

A systematic comparison of three commercial estrogen receptor assays in a single clinical outcome breast cancer cohort

Elizabeth N Kornaga¹, Alexander C Klimowicz², Natalia Guggisberg¹, Travis Ogilvie³, Don G Morris^{1,4}, Marc Webster⁴ and Anthony M Magliocco⁵

¹Translational Laboratories, Tom Baker Cancer Centre, Alberta Health Services, Calgary, AB, Canada;

²Boehringer Ingelheim Pharmaceuticals Inc., Ridgefield, CT, USA; ³Department of Pathology, Calgary

Laboratory Services, Calgary, AB, Canada; ⁴Department of Oncology, Tom Baker Cancer Centre and

University of Calgary, Calgary, AB, Canada and ⁵Department of Anatomic Pathology, H. Lee Moffitt

Cancer Center & Research Institute, Tampa, FL, USA

Breast cancers are routinely assessed for estrogen receptor status using immunohistochemical assays to assist in patient prognosis and clinical management. Specific assays vary between laboratories, and several antibodies have been validated and recommended for clinical use. As numerous factors can influence assay performance, many laboratories have opted for ready-to-use assays using automated stainers to improve reproducibility and consistency. Three commonly used autostainer vendors—Dako, Leica, and Ventana—all offer such estrogen receptor assays; however, they have never been directly compared. Here, we present a systematic comparison of three platform-specific estrogen receptor ready-to-use assays using a retrospective, tamoxifen-treated, breast cancer cohort from patients who were treated in Calgary, Alberta, Canada from 1985 to 2000. We found all assays showed good intra-observer agreement. Inter-observer pathological scoring showed some variability: Ventana had the strongest agreement followed closely by Dako, whereas Leica only showed substantial agreement. We also analyzed each estrogen receptor assay with respect to 5-year disease-free survival, and found that all performed similarly in univariate and multivariate models. Determination of measures of test performance found that the Leica assay had a lower negative predictive value than Dako or Ventana, compared with the original ligand-binding assay, while other measures—sensitivity, specificity, positive predictive value, and accuracy—were comparable between the three ready-to-use assays. When comparing against disease-free survival, the difference in negative predictive value between the vendor assays were not as extreme, but Dako and Ventana still performed slightly better than Leica. Despite some discordance, we found that all ready-to-use assays were comparable with or superior to the ligand-binding assay, endorsing their continued use. Our analysis also allowed for exploration of estrogen receptor-negative, progesterone receptor-positive cases, and we discovered that this phenotype was not consistent across the assays, suggesting this might be an artifact.

Modern Pathology (2016) 29, 799–809; doi:10.1038/modpathol.2016.74; published online 29 April 2016

Endocrine therapy is standard treatment for hormone receptor-positive breast cancers that can improve patient survival.^{1–3} Estrogen receptor and progesterone receptor analysis is routinely performed by immunohistochemical assays on breast cancer specimens to classify hormone receptor status and determine patient prognosis and management.^{1–4}

Eventually, immunohistochemical-based testing replaced the ligand-binding assay as the standard method for determining hormone receptor status.⁵ The switch from ligand-binding assay to immunohistochemistry did not occur at a specific time point, and varied by location and laboratories between the mid-90s and early 2000s. Although estrogen and progesterone receptor are both routinely evaluated, patient management decisions are typically made on the basis of estrogen receptor results alone as the role of progesterone receptor in breast cancer patient management is not clearly established.⁴ In 2009, the discovery of faulty testing and reporting of estrogen receptor levels in Newfoundland, Canada brought

Correspondence: Dr AM Magliocco, MD, FRCPC, FCAP, Department of Anatomic Pathology, H. Lee Moffitt Cancer Center & Research Institute, 12902 Magnolia Drive, Tampa, FL 33612, USA. E-mail: Anthony.Magliocco@moffitt.org

Received 13 January 2016; revised 8 March 2016; accepted 8 March 2016; published online 29 April 2016

the immunohistochemical methodologies and guidelines under scrutiny.^{6,7} In 2010, ASCO/CAP released updated guideline recommendations for immunohistochemistry-based hormone receptor testing in breast cancer to improve reproducibility of testing and reporting.⁴ This guideline provides recommendations of validated and acceptable estrogen receptor antibody clones in addition to guidelines for pre-analytical handling of the specimens; however, as the estrogen receptor testing controversy in Newfoundland demonstrated, an acceptable antibody clone is not sufficient to produce reliable results. Although laboratory-developed assays are still used for routine testing, many clinical laboratories have opted to use platform-specific, ready-to-use estrogen receptor assays developed by the major manufacturers of automated immunohistochemistry stainers including: DakoCytomation (Dako), Leica Microsystems (Leica), and Ventana Medical Systems (Ventana). Ready-to-use assays offer improved reproducibility and consistency, minimizing effects arising from reagent diversity, as well as variability that can arise from antigen retrieval.^{8–16} Surprisingly, these platform-specific ready-to-use assays have never been directly compared using the same clinical outcome series.

Here, we present a systematic comparison of three platform-specific estrogen receptor ready-to-use immunohistochemical assays using a retrospective tamoxifen-treated breast cancer cohort to: (i) evaluate observer agreement and concordance among the three ready-to-use assays; (ii) assess the prognostic ability of the ready-to-use assays, compared with the estrogen receptor ligand-binding assay, to predict response to tamoxifen; and (iii) calculate standard measures of test performance using (a) ligand-binding assay or (b) progression on tamoxifen, as the reference standards. In addition, we also did an exploratory analysis of the estrogen receptor-negative, progesterone receptor-positive phenotype to investigate whether this is a true subtype of breast cancers.

Materials and methods

Patient Cohort

The Calgary tamoxifen breast cancer cohort (Calgary cohort) is a retrospective database containing demographic, clinical, and pathological data for 819 breast cancer patients diagnosed between 1985 and 2000 at the Tom Baker Cancer Centre in Calgary, Alberta, Canada. Subjects were included in the study if they had a confirmed diagnosis of invasive breast carcinoma, received surgical intervention as their primary treatment (with or without postoperative local radiation therapy), and if they received adjuvant tamoxifen endocrine therapy (20 mg p.o./day). Subjects were excluded from the study if there was no surgical formalin-fixed paraffin-embedded archival

tissue available, if they had a prior cancer diagnosis (except non-melanoma skin cancer), or if they received primary or adjuvant chemotherapy. Of the total cohort, 532 subjects met the criteria; for these, replicate 0.6-mm cores were taken from the archival tissue blocks built into tissue microarrays.¹⁷ During the diagnosis period for this cohort, estrogen receptor ligand-binding assay was predominantly performed, and all patients were treated with tamoxifen regardless of estrogen receptor status as there was no definitive evidence at that time that estrogen receptor-negative patients did not respond to tamoxifen.^{2,3} Human epidermal growth factor receptor 2 (HER2) status was not routinely assessed at the time of diagnosis for patients in this cohort; retrospective HER2 assessment was performed by HER2 immunohistochemistry. Clinical-pathological characteristics of the patients from the Calgary cohort can be found in Table 1.

