OPEN

New Robust and Reproducible Stereological IHC Ki67 Breast Cancer Proliferative Assessment to Replace Traditional Biased Labeling Index

Gilbert Bigras, MD, PhD, Wei-Feng Dong, MD, PhD,* Sarah Canil, BSc,* Judith Hugh, MD,† Richard Berendt, MD,* George Wood, MD,* and Hua Yang, MD, PhD‡*

Abstract: There is a pressing need for an objective decision tool to guide therapy for breast cancer patients that are estrogen receptor positive and HER2/neu negative. This subset of patients contains a mixture of luminal A and B tumors with good and bad outcomes, respectively. The 2 main current tools are on the basis of immunohistochemistry (IHC) or gene expression, both of which rely on the expression of distinct molecular groups that reflect hormone receptors, HER2/neu status, and most importantly, proliferation. Despite the success of a proprietary molecular test, definitive superiority of any method has not yet been demonstrated. Ki67 IHC scoring assessments have been shown to be poorly reproducible, whereas molecular testing is costly with a longer turnaround time. This work proposes an objective Ki67 index using image analysis that addresses the existing methodological issues of Ki67 quantitation using IHC on paraffin-embedded tissue. Intrinsic bias related to numerical assessment performed on IHC is discussed as well as the sampling issue related to the "peel effect" of tiny objects within a thin section. A new nonbiased stereological parameter (V_V) based on the Cavalieri method is suggested for use on a double-stained Ki67/cytokeratin IHC slide. The assessment is performed with open-source ImageJ software with interobserver concordance between 3 pathologists being high at 93.5%. Furthermore, V_V was found to be a superior method to predict an outcome in a small subset of breast cancer patients when compared with other image analysis methods being used to determine the Ki67 labeling index. Calibration methodology is also discussed to further this IHC approach.

Key Words: MIB-1, stereology, reproducibility, breast cancer, image analysis

(*Appl Immunohistochem Mol Morphol* 2017;25:687–695)

Breast cancers (BCs) can be divided into 4 major molecular classes as follows: luminal A, luminal B, basal-like, and HER2. This subtyping has facilitated significant clinical progress in determining prognosis and appropriate therapy.[1]

The identification of basal-like and HER2 subtypes can be fairly accurately determined with current immunohistochemical (IHC) tests for estrogen receptor (ER), progesterone receptor (PR), and HER2 provided there is good laboratory practice in place.[2] However, distinction between the luminal A and luminal B subtypes has proven problematic: luminal B BCs, when compared with luminal A, are reported[3] to have lower expression of hormone receptors, higher proliferation, and are associated with a worse prognosis. Chemotherapy has been shown to have no benefit in luminal A BC patients, whereas the benefit of chemoendocrine therapies compared with endocrine therapy alone is significant for luminal B patients.[4]

Proliferative assessment can be performed with Ki67 IHC[5] as this protein's expression is limited to cells in cell cycle. Since the introduction of the Ki67 antibody, pathologists have attempted to visually quantitate tumor proliferation by estimating the percentage of positive tumor cells. A low versus high Ki67 score is used to distinguish luminal A BC from luminal B BC, respectively. Unfortunately, numerous publications[6–9] have reported low interobserver and intraobserver reproducibility of visual Ki67 assessments. This variability is attributed to several factors including pathologists' divergent definitions of what constitutes a positive nucleus, the mode of assessment (counting vs. "eyeballing"), and the selection of the area of the tumor to be evaluated. One recent international study[10] has attempted to standardize visual assessment of Ki67 by providing instructions on staining thresholds and a prescribed scoring pattern for determination of the percentage of stained tumor cells scored by

Received for publication December 10, 2015; accepted March 1, 2016.

From the *Department of Laboratory Medicine and Pathology, Cross Cancer Institute; †Department of Laboratory Medicine and Pathology, University of Alberta, Edmonton; and ‡Department of Laboratory Medicine and Pathology, University of Calgary, Calgary, AB, Canada.

G.B. and J.H. wrote the main manuscript text and all reviewed it. G.B. prepared the figures. S.C. optimized double stain IHC. G.B., W.-F.D., and H.Y. performed the reproducibility analysis on patient set #1. G.B. performed the outcome predictor analysis on patient set #2. Software available at https://github.com/gilbertbigras/Ki67.

The authors declare no conflict of interest.

Reprints: Gilbert Bigras, MD, PhD, Edmonton IHC Laboratory, Department of Laboratory Medicine and Pathology, Cross Cancer Institute, University of Alberta, 11560 University Ave NW, Edmonton, AB, Canada T6G 1Z2 (e-mail: gilbert.bigras@albertahealth services.ca).

different participants. Despite some progress, especially when compared with a previous study performed by the same group, the authors showed a significant limitation when they identified irreducible subjectivity when assessors scored faint levels of nuclear staining. An alternate methodology of visual assessment called "Eye 5,"[11] based on 5-grade ordinal scale, has recently been proposed. The authors reported a positive correlation with the labeling index (LI), rapid assessment, and good intraobserver and interobserver variability. More recently another study utilizing a 10-grade ordinal scale reported excellent concordance between observers.[7] Reducing proliferative assessment from interval to ordinal scales of measurement might improve interobserver agreement; however, demonstration of improved accuracy is still lacking.

Image analysis (IA) might theoretically eliminate interobserver variation even when using interval scale of measurement. It also significantly reduces the amount of time required to assess samples as compared with visual inspection. Numerous proprietary[12–15] and open-source software programs are available for this purpose. Immunoratio,[16] for example, is an open-source software based on ImageJ,[17] another open-source software available as either a stand-alone or Web application. Immunoratio has shown correlation with a visual LI assessment using a median Ki67 LI of 20% on a standard Ki67 IHC stain using the chromogen 3,3′-diaminobenzidine (DAB) and hematoxylin as the counterstain. Although Immunoratio obtained a hazard ratio of 2.2 in survival analysis, like all algorithms using standard IHC preparations, it is disadvantaged by the observation of lower Ki67 LI values purportedly because of normal stromal and inflammatory cells that artificially increase the denominator in LI assessments.[18] These cells cannot easily be eliminated from the analysis.

