin treatment. Likewise, exclusion of all women whose breast cancers have 0/1+ immunohistochemistry immunostaining (~75% of breast cancer cases) excludes between 9% and 17% of women whose breast cancers actually have the molecular target of *HER-2* amplification/overexpression. Because FISH is a more accurate method than immunohistochemistry for correct identification of the *HER-2* alteration (amplification/overexpression) in known molecularly characterized cases (20, 33, 34), we have used the FISH results both as an entry criterion for the clinical trial and as the standard for assessing true-positive and true-negative results.

FISH assay results are affected relatively little by tissue processing for paraffin block preparation, whereas immunohistochemistry assay results are variably affected (39). This degree of variability ranges from minimal to substantial depending on the type of fixation, the length of fixation, and the type of tissue embedding techniques used. Other investigators have already shown high false-positive rates for FDA-approved immunohistochemical methods (37, 41, 42). Given the minimal effect of tissue processing artifacts on FISH, the significantly higher accuracy of FISH to detect *HER-2* alterations, and the substantial number of immunohistochemistry false-positive results already described by other investigators and ourselves (26), we believe that it is reasonable to consider that central laboratory FISH-negative, outside/local laboratory immunohistochemistry-positive cases represent false-positive immunohistochemistry results. This presumption is supported by the high rate of immunohistochemistry positivity (>40%) observed for outside laboratories in this series compared with the expected, appropriate rate of FISH positivity (26%) observed in the

current study. These conclusions are further supported by results obtained with 10H8-IHC assays done by the BCIRG central laboratories using a method only slightly less sensitive than FISH (20). This more sensitive immunohistochemistry antibody method does not require antigen retrieval and confirms that outside or local hospital immunohistochemistry-positive, FISH-negative results are, in fact, 10H8-IHC negative, FISH negative. Finally, the 92% agreement rate between FISH results and 10H8-IHC results for assays done in the central laboratories was consistent with that expected from earlier characterization of the accuracy of these methods (20).

The current study confirms differences in *HER-2* status based on outside laboratory testing by immunohistochemistry and central laboratory testing by FISH. Tissue fixation and processing for paraffin block embedding of breast cancer tissue is associated with loss of antigenicity (43) and consequently loss of immunohistochemical staining for protein antigens, including *HER-2* (4, 39). Further, the use of antigen retrieval in paraffin-embedded tissue is associated with a variable, increased intensity of immunostaining that is not directly proportional to the original amount of protein contained in the tissue section. Thus, antigen retrieval contributes to false-positive immunostaining results. This has important clinical implications because women who falsely test positive with immunohistochemistry have little chance of responding to Herceptin-based therapy (44–46), whereas those who falsely test negative are denied this treatment.

A crucial observation in the current study is the substantially higher agreement (92%) between FISH done at the BCIRG central laboratories and FISH done at outside referring laboratories compared with the lower agreement rate (79%) observed between FISH done at the Central BCIRG Laboratories and immunohistochemistry done at outside laboratories. A total of 293 outside laboratories contributed 1,536 immunohistochemistry assay results and 64 outside laboratories contributed 131 FISH assay results to the current study. Using the central laboratory FISH results as the standard method for comparison, outside laboratory FISH results showed higher sensitivity, specificity, and accuracy than any of the outside laboratory immunohistochemistry results.

The agreement rate for FISH observed in this study is similar to the 92% agreement rate we reported previously when comparing results of FISH assays done independently at USC and at LabCorp (26). In that assessment, we used 488 paraffin-embedded tissue samples for women enrolled in the Genentech, Inc., clinical trials that led to Herceptin registration (25, 26). Women had been deemed eligible for inclusion in these trials based on immunohistochemistry testing. The accuracy of a breast cancer patient's *HER-2* status is of paramount importance when interpreting results from clinical trials. Inclusion of women in clinical trials of Herceptin who have false-positive breast cancers has the effect of diminishing the apparent response rate to treatment, whereas false-negative results exclude women who have the potential to respond to *HER-2*-directed therapy (46). There is clear evidence from the registrational studies for Herceptin that this has in fact occurred using immunohistochemical methods (25, 26).