Ligand-Binding Assay

During the diagnosis period of this cohort (1985–2000), estrogen receptor status was predominantly assessed using dextran-coated charcoal binding assay at the provincial hormone receptor laboratory in Edmonton, Alberta, Canada.¹⁸ Tumors were considered to be estrogen receptor-positive if the ligand-binding assay results were greater than 10 fmol/ml. Ligand-binding assay results are available for 411 cases (cases without ligand-binding assay results were assessed using immunohistochemistry by local laboratories with early adoption of new method).

HER2 Immunohistochemistry

HER2 immunohistochemistry was performed using the HercepTest pharmDx kit (Dako, SK001) according to kit instructions. Briefly, slides with 4- μ m sections of the tissue microarrays were deparaffinized in xylene, rinsed in ethanol, and rehydrated. Heat-induced epitope retrieval was performed by heating the slides to 97 °C for 40 min in a Dako PTLINK using the epitope retrieval solution provided in the kit. Slides were then processed on a Dako Link48 autostainer using the HercepTest pharmDx program and kit-provided reagents. Kit-provided cell line controls were run in addition to a laboratory-built reference tissue microarray that contained duplicate cores from four samples for each HER2 score (0, 1+, 2+, and 3+).

Estrogen and Progesterone Receptor Immunohistochemistry

Immunohistochemistry on the Dako platform was performed using the ER/PR pharmDx kit (Dako, SK310) according to provided FDA-approved protocol. Briefly, tissue microarray slides were

Table 1 Clinical-pathological characteristics of Calgary Tamoxifen Breast Cancer Cohort of those included in final tissue microarray vs those not included

| | In TMA (n = 532) | | Not in TMA (n = 287) | |
|----------------------------------|---------------------|-----|-------------------------|----|
| | # of Cases | % | # of Cases | % |
| Age | | | | |
| Range | 35.8–95.5 | | 32.1–95.1 | |
| Median | 66 | | 68 | |
| < 53 | 81 | 15 | 81 | 15 |
| ≥ 53 | 451 | 85 | 451 | 85 |
| Menopausal status | | | | |
| Pre-menopausal | 30 | 6 | 10 | 3 |
| Peri-menopausal | 25 | 5 | 9 | 3 |
| Post-menopausal | 385 | 72 | 223 | 78 |
| Male | 3 | 1 | 1 | 0 |
| Unknown | 89 | 17 | 44 | 15 |
| Size | | | | |
| < 2 cm | 271 | 51 | 104 | 36 |
| ≥ 2 cm | 221 | 42 | 134 | 47 |
| Unknown | 40 | 8 | 49 | 17 |
| Grade | | | | |
| 1 | 127 | 24 | 48 | 17 |
| 2 | 292 | 55 | 117 | 41 |
| 3 | 68 | 13 | 52 | 18 |
| Unknown | 45 | 8 | 70 | 24 |
| Stage | | | | |
| I | 233 | 44 | 88 | 31 |
| II | 163 | 31 | 75 | 26 |
| III | 40 | 8 | 32 | 11 |
| IV | 7 | 1 | 17 | 6 |
| Unknown | 89 | 17 | 75 | 26 |
| Lymph node status | | | | |
| Negative | 342 | 64 | 136 | 47 |
| Positive | 117 | 22 | 71 | 25 |
| Unknown | 73 | 14 | 80 | 28 |
| ER ligand-binding assay | | | | |
| Negative (< 10 fmol) | 27 | 5 | 31 | 11 |
| Positive (> 10 fmol) | 384 | 72 | 147 | 51 |
| Unknown | 121 | 23 | 109 | 38 |
| Chemotherapy | | | | |
| No | 532 | 100 | 272 | 95 |
| Yes | 0 | 0 | 15 | 5 |
| Previous cancer diagnosis | | | | |
| No | 532 | 100 | 238 | 83 |
| Yes | 0 | 0 | 49 | 17 |
| HER2 status | | | | |
| Negative | 508 | 95 | | |
| Positive | 19 | 4 | | |
| Unknown | 5 | 1 | | |
| Status at 5 years | | | | |
| Alive and free of disease | 399 | 79 | 132 | 49 |
| Alive with disease | 19 | 4 | 18 | 7 |
| Dead of disease | 85 | 17 | 116 | 43 |
| Dead of other causes | 29 | 6 | 20 | 12 |
| Lost to follow-up | 0 | 0 | 1 | 0 |

deparaffinized and rehydrated as previously described. Heat-induced epitope retrieval was performed by heating slides to 125 °C in ER/PR pharmDx epitope retrieval solution for 5 min in a decloaking chamber (Biocare Medical, DC2002). Slides were then processed on a Dako Link48 autostainer using the kit reagents and preset program, including the ready-to-use pharmDx estrogen receptor antibody cocktail (monoclonal mouse anti-human ER clones ID5 and ER-2-123) and the pharmDx progesterone receptor antibody (monoclonal mouse anti-human PR clone PgR1294).

Immunohistochemistry on the Leica BOND-III staining platform was performed using the manufacturer's recommended protocol: Onboard epitope retrieval for 20 min using Bond Epitope Retrieval 1 for estrogen receptor (Leica, AR9961) and Bond Epitope Retrieval 2 for progesterone receptor (AR9640) followed by Protocol F. Using Protocol F, slides were incubated for 15 min with Novacastra ready-to-use mouse monoclonal antibody estrogen receptor clone 6F11 (RTU-ER-6F11) or Novacastra ready-to-use mouse monoclonal progesterone receptor antibody cocktail (clones 16 and SAN27; RTU-PGR-AB) using the Bond Polymer Refine Detection Kit (DS9800).

Immunohistochemistry on the Ventana Ultra platform was performed using the manufacturer's recommended settings. Mild Cell Conditioning 1 solution was used for heat-induced epitope retrieval. Slides were incubated for 16 min with the ready-to-use CONFIRM anti-estrogen receptor antibody (monoclonal rabbit anti-human ER clone SP1, Roche Diagnostics, 790-4324) or the ready-to-use CONFIRM anti-progesterone receptor antibody (monoclonal rabbit anti-human PR clone 1E2, 790-2223) and visualized using the ultraView Universal DAB Detection Kit (760-500).

Cell line controls from the Dako ER/PR pharmDx kit were run in addition to a laboratory built reference tissue microarrays for all vendor-specific assays.

All stained tissue microarrays showed consistent staining, including cores at the edge and corners, suggesting that the antibody was evenly applied by the autostainers during the staining process.

