Standardizing Slide-Based Assays in Breast Cancer: Hormone Receptors, HER2, and Sentinel Lymph Nodes

Jeffrey S. Ross,1 W. Fraser Symmans,2 Lajos Pusztai,3 and Gabriel N. Hortobagyi3

Abstract

Despite the rapid expansion of novel diagnostics designed to personalize breast cancer care, there remain several significant unmet needs for improving the accuracy and reliability of tests that are already in common daily clinical practice. For example, although immunohistochemistry has been the predominant method for measuring estrogen receptor and progesterone receptor status for over 15 years, this assay remains unstandardized and there is a widespread concern that inaccuracy in immunohistochemistry technique and interpretation is leading to an unacceptably high error rate in determining the true hormone receptor status. Similarly, there is considerable concern that both false-negative and false-positive result rates for testing for HER2 status are unacceptably high in current clinical practice. This commentary considers a variety of factors, including preanalytic conditions and slide-scoring procedures, and other variables that may be contributing to current testing error rates and why there is a great need for the standardization of these biomarker assay procedures to further enable the highest possible quality of care for newly diagnosed breast cancer patients.

Background

In 1998, the simultaneous regulatory approvals of trastuzumab (Herceptin) and the HercepTest immunohistochemistry diagnostic for patients with HER2-overexpressing metastatic breast cancers launched the era of drug and test combinations (1, 2). The recent clinical trial results and U.S. Food and Drug Administration approval of trastuzumab-based therapy in the adjuvant setting (3, 4), the development of the dual kinase inhibitor, lapatinib (5), and the evolution of the antiangiogenesis antibody, bevacizumab (6), have maintained this enthusiasm for targeted therapies. Despite the rapid expansion of novel diagnostics designed to personalize breast cancer care, however, there remain several significant unmet needs for improving the accuracy and reliability of tests that are already in common daily clinical practice. In addition, the preanalytic conditions known to influence the final interpretation of these tests are in serious need of standardization (7–9). Thus, as we look forward to the introduction of new individualized breast cancer therapies, we must also look back and repair the defects in the personalized medicine tests that are already heavily relied on for the selection of therapy for the disease.

Hormone Receptor Assays

Measurement of estrogen and progesterone receptor (ER/PR) status (Table 1) is a standard of practice for newly diagnosed breast cancer patients and used to guide the selection of patients to receive hormonal therapies (10, 11). Although immunohistochemistry has been the predominant method for measuring ER/PR for over 15 years, this assay remains unstandardized and there is widespread concern that inaccuracy in immunohistochemistry technique and interpretation is leading to an unacceptably high error rate in determining the true hormone receptor status (10–13). Moreover, the emergence of additional hormonal-based therapeutics beyond tamoxifen, including aromatase inhibitors, selective estrogen response modulators, and selective ER down-regulators, and the possibility that these drugs may be more efficacious for patients with lower ER/PR expression levels have further intensified interest in improving the accuracy of immunohistochemistry and/or the consideration of alternative diagnostic methods (14). About 60% of ER+ tumors are also PR+ and ~75% of these ER+/PR+ tumors respond positively to endocrine therapy (15). It has been suggested that PR may be necessary for responsiveness to hormonal therapy in that ER+/PR– tumors are generally less responsive than ER+/PR+ tumors, possibly indicating that the estrogen response pathway may be deactivated in the ER+/PR– cases. ER+/PR+ tumors are rare and seem to show an intermediate response to endocrine therapy (15). Although practicing oncologists rely on the negative predictive value of the ER/PR assays, the relatively low positive predictive value of the tests (range, 30-50%) continues to contribute to the failure to achieve ideal personalized management of patients with breast cancer. Although the Seventh International Conference on Adjuvant Therapy of Primary Breast Cancer (St. Gallen, Switzerland) concluded that a precise threshold for ER/PR immunohistochemistry scoring results was difficult to achieve, a 10% positive staining...
of cells for either receptor might be considered as a reasonable threshold that could be accepted by most (16). Thus, either significant improvements and further standardization must be made in immunohistochemistry testing or novel, more reliable approaches to ER/PR testing must be introduced.

**Preanalytic issues.** Some of the most significant issues that affect the performance of tissue-based biomarker assays in breast cancer are the events that take place before the performance of the actual test (8, 9, 15–19). These preanalytic issues include the duration and conditions under which the tissue is maintained before fixation; the nature (chemical composition) and duration of fixation; the tissue processing conditions (standard versus the new non-formalin–based rapid processing techniques); the temperature of the embedding material; the length of storage of unstained slides before assessment by immunohistochemistry, fluorescent, nonfluorescent chromogenic, and silver *in situ* hybridization procedures (respectively; FISH, CISH, and SISH); and the use and nature of antigen-retrieval procedures, including both heat- and enzyme-based methods. These preanalytic variables significantly affect ER/PR results (15–19), which is highlighted by the increase in hormone receptor immunostaining as a function of length of formalin fixation (20). In one study, 200 clinical laboratories received sections from the same three tumors that showed low, moderate, or high ER expression, respectively, in a reference laboratory. Each laboratory did its own immunohistochemistry assessment and the results were scored separately, in a reference laboratory. Each laboratory did its own immunohistochemistry assessment and the results indicated considerable interlaboratory variation (21). The false-negative rates were as high as 30% to 60% (depending on the cutoff) in the low ER+ case.