This problem of contaminating normal cells can be addressed by targeting the Ki67 assessment to the epithelial compartment by using a double stain combining Ki67 and cytokeratin. This was recently investigated by Nielsen et al[19] who suggested that such a double stain could increase the accuracy of Ki67 LI in BC. However, the authors reported that Ki67-negative nuclei were difficult to separate in the double stain because they often appeared grayish. Consequently, they were unable to calculate a regular index on the basis of the number of positive and negative nuclei within the epithelial component and suggested that the area of the lesion could replace the total number of malignant nuclei. We agree with this approach but suggest an additional technical improvement for the reasons outlined below.

LI, reported as percentage of Ki67-positive nuclei out of the total number of malignant nuclei, is a mainstream in pathology practice. This approach is straightforward in cytology and flow cytometry, where the entire cell is analyzed, but becomes complex and problematic in histologic thin sections. The estimation of the number of particles per area (N_A) is biased for 2 reasons, the first of which is specific to the process of creating a thin section and we will refer to it as a "peel" effect and the second is a sampling bias because the probability for a particle (nucleus) to be sampled is proportional to its size. With respect to the "peel" effect, nuclei being tiny structures embedded in a matrix may be "peeled" at one of their extremity when a section is cut. Both the human observer and a computer algorithm might "decide" to include or ignore these tiny structures which, as a consequence, create significant variation for "N." Regarding the sampling bias, larger particles have a greater chance of being found in the thin 2-D histologic section than small ones.[20,21] Arbitrary particles of any size and shape have similar probability to be sampled only with a 3-dimensional probe. The latter can be achieved "simply" by using a 3-dimensional reference such as disector.[20–22] Disector is a stereological method that counts a particle in an imaginary 3-dimensional volume (N_V) created by a given and known distance between 2 adjacent histologic sections. It is capable of providing an assessment that is accurate to a third of a cell diameter. N_V provided by the disector method has been mainly used in neurological science and its accuracy has been verified[23] with serial reconstruction of tissue volume. However, in a routine clinical practice, obtaining serial and parallel IHC sections is impractical.

For these reasons, we developed a method to obtain an unbiased value directly from a single thin 2-D histologic section, allowing the replacement of the particle count variable "N." This value is the fraction volume V_V, where V is the volume of proliferating nuclei and v is the volume of tumor. V_V assessment is straightforward[24] and based on the Cavalieri method where V_V is directly assessed from A_A, where A represents the total area of proliferating nuclei against $_A$ that represents the total area of tumor in one 2-D histologic section assuming random sampling. A_A is a nonbiased estimator of Vv and assessment can be obtained from a double-stained IHC slide targeting both the Ki67 nuclear protein and a cytoplasmic cytokeratin protein, which demarcates the epithelial compartment. No counterstain is used in this method to increase the precision of segmentation by limiting the number of chromogens (Fig. 1).

This paper focuses on exploring this new V_V method. Three experiments were conducted demonstrating: (1) the "peel effect" (referring to the inferior robustness of N_A compared with V_V) through computational simulations; (2) reproducibility of V_V among 3 different operators, and (3) validity of V_V as an outcome predictor in ER-positive BC patients. V_V was assessed using a double-stained Ki67-cytokeratin method on BC patients with known outcome and compared with another IA method on a conventional single chromogen Ki67 preparation.

METHODS

Peel Effect and N_A Versus V_V Robustness Assessment

Thresholding an 8-bit (256 gray-level) picture is a commonly used algorithm to obtain a 1-bit (binary) picture. The threshold can be established either empirically

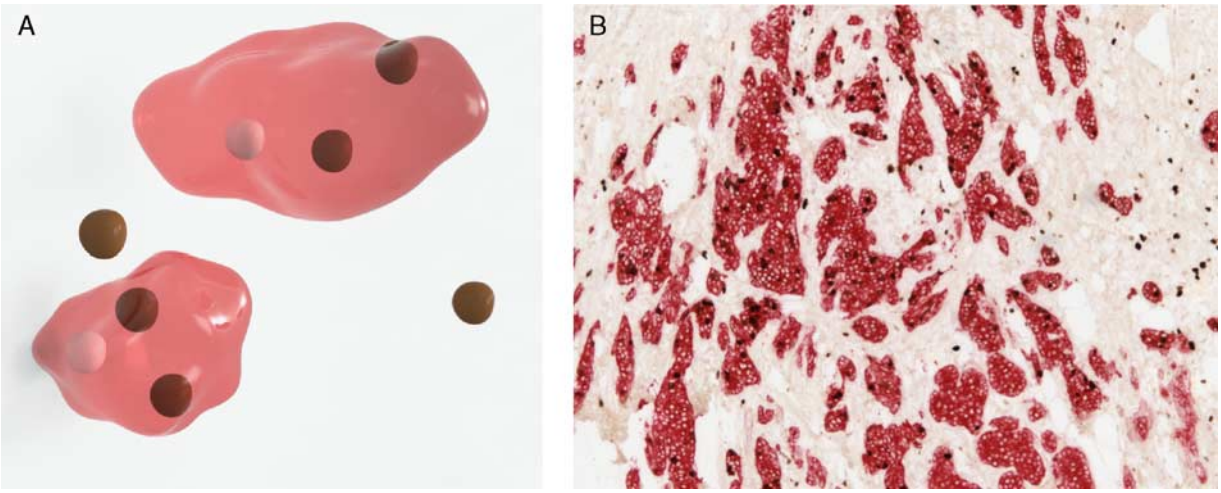

FIGURE 1. A, Representation of V_V where (V) represents the volume of nuclei (brown) involved in cell cycle within the epithelial volume in red ($_V$). B, Representation of A_A where (A) represents the area of nuclei (brown) involved in cell cycle within the epithelial area in red ($_A$).