Immunohistochemistry Scoring

HER2, estrogen receptor, and progesterone receptor immunohistochemically stained slides were manually scored following the ASCO/CAP guidelines.^{4,19} All assays were scored blinded to the ready-to-use manufacturer and clinical outcome. Specifically, HER2 was scored as either 0, 1+, 2+, or 3+, and tumors were considered positive if the HER2 average of the replicate cores for each case were greater than 2. Estrogen and progesterone receptor status was evaluated following the Allred scoring method,²⁰ and tumors were considered positive if they had an

overall Allred score of 3 or higher in at least one of the replicate 0.6-mm cores. Scoring was performed as a consensus between two trained technicians (Observer 1: AK/EK), who regularly participated in quality control testing with Canadian Immunohistochemistry Quality Control organization (cIQc) and the United Kingdom National (UK-NEQAS). Slides were also scored by two expert pathologists (Observers 2 and 3: TO and NG). Slides were rescored by Observers 1 and 3 after 3 months of the initial review to assess intra-observer reproducibility.

Statistical Analysis

Statistical analyses were performed using Stata 12 (StataCorp LP). The kappa statistic was used to measure inter- and intra-observer agreement, as well as inter-platform comparisons.²¹ The event under study was 5-year disease-free survival, defined as time from diagnosis to local recurrence, metastatic disease, or death from breast cancer. Median follow-up time for the cohort was 89.7 months. Kaplan–Meier curves were analyzed using the logrank test at 5-year disease-free survival. Cox proportional hazard regression was performed to estimate hazard ratios of estrogen receptor for 5-year disease-free survival, adjusting for lymph node status, tumor grade, tumor size, and HER2 status. Subjects were not included in the multivariate models if there were missing data for any of these variables. The proportional hazard assumptions for these models were tested by assessing log-log survival curves as well as the goodness-of-fit using the Schoenfeld residuals test. The ligand-binding assay results were used as the reference standard for calculating measures of test performance, and additional calculations were performed, comparing the ligand-binding and immunohistochemical assays, with progression on tamoxifen (5-year disease-free survival) as the reference standard.²²

Results

Agreement

Inter-observer agreement was evaluated for all three platforms by comparing observer interpretation of each core for estrogen receptor staining between two trained researchers (Observer 1) and two expert pathologists (Observers 2 and 3), and was measured using the kappa (κ) statistic. Inter-observer interpretations on the Dako platform showed strong agreement for estrogen receptor scoring with $\kappa=0.89$ between Observers 1 and 2, $\kappa=0.80$ between Observers 1 and 3, and $\kappa=0.92$ between Observers 2 and 3. Inter-observer agreement for estrogen receptor interpretation on the Leica platform showed substantial agreement with $\kappa=0.67$ between Observers 1 and 2, $\kappa=0.75$ between Observers 1 and 3, and $\kappa=0.83$ between Observers 2 and 3. Kappa values for

interpretation with the Ventana staining showed the strongest agreement, with $\kappa=0.88$ between Observers 1 and 2, $\kappa=1.00$ between Observers 1 and 3, and $\kappa=0.88$ between Observers 2 and 3 for estrogen receptor.

Tissue microarray slides were rescored by Observers 1 and 3 after 3 months of the initial review, and intra-observer agreement was calculated for the three platforms. For Observer 1, Dako and Ventana had perfect agreement ($\kappa=1.00$), while Leica had almost perfect agreement ($\kappa=0.91$). For Observer 3, Leica and Ventana had perfect agreement ($\kappa=1.00$), while Dako had almost perfect agreement ($\kappa=0.91$).

Concordance

Estrogen receptor status for the platform-specific assays was compared as a core by core analysis, and agreement was assessed using the kappa statistic. Estrogen receptor status between the Dako and Ventana platforms showed almost perfect agreement, with $\kappa=0.90$ (Figure 1a). Specifically, there were only 9 discordant cores ($n=5$, estrogen receptor-negative Dako/estrogen receptor-positive Ventana; $n=4$ estrogen receptor-positive Dako/estrogen receptor-negative Ventana). Substantial agreement was seen between Leica and Ventana ($\kappa=0.79$, Figure 1b), with 21 cases of discordance observed ($n=21$, estrogen receptor-negative Leica/estrogen receptor-positive Ventana). Substantial agreement was also reached between Dako and Leica ($\kappa=0.66$, Figure 1c), with a total of 43 discordant cores ($n=2$, estrogen receptor-negative Dako/estrogen receptor-positive Leica; $n=41$, estrogen receptor-positive Dako/estrogen receptor-negative Leica). Example images of discordance between platforms are presented in Figures 1d–f.

Univariate Analysis of Estrogen Receptor

Cases were dichotomized into positive and negative as previously described,^{4,18} and Kaplan–Meier survival curves for 5-year disease-free survival were analyzed (Figure 2). The logrank test was used to compare estrogen receptor-positive and -negative groups and hazard ratios were calculated to compare relative survival. All assays (ligand-binding and immunohistochemical) achieved significance with the logrank test and the hazard ratio estimates ($P\leq 0.0001$).

Multivariate Analysis of Estrogen Receptor

Cox proportional hazard models were analyzed for the effect of estrogen receptor status as determined by the immunohistochemical assay for each platform, as well as the ligand-binding assay, and are presented in Table 2. HER2 status and tumor size did not show significance, but were kept in the final

