Published in final edited form as:

Appl Immunohistochem Mol Morphol. 2013 March ; 21(2): 139–147. doi:10.1097/PAI.0b013e31825d73b2.

Quantitative Analysis of Estrogen Receptor Expression Shows SP1 antibody is more sensitive than 1D5

Allison W. Welsh, PhD, Malini Harigopal, MD, Hallie Wimberly, Manju Prasad, and David L. Rimm, MD, PhD

Department of Pathology, Yale University School of Medicine

Abstract

Studies comparing rabbit monoclonal SP1 antibody to 1D5 for ER immunohistochemical (IHC) testing show conflicting results. Here we use a standardized quantitative immunofluorescent (QIF) ER assay to determine the level and significance of discordance between antibodies. Both antibodies are assessed by QIF on our Index TMA of cell lines and case controls, followed by QIF and IHC on two retrospective cohorts from Yale. On the Index TMA, SP1 displayed stronger signal-to-noise than 1D5. On the patient cohorts, the range of discrepancy between the two antibodies is 8% to 16.9%, with the majority of discrepant cases being SP1-positive/1D5-negative. Kaplan Meier analysis of the discrepant cases shows outcome comparable to double positive cases, suggesting that SP1 is more sensitive than 1D5. A series of cases with high levels of ER-beta shows that neither antibody cross-reacts, suggesting equivalent specificity. Future efforts are needed to determine if response to endocrine therapies show superiority of either antibody as a companion diagnostic test.

Keywords

Immunohistochemistry; Estrogen Receptor; quantitative immunofluorescence; biomarker; endocrine therapy; companion diagnostic

INTRODUCTION

The Estrogen Receptor (ER) is arguably the most successful biomarker that exists in breast cancer today, determining both patient prognosis as well as eligibility for endocrine therapies^(1, 2). Because endocrine therapy provides a significant survival benefit, but only in ER-positive patients^(3, 4, 5), accurate ER testing is critical. While it has been shown to be equivalent or superior to the previous ligand-binding assay (LBA)^(6, 7), it is widely acknowledged that measurement of ER by the current standard, immunohistochemistry (IHC), still has many flaws, including the subjectivity involved in measurement and interpretation^(8, 9, 10), as well as variability due to pre-analytic factors, some of which are still not understood^(11, 8). Much of this variability stems from how rapidly IHC methods replaced those of LBA and ELISA in the clinical setting. The advantages of IHC (low cost, ease of analysis, and applicability on routine samples and small tumors) were so obvious that as soon as correlative studies between IHC and LBA were performed, most labs

All communication to: David L Rimm, Dept of Pathology, BML Rm 165, Yale University School of Medicine, 310 Cedar St, New Haven, CT 06520-8023, Phone: 203-737-4204, Fax: 203-737-5089, david.rimm@yale.edu.

Part of this work has previously been presented at San Antonio Breast Cancer Symposium 2010, Poster Discussion 10-09, Causes for False-Negative Estrogen Receptor Classification in Breast Cancer.

DISCLOSURE:

D.L. Rimm is a stockholder in and consultant to HistoRx Inc., the exclusive licensee to the Yale owned AQUA technology.

converted to IHC before standardized protocols, reagents, and thresholds were determined⁽¹²⁾.

A single incident in Canada, which involved re-testing ER status in 1,000 cases from 1997–2005, finding 40% misclassification between local and central laboratories⁽¹³⁾, sparked widespread awareness and concerns regarding false-negative ER classification. Our previous studies, as well as those done by other larger multi-national cooperative groups, have estimated a 10–20% false-negative rate in current U.S. clinical practice, suggesting significant potential under-treatment^(14, 15, 16, 17, 18, 19). New guidelines have recently been issued by the American Society of Clinical Oncology and the College of American Pathologists (ASCO/CAP), which aim to address the false-negative problem by lowering the threshold for ER positivity from 10% of nuclei “positive” to 1%⁽²⁰⁾. However, evidence from our recent work has suggested that false-negative ER classification is caused primarily by variability in the threshold of intensity (what constitutes a “positive” nucleus), regardless of what percentage of them are positive⁽¹⁷⁾.

We have recently described a quantitative immunofluorescence-based (QIF) assay for ER, in order to help standardize how the threshold for ER positivity is determined⁽¹⁷⁾. This assay uses a control tissue microarray (called the Index TMA) that contains a number of cell lines as well as a panel of 40 cases spanning the range of ER expression, which is stained alongside every experimental cohort in order to reproducibly determine the threshold for ER positivity. Here we examine how the use of different antibodies affects this threshold.

