Problems and Solutions in the Evaluation of Hormone Receptors in Breast Cancer

D. Craig Allred,
Department of Pathology and Immunology, Washington University School of Medicine, St Louis, Missouri

An article by Badve et al1 in this issue of the Journal of Clinical Oncology describes an interesting study comparing two methods of assessing estrogen receptor (ER) and progesterone-receptor (PR) expression in invasive breast cancer (IBC). One method measured protein by immunohistochemistry (IHC) and the other measured RNA using quantitative reverse-transcription polymerase chain reaction (RT-PCR). All assays utilized formalin-fixed, paraffin-embedded tissue samples from a case-control study cohort derived from the Eastern Cooperative Oncology Group clinical trial E2197, which randomly assigned patients with high-risk IBC to adjuvant doxorubicin/cyclophosphamide versus doxorubicin/docetaxel, with the addition of tamoxifen in receptor-positive cases.5 The cases (n = 179) represented all patients who relapsed (5 years follow-up) matched to an excess of controls (n = 597) who did not relapse, with available paraffin blocks. The IHC assays were performed twice in different laboratories. One was in the laboratories of institutions using local (and almost certainly diverse) staining protocols and definitions of positivity, most likely evaluating single 3- to 5-μm histologic sections from intact paraffin blocks of tumor. The other was in an expert central laboratory using a standardized staining protocol on single 3- to 5-μm sections from tissue microarrays (TMAs) containing two 1.0-mm cores from each tumor, with positive defined as an Allred Score of more than 2 (corresponding to 1% to 10% weakly positive cells).3,4 The RT-PCR assays were conducted with RNA extracted from multiple (three to six) 10-μm sections of intact paraffin blocks at Genomic Health Inc (GHI; Redwood City, CA) using the Oncotype DX assay,3 which evaluates 21 genes to determine a Recurrence Score (RS). The genes include ER and PR, with definitions of positive (6.5 and 5.5 units, respectively) determined in previous studies.5 Overall, there was a fairly high concordance of positive and negative results across all assays, ranging from 90% to 93% for ER, and 84% to 88% for PR. Considering a positive result by any assay to be truly positive, the discordant cases favored the central IHC results (7.2% of all cases; 91% of RT-PCR positive), central IHC over RT-PCR for PR (10% of all cases; 81% of IHC positive), and discordances were evenly balanced comparing both receptors combined (ie, ER and/or PR positive = receptor positive; 6.6% of all cases). All ER and PR assays were strongly associated (P < .0001) with relapse in all patients combined, but only marginally in the subset of patients with ER-positive tumors (P = .0144 by RT-PCR and P = .0913 by IHC). However, the RS was a highly significant (P < .0001) predictor of relapse in tamoxifen-treated ER-positive patients, consistent with previous studies.3

In current medical practice, ER and PR status are critical in determining the use and estimating the benefit of adjuvant hormonal therapy, and there is a widely (and justifiably) held perception that IHC, the present standard methodology for measuring ER and PR status, is unreliable and inaccurate in a substantial proportion of cases.6,7 The importance of the Badve et al1 study stems from its investigation of alternative methods for evaluating receptors that may be more reliable and accurate. In the limited scope of this study, it was successful by demonstrating that RT-PCR is at least equivalent to IHC in its ability to identify receptor-positive cases (considering ER and PR combined), marginally superior in predicting outcome in ER-positive patients, and superior in technical precision, which are all encouraging results.

There are many causes for the problems associated with measuring ER and PR by IHC.7 Some involve preanalytic issues unrelated to IHC itself, such as delayed fixation of tissue, which allows proteins to degrade. Others are analytic in nature, such as the use of diverse reagents with unequal sensitivities,4,8-10 or procedures which inadequately expose proteins masked during tissue fixation.7 Even postanalytic events may contribute, in the sense that tumors with low levels of receptors (eg, 1% to 10% positive cells) may respond to hormonal therapy,4,9,11 and some laboratories use arbitrary definitions of positive that are too high (eg, > 10% positive cells). With newer reagents and protocols, IHC assays for hormone receptors can even be too sensitive, resulting in bimodal results (negative or very high) that are inaccurate because breast cancers show a broad range of receptor expression.12-15 RT-PCR does not suffer this shortcoming and is able to more accurately convey this continuum. The central ER IHC results in the Badve et al1 study were essentially bimodal, suggesting that the assay was saturated, but also that insufficient sensitivity of the assay per se was not the reason that most of the discordant results were positive by RT-PCR but negative by IHC, although some of preanalytic issues alluded to could have resulted in false-negative IHC results. For example, perhaps some of the cases were poorly fixed and RT-PCR is better able to accommodate degraded RNA than IHC to accommodate degraded protein, although this has not been proven. Some of the discordances may have resulted from sampling error caused by the much smaller amount of tissue analyzed by IHC compared with RT-PCR in this study, which may have been inadequate to fully account for the heterogeneity of receptor expression in the tumors. Although some studies suggest that IHC performed on TMAs can closely approximate the results obtained by evaluating much larger
tissue sections, they are still unlikely to account for all the heterogeneity, and TMAs are not used in routine clinical practice. Many of the problems with IHC mentioned above have solutions, but some do not. For example, most IHC assays rely on enzymatic detection systems with rapid kinetics that are very difficult to control, which is one of the main reasons they are so easy to saturate. In general, it is almost impossible for enzymatic IHC to generate highly reproducible quantitative results, although even this may be surmountable with newer immunofluorescence techniques. Some apparent problems, such as discordant results between IHC and RT-PCR, may not be problems at all in the sense that protein expression and RNA expression are not the same thing, and there is no a priori reason to expect similar results or performance in every setting.