by 1 human operator or by using an automatic algorithm on the basis of histogram analysis. In either case, the threshold is set to obtain the best representation of biological structures such as nuclei to ensure optimal extraction of features such as diameter, area, and number. To test the impact of threshold variation on the number of nuclei, a computer simulation was performed on a picture illustrating positive Ki67 nuclei. An initial threshold value T was obtained from the IsoData[25] algorithm (which is the default algorithm in ImageJ) and used to create 1 binary picture. From the latter, the total area A of positive Ki67 nuclei is computed. Thereafter, the usual watershed[26] algorithm is applied to separate adjoining nuclei before counting them (N) with the "analyze particles" ImageJ facility. At the end, one obtains from the same binary picture the following 3 values: (1) A (total area of all Ki67 nuclei), (2) N (the number of nuclei), and (3) $_A$ (area of the picture). From these values one can compute V_V (which is directly proportional to A_A) and N_A. Then V_V, A_A, and N_A are recalculated after changing the threshold value in a given range around the initial T. The goal is to observe the magnitude of variation of both A and N and consequently robustness of N_A and V_V. The whole simulation is performed using macro scripting facility of ImageJ. ImageJ[17] is a public domain, Java-based image-processing program developed at the National Institutes of Health.

Double Stain Ki67-Pan Cytokeratin Preparation and Ki67 Vv Software

Ki67-pan cytokeratin staining was performed on the Ventana XT (Ventana Medical Systems Inc., Tucson, AZ) using mild CC1 antigen retrieval on a 4 μm section of formalin-fixed paraffin-embedded tissue and then stained using the Ultraview dual-stain method. Ki67 mouse monoclonal from Dako (MIB-1, M7240) was applied using a 1/50 dilution with Ventana Ultraview DAB chromogen detection. Then, a pan cytokeratin cocktail containing AE1/AE3 (DAKO, AE1/AE3, M3515) at 1/100 dilution and LMK (Leica, CK8/18, NCL-5D3) at a 1/50 dilution was applied and stained with the Ventana UltraView Red kit. All dilutions were made using Dako Envision Flex Antibody Diluent (K8006). No counterstain was utilized for this dual-stain method.

In-house developed software was used to compute V_V values of positive Ki67 nuclei (numerator V) versus volume of the tumor (denominator $_V$). The software was developed within the ImageJ environment using a combination of macro script language and available plugins. The initially acquired double stain picture (Fig. 2A) has a color deconvolution algorithm[27,28] applied to it. Color deconvolution permits one to separate colors to obtain separate Ki67 and cytokeratin information. Additional transformations, such as the Hue-Saturation-Brightness, are performed from the original picture. The operator is then requested to empirically threshold an 8-bit representation of both the tumor and positive Ki67 nuclei to obtain representative masks, as illustrated in Figures 2B, C. Figure 2D shows the final step where the total area (A) of Ki67-positive nuclei (green area) found inside the tumoral mask (black) is calculated. Tumoral mask area (A) is assessed and the ratio $A_A = V_V$ is reported. Intermediate steps of the algorithm also involve a hole-filling process (holes generated by Ki67-negative nuclei). The software to assess V_V is available online at https://github.com/gilbertbigras/Ki67.

Conventional Ki67 Preparations

Formalin-fixed paraffin-embedded (4 μm) sections were stained with Ki67 mouse monoclonal (DAKO, MIB-1, and M7240) at a 1/50 dilution after retrieval with mild CC1 on the Ventana XT. A hematoxylin counterstain was utilized. For each sample, Immunoratio software was utilized to compute a LI.

