

Automated measurement of estrogen receptor in breast cancer: a comparison of fluorescent and chromogenic methods of measurement

Elizabeth R Zarrella, Madeline Coulter, Allison W Welsh, Daniel E Carvajal, Kurt A Schalper, Malini Harigopal, David L Rimm and Veronique M Neumeister

Whereas FDA-approved methods of assessment of estrogen receptor (ER) are ‘fit for purpose’, they represent a 30-year-old technology. New quantitative methods, both chromogenic and fluorescent, have been developed and studies have shown that these methods increase the accuracy of assessment of ER. Here, we compare three methods of ER detection and assessment on two retrospective tissue microarray (TMA) cohorts of breast cancer patients: estimates of percent nuclei positive by pathologists and by Aperio's nuclear algorithm (standard chromogenic immunostaining), and immunofluorescence as quantified with the automated quantitative analysis (AQUA) method of quantitative immunofluorescence (QIF). Reproducibility was excellent ($R^2 > 0.95$) between users for both automated analysis methods, and the Aperio and QIF scoring results were also highly correlated, despite the different detection systems. The subjective readings show lower levels of reproducibility and a discontinuous, bimodal distribution of scores not seen by either mechanized method. Kaplan–Meier analysis of 10-year disease-free survival was significant for each method (Pathologist, $P=0.0019$; Aperio, $P=0.0053$, AQUA, $P=0.0026$); however, there were discrepancies in patient classification in 19 out of 233 cases analyzed. Out of these, 11 were visually positive by both chromogenic and fluorescent detection. In 10 cases, the Aperio nuclear algorithm labeled the nuclei as negative; in 1 case, the AQUA score was just under the cutoff for positivity (determined by an Index TMA). In contrast, 8 out of 19 discrepant cases had clear nuclear positivity by fluorescence that was unable to be visualized by chromogenic detection, perhaps because of low positivity masked by the hematoxylin counterstain. These results demonstrate that automated systems enable objective, precise quantification of ER. Furthermore, immunofluorescence detection offers the additional advantage of a signal that cannot be masked by a counterstaining agent. These data support the usage of automated methods for measurement of this and other biomarkers that may be used in companion diagnostic tests.

Laboratory Investigation (2016) 96, 1016–1025; doi:10.1038/labinvest.2016.73; published online 27 June 2016

For decades, the value of estrogen receptor (ER) as a prognostic and predictive marker in breast cancer has been an unparalleled example of the impact of biomarker research on patient care.^{1–3} Its importance is such that recent discoveries of high error rates in clinical testing for ER, in both Canada and the United States, spurred an immediate reaction toward improved standardization in ER assessment^{4–7} resulting in publication of the guidelines for tissue processing and analysis to optimize companion diagnostic testing of ER in breast cancer specimens. As a result, research into pre-analytical variables that may influence biomarker test results has expanded dramatically,^{8,9} although somewhat less attention has been paid to analytical variables, specifically, those

concerned with methods of ER detection and quantification/measurement.

Before the current immunohistochemistry (IHC)-based standard, ER expression was widely evaluated using the ligand-binding assay (LBA). This test incubated breast tissue lysate with radiolabeled estradiol and resulted in an absolute quantification (fmol/mg) of the ER.³ However, LBAs are limited by the large tissue requirement and their inability to provide contextual information including the capability to distinguish ER expression in benign *vs* malignant cells.¹⁰ Upon development of specific monoclonal antibodies,^{11,12} the practical ease and cost effectiveness of IHC led to rapid implementation of a new clinical standard for *in situ*

Department of Pathology, Yale University School of Medicine, New Haven, CT, USA

Correspondence: Dr VM Neumeister, MD or Professor DL Rimm, MD, PhD, Department of Pathology, Yale University School of Medicine, 310 Cedar Street, New Haven, CT 06520, USA.

E-mail: veronique.neumeister@yale.edu or david.rimm@yale.edu

Received 21 January 2016; revised 23 May 2016; accepted 23 May 2016

assessment of protein expression after demonstration of their prognostic and predictive value.^{13,14} However, the advantages of this *in situ* detection method of ER were confounded by the introduction of the human eye as a measurement tool resulting in significant reader variability.^{15,16}

Over the past few decades, many platforms have endeavored to eliminate this intra- and inter-observer variability and achieve consistent evaluation of diagnostic specimens. Systems such as the CAS-200 (ref. 10) and ChromaVision's ACIS¹⁷ function on a principle of color deconvolution; for ER and other nuclear markers, this allows optical density measurements of positive target staining within a nuclear counterstain.^{18,19} Recently, technology has allowed the development of more rapid and sophisticated methods of digital image analysis. One such platform, the Aperio ScanScope and Digital Image Analysis Suite, combines both high-resolution image capture and quantitative assessment, and is FDA-approved to assist pathologists in ER, PR, Her2, and Ki67 measurement in breast cancer.^{20–22} In spite of the FDA approval, adoption is still limited. A recent CAP survey (2014) shows that less than 25% of over 1100 labs surveyed use automated assessment for ER.

Despite these advances, any system relying on chromogenic immunostaining is subject to the inherent limitations of absorbance measurement, such as a low dynamic range and saturation of the signal intensity based on enzymatic visualization of the antibody. Most widely used is 3,3'-diaminobenzidine (DAB), a highly thermochemically stable polybenzimidazole that provides brown-colored staining.²³ The chromogen deposition occurs through a redox reaction catalyzed by an enzyme that allows direct bright-field light microscopy assessment.^{24–27} Fluorescent systems of visualization and measurement are not subject to the limitations of high density and saturation. Optical detection and quantification of fluorescent signal depends on excitation and photon emission of specific wavelengths, resulting in signal intensity directly proportional to the concentration of the target of interest.²⁸ The dynamic range of common assays with fluorophores that emit in the visible region of the spectrum is two to three times the dynamic range of chromogenic stains. Multicolor detection by using fluorescent target labeling, which can be spectrally resolved, makes it possible to examine several markers at once.^{29,30} Several methods of quantification of fluorescent staining have been described.³¹ Here we use the AQUA technology as it does not require feature-based image fractionation, but rather allows detection of biomarker expression within specific subcellular compartments, as defined by antibody-conjugated fluorophore labeling and colocalization of the target of interest with cytoplasmic or nuclear staining.³² The fluorescent intensity is measured and divided by the compartment area to yield a quantitative, continuous, and reproducible score for each field of view. This technology has been extensively previously validated in tissue microarrays (TMAs) as well as whole-tissue sections.^{33,34}

To assess the problem of user and methodological bias in quantification of ER expression in breast cancer, we chose a three-pronged experimental approach to compare both automated (Aperio) and visual (pathologist) scoring of chromogenic staining, as well as to evaluate both of these techniques against quantitative immunofluorescence (QIF)-based ER detection. Each method of staining and detection was performed with two common clinical ER antibody clones (1D5 and SP1).