**Can ER/PR immunohistochemistry assays be improved?** When immunohistochemistry was first introduced into clinical practice in the early 1990s, it was hoped that it would do in a manner equal or superior to the biochemical competitive binding assay that it replaced. After more than 15 years of experience, however, it is clear that response rates to targeted hormonal therapies have not dramatically increased in the immunohistochemistry testing era, and mRNA transcriptional profiling studies suggest that in ~10% of immunohistochemistry–determined ER+ tumors, active downstream signaling of the ER gene pathway is not observed. These facts have led to several recommendations as to how immunohistochemistry-based ER/PR testing performance can be improved: (a) test kits should be standardized to include cell line or tissue controls of known ER and PR protein expression levels determined by Western blotting; (b) the antibodies used to detect ER protein should be improved to avoid false-positive staining for tumors with ER gene mutations; (c) the automated or manual staining procedures, the preanalytic conditions (fixative, length of fixation, block or unstained slide storage time, and conditions), and the staining protocols (type and intensity of antigen retrieval, antibody concentration, and type of signal detection chemistry) should be standardized; (d) a universal immunohistochemistry slide scoring method should be used (22); (e) if a manual scoring method is used, an agreed upon cutoff for ER positivity should be established, whether this is a single variable or a more elaborate system, such as the method developed by Rhodes et al. (23); (f) if an image analysis system is used to score slides, the microscope, imaging instrument, and the software must be standardized and validated against another method of ER/PR measurement (23); and (g) widespread use of peer-reviewed quality assurance programs should be adopted by all laboratories providing ER/PR test results (24, 25).

**Will immunohistochemistry get the message?** When current immunohistochemistry procedures are used for ER/PR on well-maintained tissues, a linear relationship between the intensity and distribution of the immunostaining and the amount of nuclear ER protein is not observed (26). There is a concern, therefore, as to whether current immunohistochemistry techniques are capable of developing a sufficient dynamic range to allow for further personalization of hormone receptor–targeted therapies by guiding the actual drug selection in known ER+ patients. Thus, the 2001 U.S. NIH Consensus Conference conclusion that any level of immunohistochemistry-based expression of ER protein was sufficient to classify a tumor as ER+ may now need to be reconsidered (27). Breast cancer hormone receptor mRNA levels feature a wider dynamic range than current immunohistochemistry protocols, can be measured by both the reverse transcription-PCR and genomic microarray methods (26), and have been used both to classify and predict prognosis for the disease (28, 39). Genomic microarrays and multiplex reverse transcription-PCR assays also have the ability to assess ER pathway activation by the simultaneous determination of expression of multiple relevant genes and downstream pathways (30). Of recent interest is the question as to whether low and intermediate ER mRNA levels in ER immunohistochemistry–positive tumors predict responsiveness to aromatase inhibitors, but not to tamoxifen. Only prospective studies using large cohorts of patients in various clinical settings can satisfactorily answer this question. In addition, reverse

### Table 1. Assays for measuring ER/PR status

<table>
<thead>
<tr>
<th>Assay</th>
<th>DCC</th>
<th>IHC</th>
<th>RT-PCR</th>
</tr>
</thead>
<tbody>
<tr>
<td>Type</td>
<td>Competitive binding assay</td>
<td>Static <em>in situ</em> protein detection</td>
<td>Relative mRNA expression</td>
</tr>
<tr>
<td>Starting material</td>
<td>Fresh or frozen tissue</td>
<td>FFPE</td>
<td>Either fresh/frozen or FFPE</td>
</tr>
<tr>
<td>Preanalytic effect</td>
<td>Insufficient tumor tissue diluted by benign tissue</td>
<td>Overfixation causes nonspecific binding of antibody</td>
<td>Tissue processing prevents successful extraction of mRNA</td>
</tr>
<tr>
<td></td>
<td>Receptor denaturation by delayed preparation of cytosol</td>
<td>Excess antigen retrieval causes overstaining</td>
<td>Normalization procedure fails as housekeeping genes are not properly calibrated</td>
</tr>
</tbody>
</table>

Abbreviations: DCC, dextran-coated charcoal; IHC, immunohistochemistry; RT-PCR, reverse transcription-PCR; FFPE, formalin-fixed, paraffin-embedded tissue.
transcription-PCR and microarray-based tests for ER must also be standardized with reproducible controls and established relative mRNA expression cutoff points before they can be considered for widespread clinical use. Finally, additional potential biomarkers of hormone receptor function, including PR status and others, such as BCL-2 expression, must also be evaluated for their ability to customize hormone receptor–targeted therapy for both newly diagnosed and recurrent tumors (31). Therefore, whether by improvement and standardization of immunohistochemistry, transition to mRNA-based testing, or the introduction of additional validated biomarkers, it is now widely anticipated that hormone receptor testing in breast cancer must and will be enhanced to further personalize breast cancer care in an emerging era of increasing hormonal targeted therapy choices.