Comparisons of immunohistochemistry results from accruing institutions for two clinical trials with FISH or immunohistochemistry done by a central laboratory show poor agreement. Only 79% of the first 104 women with 3+

immunohistochemical staining as done by the accruing institutions who were entered in the National Surgical Adjuvant Breast and Bowel Project protocol B-31 had gene amplification by FISH (35). In addition, repeat immunohistochemistry assays done at a central laboratory confirmed the strongly positive (3+) immunohistochemistry result by HercepTest in only 82 of 104 (79%) of these cases (35). Similarly, a Breast Intergroup report for protocol N9831 describing the first 119 women with 3+ immunohistochemistry immunostaining at the accruing institution shows that only 74% were 3+ by HercepTest done centrally.

Additionally, the College of American Pathologists (CAP) has reported higher agreement among clinical laboratories performing FISH assays than among clinical laboratories performing immunohistochemistry assays to assess HER-2 status (47). CAP conducted proficiency surveys in 2000 and 2001 with clinical laboratories that provide HER-2 testing. The same two cases, one with amplification/overexpression and one with no amplification/low expression, were provided to all clinical laboratories. Those laboratories using FISH assays showed complete (100%) agreement for the assessment of HER-2 status, whereas those laboratories performing immunohistochemistry assays showed a 72.3% agreement rate for HER-2 positivity based on immunostaining versus no immunostaining.

Several factors may account for the higher agreement rates between the BCIRG central laboratories and outside laboratories performing FISH. Similar techniques are being compared, although outside FISH was not prepared by any standardized protocol. Normal cells included in each biopsy specimen, which represent a critically important quantitative internal control for FISH, can be used as internal controls to confirm that the normal number of *HER-2* gene copies are identified in normal cells. In addition, FISH is easier to standardize because it is a semiquantitative assay method, whereas immunohistochemistry is a subjective assay method and subjective methods are more difficult to standardize from one laboratory to another. Finally, 64 clinical sites contributed FISH assay results (average of 2.05 assays per site), whereas 293 clinical sites contributed the immunohistochemistry assay results (average of 5.24 assays per site).

Not all investigators report high agreement rates between laboratories performing FISH assays. In the N9831 Intergroup Breast Cancer clinical trial, only six of nine cases tested locally by FISH had gene amplification confirmed in the central laboratory (36). A note added in proof to this report states that among 36 women enrolled in the trial with local FISH testing only 23 (64%) had tumors that were amplified by central FISH testing (36). The Intergroup experience with local versus central FISH testing contrasts significantly with the experience reported here (131 cases from 64 different laboratories) and elsewhere (488 cases; ref. 26) as well as the experience of the CAP (47, 48) and others (49). The potential for drawing erroneous conclusions from the analysis of small numbers of cases has already been clearly shown in the field of HER-2 testing (8) and is also a concern for interpretation of FISH results.

The BCIRG collection of a large group of HER-2 test results from referring institutions screened prospectively for randomized therapeutic trials clearly reflects use of these assays in the "general medical community" with relatively few breast cancer patients (131 of 1,617 or 8.1%) evaluated by FISH. Most of these outside laboratory FISH assays (123 of 131 or 94%) were done in the United States. Data from CAP show that 90.3% of

surveyed laboratories in their study were not performing FISH to determine HER-2 status but instead using immunohistochemistry (47). In our study, >90% of the outside HER-2 analyses were also done by immunohistochemistry with the majority of these tests (862 of 1,617 or 53%) done by the DAKO immunohistochemistry HercepTest.

Based on previous studies and the current investigation, FISH seems to be more accurate and more reproducible across laboratories. However, some have questioned the feasibility of routinely performing FISH as a first-line test in general practice, suggesting that "practical barriers" to primary FISH testing exist. The most important component of the FISH evaluation is the ability to distinguish tumor cells from normal reactive cells in 4',6-diamidino-2-phenylindole-stained tissue sections. Board-certified pathologists are the most qualified professionals to make this distinction. Perhaps some are concerned that the technical aspects of FISH, which are somewhat more demanding than immunohistochemistry, cannot be done in all local hospital laboratories. FISH is, however, routinely done in local cytogenetic laboratories and these laboratories could assist local surgical pathology laboratories. Where FISH cannot be done, tissue blocks can be submitted by overnight mail to a reference laboratory that is equipped to perform the FISH assay. There are ~146 laboratories in the United States performing HER-2 FISH assays that are also participants in the CAP quality assurance program for FISH.⁷ Others have raised issues of cost as an objection to performing primary FISH testing of breast cancer specimens (40); however, cost analyses from two independent studies have favored primary FISH testing as an economically superior strategy when the cost of incorrectly using trastuzumab is considered (50, 51).