MATERIALS AND METHODS

Patient Cohorts and TMA Construction

Two retrospective breast cancer cohorts were constructed consisting of tissue obtained from the Archives of the Pathology Department at Yale University (New Haven, CT) and used to create two representative TMAs, as previously described. Briefly, YTMA 49 consists of 621 patients diagnosed between 1962 and 1982. This cohort is completely annotated with clinicopathological and follow-up information. YTMA 128 contains 235 patients diagnosed between 2003 and 2008. Cohort characteristics are summarized in Supplementary Tables 1 and 2. For both cohorts, 0.6 mm cores were taken from each specimen and combined into randomized TMAs, which were cut into 5- μ m sections and adhered to glass slides for immunostaining. An Index TMA consisting of cell lines with known concentration of ER and of patient samples with variable ER expression pattern (described previously in Welsh *et al.*³⁵) was run alongside each experiment for standardization and reproducibility purposes and to determine the threshold of detection for ER positivity for the different staining and reading methods described here.

Immunostaining with SP1

To visualize ER expression with the rabbit monoclonal SP1 antibody (ThermoScientific, Waltham, MA), slides were baked at 60 °C for 30 min to remove excess paraffin. Deparaffinization was performed in xylenes for two periods of 20 min each, after which slides were transferred to 100% ethanol and rehydrated to water in grades of ethanol. Heat-induced antigen retrieval took place in a PT module (LabVision, Kalamazoo, MI), where slides were immersed in sodium citrate buffer (pH 6) for 20 min at 97 °C. Slides were then rinsed in distilled water, transferred to a solution of 0.75% H₂O₂ in methanol for 30 min at room temperature to block endogenous peroxidases, and rinsed again in distilled water. They were then transferred to a Labvision autostainer, where the remaining staining steps were performed at room temperature and rinsed with tris-buffered saline/0.05% Tween-20 (TBST) between each stage. Nonspecific antigens were blocked by 30 min in 0.3% bovine serum albumin (BSA) diluted in TBST.

For chromogenic visualization, slides were incubated for 1 h with SP1 antibody (1:100) in BSA-TBST, and then with anti-rabbit EnVision (Dako, Carpinteria, CA) for 1 h. Signal

was developed for 5 min in DAB solution (Dako; prepared according to the manufacturer's instructions), followed by counterstaining for 1 min with hematoxylin (Tacha's automated hematoxylin, BioCare Medical, Concord, CA). Slides were removed from autostainer and coverslipped with Prolong Gold mounting medium (Life Technologies).

For slides to be visualized with fluorescence, a cocktail of SP1 antibody (1:100) and mouse pan-cytokeratin (Dako; 1:100) in 0.3% BSA-TBST were added for 1 h. The slides were then incubated with a secondary antibody cocktail of goat anti-mouse AlexaFluor 546 (Life Technologies) diluted 1:100 in anti-rabbit EnVision (Dako) for 1 h. Signal was amplified with Cy5-tyramide (Perkin Elmer, Waltham, MA) for 10 min, and nuclear staining was accomplished with 10 μ g/ml DAPI (Life Technologies) in BSA-TBST for 20 min. Slides were then removed from the autostainer and coverslipped using the Prolong Gold mounting medium (Life Technologies).

Immunostaining with 1D5

For ER visualization with the 1D5 antibody (Dako) and subsequent analysis with Aperio's FDA-approved nuclear algorithm, slides were stained according to the clinical site protocol for 1D5 as described previously.²²

ER 1D5 slides intended for fluorescent visualization were immunostained according to the same protocol as described for ER SP1. Slides were incubated in a primary antibody cocktail containing 1D5 (1:50) and pan-cytokeratin (rabbit polyclonal, Dako) at 1:100 in BSA-TBST for 30 min, followed by a secondary cocktail of goat anti-rabbit Alexa-Fluor 546 (1:100) in anti-mouse EnVision (Dako) for 30 min, as well as signal amplification with Cy5 and DAPI staining.

Aperio Nuclear Algorithm

For analysis with Aperio's nuclear algorithms, chromogenic slides were scanned to create bright-field digital images using the ScanScope CS (Aperio, Vista, CA). All digital images were viewed in ImageScope and analysis performed in Spectrum, elements of the Aperio image review and analysis suite. Slide images were first segmented to obtain a single image for each TMA spot, after which the pen tool was used to circle ('annotate') tumor areas for each spot. This was refined by use of a negative pen tool to subtract stromal areas enclosed by tumor to ensure that analysis would be restricted to tumor only.

For ER 1D5 scoring with the FDA-approved nuclear algorithm on YTMA 128, TMA spot images were first annotated to exclude stroma and restrict analysis to tumor areas only. The algorithm was then run on each spot to generate both a markup image (showing scoring for individual nuclei) and a percent-positive nucleus score for each spot.

For ER SP1 scoring, the unlocked nuclear algorithm was modified to take into account a darker counterstain and improve color deconvolution, but was otherwise not altered from the settings of the FDA-approved nuclear algorithm.

The nuclear algorithm input includes a section for red, green, and blue absorbance (OD) values for the hematoxylin counterstain in order to facilitate deconvolution from the nuclear stain, which has its own set of OD values. ImageScope's Image Quality feature was used to measure the RGB OD values within negative control spots. These were then averaged for the slide, substituted for the defaults, and the resultant algorithm saved and used to generate ER scores as percent-positive nuclei in annotated spot images. The counterstain RGB values were determined separately for each slide stained with SP1 to account for subtle variations in hematoxylin counterstaining between slides.

Pathologist Scoring

YTMA 49 and YTMA 128 slides with ER staining visualized by DAB were submitted to three board-certified pathologists (Path1, Path2, and Path3), who estimated percent-positive nuclei using the digital images acquired by Aperio's ScanScope CS. TMA spots denoted by a pathologist to contain no invasive breast cancer were excluded from further analysis in all three ER assessment methods, as were spots with diffuse cytoplasmic staining instead of specific nuclear signal.