Three ER antibodies are currently clinically-validated and approved by ASCO/CAP for ER testing (1D5, SP1 and 6F11)^(20, 21, 6, 22, 14, 5). Studies have suggested that compared to 1D5, SP1 has an 8x greater affinity for ER by ELISA, and is more sensitive (but still as specific) on formalin-fixed, paraffin-embedded (FFPE) tissue^(23, 24). The existing data comparing the antibodies on large patient cohorts is minimal, but conflicting. One large study performed on over 4,000 cases of frozen tissue on TMAs suggested that SP1 is more sensitive than 1D5 (using biochemical assays as the gold standard), that 8% of cases were SP1-positive but 1D5-negative, and that these cases were associated with better outcomes⁽²²⁾. A later prospectively-designed study was performed on consecutive cases analyzed for routine clinical testing (fresh whole sections that were formalin-fixed and paraffin-embedded), and found only 2 of 508 carcinoma cases to be discrepant when comparing the two antibodies⁽²⁵⁾.

Given our ability to objectively assess the threshold for ER positivity using the Index TMA, we first sought to determine if SP1 and 1D5 showed similar sensitivity using QIF on this panel of cell line and patient controls. We then examined two retrospective cohorts from Yale (YTMA 49 and YMA 128) using both antibodies, and compared the level and significance of their discordance in determining ER status, both by traditional IHC and our novel QIF assay.

MATERIALS & METHODS

Cell Line Panel & Culture

A panel of ATCC breast cancer cell lines (see figure 1b) was chosen to span a range of ER expression. All cells were maintained at 37°C and 5% CO₂, and grown either in suggested media, or in RPMI 1640 culture medium (Gibco) supplemented with 10% fetal bovine serum (FBS, Gemini BioProducts), 100units/mL penicillin G and 100µg/mL streptomycin (Gibco), 1mM sodium pyruvate (Gibco), and 2mM L-glutamine (Gibco).

Western blotting

Whole-cell lysates were prepared in buffer containing 1% Nonidet P-40, 20mM TrisHCl pH8.0, 137mM NaCl, 10% glycerol, 2mM EDTA, 1mM DTT, 1mM NaVO₃, and complete mini EDTA-free protease inhibitor cocktail (Roche) in dH₂O. 15μg of each lysate was resolved by SDS-PAGE on a 4–12% Bis-Tris gel (NuPage), using NuPage MOPS SDS Running Buffer at 45mA. Resolved protein was transferred using NuPage Transfer Buffer at 50V for 2h. Western blotting was performed according to standard procedures, using ER using rabbit monoclonal SP1 antibody (Thermo), diluted 1:500. β-tubulin (Cell Signaling Technology, 2146), diluted 1:4000, was used as a loading control.

Construction of the Index TMA

Whole cell pellets (fixed in formalin and paraffin-embedded) were created from each of the cell lines in the panel (for a detailed protocol, see *Dolled-Filhart et al*²⁶, and *McCabe et al*²⁷). The Index TMA was created from cores of the 12 cell lines in the panel as well as an Index of 40 patient controls (random selection of patients spanning the range of ER expression). ER was previously quantified on this Index TMA (see Welsh et al 2011¹⁷), and it was ran alongside each cohort during IF staining, in order to reproducibly and objectively determine the threshold for ER positivity.

Patient Cohorts & construction of TMAs

Two retrospective cohorts of archival breast cancer cases from Yale were used: YTMA 49 (diagnosed 1962–1982, n = 619) and YTMA 128 (diagnosed 2003–2006, n = 257). Clinicopathologic characteristics of cases with valid assay results using both antibodies are found in Table 1. Both cohorts were constructed from single blocks of each patient, which had been formalin-fixed and paraffin-embedded from fresh tissue (never frozen), and constructed in a tissue microarray (TMA) as previously described (see *McCabe et al*²⁷), before cutting into 5μm sections. Because YTMA 128 is a newer cohort, we do not yet have sufficient followup information to perform survival analyses, and thus only used the cohort to compare ER status using the two different antibodies.

Immunofluorescent (IF) Staining & AQUA® Analysis

IF staining for ER was performed on the Index TMA as well as on sequential cuts of both YTMA 49 and YTMA 128. Slides were deparaffinized by melting at 60°C for 20min, followed by soaking twice for 20min in xylene (JT Baker). Rehydration was performed twice in 100% EtOH for 1min, followed by 70% EtOH for 1min, and tap water for 5min. Antigen retrieval was performed in citrate buffer (3.84g sodium citrate dihydrate in 2L ddH₂O, brought to pH 6.0 with 1M citric acid) using the PT module from LabVision. Endogenous peroxidases were blocked by 30 min incubation in 2.5% hydrogen peroxide in methanol at room temperature (RT). After washing, non-specific antigens were blocked by incubation in 0.3% bovine serum albumin in TBST for 30min at RT in humidity chamber. Rabbit Cytokeratin (Dako), diluted 1:100 in block (BSA in TBST above), and was incubated overnight at 4°C. ER was stained using 1D5 antibody (Dako, 1:50 in block, incubated 1h at RT), or SP1 antibody (Thermo, 1:1000 in block, incubated overnight at 4°C). ER-beta expression was assessed using PPG5/10 (Thermo-Scientific) for ERβ1 and Clone 5/25 (Serotec) for ERβ5. Primary antibodies were followed by Alexa 546-conjugated Goat anti-Rabbit or anti-Mouse secondary antibody (Molecular Probes) diluted 1:100 in mouse or rabbit EnVision reagent (Dako) for 1h at RT. Signal was amplified using Cyanine 5 (Cy5)-tyramide (Perkin-Elmer) at a dilution of 1:50 for 10min at RT. Nuclei were stained using 10μg/mL DAPI (Molecular Probes) in block for 20min at RT, and coverslips mounted with Prolong mounting medium (ProLong Gold, Molecular Probes).