We describe the experience of the BCIRG with the establishment of two centralized laboratories to evaluate the status of a molecular target, the HER-2, for entry of women to prospective clinical trials of trastuzumab in the adjuvant treatment of breast cancer. This is, to our knowledge, the first large, prospective clinical trial of a targeted therapeutic to make use of centralized laboratories to screen for the presence of the HER-2 alteration as an entry criterion for selection to the clinical trial. Usually, the status of molecular targets is evaluated in many local or hospital laboratories to permit entry to clinical trials. The current study, therefore, provided an opportunity to compare the status of HER-2 determined in two central laboratories with the status of the molecular target assessed in many local/hospital laboratories. Our current findings, as well as those of others, including CAP, indicate substantial problems with immunohistochemistry-based HER-2 assay methods used in the general medical community. The data from this study support the importance of using accurate, reliable, and reproducible diagnostic testing for molecular alterations that are to be the molecular target for therapeutics, such as trastuzumab.

Acknowledgments

We thank Jennifer Stepic (USC), Armen Gasparyan (USC), Hedvika Novotny (Basel), Sandra Schneider (Basel), Rosemarie Chaffard (Basel), and Dr. Maria Sibug-Saber (USC) for technical assistance; David J. Press (USC) for clerical assistance; and Camille Lorent (BCIRG), Valerie Bee (BCIRG), and Henry Taupin (BCIRG) for logistical support.

⁷ R. Tubbs, personal communication, 2003.