Automated Quantitative Analysis

Immunofluorescence staining for both SP1 and 1D5 antibodies was quantified using automated quantitative analysis (AQUA) as previously described.³² Briefly, monochromatic images for each of the DAPI, Cy3, and Cy5 channels were captured after for each TMA spot using an automated PM-2000 microscope platform (Genoptix/Novartis). The cytokeratin expression (Cy3) was used to binarize pixels to create an epithelial tumor mask. DAPI staining within this tumor mask was used to create a nuclear compartment, in which ER expression (Cy5) was measured as the sum of all pixel intensities, divided by the area of the nuclear compartment. Scores were then individually normalized according to exposure time, bit depth, and lamp hours to allow direct comparison between spots on the same slide.

Statistical Analysis

Regression analysis to assess method and assay reproducibility was performed in Microsoft Excel 2010, and results were confirmed in the StatView software platform (SAS Institute, Cary, NC) by means of Pearson coefficients and ANOVA testing. Kaplan–Meier survival analysis was performed using StatView for each ER scoring method, and statistical significance was assessed using the log-rank test.

RESULTS

Fluorescent and Chromogenic Assessment

To evaluate methods of ER visualization and measurement, immunostaining was performed on serial sections of two breast cancer TMA cohorts collected at Yale, as previously described.³⁵ Figure 1 shows examples of low and high ER

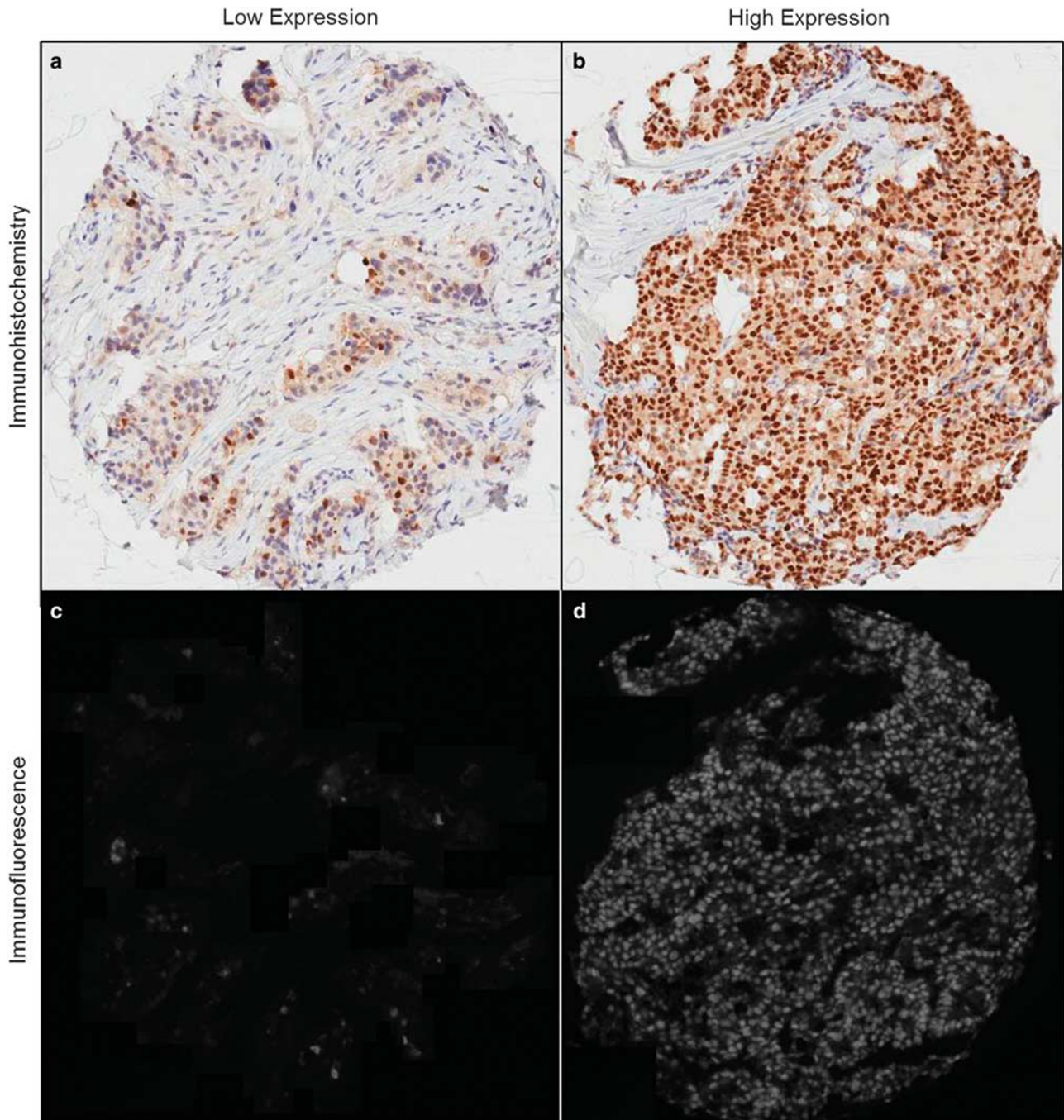


Figure 1 Examples of estrogen receptor staining in breast cancer tissue microarrays by both chromogenic and fluorescent methods. (a) Low and (b) high expression as visualized by 3-diaminobenzidine; corresponding on serial sections (c) low and (d) high expression as seen via conjugation with Cy5-tyramide.

expression with both chromogenic and fluorescent detection methods on serial sections. Digital images of each slide were then captured for further analysis (Figure 2).

Fluorescent detection slides were scanned at $\times 20$ to collect images from the DAPI, Cy3 (cytokeratin), and Cy5 (ER) channels (Figure 2a). These images were then analyzed with the AQUA software, which created an epithelial tumor mask from cytokeratin expression, and then used

DAPI expression within this mask to form a nuclear compartment. ER signal was quantified as the sum of pixel intensities divided by the nuclear compartment area and normalized to generate a Nuclear AQUA Score for each patient.