ER immunofluorescence (IF) was quantified using AQUA. Briefly, a series of high-resolution monochromatic images were captured by the PM-2000 microscope (HistoRx) using AQUAsition 2.2 software (HistoRx). Images were collected for each histospot after auto-focus and auto-exposure. Fluorophores included DAPI (to create nuclear compartment), Cy3 (Alexa 546-cytokeratin to distinguish tumor from stroma and create cytoplasmic compartment), and Cy5 for the target (ER). Image analysis was performed using AQUAnalysis 2.2 software (HistoRx), which binarizes the cytokeratin stain (each pixel being “on” or “off”) to create an epithelial tumor “mask”. It uses a clustering algorithm to assign each pixel, with 95% confidence, to either a nuclear or cytoplasmic compartment. The AQUA score of ER in each subcellular compartment (nuclear, cytoplasmic, and whole tumor mask) is calculated by dividing the ER pixel intensities by the area of the compartment within which they were measured. AQUA scores are normalized to the exposure time, bit depth, and lamp hours at which the images were captured, allowing scores collected at different exposure times to be directly comparable. For the purpose of this study, only nuclear AQUA scores were used in the analyses. The Index TMA contains both positive and negative controls, and is used to determine the threshold AQUA score for ER positivity with each antibody (see *Welsh et al*¹⁷ for more information on this standardization method).

Immunohistochemical (IHC) Staining

IHC staining was performed on sequential cuts of YTMA 49 in a CLIA-certified laboratory, using either the 1D5 (Dako) or SP1 (Ventana) staining system for ER. Stained slides were digitally scanned using BioImage and visually assessed using ImageViewer software (BioImage). Each slide was scored by three blinded individuals (two board-certified pathologists, DLR and MH), for both intensity (0–3) and %-positivity (0–100). Scores for each case were then binarized, using the new 1% threshold (*Hammond et al*²⁰), to be classified as either ER positive or negative.

Statistical Analysis

All statistical analyses (bivariate regressions, spearman-rho correlations, univariate and multivariate analyses, Kaplan-Meier survival analyses) were performed using StatView analysis software. Disease-specific survival was used as an endpoint.

RESULTS

SP1 antibody shows higher signal-to-noise ratio than 1D5 on Index TMA

In order to determine if rabbit monoclonal antibody SP1 and mouse monoclonal 1D5 showed the same sensitivity by immunofluorescent (IF) staining, we compared both antibodies on a panel of cell line and patient controls specifically designed for measurement of ER. We call this array the Index TMA, and it contains a panel of cell lines with known ER status that were cultured under normal conditions, then pelleted, formalin-fixed, paraffin-embedded (FFPE), and cored as if a tissue block, for placement in duplicate on the TMA (see Methods section). The TMA was stained using IF with both antibodies, and nuclear immunoreactivity was quantified using AQUA software, which determines ER expression on a continuous scale represented as an AQUA score. IF measurement of ER in the cell line panel using SP1 produced AQUA scores with a range of 63–1432, with a 142-unit jump in score between the last negative cell line (MB 468, AQUA score 89) and the first positive cell line (ZR 751, AQUA score 231) (Figure 1A). Western blot analysis of ER using SP1 on lysates from the same cell lines (prepared in parallel with the FFPE cores) confirmed the expected ER-positivity of the four cell lines with highest AQUA scores by IF (Figure 1B). However, IF measurement of ER in the cell line panel using 1D5 revealed AQUA scores with a smaller range than SP1 (145 – 871), with roughly a 90-unit jump in

score between the last negative cell line (MDA-MB-231 in this case, AQUA score 284) and the first positive one (ZR 751, AQUA score 376) (Figure 1C). There was an overall higher level of background (nuclear immunoreactivity in ER-negative cell lines) with 1D5 (scores ranging from 145–284) than SP1 (scores ranging from 63–89).

Both antibodies were also used to measure ER by IF on the panel of 40 control cases, which were also part of the Index TMA, and were chosen to span the full range of ER expression seen in patients. In this panel, SP1 showed a range of AQUA scores from 45 – 12,417 (Figure 1D), which were well-correlated (pearson's $r^2 = 0.85$, spearman rho rank-correlation = 0.975, Figure 1E), with the scores using 1D5, ranging from 107 – 10,635 (Figure 1F). In these 40 patient controls, the same cases were considered ER positive or negative with SP1 and 1D5, however, the difference between the last negative case and the first positive case was more than three times as robust with SP1 (250-unit jump in score from 74 to 323, see red arrows and inset in Figure 1D) than with 1D5 (70-unit jump in score from 198 to 268, see red arrows and inset in Figure 1F). When visually validating the ER status of these cases right at the threshold for ER-positivity, this difference is clearly visible by eye with SP1 (Figure 1G, left panel), but much more difficult to distinguish specific nuclear immunoreactivity against non-specific background with 1D5 (Figure 1G, right panel).