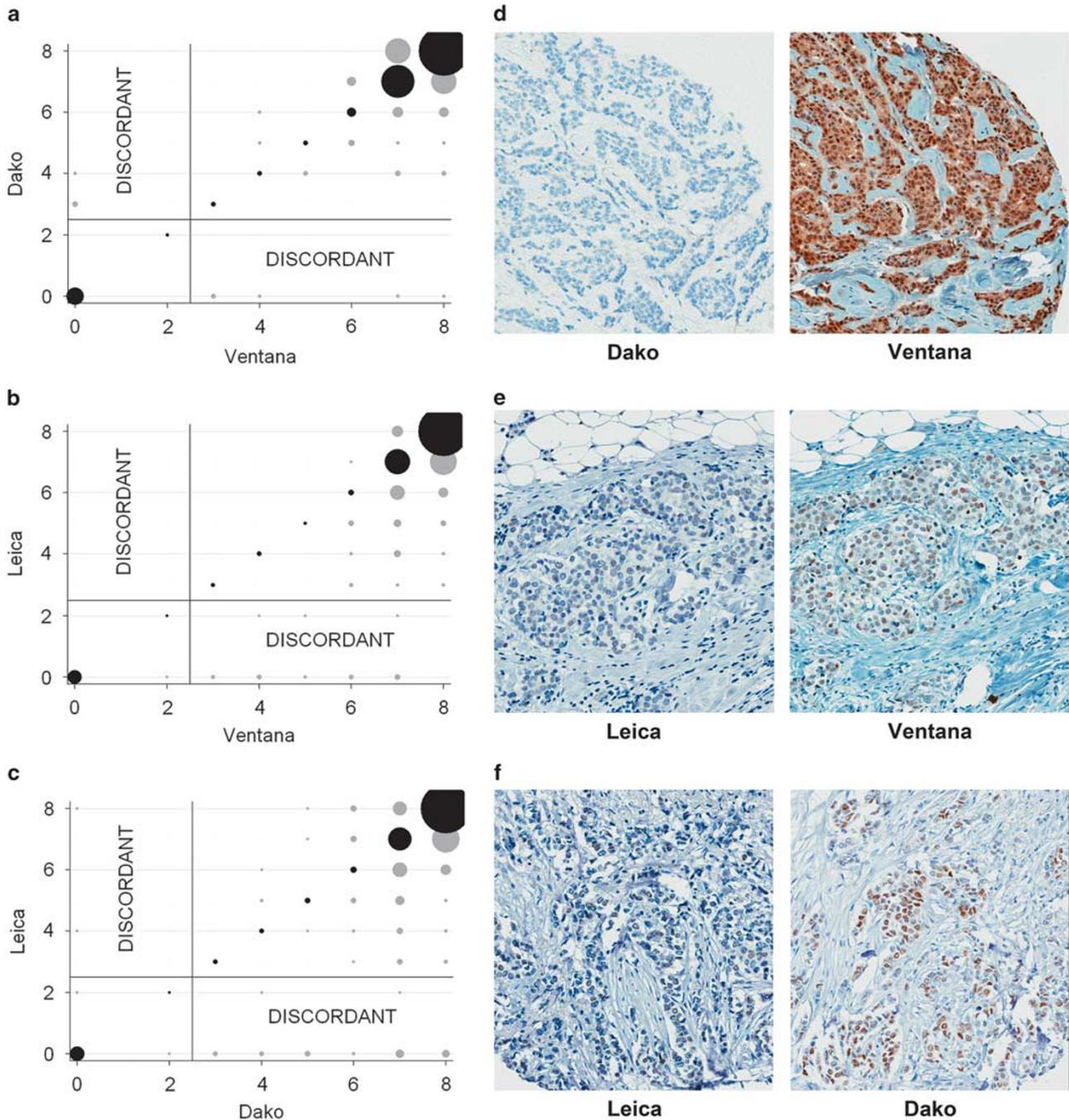


Figure 1 Estrogen receptor immunohistochemical discordance between ready-to-use assays. Figures (a–c) show weighted scatterplots with discordance highlighted by text and perfect concordance in black. (d–f) are representative cases of discordance. (a and d) Dako [1D5 and ER-2-123] and Ventana [SP1]; (b and e) Leica [6F11] and Ventana; and (c and f) Dako and Leica.

multivariate model owing to clinical relevance. Estrogen receptor status determined by all assays maintained significance in the multivariate models.

Measures of Test Performance

Sensitivity, specificity, positive predictive value, negative predictive value, and accuracy were

calculated for all platforms, using the ligand-binding assay results as the reference standard (Table 3). All ready-to-use assays were comparable with the ligand-binding assay in regards to their sensitivity, positive predictive value, and accuracy; however, the Dako assay had the strongest negative predictive value (83.3% compared with 52.4% and 70.0% for Leica and Ventana, respectively) and Ventana showed the weakest specificity (31.8%

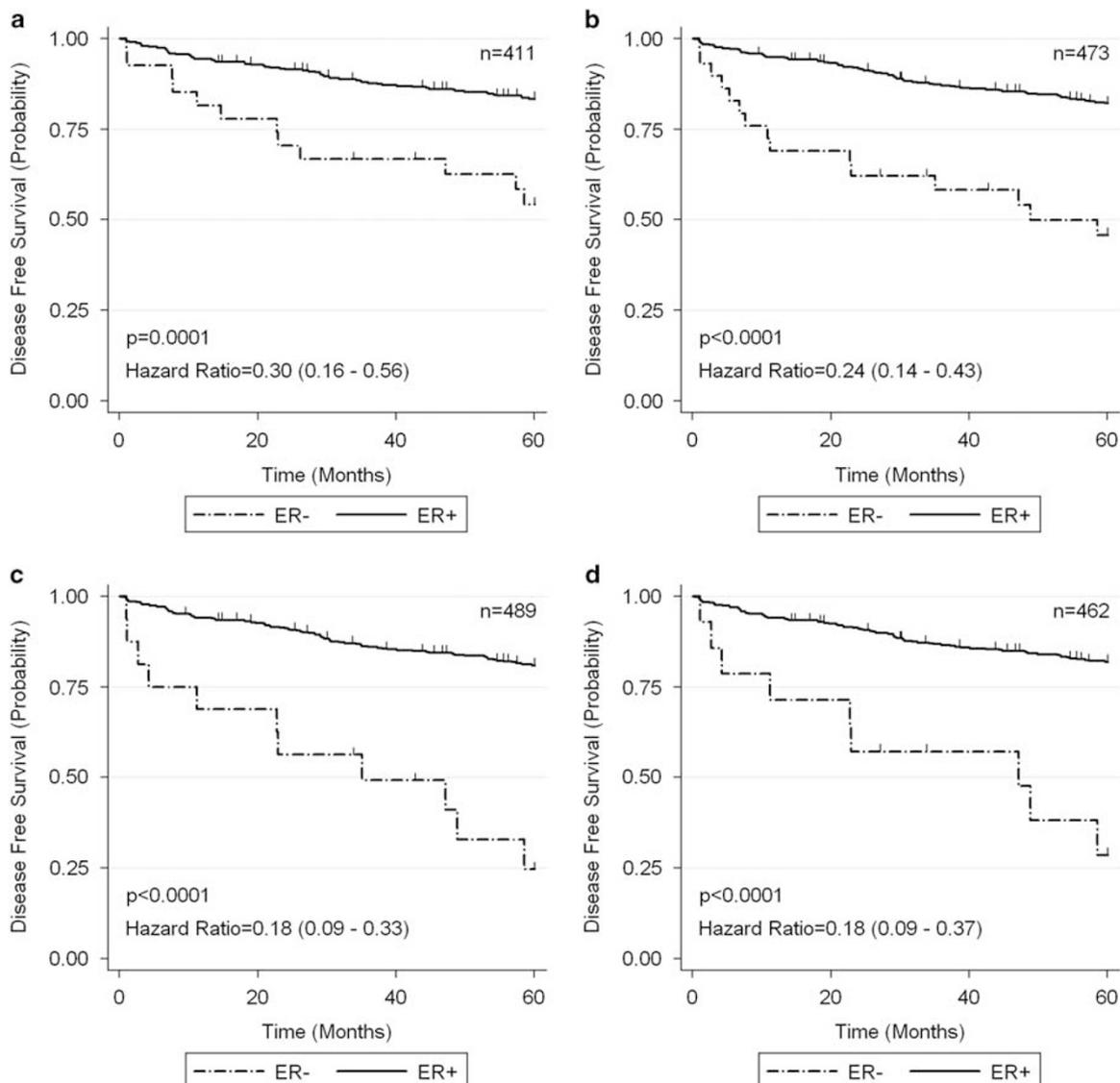


Figure 2 Univariate analysis, including hazard ratios and 95% confidence intervals of estrogen receptor assays: (a) ligand-binding assay; (b) Dako; (c) Leica; and (d) Ventana. Total sample size for univariate analysis of each assay is indicated in the upper right corner. (ER+, Estrogen Receptor Positive; ER-, Estrogen Receptor Negative).

compared with 40.0% and 44.0% for Dako and Leica, respectively).