Chromogenic detection slides were scanned using Aperio's ScanScope CS digital image acquisition system, and board-certified pathologists scored percent-positive nuclei for

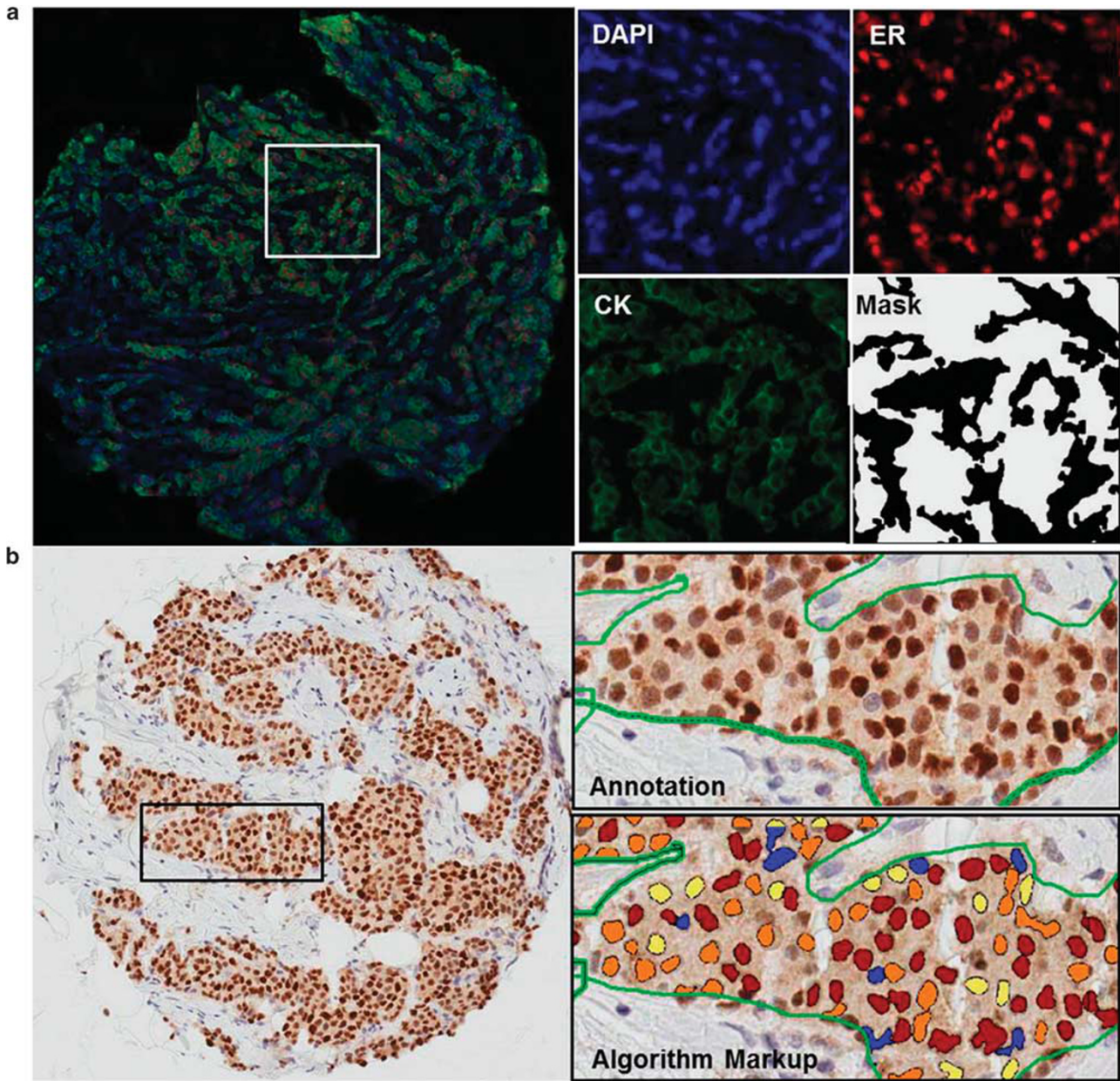


Figure 2 A demonstration of the components of fluorescent and chromogenic quantification as utilized by automated quantitative analysis (AQUA) and Aperio’s nuclear algorithm, respectively. **(a)** Simultaneous visualization of nuclei (blue, DAPI), pan-cytokeratin (green, AlexaFluor 546), and estrogen receptor (red, Cy5-tyramide) in a single tissue microarray spot. The AQUA program generates a tumor mask compartment from cytokeratin expression, further refines it into a nuclear compartment using DAPI positivity, and measures target signal intensity in the nuclear compartment. **(b)** Typical chromogenic staining for estrogen receptor in a strongly positive case, as visualized by diaminobenzidine (DAB) and counterstained with hematoxylin. The tumor areas are manually outlined (annotated; green line) by the user to exclude stromal nuclei. Aperio’s nuclear algorithm then uses morphological characteristics and the hematoxylin counterstain to identify nuclei. DAB intensity is then measured on a per-cell basis to determine positivity, and a markup image generated to illustrate results. Nuclei are binned into four categories to mimic pathologist intensity scoring: negative (blue = 0), weak positive (yellow = 1), positive (orange = 2), and strong positive (red = 3).

each TMA spot using these digital images. The images were then manually annotated by a trained technician to exclude stromal areas, and were analyzed with Aperio’s nuclear algorithm. Nuclei are binned into four categories (negative nuclei or weak, medium, and strong positive

nuclei), and a markup image created to reflect scoring results (Figure 2b). Aperio’s nuclear algorithm quantifies the annotated tissue for percent-positive nuclei as well as staining intensity according to predefined four categories resulting in a semiquantitative scoring system.