Comparison of SP1 to 1D5 on YTMA 49 using traditional IHC

In order to assess the clinical relevance of this observed difference in sensitivity, we next compared both antibodies using traditional immunohistochemical (IHC) staining on a retrospective cohort of breast cancer cases from Yale (YTMA 49, clinicopathological characteristics shown in Table 1). Two sequential 5 μ m cuts of the TMA were stained in a CLIA-certified lab, using either the SP1 (Ventana) or 1D5 (Dako) system. The TMAs were then digitally scanned, and both independently scored by three blinded individuals (including two board-certified pathologists) for both intensity (0–3) and %-positive nuclei (0–100). Cases were then binarized as either ER positive or negative according to the current ASCO-CAP guidelines, which define a 1% threshold for positivity.

Comparison of ER status using both antibodies revealed 83.3% agreement (20.9% double-negative, 62.4% double-positive), but 16.7% disagreement, with the majority (65 of 73 total discrepant cases) ER positive by SP1, but negative by 1D5 (Figure 2A). To determine if the discrepancy was due to a difference in sensitivity of the antibodies, we examined the individual scores for intensity and %-positivity in these cases. The scores for %-positive nuclei showed even distribution across the spectrum from 0 to 100 (Figure 2B, top panels). In contrast, the distribution of intensity scores were significantly skewed to the low end of the spectrum (Figure 2B, bottom panels), showing that the majority of discrepant cases had low levels of ER (given an intensity score of 1) that were detected with SP1, but missed with 1D5 (instead given an intensity score of 0). Kaplan Meier analysis revealed that the discrepant cases (SP1-positive/1D5-negative) showed disease-specific survival similar to “true” ER-positives (positive with both antibodies), and trending towards significantly different than “true” ER-negatives (log-rank $p = 0.101$, Figure 2C). Representative examples of discrepant cases are shown in Figure 2D (1D5-positive/SP1-negative in left panels, SP1-positive/1D5-negative in right panels).

Comparison of SP1 to 1D5 on YTMA 49 and YTMA 128 using immunofluorescence (IF)

In order to confirm these findings using a second method, we also compared both antibodies in assessment of ER status using quantitative IF. Both antibodies were stained on two retrospective cohorts from Yale (YTMA 49 and YTMA 128) alongside the Index TMA, in order to standardize the AQUA score threshold for positivity using each antibody (see *Welsh et al*¹⁷ for more information on standardization). Analysis of ER status for each case

on YTMA 49 using both antibodies revealed 91.2% agreement and 8.8% discrepant cases, the majority of which were SP1-positive/1D5-negative (Figure 3A). Kaplan Meier analysis revealed that these discrepant cases showed disease-specific survival similar to “true” ER positives (positive with both antibodies), and trending towards significantly different than “true” ER-negatives (log-rank $p = 0.159$, Figure 3B). On this cohort, both antibodies are significant prognostic factors in a Univariate Cox Proportional Hazards Model; however, SP1 is stronger ($p = 0.0027$, HR = 1.965; compared to $p = 0.01$, HR = 1.72 with 1D5) (Table 2). Both antibodies also hold up on a Multivariate Cox Model, adjusted for age, nodal status, tumor size, and nuclear grade (Table 3).

In order to confirm this level of discrepancy on a more recent cohort, we performed the same analysis on YTMA 128. Comparison of both antibodies on YTMA 128 revealed 11% discrepancy, this time with all discrepant cases positive by SP1, but negative by 1D5 (Figure 3C). This cohort is too recent to have sufficient follow-up, and thus Kaplan Meier and Cox Univariate analysis of disease-specific survival were not performed. Representative IF images for discrepant cases, as well as cases in agreement, are shown in figure 4 for each antibody, demonstrating that the difference in signal-to-noise ratio with both antibodies is visible by eye.

Finally, given the data above on sensitivity, we assessed antibody cross-reactivity to determine if there is a difference in antibody specificity. Since ER-beta is the most closely related isoform, we measured ER β 1 and ER β 5 using validated antibodies PPG5/10 and Clone 5/25 respectively on YTMA-49. If there was cross reactivity between SP1 or 1D5 and ER-beta, then all cases positive for either ER-beta antibody should also be positive SP1 or 1D5. For each ER-beta antibody, there were over 50 cases that were strongly positive for ER-beta but completely negative for both SP1 and 1D5. Thus there appears to be no difference in specificity, at least with respect to ER-beta.

DISCUSSION

In a panel of cell line and patient controls, we show that SP1 displays a stronger signal-to-background ratio than 1D5. On two retrospective cohorts, by IHC and IF, we see discrepancy between the two antibodies ranging from 8% to 16.9%, with the majority of (or in one cohort, all) discrepant cases positive by SP1, but negative by 1D5. Survival analysis of these cases shows they are correlated with better outcomes than “true” ER negative patients, suggesting that SP1 may be at least 8% more sensitive than 1D5. Furthermore, Univariate Cox Models suggest that while both antibodies are significant predictors of disease-specific survival, SP1 is slightly stronger.