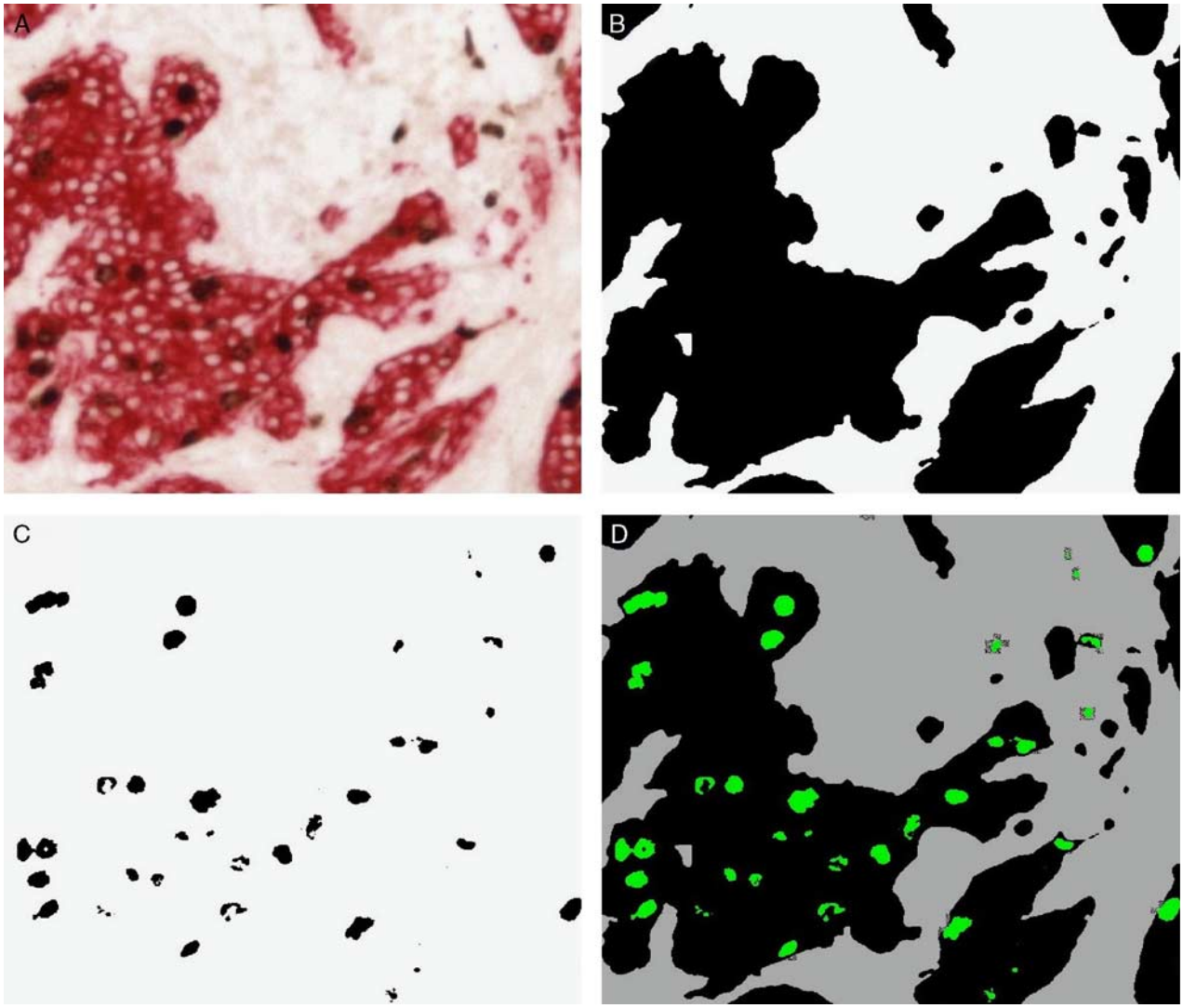

FIGURE 2. (A) Original double stain cytokeratin (red) Ki67 (brown)—without counterstain. (B) Tumor mask. (C) Ki67-positive mask. (D) Superimposition of masks permitting V_V assessment; nuclei found outside the tumoral mask are excluded from the calculation. full color online

Sample Selection

Two separate sets of samples of BC were collected. The first set being used to investigate the reproducibility of the assessment among observers. It included 115 consecutive cases of invasive BC collected between January 2014 and June 2014. As the aim of this set of samples was to see whether concordance could be demonstrated among different observers, clinical data were not collected on this subset. The second sets of samples were pulled from a patient database with known outcomes collected in the Edmonton area between September 1991 and June 1993. This subset consisted of 94 patients with clinical characteristics detailed in Table 1, with either a good or poor outcome. For this study, poor outcome was defined as recurrence or death because of BC. Previously stained Ki67 IHC slides were the only material available on this second subset of patients. Images of the Ki67 stain were obtained for Immunoratio analysis. These slides were then reprocessed by simply removing the coverslip and staining it with the pan cytokeratin cocktail using the UltraView Red detection system on the Ventana XT with no antigen retrieval.

Ki67 Vv and Reproducibility Performance

Ki67 Vv was assessed on double-stained slides on 115 patients by 3 different pathologists (G.B., W.-F.D., and H.Y.). The 3 pathologists followed the same assessment protocol and used the same software. Each pathologist had to acquire three $\times 10$ microscopic fields which represented, according to each pathologist, the highest proliferative area. Each acquired image had to be processed with direct

TABLE 1. Clinical Characteristics of the Breast Cancer Cases With Known Outcome (n = 94)

	n (%)	
	Good Outcome (69)	**Bad Outcome (25)**
Age mean	59.7	55.8
< 50	19 (28)	9 (36)
≥ 50	50 (72)	16 (64)
Carcinoma subtype		
Ductal NOS	58 (84)	23 (92)
Ductal mucinous	4 (5.7)	0
Ductal tubular	3 (4.3)	0
Lobular	4 (5.7)	2 (8)
Tumor size average (cm)	1.58	1.8
Angiolymphatic invasion	15 (21.7)	9 (36)
MBR grade		
I	29 (42)	1 (4)
II	29 (42)	14 (56)
III	11 (15.9)	10 (40)
Allred score ER (mean)	7.5/8	7.48/8
Positive	69 (100)	25 (100)
Negative	0	0
Allred score PR (mean)	5.17/8	5.7/8
Positive	55 (80)	20 (80)
Negative	14 (20)	5 (20)
Her2		
Positive	0	0
Negative	69 (100)	25 (100)
Chemotherapy	6 (9)	5 (20)
Radiotherapy	37 (53)	12 (48)
Hormone therapy	18 (26)	10 (40)
Ki67 Immunoratio	4.3	6.88
Ki67 V_V	4.1	8.4

ER indicates estrogen receptor; MBR, modified Bloom–Richardson; NOS, not otherwise specified; PR, progesterone receptor.

thresholding supervision to obtain representative tumor and nuclear masks as described in Figure 2.

Image Acquisition Performed on First Set (V_V Reproducibility Performance)

Each of the 3 pathologists (G.B., W.-F.D., and H.Y.) digitized 3 fields on each of 115 double-stained slide at ×10 objective. G.B. utilized a Nikon Eclipse E600 microscope, 0.25 aperture, whereas W.-F.D. and H.Y. utilized Nikon Eclipse 80i microscope, 0.25 aperture (Nikon Instruments Inc., Melville, NY). All 3 pathologists used the same QImaging Micropublisher 5.0 RTV camera (QImaging Corp., Surrey, BC) equipped with Sony ICX282 progressive scan interline CCD producing 24-bit color pictures with a resolution of 2560 × 1920 pixels. A priori background correction[29] was applied using the ImageJ image processing software (US National Institutes of Health, Bethesda, MD).

Image Acquisition Performed on Second Set (V_V and Immunoratio as Outcome Predictors)

One pathologist (G.B.) digitized 3 fields on each of the 94 conventional Ki67 slide at ×20 objective for Immunoratio analysis to obtain an LI. Thereafter, the same operator digitized 3 fields on each of the 94 subsequently double-stained Ki67-cytokeratin slide at ×10 objective for V_V. The same equipment was utilized for image acquisition on all study sets. In addition to the V_V assessment, alternative formulations of V_V were tested to explore for their effect on accuracy. These modifications were as follows: (i) including all positive nuclei in the calculation of area for the numerator; (ii) including all positive nuclei in the calculation of area for the numerator and using the whole microscopic field area instead of tumor area; and (iii) using the whole microscopic field area instead of tumor area.