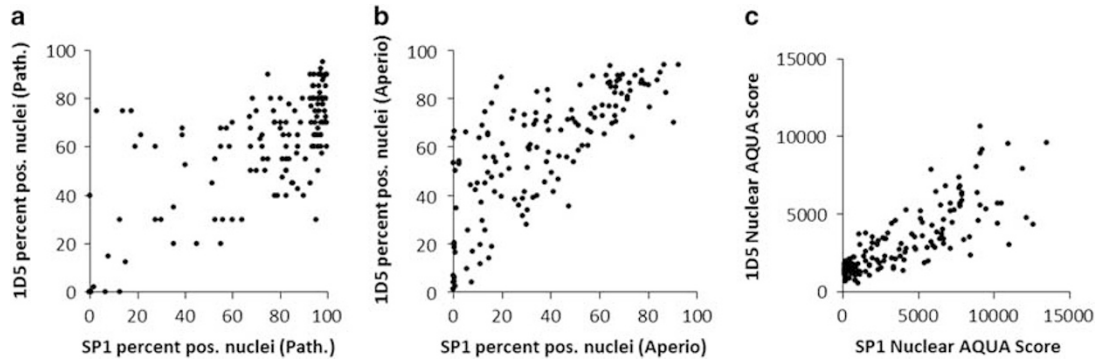


Figure 3 A comparison of antibody clones 1D5 and SP1 as they affect manual and automated assessment of estrogen receptor expression on YTMA 128. (a) Pathologist scoring of 1D5 vs pathologist scoring of SP1. (b) Aperio's FDA-approved nuclear algorithm scoring of 1D5 vs scoring of SP1 by a modified version of Aperio's nuclear algorithm. (c) Automated quantitative analysis (AQUA) scoring of 1D5 vs SP1 in the nuclear compartment.

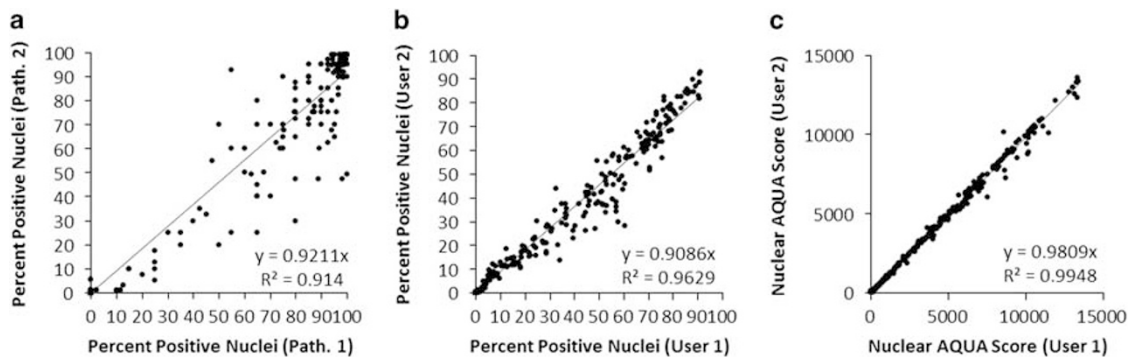


Figure 4 Interuser reproducibility for methods used to quantify estrogen receptor expression. (a) Pathologist scoring and (b) Aperio nuclear algorithm assessment of ER positivity were reported as percent-positive nuclei (chromogenic visualization), and (c) automated quantitative analysis (AQUA) quantification as the Nuclear AQUA Score (fluorescent visualization).

Antibody and User Variability

Our first step was to examine the relationship between ER 1D5 and ER SP1 scoring on YTMA 128 by all three methods of assessment (Figure 3). Whereas all methods show a correlation between the 1D5 and SP1 scores (Figure 3c), the relationship changes as a function of the method. Despite following the clinical site protocol precisely, we observed a titration independent, light brown haze over the tissue stained with the 1D5 antibody that was not present with SP1. As we wished to omit antibody-specific variables confounding reading and interpretation of the slides, all further analyses were performed using the ER SP1 clone.

To assess operator-based reproducibility, each assay analysis method was completed by two different operators allowing assessment of the subjective component of each scoring method (Figure 4). The Pearson coefficients (R^2) were above 0.9 for all methods; however, both automated scoring methods had higher reproducibility ($R^2 > 0.95$) between different operators. The regression R^2 between pathologists 1 and 2, as assessed by traditional visual scoring methods, was 0.92. The non-continuity of the scores can also be seen in

Figure 4a. The regression between the Aperio scores for two users was 0.96, showing better performance than traditional scoring but still suggesting some element of subjectivity. When two different users completed the AQUA scoring, the regression was nearly perfect (0.995), suggesting minimal user variation.

Assessment Methods Comparison

We then examined variability between methods using a linear regression analysis for continuous data (Figure 5). Although the pathologist data are not truly continuous, the estimations of percentage of positive nuclei were assumed to be continuous for the purposes of this assay. The regression between either pathologists' percent-positive nucleus scores and the score from Aperio's nuclear algorithm showed a nonlinear relationship where the pathologist scores were consistently higher than those generated by the Aperio nuclear algorithm (Figure 5a). There were essentially no cases where the pathologist's estimate was below the Aperio score. A similar pattern was seen with AQUA scores. Although AQUA measures pixel intensity of the target of

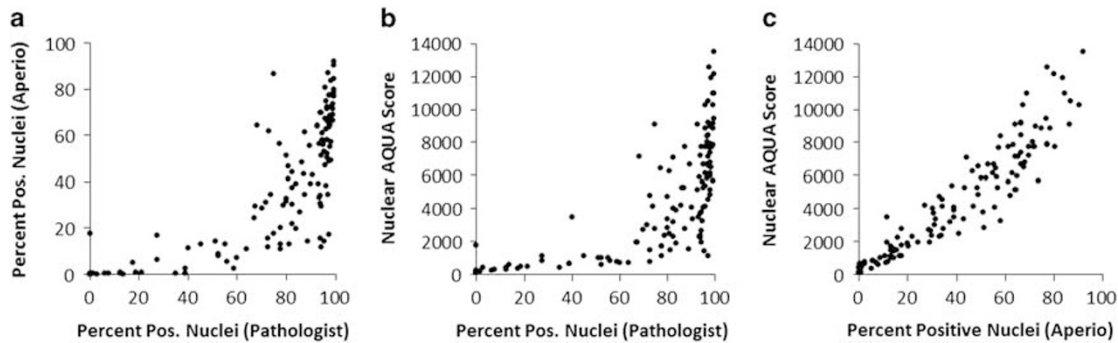


Figure 5 Relationships between methods used to assess estrogen receptor. (a) Aperio’s nuclear algorithm vs pathologist scoring; (b) automated quantitative analysis (AQUA) vs pathologist scoring; and (c) AQUA vs Aperio’s nuclear algorithm.