These results agree with *Cheang et al* (22), who found that there is an 8% discrepancy between the two antibodies, and these discrepant cases (again SP1-positive/1D5-negative) correlate with better outcome and better response to Tamoxifen. Unfortunately, on our cohorts we could only examine disease-specific survival as a surrogate for ER “positivity” (as opposed to response to Tamoxifen), since one cohort pre-dated the routine use of Tamoxifen (YTMA 49), and the other was too recent to have follow-up information. We have also observed a higher level of discrepancy with IHC than QIF, which could be due to the added variability associated with the subjectivity of scoring IHC and variability of the chromogen. Regardless, the 8% (at least) level of discrepancy we have observed is contrary to the study by *Brock et al* (25), who found that the two antibodies are equivalent on FFPE tissue.

There are a number of possible reasons for the level of discrepancy we have observed. Most obviously, 1D5 is a mouse, and SP1 a rabbit, monoclonal. A number of published studies

suggest that rabbit monoclonals may display greater affinity for their epitope^(23, 24), and that SP1 itself is more sensitive than 1D5 on patient tissue^(28, 22). While the epitopes themselves are different (SP1 is C-terminal, 1D5 is N-terminal), there is no known evidence to date of a prevalence of C-terminal ER isoforms in breast carcinoma cases. It has also been suggested that the two antibodies have different sensitivities to pre-analytic variables, specifically that delays in fixation can affect loss of antigenicity for 1D5 more than for SP1⁽²⁹⁾. Finally it is possible that there may be a difference due to structurally defective ER⁽³²⁾.

Another key issue, which was raised in the study by *Brock et al*⁽²⁵⁾, is the use of TMAs, given their limitations with regards to heterogeneity. The use of TMAs was deliberate in this study since we were interested in comparing sensitivity of both antibodies at the threshold for positivity, and in order to minimize variability in threshold, wanted to compare all cases on a single slide. However, we recognize that this is a limitation of this work. Previous data has suggested that two tissue cores on a TMA is enough to achieve >95% reproducibility⁽³⁰⁾, while more recent studies suggest a large number of “fields of view” are required to address issues of heterogeneity. Here, we only tested a single core from each patient. If we assume SP1 to be more sensitive than 1D5, as our data suggests, this problem of heterogeneity and representation would explain the small proportion of 1D5-positive/SP1-negative cases we have observed (perhaps these are true ER positive cases, and on a whole tissue section, would be positive with SP1). However, since this was a comparison study, both 1D5 and SP1 were subject to the same limitations imposed by the use of a single TMA core and thus we believe this limitation does not dramatically affect our conclusion.

Overall, our data, using both IHC and standardized QIF on fresh FFPE tissue, supports previous findings that SP1 is more sensitive than 1D5, and displays a stronger signal-to-background ratio. This would suggest that, in addition to its benefits for cost-efficiency⁽³¹⁾, the use of SP1 in a clinical setting may help reduce the false-negative rate. Further studies on response to endocrine therapies in patients with low levels of ER (cases just above the threshold that may be caught with SP1, but not with 1D5) are a critical next step, and these will provide the ultimate insight into whether one antibody is superior in the clinical setting.

Acknowledgments

Support for this research comes from NIH R33 CA 106709 (to DLR) and a US Army CDMRP pre-doctoral fellowship (AWW).

The authors thank Dr. Manju Prasad for allowing staining of the YTMA-49 TMAs in the Yale IHC CLIA lab, along with the clinical work. We also acknowledge the outstanding work of Lori Charette and her colleagues in the Yale Pathology Tissue Services TMA facility.

References

1. Hahnel R, Woodings T, Vivian AB. Prognostic value of estrogen receptors in primary breast cancer. *Cancer*. 1979; 44:671–675. [PubMed: 476576]
2. Allred DC, Harvey JM, Berardo M, et al. Prognostic and predictive factors in breast cancer by immunohistochemical analysis. *Mod Pathol*. 1998; 11:155–168. [PubMed: 9504686]
3. Chia SK, Wolff AC. With maturity comes confidence: EBCTCG tamoxifen update. *Lancet*. 2011; 378:747–749. [PubMed: 21802719]
4. Clark GM, McGuire WL, Hubay CA, et al. The importance of estrogen and progesterone receptor in primary breast cancer. *Prog Clin Biol Res*. 1983; 132E:183–190. [PubMed: 6647468]
5. Dowsett M, Allred C, Knox J, et al. Relationship between quantitative estrogen and progesterone receptor expression and human epidermal growth factor receptor 2 (HER-2) status with recurrence in the Arimidex, Tamoxifen, Alone or in Combination trial. *J Clin Oncol*. 2008; 26:1059–1065. [PubMed: 18227529]