Measures of test performance were also calculated for the ligand-binding and immunohistochemical assays with disease-free survival at 5 years used as the reference standard (Table 4). All three immunohistochemical assays performed similar to the ligand-binding assay for sensitivity, specificity, positive predictive value, and accuracy. Interestingly, the ligand-binding assay showed the lowest ability for predicting patients who are most likely to recur (local or metastatic) or die from breast cancer (negative predictive value for ligand-binding assay 44.4% vs 51.7–68.8% for the immunohistochemical assays).

Estrogen Receptor-Negative, Progesterone Receptor-Positive Phenotype

Cases identified as estrogen receptor-negative and progesterone receptor-positive by either of the three methods evaluated in this study were identified and results from the other vendors assays, in addition to the ligand-binding assay or original immunohistochemical clinical test results, were compared (Table 5). Leica had the highest rate of estrogen receptor-negative, progesterone receptor-positive cases (1.2%), as compared with the 0.5% seen with the Dako and Ventana assays. Of note, the estrogen receptor-negative, progesterone receptor-positive phenotype was not consistent across the various methods.

Table 2 Cox proportional hazard models for ligand-binding assay (LBA), Dako, Leica, and Ventana ready-to-use immunohistochemical assays

| | LBA | | | Dako | | | Leica | | | Ventana | | |
|--------------------------|------|-------------|---------|------|-------------|---------|-------|-------------|---------|---------|-------------|---------|
| | HR | 95% CI | P-value | HR | 95% CI | P-value | HR | 95% CI | P-value | HR | 95% CI | P-value |
| Estrogen receptor status | 0.31 | (0.15–0.64) | 0.002 | 0.30 | (0.14–0.68) | 0.004 | 0.39 | (0.19–0.78) | 0.008 | 0.26 | (0.11–0.63) | 0.003 |
| Lymph node status | 4.59 | (2.52–8.37) | < 0.001 | 3.70 | (2.23–6.14) | < 0.001 | 3.18 | (1.88–5.37) | < 0.001 | 3.21 | (1.88–5.48) | < 0.001 |
| Tumor grade | 2.5 | (1.34–4.65) | 0.004 | 2.90 | (1.70–4.93) | < 0.001 | 3.15 | (1.84–5.38) | < 0.001 | 2.95 | (1.71–5.08) | < 0.001 |
| Tumor size | 1.54 | (0.81–2.95) | 0.191 | 1.66 | (0.96–2.89) | 0.072 | 1.67 | (0.94–2.98) | 0.083 | 1.71 | (0.95–3.08) | 0.076 |
| HER2 status | 2.08 | (0.64–6.80) | 0.344 | 1.52 | (0.55–4.20) | 0.417 | 1.13 | (0.38–3.33) | 0.823 | 1.16 | (0.39–3.46) | 0.794 |

Table 3 Measures of test performance comparing ready-to-use immunohistochemical assays to the ligand-binding assay

| | DAKO | LEICA | VENTANA |
|-----------------|-------|-------|---------|
| Sensitivity (%) | 99.4% | 97.1% | 99.1% |
| TP | 351 | 329 | 331 |
| TP+FN | 353 | 339 | 334 |
| Specificity (%) | 40.0% | 44.0% | 31.8% |
| TN | 10 | 11 | 7 |
| TN+FP | 25 | 25 | 22 |
| PPV (%) | 95.9% | 95.9% | 95.7% |
| TP | 351 | 329 | 331 |
| TP+FP | 366 | 343 | 346 |
| NPV (%) | 83.3% | 52.4% | 70.0% |
| TN | 10 | 11 | 7 |
| TN+FN | 12 | 21 | 10 |
| Accuracy (%) | 95.5% | 93.4% | 94.9% |
| TP+TN | 361 | 340 | 338 |
| TP+FP+TN+FN | 378 | 364 | 356 |

Abbreviations: FN, false negative; FP, false positive; TN, true negative; TP, true positive.

Table 4 Measure of test performance comparing ligand-binding (LBA) and ready-to-use immunohistochemical assays to 5-year disease-free survival

| | LBA | DAKO | LEICA | VENTANA |
|-----------------|-------|-------|-------|---------|
| Sensitivity (%) | 95.5% | 98.7% | 96.3% | 98.7% |
| TP | 321 | 384 | 366 | 368 |
| TP+FN | 336 | 389 | 380 | 373 |
| Specificity (%) | 16.0% | 11.0% | 16.1% | 10.1% |
| TN | 12 | 11 | 15 | 9 |
| TN+FP | 75 | 100 | 93 | 89 |
| PPV (%) | 83.6% | 81.2% | 82.4% | 82.1% |
| TP | 321 | 384 | 366 | 368 |
| TP+FP | 384 | 473 | 444 | 448 |
| NPV (%) | 44.4% | 68.8% | 51.7% | 64.3% |
| TN | 12 | 11 | 15 | 9 |
| TN+FN | 27 | 16 | 29 | 14 |
| Accuracy (%) | 81.0% | 80.8% | 80.5% | 81.6% |
| TP+TN | 333 | 395 | 381 | 377 |
| TP+FP+TN+FN | 411 | 489 | 473 | 462 |

Abbreviations: FN, false negative; FP, false positive; TN, true negative; TP, true positive.

Discussion

This study presents a systematic comparison of three estrogen receptor ready-to-use immunohistochemical assays using a retrospective, tamoxifen-treated, breast cancer cohort. Opting to use a ready-to-use assay for immunohistochemical testing offers increased reproducibility and standardization, minimizing potential analytical errors that can lead to incorrect test results.^{8,12,23} In particular, ready-to-use antibodies are titrated by the vendor to ensure optimal results. This negates the need for laboratory personnel to perform rigorous titrations for each new antibody lot, and minimizes potential inter- and intra-laboratory variability (lot to lot evaluations should still be performed). Furthermore, the use of autostainers offers a more consistent and reproducible protocol than can be obtained by manual assays. Taken together, this leads to improved reliability and performance, which is extremely important for a clinically utilized assay such as the estrogen receptor, where the results are used to assist in patient management.

Estrogen receptor assays are performed by histotechnicians, whereas estrogen receptor test results are ultimately reviewed and reported by trained

pathologists following established guidelines.⁴ The use of ready-to-use assays provides a more consistent assay performance, which subsequently allows for increased reproducibility in reporting the results between observers. Our study assessed agreement among three observers and found that the Dako and Ventana estrogen receptor immunohistochemical assays had strong agreement ($\kappa \geq 0.80$), while the Leica platform showed substantial inter-observer agreement ($\kappa = 0.67, 0.75$ and 0.83). These results are comparable with other groups that have evaluated inter-observer agreement for estrogen receptor immunohistochemical assays.^{24,25} To our knowledge, ours is the only study that also evaluated intra-observer agreement. Three months after initial review, Observers 1 and 3 rescored the tissue microarray slides and we found strong agreement across all platforms ($\kappa \geq 0.91$).