Statistical Analysis

R language[30] (version 3.2.2; R Foundation, Vienna, Austria) was used for statistical analysis and creation of several figures. Concordance among observers was tested using Kendall's W coefficient of concordance.[31] Receiver operating characteristic (ROC) and accuracy performance of V_V and Immunoratio were computed using the ROCR package.[32] A P-value of ≤ 0.05 was selected as the level of significance in all analyses. Figures 2 and 3 were created with the FigureJ plugin.[33] This study was approved by the Alberta Cancer Research Ethics Committee (project #25,861.00).

RESULTS

Peel Effect and N_A Versus V_V Robustness Assessment

Figure 3 illustrates the transformation of an 8-bits picture representing Ki67-positive nuclei from a binary picture (using a threshold of 167) to a watershed picture with random colors as a label of individual nuclei. The same transformation was repeated 38 times using threshold values ranging from 151 to 188 as illustrated in Figure 4, which plots the variation in percentage of A (total area) and N (number of particles, nuclei) in the watershed image. Although both A and N increase the former shows stable and slow progression (from 2% to 3%), whereas the latter is erratic (between 0% and 7%).

Reproducibility of V_V Among 3 Pathologists Assessing the Same Slides (115 Patients)

The Kendall's W coefficient of concordance among the 3 pathologists was 93.5% ($P = 2.06^{-21}$). As perfect concordance is 100%, this value indicates a high degree of concordance as illustrated in Figure 5 where a linear projection in space (forming a tight cone) can be seen. The average time spent per case ranged between 3 and 5 minutes. Very high and very low levels of cytokeratin staining intensity slowed the analysis because of the requirement for operator thresholding to exclude artifactual Ki67 positivity or exclusion of the epithelial compartment, respectively. High cytokeratin staining was seen in some well-differentiated tumors, whereas low cytokeratin staining was seen in some poorly differentiated tumors.

V_V and Immunoratio as Outcome Predictors (94 Patients)

Table 1 summarizes the clinicopathologic data for the good and poor outcome patients. As these patients

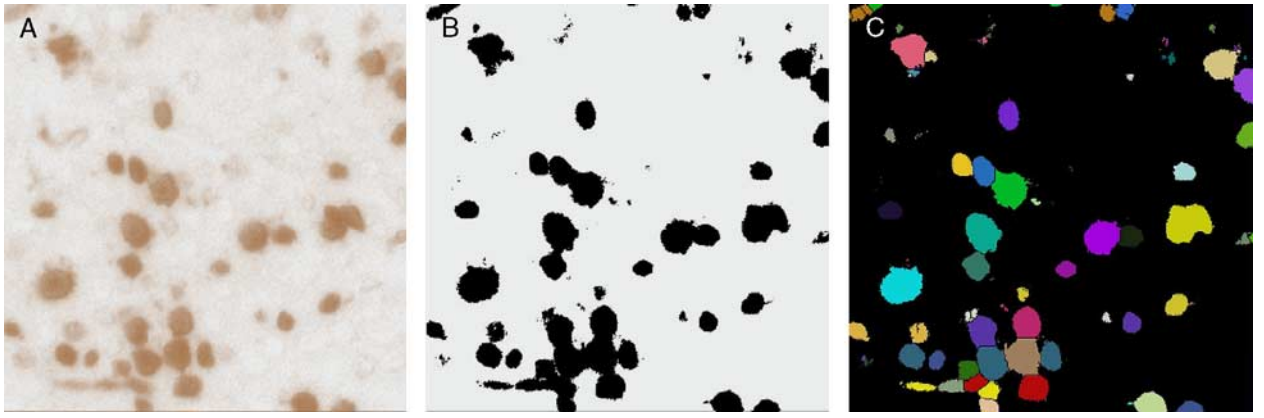

FIGURE 3. Algorithmic transformation from 8 bits (A) to binary picture (B) providing A and watershed (C) providing N. full color online

were managed almost 25 years ago, adjuvant treatment regimens differ from today's standard. Figure 6 shows distribution of all results according to the known outcome (red circle, poor outcome; gray circle, good outcome). Because the Ki67 data are skewed with a large number of results being very low, the raw data were transformed using a logarithmic function and thereafter normalized from 0 to 1. The very first strip chart (A) illustrates the V_V results, whereas the last one (E) is the Immunoratio results. Visually both achieve some segregation between good and poor outcomes. This segregation appears more obvious for V_V with no poor outcomes found between 0.0 and 0.2. The 3 additional strip charts are numerator and/or denominator variations of V_V as follows:

Strip chart (B): numerator includes all nuclei positive for Ki67.

Strip chart (C): numerator includes all nuclei positive for Ki67 and denominator is the whole microscopic area.

Strip chart (D): denominator is the whole microscopic area.

Visually, there is a loss of segregation in strip charts (C) and (D) when the denominator does not restrict the area of reference to the tumor area but includes the whole microscopic field. The same results are presented with ROC in Figure 7. The greatest area under the curve (76.9%) is associated with the V_V method, whereas the smallest (69.5%) is associated to the Immunoratio method. Modified V_V are found in between. The highest accuracy found for V_V against Immunoratio is also depicted in Figure 8. Note that the logarithm transformation does not modify either the ROCs or area under the curves.