6. Harvey JM, Clark GM, Osborne CK, et al. Estrogen receptor status by immunohistochemistry is superior to the ligand-binding assay for predicting response to adjuvant endocrine therapy in breast cancer. *J Clin Oncol.* 1999; 17:1474–1481. [PubMed: 10334533]
7. Elledge RM, Green S, Pugh R, et al. Estrogen receptor (ER) and progesterone receptor (PgR), by ligand-binding assay compared with ER, PgR and pS2, by immuno-histochemistry in predicting response to tamoxifen in metastatic breast cancer: a Southwest Oncology Group Study. *Int J Cancer.* 2000; 89:111–117. [PubMed: 10754487]
8. Gown AM. Current issues in ER and HER2 testing by IHC in breast cancer. *Mod Pathol.* 2008; 21 (Suppl 2):S8–S15. [PubMed: 18437174]
9. Allred DC. Commentary: hormone receptor testing in breast cancer: a distress signal from Canada. *Oncologist.* 2008; 13:1134–1136. [PubMed: 18987048]
10. Allison KH. Estrogen receptor expression in breast cancer: we cannot ignore the shades of gray. *Am J Clin Pathol.* 2008; 130:853–854. [PubMed: 19019759]
11. Goldstein NS, Ferkowicz M, Odish E, et al. Minimum formalin fixation time for consistent estrogen receptor immunohistochemical staining of invasive breast carcinoma. *Am J Clin Pathol.* 2003; 120:86–92. [PubMed: 12866377]
12. Hammond ME, Hayes DF, Dowsett M, et al. American Society of Clinical Oncology/College of American Pathologists guideline recommendations for immunohistochemical testing of estrogen and progesterone receptors in breast cancer (unabridged version). *Arch Pathol Lab Med.* 2010; 134:e48–72. [PubMed: 20586616]
13. Hede K. Breast cancer testing scandal shines spotlight on black box of clinical laboratory testing. *J Natl Cancer Inst.* 2008; 100:836–837. 844. [PubMed: 18544732]
14. Regan MM, Viale G, Mastropasqua MG, et al. Re-evaluating adjuvant breast cancer trials: assessing hormone receptor status by immunohistochemical versus extraction assays. *J Natl Cancer Inst.* 2006; 98:1571–1581. [PubMed: 17077359]
15. Badve SS, Baehner FL, Gray RP, et al. Estrogen- and progesterone-receptor status in ECOG 2197: comparison of immunohistochemistry by local and central laboratories and quantitative reverse transcription polymerase chain reaction by central laboratory. *J Clin Oncol.* 2008; 26:2473–2481. [PubMed: 18487567]
16. Viale G, Regan MM, Maiorano E, et al. Prognostic and predictive value of centrally reviewed expression of estrogen and progesterone receptors in a randomized trial comparing letrozole and tamoxifen adjuvant therapy for postmenopausal early breast cancer: BIG 1–98. *J Clin Oncol.* 2007; 25:3846–3852. [PubMed: 17679725]
17. Welsh AW, Moeder CB, Kumar S, et al. Standardization of estrogen receptor measurement in breast cancer suggests false-negative results are a function of threshold intensity rather than percentage of positive cells. *J Clin Oncol.* 2011; 29:2978–2984. [PubMed: 21709197]
18. Rhodes A, Jasani B, Balaton AJ, et al. Immunohistochemical demonstration of oestrogen and progesterone receptors: correlation of standards achieved on in house tumours with that achieved on external quality assessment material in over 150 laboratories from 26 countries. *J Clin Pathol.* 2000; 53:292–301. [PubMed: 10823126]
19. Francis GD, Dimech M, Giles L, et al. Frequency and reliability of oestrogen receptor, progesterone receptor and HER2 in breast carcinoma determined by immunohistochemistry in Australasia: results of the RCPA Quality Assurance Program. *J Clin Pathol.* 2007; 60:1277–1283. [PubMed: 17259294]
20. Hammond ME, Hayes DF, Dowsett M, et al. American Society of Clinical Oncology/College Of American Pathologists guideline recommendations for immunohistochemical testing of estrogen and progesterone receptors in breast cancer. *J Clin Oncol.* 2010; 28:2784–2795. [PubMed: 20404251]
21. Barnes DM, Harris WH, Smith P, et al. Immunohistochemical determination of oestrogen receptor: comparison of different methods of assessment of staining and correlation with clinical outcome of breast cancer patients. *Br J Cancer.* 1996; 74:1445–1451. [PubMed: 8912543]
22. Cheang MC, Treaba DO, Speers CH, et al. Immunohistochemical detection using the new rabbit monoclonal antibody SP1 of estrogen receptor in breast cancer is superior to mouse monoclonal antibody 1D5 in predicting survival. *J Clin Oncol.* 2006; 24:5637–5644. [PubMed: 17116944]