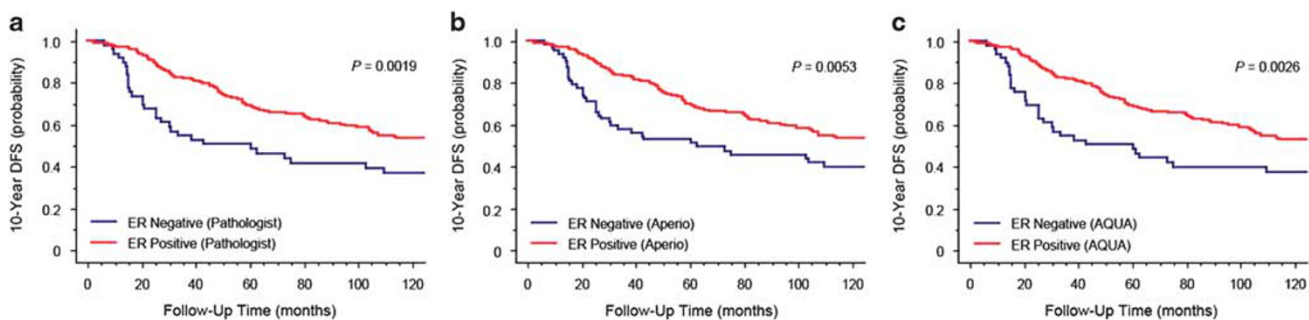


Figure 6 Kaplan–Meier survival analysis of breast cancer patients on YTMA 49 with estrogen receptor-negative (blue) and -positive (red) tissue, as measured by (a) pathologist, (b) Aperio’s nuclear algorithm, and (c) automated quantitative analysis (AQUA). The cutoff used for pathologist scoring and Aperio’s nuclear algorithm was 1% positive nuclei, as per the ASCO-CAP guidelines. The ER positivity threshold for AQUA was determined using an Index TMA with positive and negative cell lines stained alongside YTMA 49. Number of positive and negative cases in each group are summarized in Table 1.

interest (ER in this study) as opposed to percent positivity, it has a similar relationship when compared with pathologist scoring (Figure 5b). The closest relationship between any two methods is clearly between the two types of automated scoring, despite the different detection techniques (Figure 5c). However, comparing the two automated scoring methods reveals the lower dynamic range and enzymatic saturation of the DAB signal as compared with fluorescent measurement.

Survival Analysis and Discordance

Although regressions help us examine the similarities and differences in ER quantification methods, they do not provide any case-specific information on patient classification into the ER-negative or ER-positive groups. Furthermore, comparison of tests is more valuable when the test comparison can be assessed as a function of patient outcome. To see how the three assessment methods compared on this basis, we looked at their determination of ER status for patients on YTMA 49, a large, historic cohort collected at Yale between 1962 and 1982. The 10-year disease-free survival Kaplan–Meier curves are very similar between all three methods (Figure 6); however, their differences can be seen in the summary tables (Tables 1 and 2). When the continuous scores are binarized to

generate positive or negative output, only 19 of 233 total cases were discordant. There was only one case that was positive by pathologist and Aperio scoring, but negative by AQUA. In contrast, there were 10 cases that positive by pathologist and AQUA, but negative by Aperio (Table 1). There were three cases that were positive by pathologist, and negative by the AQUA and Aperio methods; and finally, five cases were positive by AQUA, and negative by pathologist and Aperio scoring. The number of discordant cases is too small to evaluate which method better correlates with outcome.

These discordant cases were carefully reviewed by an independent pathologist, who was not involved in previous readings, to determine reasons for discordance (images not shown). In the one case positive by the pathologist and Aperio, but negative by AQUA, there was a clear nuclear fluorescent staining visual by eye, but the nuclear AQUA score for that case was 107, just barely below the threshold of 110 (in a set of scores that ranged from 0 to 12 500). In contrast, for the five cases positive by AQUA and negative by pathologist and Aperio scoring, low but clearly positive fluorescent nuclear staining can be seen by eye, whereas by chromogenic detection no nuclear staining is detectable. This may be because of masking by the hematoxylin counterstain

Table 1 Summary of ER assessment method discordance on YTMA 49

Pathologist	Aperio	AQUA		Total
		Positive	Negative	
Positive	Positive	170	1	171
	Negative	10	3	13
	Total	180	4	184
Negative	Positive	0	0	0
	Negative	5	44	49
	Total	5	44	49

on these particular spots. Similarly, the 10 cases positive by pathologist and AQUA, but negative by Aperio, have clearly visual nuclear staining on both the fluorescent and chromogenic detection slides, but, for unknown reasons, the hematoxylin counterstain appears somewhat darker than most spots on the slide and was not detected by the Aperio algorithm. Finally, in the three cases that were positive by pathologist scoring and negative by the AQUA and Aperio algorithms, closer pathologist examination was unable to determine whether the cells considered positive contained extremely strong hematoxylin or were, in fact, positive for diaminobenzidine (spots appeared black).

In an effort to test the flexibility and performance of the Aperio nuclear algorithm, we attempted to further adjust the RGB values for the counterstain levels to see whether the algorithm would pick up the 10 false-negative cases. However, we were unable to find a set of values that would satisfy all cases. When settings were changed that allowed the algorithm to recognize these 10 cases as positive, the altered algorithm classified clearly negative nuclei as positive in other cases, or picked up far fewer nuclei than were actually present.

DISCUSSION

The 2010 ASCO-CAP guidelines for ER assessment recommend image analysis to quantify percent-positive tumor cells,⁵ especially as it is difficult to reliably score to a 1% threshold without laboriously counting individual cells. Aside from assisting pathologists, automated analysis systems such as the Aperio ScanScope XT and its associated algorithms have also been shown to be useful in discovery of more complex relationships between biomarkers.³⁶ Here we show that one method of automated chromogenic assessment shows good reproducibility and prognostic value, but, compared with fluorescence, is limited by the nature of chromogenic staining itself. Chromogenic staining requires a counterstain to provide context; however, this counterstain introduces inherent complications to objective scoring. It is well known that the quality and intensity of hematoxylin counterstaining varies among preparations, vendors and

Table 2 Hazard ratios for ER positivity in unselected breast cancer cohort YTMA 49 as diagnosed by different reading methods

Reading method	HR (95% CI)	P-value
AQUA	0.53 (0.35–0.82)	0.005
Pathologist	0.52 (0.35–0.80)	0.004
Aperio	0.57 (0.39–0.86)	0.008

protocols, over the lifetime of the reagent, and also between cell and tissue types. The CAS-200 platform is an example of a system that required adjustments to account for counterstain differences between slides and batches.³⁷ In the clinic, when a patient case has an obvious problem with the counterstain, the slide can be sent back and another stain requested. However, there is still a chance that even ‘acceptable’ counterstaining can mask low-level chromogenic staining, whether by eye or by automated color-deconvolution (or spectral unmixing) analysis, as occurred in five cases in this study.³⁸ Previous unpublished work from our lab suggests that there are a number of cases where dark staining with hematoxylin, due either to tissue variation or pathologist preference, has obscured low-level ER expression to generate a false-negative test.