**HER2 testing**

*Will CISH or SISH replace FISH and immunohistochemistry?* Today, as anti-HER2–targeted therapy expands to the adjuvant setting, the importance of HER2 testing accuracy has never been more important. Worldwide, immunohistochemistry is the most often used primary test, with FISH testing continuing to grow incrementally (Fig. 1; Table 2; ref. 2). The transition to primary FISH testing has been triggered, in part, by concerns over the accuracy of immunohistochemistry results, especially when originating from small volume laboratories (32, 33), and a retrospective study that showed superiority of FISH-based testing over immunohistochemistry for the prediction of trastuzumab response (34). Recently, a task force report from the National Comprehensive Cancer Network concluded that properly done immunohistochemistry testing was as accurate as FISH in predicting trastuzumab response (35). The relative strengths and weaknesses of immunohistochemistry and FISH have been widely discussed in the literature (2). The advantages of immunohistochemistry, such as advantages of low cost, ease of performance, and widespread familiarity, must be weighed against a potential 10% false-negative rate that might prevent thousands of newly diagnosed women from being considered candidates for treatment with a drug that has shown substantial efficacy in the treatment of their disease, both in early and late stages (1, 3, 4). One strategy that can conceivably reduce the false-negative rate for immunohistochemistry is to make certain that a known 3+ positive cell line or tissue section is placed on the same slide as the patient’s tumor to prevent a technique error (e.g., failure to apply reagents) from causing a false-negative result. Similarly, stringent adherence to published slide scoring criteria and avoiding overscoring of cases with thermal artifacts, edge staining only, and especially, improper scoring of cases with staining of benign breast epithelium can similarly reduce the false (3+)-positive scoring rate. For FISH-based testing, despite the advantages of a more stable DNA target and the built-in internal control HER2 (ERBB2) gene signals in adjacent benign epithelium (or unamplified tumors), the use of the technique has been limited by its significantly higher cost in equipment, reagents, and personnel and the general lack of familiarity with the technique in smaller laboratories and community hospitals. Although discordance among laboratories in FISH testing has

---

**Fig. 1.** HER2 testing. *A.* 3+ membrane staining by immunohistochemistry using the DAKO HercepTest. *B.* Gene amplification detected by FISH with ratio of HER2 signals to chromosome 17 centromere signals of 5.7 using the Abbott-Vysis Pathview System. *C.* HER2 gene amplification detected by CISH using the Invitrogen-Zymed system.
been lower than immunohistochemistry in some studies (32), this has not been the case for others (33). In addition, some investigators have recommended simultaneous measurement of HER2 gene amplification and protein overexpression (36). Thus, as described previously for ER/PR testing, the most important approach to achieving improved accuracy for both immunohistochemistry and FISH-based assays is for all laboratories that provide this type of testing to participate in peer reviewed quality assurance programs that include proficiency testing (33).

Additional approaches to HER2 gene amplification testing that potentially feature the low cost and convenience of immunohistochemistry and the higher accuracy and built-in internal control advantages of FISH are the CISH (Invitrogen, Inc.) and the SISH (Ventana Medical Systems, Inc.) techniques. These methods use a complementary probe to the HER2 gene sequence similar to the FISH assay, but use immunohistochemistry-type signal detection in place of fluorescence and can be scored with a conventional light microscope while preserving the background tumor histology. Despite these potential advantages, the CISH (37) and SISH (38) techniques have several potential drawbacks, including hybridization failures, difficulty in counting overlapping gene signals, and classifying cases whose average gene copy number is close to the threshold of a positive result. A commercial version of the CISH assay is currently under review by the U.S. Food and Drug Administration seeking a similar indication for selection of patients for eligibility to receive trastuzumab as exists for previous Food and Drug Administration–approved immunohistochemistry and FISH assays. CISH has shown high concordance with FISH (37) and has the potential to replace both immunohistochemistry and FISH techniques for primary HER2 testing in that it would have the total cost, convenience, and familiarity advantages associated with immunohistochemistry and eliminate the need for a triage system that requires 2+ immunohistochemistry results to be retested by FISH. Undoubtedly, there will be great interest among oncologists and pathologists to see if the HER2 testing landscape actually does change in this way over the next several years.