23. Huang Z, Zhu W, Szekeres G, et al. Development of new rabbit monoclonal antibody to estrogen receptor: immunohistochemical assessment on formalin-fixed, paraffin-embedded tissue sections. *Appl Immunohistochem Mol Morphol*. 2005; 13:91–95. [PubMed: 15722800]
24. Rossi S, Laurino L, Furlanetto A, et al. Rabbit monoclonal antibodies: a comparative study between a novel category of immunoreagents and the corresponding mouse monoclonal antibodies. *Am J Clin Pathol*. 2005; 124:295–302. [PubMed: 16040303]
25. Brock JE, Hornick JL, Richardson AL, et al. A comparison of estrogen receptor SP1 and 1D5 monoclonal antibodies in routine clinical use reveals similar staining results. *Am J Clin Pathol*. 2009; 132:396–401. [PubMed: 19687316]
26. Dolled-Filhart M, McCabe A, Giltane J, et al. Quantitative in situ analysis of beta-catenin expression in breast cancer shows decreased expression is associated with poor outcome. *Cancer Res*. 2006; 66:5487–5494. [PubMed: 16707478]
27. McCabe A, Dolled-Filhart M, Camp RL, et al. Automated quantitative analysis (AQUA) of in situ protein expression, antibody concentration, and prognosis. *J Natl Cancer Inst*. 2005; 97:1808–1815. [PubMed: 16368942]
28. Treaba, DO.; Hing, AW.; Tse, CC., et al. Significantly improved sensitivity for ER detection in breast cancer using a new rabbit monoclonal anti-ER antibody (SP1). 2005.
29. Qiu J, Kulkarni S, Chandrasekhar R, et al. Effect of delayed formalin fixation on estrogen and progesterone receptors in breast cancer: a study of three different clones. *Am J Clin Pathol*. 2010; 134:813–819. [PubMed: 20959666]
30. Camp RL, Charette LA, Rimm DL. Validation of tissue microarray technology in breast carcinoma. *Lab Invest*. 2000; 80:1943–1949. [PubMed: 11140706]
31. Gown AM, Barry TS, Kandalaft P, et al. A new anti-ER Rabbit Monoclonal Antibody improves efficiency of immunohistochemical evaluation of ER status in breast cancer. *Modern Pathology*. 2005; 18(suppl1):35A.
32. Traish AM, al-Fadhli S, Klinge, et al. Identification of structurally altered estrogen receptors in human breast cancer by site-directed monoclonal antibodies. *Steroids*. 1995; 60:467–474. [PubMed: 7482632]

Figure 1. Comparison of SP1 to 1D5 using QIF (AQUA) on a panel of cell line and patient controls

ER expression was measured by QIF (output is AQUA score) in the panel of cell lines on the Index TMA using both SP1 (AQUA score distribution shown in A) as well as 1D5 (AQUA scores in C). Red bars represent the jump in AQUA score between the last ER-negative cell line and the first ER-positive cell line (ZR751). ER status of cell lines was confirmed by western blot analysis of lysates prepared from the same cultures used to construct the Index TMA (B). ER was also quantified in the patients on the Index TMA using SP1 (AQUA score distribution in D) and 1D5 (AQUA score distribution in F), revealing strong correlation between both antibodies (E, pearson's $r^2 = 0.853$, spearman rho rank = 0.975). On these 39 cases there was no discordance in ER status between the two antibodies, however the difference between the highest negative case and lowest positive case is much greater with SP1 (red arrows and inset, D) than with 1D5 (red arrows and inset, F). QIF images of ER staining in these cases (designated with red arrows) are shown in G along with their AQUA scores, with SP1 in the left panels and 1D5 in the right panels. In these images, we expanded the dynamic range of the grayscale around the threshold (adjusted maximum RGB input level from 255 to 17 using Adobe Photoshop) in order to visualize very low levels of specific nuclear staining as well as non-specific background.

Figure 2. Discordance in ER status on YTMA 49 with SP1 versus 1D5 using traditional IHC ER was stained on sequential cuts of YTMA 49 in a CLIA-certified lab, using routine IHC protocol, with either SP1 (Ventana) or 1D5 (Dako) antibody systems. Stained slides were digitally scanned in a central laboratory and scored by three individual pathologists for Intensity (0–3) and %-positive (0–100), and then binarized into ER positive or negative according to the recent 1% threshold guidelines. A) The 375 cases with valid tissue for analysis by both antibodies are shown grouped according to ER status by both SP1 and 1D5. B) Frequency distributions of scores for %-positive (top panels) and Intensity (bottom panels) are shown for the cases with discordant ER status (SP1-negative/1D5-positive in turquoise in A and B, SP1-positive/1D5-negative in red in A and B). C) Kaplan-Meier analysis of 10-year DFS in cases grouped according to ER status by both antibodies (color-coded in A), with log-rank p-values (p_1 = SP1-positive/1D5-negative versus double-negative cases, p_2 = double-positive versus double-negative cases). D) Representative images of an SP1-negative/1D5-positive case (left panel) as well as an SP1-positive/1D5-negative case (right panel). (IHC = immunohistochemistry, DFS = disease-free survival).