References

- Coussens L, Yang-Feng TL, Chen YC, et al. Tyrosine kinase receptor with extensive homology to EGF receptor shares chromosomal location with *neu* oncogene. *Science* 1985;230:1132–9.
- King CR, Kraus MH, Aaronson SA. Amplification of a novel *v-erb B*-related gene in a human mammary carcinoma. *Science* 1985;229:974–6.
- 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;235:177–82.
- Slamon DJ, Godolphin W, Jones LA, et al. Studies of the HER-2/*neu* proto-oncogene in human breast and ovarian cancer. *Science* 1989;244:707–12.
- Press MF, Bernstein L, Thomas PA, et al. HER-2/*neu* gene amplification by fluorescence *in situ* hybridization: evaluation of archival specimens and utility as a marker of poor prognosis in node-negative invasive breast carcinomas. *J Clin Oncol* 1997;15:2894–904.
- Press MF, Pike MC, Chazin VR, et al. HER-2/*neu* expression in node-negative breast cancers: direct tissue quantitation by computerized image analysis and association of overexpression with increased risk of recurrent disease. *Cancer Res* 1993;53:4960–70.
- Ro J, El-Naggar A, Ro JY, et al. c-erb B-2 amplification in node-negative breast cancer. *Cancer Res* 1989;49:6941–4.
- Gullick WJ, Love SB, Wright C, et al. c-erb B-2 protein over-expression in breast cancer is a risk factor in patients with involved and uninvolved lymph nodes. *Br J Cancer* 1991;63:434–8.
- Slamon DJ, Press MF, Godolphin W, et al. Studies of the HER-2/*neu* proto-oncogene in human breast cancer. *Cancer Cells* 1989;7:371–80.
- Winstanley J, Cooke T, Murray GD, et al. The long term prognostic significance of c-erb B-2 in primary breast cancer. *Br J Cancer* 1991;63:447–50.
- Kallioniemi O-P, Hollt K, Visakorpi T, Koivula T, Helin H, Isola J. Association of c-erb B-2 oncogene overexpression with high rate of cell proliferation, increased risk for visceral metastasis and poor long-term survival in breast cancer. *Int J Cancer* 1991;49:650–5.
- Gusterson BA, Gelber RD, Goldhirsch A, et al. Prognostic importance of c-erbB-2 expression in breast cancer. International (Ludwig) Breast Cancer Study Group. *J Clin Oncol* 1992;10:1049–56.
- Borg A, Tandon AK, Sigurdsson H, et al. HER-2/*neu* amplification predicts poor survival in node-positive breast cancer. *Cancer Res* 1990;50:4332–7.
- Carlomagno C, Perrone F, Gallo C, et al. c-erb B2 overexpression decreases the benefit of adjuvant tamoxifen in early-stage breast cancer without axillary lymph node metastases. *J Clin Oncol* 1996;14:2702–8.
- Borg A, Baldetorp B, Ferno M, et al. erb B2 amplification is associated with tamoxifen resistance in steroid-receptor positive breast cancer. *Cancer Lett* 1994;81:137–44.
- Muss HB, Thor AD, Berry DA, et al. c-erb B-2 expression and response to adjuvant therapy in women with node-positive early breast cancer. *N Engl J Med* 1994;330:1260–6.
- Thor AD, Berry DA, Budman DR, et al. erb B-2, p53, and efficacy of adjuvant therapy in lymph node-positive breast cancer. *J Natl Cancer Inst* 1998;90:1346–60.
- Tetu B, Brisson J, Plante V, Bernard P. p53 and c-erb B-2 as markers of resistance to adjuvant chemotherapy in breast cancer. *Mod Pathol* 1998;11:823–30.
- Stal O, Sullivan S, Wingren S, et al. c-erb B-2 expression and benefit from adjuvant chemotherapy and radiotherapy of breast cancer. *Eur J Cancer* 1995;31A:2185–90.
- Press MF, Slamon DJ, Flom KJ, Park J, Zhou J-Y, Bernstein L. Evaluation of HER-2/*neu* gene amplification and overexpression: comparison of frequently used assay methods in a molecularly characterized cohort of breast cancer specimens. *J Clin Oncol* 2002;20:3095–105.
- Pauletti G, Godolphin W, Press MF, Slamon DJ. Detection and quantitation of HER-2/*neu* gene amplification in human breast cancer archival material using fluorescence *in situ* hybridization. *Oncogene* 1996;13:63–72.
- Petrosyan K, Press MF. Multispecimen tissue blocks in pathology: an improved technique of preparation. *Lab Invest* 1997;77:541–2.
- Birner P, Oberhuber G. Is fluorescence *in situ* hybridization really superior to HercepTest? *J Clin Oncol* 2002;20:4607.
- Park JM, Yang X, Park J, Press O, Press MF. Assessment of novel anti-p185HER-2 monoclonal antibodies for internalization-dependent therapies. *Hybridoma* 1999;18:487–95.
- Mass RD, Press MF, Anderson S, et al. Evaluation of clinical outcomes according to HER2 detection by fluorescence *in situ* hybridization in women with metastatic breast cancer treated with trastuzumab. *Clin Breast Cancer*. In press 2005.
- Dybdal N, Lieberman G, Anderson S, et al. Determination of HER2 gene amplification by fluorescence *in situ* hybridization and concordance with the clinical trials immunohistochemical assay in women with metastatic breast cancer evaluated for treatment with trastuzumab. *Breast Cancer Res Treat*. In press 2005.