DISCUSSION

The general focus of this paper is on predicting the outcome in ER-positive BC patients by establishing an accurate and reproducible proliferative status using Ki67

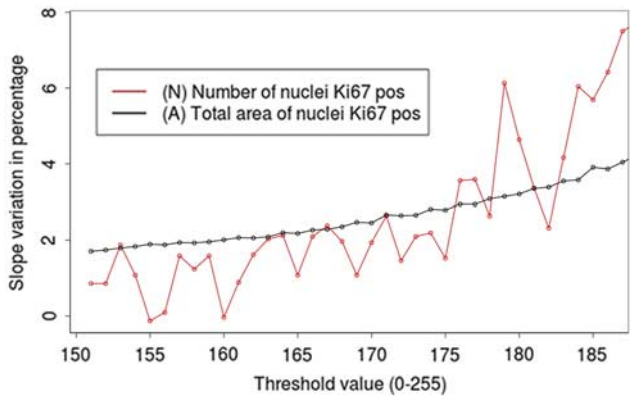

FIGURE 4. Variation in the percentage of values N (number of nuclei positive for Ki67) and A (total area of positive nuclei for Ki67) against a moving threshold. Figure 3 represents the initial scenario when threshold (T) = 167. Overall both A and N increase when threshold increases but the former variation is regular, whereas the latter is erratic. full color online

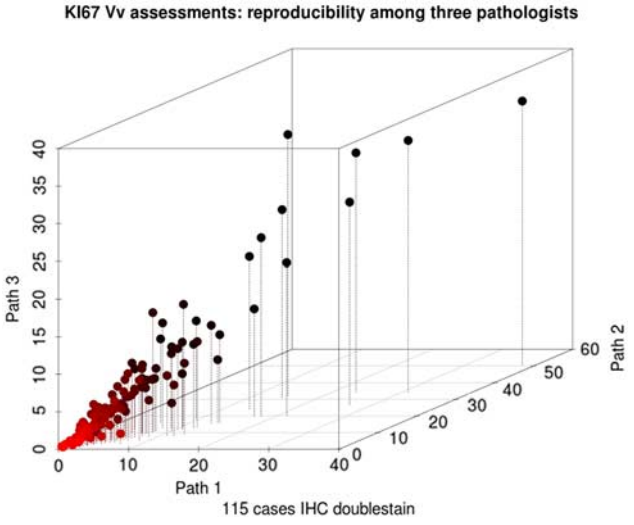

FIGURE 5. Three-dimensional projections of V_V results from 115 double stained assessed by 3 different pathologists. Visually high concordance is suggested by creation of a tight cone in space by the 115 dots. This concordance is confirmed with a Kendall's *W* coefficient of 93.5%. full color online

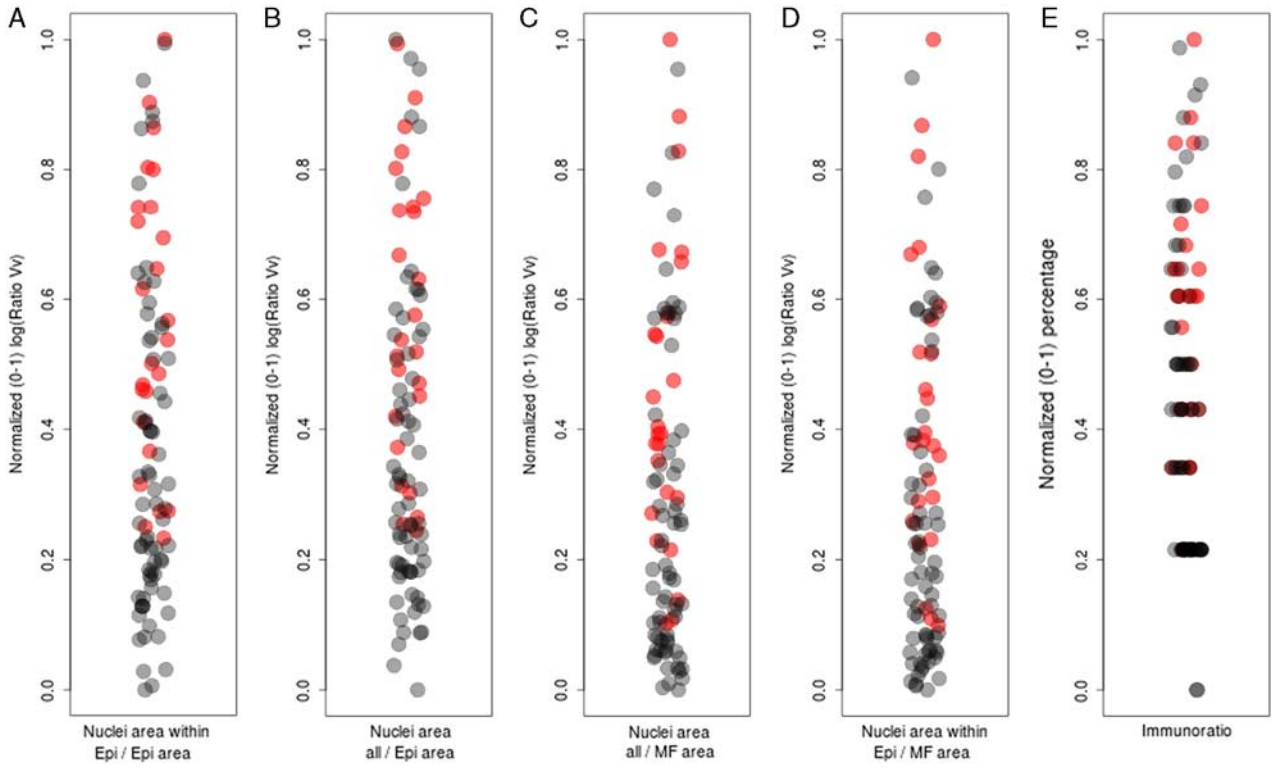

FIGURE 6. Overall, 94 patients with known outcome: gray represents good outcome and red represents poor outcome. Five strip charts with jitter effect representing the same set. All results are normalized values of logarithmic raw data. On the left, strip chart (A) represents V_V; on the right strip chart (E) represents Immunoratio. Strip charts (B), (C), and (D) are variation of V_V as explained in the text. full color online

IHC. This information can then be combined with semi-quantitative assessments of ER, PR, and HER2/*neu* to generate the IHC4 to predict the outcome and determine the need for adjuvant therapy. Cuzick et al[34] have demonstrated that the IHC4 contains as much information as the Genomic Health Recurrence Score (GHrs).[35] GHrs is widely accepted and recently published.[36] The first results of their TAILORx trial showed a 5-year rate of an invasive disease-free survival of 93.8% and freedom from distant recurrence of 99.3% (n = 1626 patients) in 15.9% of patients with a newly defined low-risk RS of ≤ 10. Although impressive, these results suggest that there may be a new intermediate risk category with RS scores of 11 to 25 (previously 18 to 30) that now contain the majority of BC patients (67.3%) for which therapy decisions will have to rely on other considerations. Other reports have highlighted high false-negative rates of HER2 quantitative reverse transcription provided by GHrs,[37] significant numbers of false-negative ER and PR,[38] and questionable proliferative assessments by GHrs[39] in well-differentiated low-grade invasive carcinoma that show mitotically active cellular stroma. Finally, the provision of the test by a single commercial laboratory precludes peer assessment and external measures of accuracy and reproducibility. For these reasons, a reproducible, inexpensive IHC proliferative assessment tool is proposed.