Fluorescent detection avoids the disadvantages and limitations of the hematoxylin counterstain, but has other limitations. Specifically, the absence of hematoxylin makes it challenging to generate the cellular context with a conventional IHC appearance. Whereas additional fluorophores can be used to visualize other tissue features, the image is still quite different from conventional IHC. QIF is also generally costlier than traditional IHC. Unfortunately, the cost analysis of automated ER evaluation in clinical lab settings is beyond the scope of this manuscript, and this information is not accessible to us. One could imagine though that routine ER assessment might be performed using regular DAB-based IHC as established, and just the cases that are negative by this assay could be sent out to laboratories that offer fluorescence-based assays, taking advantage of increased sensitivity of this assay for low expressing biomarkers. Other advantages of QIF consist of broader dynamic range, dynamic adjustment of exposure time, and decreased requirement for human interface for tumor selection.

Perhaps the greatest advantage of QIF lies in the potential to generate a standard curve that can be used to establish a defined, reproducible cutoff for every assay. This method also has the potential to enable more accurate quantification of biomarker expression.³⁸ Recent studies have demonstrated that quantification by ELISA can provide more accurate assessment of patient outcome than qualitative IHC, and may even demonstrate a distinct benefit between negative, moderate, and strong ER positivity rather than just between positive and negative groups.³⁹ This advantage extends

beyond analysis of ER in breast cancer to most accurate quantification of biomarker expression levels in various cancer and tissue types.

Whereas this study of comparison of different methods of ER analysis was performed in a rigorous and tightly controlled manner, it is subject to a number of limitations. Evaluation of ER expression was performed on TMAs, which allows a high-throughput approach, but does not truly represent the clinical setting where biopsies or whole-tissue sections are routinely stained and evaluated for the biomarker in question. One can argue that discordances in ER assessment are because of the small amount of tumor represented in a 0.6-mm TMA core. This might be a valid argument regarding ER heterogeneity, as 0.6 mm might not always represent the ER status of whole-tissue sections. However, the different staining methods were performed on serial sections, reducing heterogeneity between methods to a minimum. In addition, it does not resolve the issue of false-negative reading due to variability in hematoxylin staining intensity. Moreover, the three methods of ER analysis were also compared on a number of whole-tissue sections (around 25 samples for this study). These data were not shown in the manuscript because they did not render additional information. The results of ER analysis on whole-tissue sections using the different methods of assessment did not show any discrepancies, probably because of the low number of cases. Another limitation of this study is that staining and analysis were performed within a single institution. Whereas this approach guarantees consistency for pre-analytical tissue processing and analytical procedures, these results would be more robust if more than one laboratory participated in the study.

In addition, this study does not reveal a significant difference of ER reading methods in regards to survival analysis. However, this observation might be because of the relative small number of patients included in survival analysis. To determine the best prognostic and predictive value of these tests by Kaplan–Meier analysis, a larger number of patients would need to be analyzed with all the three methods.

In summary, each of the methods of *in situ* protein detection in FFPE tissue samples has its strengths and weaknesses. Whereas conventional DAB-based IHC is a well-established and inexpensive procedure, reproducibility and sensitivity of the scoring are dependent on the counterstain and the reading method—by eye or automated. QIF on the other hand offers an automated and standardized approach to biomarker evaluation. Higher sensitivity of the assay and broader dynamic range facilitate more exact measurements of protein concentrations. Increased costs of QIF and the absence of hematoxylin generating the cellular context with a conventional IHC appearance need to be considered.

In theory, QIF can combine the best of both worlds—*in situ* evaluation of a biomarker and rigorous quantification. Our data here and in previous work by others and us suggest

that patient care may be improved with quantitative assessment. Whereas the percentage of discordant cases in this study (8.2%) is relatively low, and in keeping with expected variability compared with other studies,⁴⁰ a more objective estimate of ER positivity could benefit hundreds of thousands of women worldwide.

Supplementary Information accompanies the paper on the Laboratory Investigation website (<http://www.laboratoryinvestigation.org>)

ACKNOWLEDGMENTS

This project has been funded in whole or in part with federal funds from the National Cancer Institute, National Institutes of Health, under Contract No. HHSN261200800001E. The content of this publication does not necessarily reflect the views or policies of the Department of Health and Human Services, nor does mention of trade names, commercial products, or organizations imply endorsement by the US Government.

DISCLOSURE/CONFLICT OF INTEREST

Within the last 12 months, DR has served as a paid consultant or advisor to Genoptix/Novartis, Cernostics, BMS, Bioccept, Perkin Elmer, and Metamark Genetics. The remaining authors declare no conflict of interest.

1. 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.
2. Osborne CK, Fisher E, Redmond C, *et al*. Estrogen receptor, a marker for human breast cancer differentiation and patient prognosis. *Adv Exp Med Biol* 1981;138:377–385.
3. McGuire WL, De La Garza M, Chamness GC. Evaluation of estrogen receptor assays in human breast cancer tissue. *Cancer Res* 1977;37:637–639.
4. Hede K. Breast cancer testing scandal shines spotlight on black box of clinical laboratory testing. *J Natl Cancer Inst* 2008;100:836–837 44.
5. 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.
6. Allison KH. Estrogen receptor expression in breast cancer: we cannot ignore the shades of gray. *Am J Clin Pathol* 2008;130:853–854.
7. Allred DC. Commentary: hormone receptor testing in breast cancer: a distress signal from Canada. *Oncologist* 2008;13:1134–1136.
8. De Cecco L, Musella V, Veneroni S, *et al*. Impact of biospecimens handling on biomarker research in breast cancer. *BMC Cancer* 2009;9:409.
9. Hicks DG, Boyce BF. The challenge and importance of standardizing pre-analytical variables in surgical pathology specimens for clinical care and translational research. *Biotechnol Histochem* 2012;87:14–17.
10. Cavaliere A, Bucciarelli E, Sidoni A, *et al*. Estrogen and progesterone receptors in breast cancer: comparison between enzyme immunoassay and computer-assisted image analysis of immunocytochemical assay. *Cytometry* 1996;26:204–208.
11. Miller RT, Hapke MR, Greene GL. Immunocytochemical assay for estrogen receptor with monoclonal antibody D753P gamma in routinely processed formaldehyde-fixed breast tissue. Comparison with frozen section assay and with monoclonal antibody H222. *Cancer* 1993;71:3541–3546.
12. Snead DR, Bell JA, Dixon AR, *et al*. Methodology of immunohistological detection of oestrogen receptor in human breast carcinoma in formalin-fixed, paraffin-embedded tissue: a comparison with frozen section methodology. *Histopathology* 1993;23:233–238.
13. 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.
14. Pertschuk LP, Feldman JG, Kim YD, *et al*. Estrogen receptor immunocytochemistry in paraffin embedded tissues with ER1D5 predicts breast cancer endocrine response more accurately than H222Sp