- Kononen J, Bubendorf L, Kallioniemi A, et al. Tissue microarrays for high-throughput molecular profiling of tumor specimens. *Nat Med* 1998;4:844–7.
- Simon R, Nocito A, Hubscher T, et al. Patterns of HER-2/*neu* amplification and overexpression in primary and metastatic breast cancer. *J Natl Cancer Inst* 2001;93:1141–6.
- Press MF, Cordon-Cardo C, Slamon DJ. Expression of the HER-2/*neu* proto-oncogene in normal adult and fetal tissues. *Oncogene* 1990;5:953–62.
- Fleiss JL. Statistical methods for rates and proportions. New York: John Wiley & Sons; 1981. p. 212–36.
- Landis JR, Koch GG. The measurement of observer agreement for categorical data. *Biometrics* 1977;33:159–74.
- Potter HG, Linklater JM, Allen AA, Hannafin JA, Haas SB. Magnetic resonance imaging of articular cartilage in the knee: an evaluation with use of fast-spin-echo imaging. *J Bone Joint Surg Am* 1998;80-A:1276–84.
- Press MF, Bernstein L, Flom KJ, Slamon DJ. In reply: FDA-approved HercepTest is far less sensitive for detection of breast cancer. *J Clin Oncol* 2002;20:4607–9.
- Bartlett JM, Going JJ, Mallon EA, et al. Evaluating HER2 amplification and overexpression in breast cancer. *J Pathol* 2001;195:422–8.
- Paik S, Bryant J, Tan-Chiu E, et al. Real-world performance of HER2 testing—National Surgical Adjuvant Breast and Bowel Project Experience. *J Natl Cancer Inst* 2002;94:852–4.
- Roche PC, Suman VJ, Jenkins RB, et al. Concordance between local and central laboratory HER2 testing in the Breast Intergroup Trial N9831. *J Natl Cancer Inst* 2002;94:855–7.
- Tubbs RR, Pettay JD, Roche PC, Stoler MH, Jenkins RB, Grogan TM. Discrepancies in clinical laboratory testing of eligibility for trastuzumab therapy: apparent immunohistochemical false-positives do not get the message. *J Clin Oncol* 2001;19:2714–21.
- Press MF, Anderson S, Dybdal N, Lieberman G, Mass R. Fluorescence *in situ* hybridization (FISH) is superior to immunohistochemistry (IHC) for determining HER2 status: the Herceptin experience [abstract 186]. *Lab Invest* 2002;82:47A.
- Press MF, Hung G, Godolphin W, Slamon DJ. Sensitivity of HER-2/*neu* antibodies in archival tissue samples: potential source of error in immunohistochemical studies of expression. *Cancer Res* 1994;54:2771–7.
- Yaziji H, Goldstein LC, Barry TS, et al. HER-2 testing in breast cancer using parallel tissue-based methods. *J Am Med Assoc* 2004;291:1972–7.
- Jacobs TW, Gown AM, Yaziji H, Barnes MJ, Schnitt SJ. Specificity of HercepTest in determining HER-2/*neu* status of breast cancers using the United States Food and Drug Administration-approved scoring system. *J Clin Oncol* 1999;17:1983–7.
- Jacobs TW, Gown AM, Yaziji H, Barnes MJ, Schnitt S. Comparison of fluorescence *in situ* hybridization and immunohistochemistry for the evaluation of HER-2/*neu* in breast cancer. *J Clin Oncol* 1999;17:1974–82.
- Werner M, Chott A, Fabiano A, Battifora H. Effect of formalin tissue fixation and processing on immunohistochemistry. *Am J Surg Pathol* 2000;24:1016–9.
- Mass RD, Press MF, Anderson S, Murphy M, Slamon DJ. Improved survival benefit from Herceptin (trastuzumab) in patients selected by fluorescence *in situ* hybridization (FISH) [abstract 85]. *Proc Am Soc Clin Oncol* 2001;20:22a.
- Press MF, Slamon DJ, Cobleigh M, et al. Improved clinical outcomes for Herceptin-treated patients selected by fluorescence *in situ* hybridization (FISH) [abstract 185]. *Lab Invest* 2002;82:47A.
- Vogel CL, Cobleigh MA, Tripathy D, et al. Efficacy and safety of trastuzumab (Herceptin) as a single agent in first-line treatment of HER2-overexpressing metastatic breast cancer. *J Clin Oncol* 2002;20:719–26.
- Cell Markers and Cytogenetics Committees College of American Pathologists clinical laboratory assays for HER-2/*neu* amplification and overexpression: quality assurance, standardization, and proficiency testing. *Arch Pathol Lab Med* 2002;126:803–8.
- Mascarello JT, Brothman AR, Davison K, et al. Proficiency testing for laboratories performing fluorescence *in situ* hybridization with chromosome-specific DNAProbes. *Arch Pathol Lab Med* 2002;126:1458–62.
- Persons DL, Bui MM, Lowery MC, et al. Fluorescence *in situ* hybridization (FISH) for detection of HER-2/*neu* amplification in breast cancer: a multicenter portability study. *Ann Clin Lab Sci* 2000;30:41–8.
- Elkin EB, Weinstein M, Winer EP, Kuntz KM, Schmitt SJ, Weeks JC. HER2 testing and trastuzumab therapy for metastatic breast cancer: a cost-effectiveness analysis. *J Clin Oncol* 2004;22:854–63.
- Kamath T. Cost-effectiveness of alternative testing strategies to select women with metastatic breast cancer for treatment with trastuzumab [in review]. *J Clin Oncol* 2005.