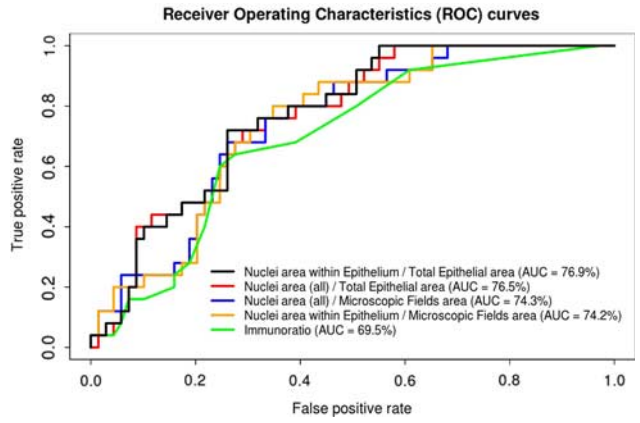

FIGURE 7. Overall, 94 patients with known outcome: receiver operating characteristics curves. The most accurate predictor is V_V (continuous black line) with a 76.9% area under the associated curve and the less accurate is Immunoratio with 69.5% (green line). Red, blue and orange curves are V_V variations as explained in the text. full color online

This method removes the "peel" artifact engendered by thin sections and compensates for the intrinsic bias related to particle size (N_A) assessment. This straightforward nonbiased approach had a very high concordance (93.5%) among the 3 observers and was able to predict the outcome in 94 ER-positive BC patients more accu-

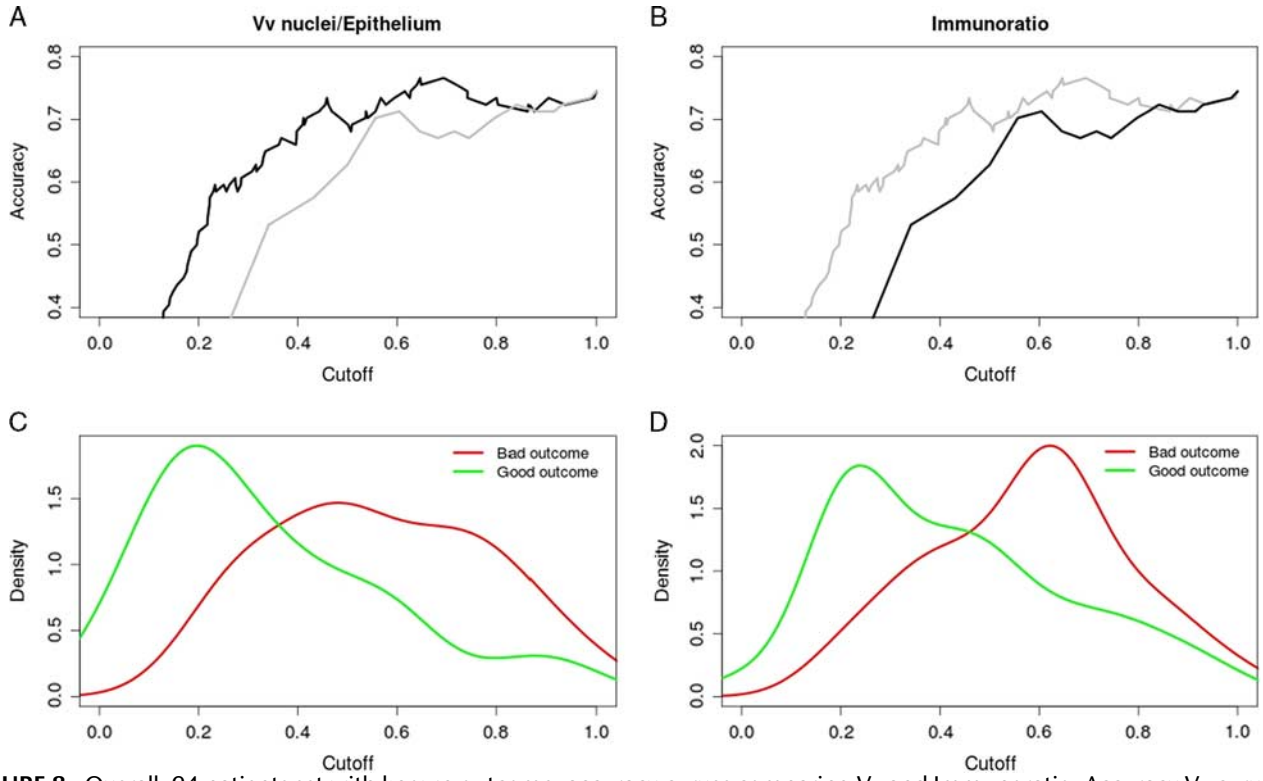

FIGURE 8. Overall, 94 patients set with known outcome: accuracy curves comparing V_V and Immunoratio. Accuracy V_V curve is in black and Immunoratio in gray (A), whereas it is the opposite in (B). In all possible cutoffs V_V accuracy is higher. C and D, Density distribution of poor outcome patients (red) versus good outcome (green) for V_V (C) and Immunoratio (D).V_V variations are not illustrated. full color online

rately than other predictors. The method relies on experienced pathologists to choose representative tumor fields and apply appropriate thresholds to obtain appropriate masks of the nucleus and cytoplasm. This requires approximately 4 minutes per BC case.