We also evaluated platform concordance by core to core analysis, and saw that concordance varied between platforms (Figure 1). Interestingly, the highest discordance rates were seen with the Leica assay (Leica and Ventana, $n = 21$; Dako and Leica, $n = 43$), whereas Dako and Ventana had only 9 discordant cases. Furthermore, the Leica ready-to-use assay resulted in a greater number of estrogen

Table 5 All immunohistochemical and ligand-binding assay (LBA) results for estrogen receptor (ER) and progesterone receptor (PgR) for any case identified as estrogen receptor-negative, progesterone receptor-positive by a ready-to-use immunohistochemical assay

| Study ID | LBA | | Original IHC | | Dako | | Leica | | Ventana | |
|-----------|-----|-----|--------------|-----|------|-----|-------|-----|---------|-----|
| | ER | PgR | ER | PgR | ER | PgR | ER | PgR | ER | PgR |
| | 114 | N/A | N/A | 2 | 3 | 1 | 1 | 0 | 1 | 1 |
| 277 | 39 | 236 | N/A | N/A | 1 | 0 | 0 | 0 | 0 | 1 |
| 302 | 213 | 362 | N/A | N/A | 1 | 1 | 0 | 1 | 1 | 1 |
| 421 | N/A | N/A | 0 | 0 | 0 | 1 | 0 | 0 | 0 | 1 |
| 496 | 77 | 149 | N/A | N/A | 1 | 1 | 0 | 1 | N/A | 1 |
| 642 | 0 | 0 | N/A | N/A | 0 | 1 | 0 | 1 | 0 | 0 |
| 660 | 120 | 202 | N/A | N/A | 1 | N/A | 0 | 1 | 1 | 1 |
| ER- /PgR+ | | | | | | | | | | |

receptor-negative cases (~6%), compared with the Dako and Ventana assays (~3%). The higher discordant rates with the Leica assay might be due, in part, to differences in counterstain between the platforms. It was our observation that the Leica assay had a darker overall counterstain making it difficult to notice low intensities of estrogen receptor. All observers who marked the tissue microarrays agreed that the Dako assay was preferred with regards to ease of scoring owing to the lighter counterstain, followed by the Ventana assay. The ready-to-use immunohistochemical assays for each vendor can allow for some flexibility in performing the counterstain, and we recommend that each laboratory optimize the counterstain to minimize the potential for reporting false negatives, which can occur with excessive bluing and/or hematoxylin.

In addition, each platform utilized a different polymer-based detection system for visualization: the Leica platform had more intense brown staining as compared with the Dako and Ventana platforms, whose staining was generally lighter. As we did not look at the assay under different counterstain conditions, it is unknown as to whether the darker appearance of the Leica assay was a result of the detection system or the strong counterstain, or potentially both. Moreover, observer experience with the Leica assay was minimal for two of the three observers, and this may have influenced the overall visual assessment and scoring. Thus, we recommend that pathologists have ample experience and clearly defined guidelines for the clinical assay utilized in their laboratory prior to reporting results in a clinical setting.

Furthermore, the tissue fixation process, while preserving tissue morphology, alters the structure of proteins. Tissues for this study were processed in 10% neutral-buffered formalin, which leads to cross-linking as the formalin reacts with amino acids with available hydrogen sites and forms methylene bridges.²⁶ These cross-links can alter epitope

structure, ultimately affecting antibody binding. This alteration can be reversed using heat-induced epitope retrieval,²⁷ and this technique is affected by pH of the retrieval solution as well as time and temperature.²⁸ All platforms investigated had unique heat-induced epitope retrieval methods optimized for their specific assay; however, this introduces another potential source of discordance.

Finally, all vendors used different antibodies, which is another potential source of discordance in our study. Each antibody recognized a different epitope: 6F11 (Leica) and 1D5 and ER-2-123 (Dako) recognizes the N-terminus, whereas SP1 (Ventana) recognizes the C-terminus.^{5,23,29–32} Specifically, 6F11 and ER-2-123 binds within the sequence of amino acids 15-23, 1D5 within the amino acid region 127-130, and SP1 binds an epitope within the sequence of amino acids 578-595. Despite some discordance between the assays (Figure 1), all assays performed similarly on univariate analysis (Figure 2) and maintained significance in the multivariate models (Table 2), suggesting functional equivalence of the ready-to-use assays.

Our final assessment involved determining standard measures of test performance—sensitivity, specificity, positive predictive value, negative predictive value, and accuracy. All assays had strong sensitivity, positive predictive value, and accuracy when the ligand-binding assay was set as the reference measure (Table 3). Leica had a substantially lower negative predictive value, in accordance with the higher number of estrogen receptor-negative cases seen compared with the Dako or Ventana assays. This has potential clinical significance as use of this assay may result in endocrine treatment not being offered to patients who may yield some benefit. To obtain a better understanding of the clinical relevance of this finding, we repeated the measures of test performance using 5-year disease-free survival as the reference standard. All assays (ligand-binding and the ready-to-use immunohistochemical assays) had high sensitivity, positive predictive value, and accuracy (Table 4). In addition, all had relatively low specificity. This result is not surprising, as it is known that one-third of patients who are estrogen receptor-positive are inherently resistant to endocrine therapy, and furthermore, 30–40% of initial responders develop endocrine-resistant disease.³³ The differences between the assays for negative predictive value are not as dramatic when 5-year disease-free survival is utilized as the reference standard compared with the ligand-binding assay; however, Dako and Ventana notably perform better than the ligand-binding or Leica assay (68.8% and 64.3% vs 44.4% and 51.7%, respectively). The Calgary cohort includes patients who were given tamoxifen despite being hormone-negative, and the negative predictive value results suggests that these assays may miss patients who would benefit from tamoxifen treatment. It would be interesting to investigate whether alternate methods,

such as RT-PCR, for assessing estrogen receptor may improve the negative predictive value, and this warrants further investigation.

The discordance seen between the varying estrogen receptor assays brought forward an important clinical implication for patient management: are we potentially undertreating patients based on current methodology for assessing estrogen receptor status? Alternatively, are we potentially exposing patients who are falsely classified as estrogen receptor-negative to unnecessary toxicity with chemotherapy? To explore this further, we looked at the 5-year disease-free survival status of all cases where estrogen receptor status was classified as negative by at least one of the assays, and conversely was found to be estrogen receptor-positive by an alternate assay. What we discovered was that only 21% of cases had poor outcome at 5 years (data not shown), suggesting that current methods may indeed be missing patients who would benefit from an endocrine treatment. Our sample size for this exploration was small and further studies should be completed; nonetheless, these findings argue that estrogen receptor-negative patients may benefit from retesting their estrogen receptor status using an alternate method.

It is also important to note that cytoplasmic/background staining was noticed in some of the tissue cores (eg, Figure 1d, Ventana), often in highly estrogen receptor-positive cases (Allred = 8). Notes were taken during the scoring process and we subsequently explored whether this was more pronounced in one system over another. We found that cytoplasmic/background staining was consistent between cases across the various assays investigated, and not unique to one specific ready-to-use assay. The potential implications of cytoplasmic/background staining was not explored further and is beyond the scope of this study.