Figure 3. Discordance in ER status on YTMA 49 and 128 with SP1 versus 1D5 using QIF
ER status was determined by AQUA analysis in two Yale retrospective breast cancer cohorts, YTMA 49 and YTMA 128 (clinicopathological characteristics in Table 1) using both SP1 and 1D5 antibodies. A distribution of ER by AQUA score (standardized threshold determined using the Index TMA) is shown where each case is color-coded in the bar below, according to its ER status by both antibodies (YTMA 49 in A, YTMA 128 in C). Kaplan-Meier curves (B) show 10-year DFS in YTMA 49, where patients are grouped according to the classifications shown in A. Log-rank p-values are shown inset (p_1 = double-positive versus double-negative cases, p_2 = 1D5-negative/SP1-positive versus double-negative cases). The 1D5-positive/SP1-negative group (n=5) was excluded from survival analysis on account of small size and insufficient power. Survival analysis could not be performed on YTMA 128 because it is too recent to have sufficient follow-up information. Representative images of QIF staining using 1D5 and SP1 in discordant and concordant cases can be seen in Supplementary Figure 1. (QIF = quantitative immunofluorescence, DFS = disease-free survival).

Figure 4. Representative QIF images with 1D5 and SP1

ER status was determined by AQUA analysis in two Yale retrospective breast cancer cohorts, YTMA 49 and YTMA 128 (clinicopathological characteristics in Table 1) using both SP1 and 1D5 antibodies. Representative images of IF staining in discordant and concordant cases are shown (1D5, top panel; SP1, bottom panel), with corresponding AQUA scores inset. For each of these images, we contracted the dynamic range of the grayscale (adjusted maximum RGB input level from 255 to 17 using Adobe Photoshop) in order to visualize very low levels of specific nuclear staining as well as non-specific background.

Table 1

Clinicopathological Characteristics of cases in YTMA 49 and YTMA 128 with valid assay results using both SP1 and 1D5

	YTMA 49 Cohort	YTMA 128 Cohort
Characteristic	N (%)	N (%)
All patients	388	163
Age (y)		
<50	106 (27.3)	54 (33.1)
50	282 (72.7)	94 (57.7)
unknown	0 (0)	15 (9.2)
Nodal Status		
positive	217 (55.9)	59 (36.2)
negative	170 (43.8)	93 (57.1)
unsampled/unknown	1 (0.3)	11 (6.7)
Tumor Size (mm)		
2	170 (43.8)	97 (59.5)
2–5	127 (32.7)	46 (28.2)
5	60 (15.5)	9 (5.6)
unknown	31 (8.0)	11 (6.7)
ER (IHC)		
positive (1–3)	200 (51.6)	114 (69.9)
negative (0)	179 (46.1)	21 (12.9)
unknown	9 (2.3)	28 (17.2)
PR (IHC)		
positive (1–3)	193 (49.8)	101 (62.0)
negative (0)	179 (46.1)	34 (20.9)
unknown	16 (4.1)	28 (17.1)
HER2 (IHC)		
positive (2–3)	64 (16.5)	34 (20.9)
negative (0–1)	310 (79.9)	97 (59.5)
unknown	14 (3.6)	32 (19.6)
Follow-up (m)		
median (range)	90.2 (2.4 – 498.0)	52.0 (0.5 – 340.0)

Abbreviations: ER = estrogen receptor, IHC = immunohistochemistry, PR = progesterone receptor, y = years, m = months.

Univariate Cox Proportional Hazard Model for 10-year disease-free survival (DFS) using 1D5 or SP1 to determine ER Status

Table 2

Characteristic	ER Status determined by 1D5 n = 252			ER Status determined by SP1 n = 252		
	HR	95% CI	P	HR	95% CI	P
ER status						
Positive vs. negative	1.718	1.133 – 2.604	0.0109	1.965	1.265 – 3.054	0.0027

Abbreviations: ER = estrogen receptor, HR = adjusted hazard ratio, DFS = disease-free survival

Table 3

Multivariate Cox Proportional Hazard Model for 10-Year Disease-Free Survival (DFS) using ID5 or SP1 to determine ER Status

Characteristic	ER Status determined by ID5 n = 225			ER Status determined by SP1 n = 225		
	HR	95% CI	P	HR	95% CI	P
Age at diagnosis, years						
<50 vs. 50	0.990	0.607–1.615	0.9676	0.976	0.597–1.595	0.9220
Nodal status						
Positive vs. negative	0.481	0.308–0.752	0.0013	0.475	0.303–0.744	0.0012
Tumor size, mm						
2 vs. 5	0.287	0.160–0.516	<0.0001	0.308	0.171–0.556	<0.0001
2–5 vs. 5	0.483	0.292–0.799	0.0046	0.488	0.294–0.810	0.0055
Nuclear Grade						
1 vs. 3	0.855	0.430–1.700	0.6547	0.887	0.441–1.783	0.7365
2 vs. 3	0.983	0.617–1.564	0.9413	0.963	0.603–1.538	0.8745
ER status						
Positive vs. negative	1.943	1.230–3.069	0.0044	1.994	1.220–3.259	0.0059

Overall model is significant with ID5 (Chi-Square = 44.169, Wald P-value = <0.0001) and with SP1 (Chi-Square = 44.319, Wald P-value = <0.0001). ER, estrogen receptor; HR, adjusted hazard ratio; DFS, disease-free survival