**Sentinel Lymph Nodes.** Over the last 10 years, the sentinel lymph node (SLN) procedure has become a routine approach to the management of newly diagnosed breast cancers (39). There is significant lack of standardization, however, as to how the removed lymph nodes will be assessed by pathology departments (40). Variations in procedures include the following: (a) whether intraoperative frozen sections are used; (b) whether touch preparations with cytologic interpretations are incorporated; (c) the extent of serial sectioning of subsequent permanent paraffin blocks; and (d) the use and conditions associated with ancillary immunohistochemistry procedures (Fig. 2). In addition, nonmorphologic molecular methods for detecting metastases and microdeposits in SLNs have been developed and are undergoing clinical trials (41). Thus, given the multiple methods of tissue assessment available, it is no surprise that the SLN evaluation is far from standardized. A major concern is the use of the 0.2- and 2.0-mm cutoffs for determining the final pathologic stage, which were originally determined by a retrospective analysis of lymph node metastasis size and long-term disease outcome (42). Currently, SLNs with isolated tumor cells, small aggregates, or tumor cell clusters measuring <0.2 mm are given the designation N0I+, to indicate that the microdeposits identified only by immunohistochemistry were <0.2 mm in greatest diameter. There is concern whether these patients can safely be considered SLN negative and whether a better assessment than on slide diameter measurements could be developed for these cases. In ~18% of SLN biopsies, immunohistochemistry will detect occult tumor cells and clusters not identified on routine microscopy (43). Although it is widely accepted that the N0I+ status, also known as nanometastasis, imparts a significantly poorer prognosis than for N0 patients (44), confirmatory data from large prospective studies are lacking. Consequently, both the International Society of Cell Therapy and the National Cancer Institute have recognized the need for standardization.

---

**Table 2. Assays for determining HER2 status**

<table>
<thead>
<tr>
<th>Assay</th>
<th>IHC</th>
<th>FISH</th>
<th>CISH/SISH</th>
</tr>
</thead>
<tbody>
<tr>
<td>Starting material</td>
<td>FFPE</td>
<td>FFPE</td>
<td>FFPE</td>
</tr>
<tr>
<td>Target of test</td>
<td>Protein</td>
<td>DNA</td>
<td>DNA</td>
</tr>
<tr>
<td>Routine microscopy</td>
<td>Yes</td>
<td>No</td>
<td>Yes</td>
</tr>
<tr>
<td>Cost of reagents</td>
<td>Low</td>
<td>High</td>
<td>Low*</td>
</tr>
<tr>
<td>Time to complete</td>
<td>Short (5 min)</td>
<td>Long (40 min)</td>
<td>Short (10 min)</td>
</tr>
<tr>
<td>Built-in internal control</td>
<td>No</td>
<td>Yes</td>
<td>Yes</td>
</tr>
<tr>
<td>Effect of fixation, storage, and preanalytic factors</td>
<td>Significant</td>
<td>Less significant</td>
<td>Less significant</td>
</tr>
</tbody>
</table>

*It is estimated that the cost of CISH is only ~20% to 30% higher than immunohistochemistry.
of the SLN immunohistochemistry assay. Thus, the clinical management of patients with NoSLN disease remains controversial (45, 46). One emerging approach to this dilemma is the breast cancer stem cell concept that holds that the spread of stem cells to the SLN and bone marrow sinusues indicates a high risk for subsequent relapse and metastatic disease development after the completion of primary treatment (47, 48). Thus, if the phenotypic markers of breast cancer stem cells could be determined and applied to the <0.2-mm microdeposits in SLNs, this might lead to a more rational and biologically sound approach to the pathologic staging procedure. As further information on breast cancer stem cells emerges, it will be of great interest how these findings can be applied to this continuing clinical problem.

References

Standardizing Slide-Based Assays in Breast Cancer: Hormone Receptors, HER2, and Sentinel Lymph Nodes

Jeffrey S. Ross, W. Fraser Symmans, Lajos Pusztai, et al.


Updated version: Access the most recent version of this article at:
http://clincancerres.aacrjournals.org/content/13/10/2831

Cited articles: This article cites 47 articles, 15 of which you can access for free at:
http://clincancerres.aacrjournals.org/content/13/10/2831.full.html#ref-list-1

Citing articles: This article has been cited by 6 HighWire-hosted articles. Access the articles at:
/content/13/10/2831.full.html#related-urls

E-mail alerts: Sign up to receive free email-alerts related to this article or journal.

Reprints and Subscriptions: To order reprints of this article or to subscribe to the journal, contact the AACR Publications Department at pubs@aacr.org.

Permissions: To request permission to re-use all or part of this article, contact the AACR Publications Department at permissions@aacr.org.