As this was the first study to assess estrogen and progesterone receptor immunohistochemistry using three different ready-to-use assays on the same cohort of patients, we also wanted to further explore the estrogen receptor-negative, progesterone receptor-positive phenotype. We found that incidence of estrogen receptor-negative, progesterone receptor-positive varied across the platforms, with a rate of 0.5% (Dako and Ventana) vs 1.2% (Leica). Previous studies have reported an estrogen receptor-negative, progesterone receptor-positive rate between 1.5 and 10%.^{3,20,34–44} The methods used to determine estrogen and progesterone receptor status varied between these studies, with ligand-binding and laboratory-developed assays predominantly utilized. Of the more recent studies, the range of estrogen receptor-negative, progesterone receptor-positive is 1.5–4.0%; however, almost all of these studies were based on laboratory-developed assays. Interestingly, we found that there was not one case in which estrogen receptor-negative, progesterone receptor-positive expression was consistent across

all three assays investigated (Table 5). This may be a consequence of all antibodies having different recognition areas on the receptor, as previously discussed. There was one study in which no cases of estrogen receptor-negative, progesterone receptor-positive were reported;⁴⁵ this study utilized the 1D5 antibody clone for estrogen receptor and they suggest that estrogen receptor expression using this monoclonal antibody is an ‘all-or-none phenomenon’. Although we also investigated 1D5, it was cocktailed along with ER-2-123 (Dako), and we are unable to deny or verify this statement. More recently, another group that found a rate of estrogen receptor-negative, progesterone receptor-positive of 1.8% in a retrospective review retested the 43 samples in question and found that repeating the testing showed that no cases were confirmed to carry this hormone receptor expression pattern.⁴⁰ 24 samples were found to be estrogen receptor-positive, progesterone receptor-positive; 4 estrogen receptor-positive, progesterone receptor-negative; and 15 estrogen receptor-negative, progesterone receptor-negative after retesting the archival specimens. We found that of the seven cases in our cohort identified as estrogen receptor-negative, progesterone receptor-positive by one of the assays, five were estrogen receptor-positive, progesterone receptor-positive and two cases were estrogen receptor-negative, progesterone receptor-negative by alternate methods. Our results, in conjunction with Maleki *et al*,⁴⁰ suggest that any clinical specimen reported to be estrogen receptor-negative, progesterone receptor-positive should be retested by an alternate assay that utilizes different antibody clones.

This study was based on tissue microarray analysis, which could lead to higher rates of false negatives due to sampling errors, potentially affecting the univariate and multivariate statistical analysis, as well as the measures of test performance. In addition, this was a retrospective analysis, resulting in missing data that reduced the overall sample size for analysis. The highest rate of missing data was 28%; however, most variables had less than 15% (Table 1). Nevertheless, we were able to obtain high quality data that have undergone two separate audits to ensure accuracy and minimize transcriptional and logical errors. Also, the cohort was not a representative sample of all Calgary patients during the time period. Selection was initially based on patient treatment (tamoxifen), and was enriched for patients with poor outcomes and estrogen receptor positivity; however, this selection criteria allows for greater power for detection of response to tamoxifen treatment. Also, this cohort does contain some estrogen receptor-negative patients that received tamoxifen, increasing its utility and value. Furthermore, final tissue microarray construction was based on the availability of archival tissue blocks, and pre-analytical variables could not be accounted for. All cases in this cohort were collected prior to the

development of the ASCO guidelines for controlling pre-analytical variation in fixation.

In conclusion, despite some discordance seen between the vendor-specific estrogen receptor assays, all immunohistochemical assays were comparable with or superior to the ligand-binding assay, validating their continued use in the clinical laboratory. It is still important to remember that adoption of ready-to-use assays still requires verification in accordance to current ASCO/CAP guidelines,^{4,46} and furthermore, that counterstain should be optimized for each assay. Optimization of counterstaining is not addressed in current guidelines, and should be performed by the laboratory in collaboration with site pathologists. In addition, involvement in high quality proficiency testing programs, such as UK-NEQAS, which evaluate the counterstaining of the assay, can assist in ensuring optimal counterstaining appropriate for a given assay.

Acknowledgments

We would like to acknowledge Calgary Laboratory Services for financial support in the construction of the Calgary cohort, as well as Dr Annie Yau and Ms Mie Konno for their assistance in the creation of the database. We are grateful to the Breast Cancer Society of Canada for their financial support of this project. In addition, we would like to thank all the vendors—Dako, Leica, and Ventana—for providing us with free reagents to complete this study. All vendors were aware of the study objectives and the participation of their competitors' involvement.

Disclosure/conflict of interest

ACK was provided reimbursement from Leica for attending the Leica Breast Cancer Symposia in London, UK April 2013. AMM has the following disclosures: Consultant and grant recipient (Ventana Medical Systems). The remaining authors declare no conflict of interest. We would also like to disclose that a portion of these results have been presented in poster format at the 2012 San Antonio Breast Cancer Symposium (P1-07-10) as well as at the 2013 Leica Breast Cancer Symposia in London, UK.

References

- 1 EBCTC Group. Effects of chemotherapy and hormonal therapy for early breast cancer on recurrence and 15-year survival: an overview of the randomised trials. *Lancet* 2005;365:1687–1717.
- 2 EBCTC Group. Systemic treatment of early breast cancer by hormonal, cytotoxic, or immune therapy. *Lancet* 1992;339:1–15.
- 3 EBCTC Group. Tamoxifen for early breast cancer: an overview of the randomised trials. *Lancet* 1998;351:1451–1467.