Nonetheless, assessing ER by RT-PCR in the Badve et al study was statistically superior to IHC in predicting relapse in tamoxifen-treated, ER-positive patients, which is a potentially significant achievement. The superiority of RT-PCR over IHC for measuring ER may indeed be generally true, but additional validation studies should be required before changing practice standards because the results were so close, the study design was imperfect to address this issue, and some IHC assays may perform better. Furthermore, receptor status in clinical practice today is not determined by ER alone, and a detailed comparison of the predictive abilities of PR alone, or ER/PR combined, was not provided in this study. On the basis of the concordance results, IHC performed better in assessing PR than ER, and may have fared better as a predictor of clinical outcome. Similar to IHC, RT-PCR is also difficult and fraught with potential problems, as Badve et al acknowledged. Although the assay is undoubtedly reliable and accurate in their hands, this may not be true in the hands of all of the thousands of laboratories around the world testing for ER and PR. Given the current need for accurate ER and PR results in all patients with breast cancer, an obvious short-term priority should be to improve the quality of widespread IHC testing, in addition to encouraging new studies intended to develop better tests, such as the laudable effort by Badve. Although enzymatic IHC is not well suited for analyses requiring a high degree of analytic precision or quantification, it is still very useful in many settings. In the setting of hormone receptors in breast cancer, a number of studies have demonstrated that the IHC procedures used in some laboratories are capable of providing results that are as reliable and useful as the previous gold standard ligand binding assays (LBAs), which it replaced, and LBAs are precise and highly quantitative (RT-PCR may be better than either, and only time will tell). Improving the quality of IHC assays on a global scale will require that all laboratories implement such rigorously standardized and validated assays, which will be difficult. The College of American Pathologists and the American Society of Clinical Oncology are aware of this need, and are initiating an effort to develop and publish enforceable guidelines for ER and PR testing, similar to recent College of American Pathologists/American Society of Clinical Oncology guidelines for human epidermal growth factor receptor 2 testing. Ultimately, however, it is unrealistic to expect that even perfect tests for ER and PR alone, regardless of the technology, will be sufficiently powerful to accurately predict the response of all patients because the biology is so complex. New, more powerful predictors are needed, most likely based on multiple biomarkers. In this regard, the Badve et al study provided further validation that the Oncotype DX RS is powerful in further stratifying response to hormonal therapy in receptor-positive patients, although including individual ER and PR results would be a welcome addition to the report. Other strategies under investigation also look promising, including the HOXB13/L17BR gene ratio determined by RT-PCR, and estrogen-regulated gene signatures determined by microarrays. Hopefully, these and other approaches will lead to significant improvements in predicting responses to hormonal therapies in the near future.

AUTHOR’S DISCLOSURES OF POTENTIAL CONFLICTS OF INTEREST
Although all authors completed the disclosure declaration, the following author(s) indicated a financial or other interest that is relevant to the subject matter under consideration in this article. Certain relationships marked with a “U” are those for which no compensation was received; those relationships marked with a “C” were compensated. For a detailed description of the disclosure categories, or for more information about ASCO’s conflict of interest policy, please refer to the Author Disclosure Declaration and the Disclosures of Potential Conflicts of Interest section in Information for Contributors.

Employment or Leadership Position: None Consultant or Advisory Role: D. Craig Allred, Genomic Health Inc (C) Stock Ownership: None Honoraria: None Research Funding: None Expert Testimony: None Other Remuneration: None

REFERENCES