- gamma in frozen sections or cytosol-based ligand-binding assays. *Cancer* 1996;77:2514–2519.
15. Parker RL, Huntsman DG, Lesack DW, *et al*. Assessment of interlaboratory variation in the immunohistochemical determination of estrogen receptor status using a breast cancer tissue microarray. *Am J Clin Pathol* 2002;117:723–728.
 16. Diaz LK, Sneige N. Estrogen receptor analysis for breast cancer: current issues and keys to increasing testing accuracy. *Adv Anat Pathol* 2005;12:10–19.
 17. Gokhale S, Rosen D, Sneige N, *et al*. Assessment of two automated imaging systems in evaluating estrogen receptor status in breast carcinoma. *Appl Immunohistochem Mol Morphol* 2007;15:451–455.
 18. Esteban JM, Battifora H, Warsi Z, *et al*. Quantification of estrogen receptors on paraffin-embedded tumors by image analysis. *Mod Pathol* 1991;4:53–57.
 19. Makkink-Nombrado SV, Baak JP, Schuurmans L, *et al*. Quantitative immunohistochemistry using the CAS 200/486 image analysis system in invasive breast carcinoma: a reproducibility study. *Anal Cell Pathol* 1995;8:227–245.
 20. Administration UFaD. 510(k) Summary of Substantial Equivalence, Aperio Technologies. (ScanScope XT System) www.accessdata.fda.gov/cdrh_docs/pdf7/K073677.pdf. Accessed on 1 March 2014.
 21. Nassar A, Cohen C, Agersborg SS, *et al*. A multisite performance study comparing the reading of immunohistochemical slides on a computer monitor with conventional manual microscopy for estrogen and progesterone receptor analysis. *Am J Clin Pathol* 2011;135:461–467.
 22. Nassar A, Cohen C, Agersborg SS, *et al*. A new immunohistochemical ER/PR image analysis system: a multisite performance study. *Appl Immunohistochem Mol Morphol* 2011;19:195–202.
 23. Graham RC Jr, Karnovsky MJ. The early stages of absorption of injected horseradish peroxidase in the proximal tubules of mouse kidney: ultrastructural cytochemistry by a new technique. *J Histochem Cytochem* 1966;14:291–302.
 24. Ramos-Vara JA, Miller MA. When tissue antigens and antibodies get along: revisiting the technical aspects of immunohistochemistry—the red, brown, and blue technique. *Veterinary Pathol* 2014;51:42–87.
 25. Tubbs RR, Sheibani K, Deodhar SD, *et al*. Enzyme immunohistochemistry: review of technical aspects and diagnostic applications. *Cleve Clin Q* 1981;48:245–281.
 26. Nakane PK, Pierce GB Jr. Enzyme-labeled antibodies: preparation and application for the localization of antigens. *J Histochem Cytochem* 1966;14:929–931.
 27. Speel EJ. Robert Feulgen Prize Lecture 1999. Detection and amplification systems for sensitive, multiple-target DNA and RNA in situ hybridization: looking inside cells with a spectrum of colors. *Histochem Cell Biol* 1999;112:89–113.
 28. Waggoner A. Fluorescent labels for proteomics and genomics. *Curr Opin Chem Biol* 2006;10:62–66.
 29. Ioannou D, Tempest HG, Skinner BM, *et al*. Quantum dots as new-generation fluorochromes for FISH: an appraisal. *Chromosome Res* 2009;17:519–530.
 30. Xu H, Xu J, Wang X, *et al*. Quantum dot-based, quantitative, and multiplexed assay for tissue staining. *ACS Appl Mater Interfaces* 2013;5:2901–2907.
 31. Rojo MG, Bueno G, Slodkowska J. Review of imaging solutions for integrated quantitative immunohistochemistry in the Pathology daily practice. *Folia Histochem Cytobiol* 2009;47:349–354.
 32. Camp RL, Chung GG, Rimm DL. Automated subcellular localization and quantification of protein expression in tissue microarrays. *Nat Med* 2002;8:1323–1327.
 33. Camp RL, Charette LA, Rimm DL. Validation of tissue microarray technology in breast carcinoma. *Lab Invest* 2000;80:1943–1949.
 34. Giltneane JM, Rimm DL. Technology insight: Identification of biomarkers with tissue microarray technology. *Nat Clin Pract Oncol* 2004;1:104–111.
 35. Welsh AW, Harigopal M, Wimberly H, *et al*. Quantitative analysis of estrogen receptor expression shows SP1 antibody is more sensitive than 1D5. *Appl Immunohistochem Mol Morphol* 2013;21:139–147.
 36. Laurinavicius A, Laurinaviciene A, Ostapenko V, *et al*. Immunohistochemistry profiles of breast ductal carcinoma: factor analysis of digital image analysis data. *Diagn Pathol* 2012;7:27.
 37. Johnsson A, Olsson C, Anderson H, *et al*. Evaluation of a method for quantitative immunohistochemical analysis of cisplatin-DNA adducts in tissues from nude mice. *Cytometry* 1994;17:142–150.
 38. 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.
 39. Mazouni C, Bonnier P, Goubar A, *et al*. Is quantitative oestrogen receptor expression useful in the evaluation of the clinical prognosis? Analysis of a homogeneous series of 797 patients with prospective determination of the ER status using simultaneous EIA and IHC. *Eur J Cancer* 2010;46:2716–2725.
 40. 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.