- 4 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.
- 5 Harvey JM, Clark GM, Osborne CK, Allred DC. 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.
- 6 Allred DC. Commentary: Hormone receptor testing in breast cancer: a distress signal from Canada. *Oncologist* 2008;13:1134–1136.
- 7 Hede K. Breast cancer testing scandal shines spotlight on black box of clinical laboratory testing. *J Natl Cancer Inst* 2008;100:836–837.
- 8 Taylor CR. The total test approach to standardization of immunohistochemistry. *Arch Pathol Lab Med* 2000;124:945–951.
- 9 Taylor CR. Focus on biospecimens: the issue is the tissue. Observations on the "NCI-NIST fitness-for-purpose quality assessment and standards development workshop," October 2010. *Appl Immunohistochem Mol Morphol* 2011;19:95–98.
- 10 Sompuram SR, Vani K, Messana E, Bogen SA. A molecular mechanism of formalin fixation and antigen retrieval. *Am J Clin Pathol* 2004;121:190–199.
- 11 Gown AM. Current issues in ER and HER2 testing by IHC in breast cancer. *Mod Pathol* 2008;21:S8–S15.
- 12 Yaziji H, Taylor CR, Phil D *et al*. Consensus recommendations on estrogen receptor testing in breast cancer by immunohistochemistry. *Appl Immunohistochem Mol Morphol* 2008;16:513–520.
- 13 Pekmezci M, Szpaderska A, Osipo C *et al*. The effect of cold ischemia time and/or formalin fixation on estrogen receptor, progesterone receptor, and human epidermal growth factor receptor-2 results in breast carcinoma. *Patholog Res Int* 2012;2012:947041.
- 14 Tong LC, Nelson N, Tsourigiannis J *et al*. The effect of prolonged fixation on the immunohistochemical evaluation of estrogen receptor, progesterone receptor, and HER2 expression in invasive breast cancer: a prospective study. *Am J Surg Pathol* 2011;35:545–552.
- 15 Yildiz-Aktas IZ, Dabbs DJ, Bhargava R. The effect of cold ischemic time on the immunohistochemical evaluation of estrogen receptor, progesterone receptor, and HER2 expression in invasive breast carcinoma. *Mod Pathol* 2012;25:1098–1105.
- 16 Fitzgibbons PL, Bradley LA, Fatheree LA *et al*. Principles of analytic validation of immunohistochemical assays: guidelines from the College of American Pathologists pathology and laboratory quality center. *Arch Pathol Lab Med* 2014;138:1432–1443.
- 17 Remotti H. Tissue microarrays: construction and use. *Methods Mol Biol* 2013;980:13–28.
- 18 Molnar LA, Dhirani N, Tran K *et al*. Immunoassay for estrogen receptor does not detect inactivated receptor. *Clin Chem* 1989;35:824–828.
- 19 Wolff AC, Hammond ME, Schwartz JN *et al*. American Society of Clinical Oncology/College of American Pathologists guideline recommendations for human epidermal growth factor receptor 2 testing in breast cancer. *J Clin Oncol* 2007;25:118–145.
- 20 Allred DC, Harvey JM, Berardo M *et al*. Prognostic and predictive factors in breast cancer by immunohistochemical analysis. *Mod Pathol* 1998;11:155–168.

- 21 Viera AJ, Garrett JM. Understanding interobserver agreement: the kappa statistic. *Fam Med* 2005;37:360–363.
- 22 Akobeng AK. Understanding diagnostic tests 1: sensitivity, specificity and predictive values. *Acta Paediatr* 2007;96:338–341.
- 23 Phillips T, Murray G, Wakamiya K *et al*. Development of standard estrogen and progesterone receptor immunohistochemical assays for selection of patients for antihormonal therapy. *Appl Immunohistochem Mol Morphol* 2007;15:325–331.
- 24 Cohen DA, Dabbs DJ, Cooper KL *et al*. Interobserver agreement among pathologists for semiquantitative hormone receptor scoring in breast carcinoma. *Am J Clin Pathol* 2012;138:796–802.
- 25 Reisenbichler ES, Lester SC, Richardson AL *et al*. Interobserver concordance in implementing the 2010 ASCO/CAP recommendations for reporting ER in breast carcinomas. *Am J Clin Pathol* 2013;140:487–494.
- 26 Miller RT, Swanson PE, Wick MR. Fixation and epitope retrieval in diagnostic immunohistochemistry: a concise review with practical considerations. *Appl Immunohistochem Mol Morphol* 2000;8:228–235.
- 27 Shi S-R, Key ME, Kalra KL. Antigen retrieval in formalin-fixed, paraffin-embedded tissues: an enhancement method for immunohistochemical staining based on microwave oven heating of tissue sections. *J Histochem Cytochem* 1991;39:741–748.
- 28 Gown AM. Unmasking the mysteries of antigen or epitope retrieval and formalin fixation. *Am J Clin Pathol* 2004;121:172–174.
- 29 Xia H, Huang Z. High affinity Monoclonal Antibody for Recognizing the Estrogen Receptor (ER) and Method for Creating the Antibody. Spring Bioscience, assignee: Fremont, CA (US), 2006.
- 30 Dako., Dako ER/PR pharmDx kit (Link), 2009.
- 31 Bevitt DJ, Milton ID, Piggot N *et al*. New monoclonal antibodies to oestrogen and progesterone receptors effective for paraffin section immunohistochemistry. *J Pathol* 1997;183:228–232.
- 32 Kaplan PA, Frazier SR, Loy TS *et al*. 1D5 and 6F11: an immunohistochemical comparison of two monoclonal antibodies for the evaluation of estrogen receptor status in primary breast carcinoma. *Anat Pathol* 2005;123:276–280.
- 33 Nicholson RI, Johnston SR. Endocrine therapy - current benefits and limitations. *Breast Cancer Res Treat* 2005;93:3–10.
- 34 Bardou V-J, Arpino G, Elledge RM *et al*. Progesterone receptor status significantly improves outcome prediction over estrogen receptor status alone for adjuvant endocrine therapy in two large breast cancer databases. *J Clin Oncol* 2003;21:1973–1979.
- 35 Bauer K, Parise C, Caggiano V. Use of ER/PR/HER2 subtypes in conjunction with the 2007 St Gallen consensus statement for early breast cancer. *BMC Cancer* 2010;10:228.
- 36 Colomer R, Beltran M, Dorcas J *et al*. It is not time to stop progesterone receptor testing in breast cancer. *J Clin Oncol* 2005;23:3868–3869.
- 37 Dowsett M, Cuzick J, Wale C *et al*. Retrospective analysis of time to recurrence in the ATAC trial according to hormone receptor status: an hypothesis-generating study. *J Clin Oncol* 2005;23:7512–7517.
- 38 Grann VR, Troxel AB, Zojwalla NJ *et al*. Hormone receptor status and survival in a population-based cohort of patients with breast carcinoma. *Cancer* 2005;103:2241–2251.
- 39 Liu S, Chia SK, Mehl E *et al*. Progesterone receptor is a significant factor associated with clinical outcomes and effect of adjuvant tamoxifen therapy in breast cancer patients. *Breast Cancer Res Treat* 2010;119:53–61.
- 40 Maleki Z, Shariat S, Mokri M *et al*. ER-negative/PR-positive breast carcinomas or technical artifacts in immunohistochemistry? *Arch Iran Med* 2010;15:366–369.
- 41 Osborne CK, Yochmowitz MG, Knight WA *et al*. The value of estrogen and progesterone receptors in the treatment of breast cancer. *Cancer* 1980;46:2884–2888.
- 42 Tesch M, Shawwa A, Henderson R. Immunohistochemical determination of estrogen and progesterone receptor status in breast cancer. *Am J Clin Pathol* 1993;99:8–12.
- 43 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.
- 44 Viale G, Regan MM, Maiorano E *et al*. Chemoendocrine compared with endocrine adjuvant therapies for node-negative breast cancer: predictive value of centrally reviewed expression of estrogen and progesterone receptors - international breast cancer study group. *J Clin Oncol* 2008;26:1404–1410.
- 45 Nadji M, Gomez-Fernandez C, Ganjei-Azar P *et al*. Immunohistochemistry of estrogen and progesterone receptors reconsidered: experience with 5,993 breast cancers. *Am J Clin Pathol* 2005;123:21–27.
- 46 Hammond ME. ASCO-CAP guidelines for breast predictive factor testing: an update. *Appl Immunohistochem Mol Morphol* 2011;19:499–500.