Before this technique is widely implemented, it requires calibration and validation of clinical significance. Even though the current work has demonstrated high concordance among 3 observers, all double-stained Ki67-cytokeratin slides were created in 1 laboratory. For a given Ki67 stain, 2 different laboratories can have 2 very different protocols in which antigen retrieval, antibody clone and vender, dilution, diluents, detection, instrumentation, and environmental factors can vary. As it is virtually impossible to replicate every condition between laboratories, it is possible that 2 Ki67-cytokeratin slides (from 2 contiguous thin sections) performed in 2 laboratories would provide different V_V values. To address this interlaboratory variation, we suggest that similar to existing external quality[40] assurance testing protocols that use tissue microarray, calibration of threshold Ki67-positive nuclei area could be performed at the software level utilizing common tissue microarray with BC expressing a variety of proliferative intensities (from very low to very high). Reproducibility scales built with normalized log V_V values could be established.

In addition, the clinical significance of V_V still needs to be validated with a larger number of patients. Nonetheless, these preliminary results showing V_V as a significant predictor of outcome suggests that further investigation is warranted.

In conclusion, this paper has discussed issues related to Ki67 assessment and presented methodological considerations. A novel, interactive IA V_V IHC proliferative assessment has been developed that shows almost perfect concordance between 3 observers and clinical utility in a small group of patients. As current molecular signatures are costly and suffer because of stromal contamination, tumor-targeted IA on IHC-stained slides is a viable alternative.

REFERENCES

1. Brenton JD, Carey LA, Ahmed AA, et al. Molecular classification and molecular forecasting of breast cancer: ready for clinical application? *J Clin Oncol*. 2005;23:7350–7360.
2. Hammond MEH, 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.
3. Felipe Ades DZ. Luminal B breast cancer: molecular characterization, clinical management, and future perspectives A B. *J Clin Oncol*. 2014;32:2794–2803.

4. Esposito A, Criscitiello C, Salè EO, et al. Optimal adjuvant chemotherapy in breast cancer: selection of agents. *Expert Rev Clin Pharmacol*. 2014;7:605–611.
5. Pathmanathan N, Balleine RL. Ki67 and proliferation in breast cancer. *J Clin Pathol*. 2013;66:512–516.
6. Jonat W, Arnold N. Is the Ki-67 labelling index ready for clinical use? *Ann Oncol*. 2011;22:500–502.
7. Shui R, Yu B, Bi R, et al. An interobserver reproducibility analysis of Ki67 visual assessment in breast cancer. *PLoS One*. 2015;10: e0125131.
8. Varga Z, Diebold J, Dommann-Scherrer C, et al. How reliable is Ki-67 immunohistochemistry in grade 2 breast carcinomas? A QA study of the Swiss Working Group of breast- and gynecopathologists. *PLoS One*. 2012;7:e37379.
9. Harvey J, Thomas C, Wood B, et al. Practical issues concerning the implementation of Ki-67 proliferative index measurement in breast cancer reporting. *Pathology*. 2015;47:13–20.
10. Polley M-Y C, Leung SCY, Gao D, et al. An international study to increase concordance in Ki67 scoring. *Mod Pathol*. 2015;28:778–786. Doi:10.1038/modpathol.2015.38.
11. Hida AI, Bando K, Sugita A, et al. Visual assessment of Ki67 using a 5-grade scale (Eye-5) is easy and practical to classify breast cancer subtypes with high reproducibility. *J Clin Pathol*. 2015;68: 356–361.
12. Laurinavicius A, Plancoulaine B, Laurinaviciene A, et al. A methodology to ensure and improve accuracy of Ki67 labelling index estimation by automated digital image analysis in breast cancer tissue. *Breast Cancer Res*. 2014;16:R35.
13. Mohammed ZMA, McMillan DC, Elsberger B, et al. Comparison of Visual and automated assessment of Ki-67 proliferative activity and their impact on outcome in primary operable invasive ductal breast cancer. *Br J Cancer*. 2012;106:383–388.
14. Santisteban M, Reynolds C, Fritcher EGB, et al. Ki67: a time-varying biomarker of risk of breast cancer in atypical hyperplasia. *Breast Cancer Res Treat*. 2010;121:431–437.
15. Brown JR, DiGiovanna MP, Killelea B, et al. Quantitative assessment Ki-67 score for prediction of response to neoadjuvant chemotherapy in breast cancer. *Lab Invest*. 2014;94:98–106.
16. Tuominen VJ, Ruotoistenmäki S, Viitanen A, et al. ImmunoRatio: a publicly available web application for quantitative image analysis of estrogen receptor (ER), progesterone receptor (PR), and Ki-67. *Breast Cancer Res*. 2010;12:R56.
17. Abramoff MD, Magalhães PJ, Ram SJ. Image processing with ImageJ. *Biophotonics Int*. 2004;11:36–42.
18. Fasanella S, Leonardi E, Cantaloni C, et al. Proliferative activity in human breast cancer: Ki-67 automated evaluation and the influence of different Ki-67 equivalent antibodies. *Diagn Pathol*. 2011;6(suppl 1):S7.
19. Nielsen PS, Bentzer NK, Jensen V, et al. Immunohistochemical Ki-67/KL1 double stains increase accuracy of Ki-67 indices in breast cancer and simplify automated image analysis. *Appl Immunohistochem Mol Morphol*. 2014;22:568–576.
20. Gundersen HJG. Stereology of arbitrary particles. *J Microsc*. 1986; 143:3–45.
21. Sterio DC. The unbiased estimation of number and sizes of arbitrary particles using the disector. *J Microsc*. 1984;134:127–136.
22. Kaplan S, Odacı E, Canan S, et al. The disector counting technique. *NeuroQuantology*. 2011;10:44–53.
23. Pover CM, Coggeshall RE. Verification of the disector method for counting neurons, with comments on the empirical method. *Anat Rec*. 1991;231:573–578.
24. Volume. Stereology Information Center. stereology.info. Available at: http://www.stereology.info/volume/. Accessed September 26, 2015.
25. Ridler TW, Calvard S. Picture thresholding using an iterative selection method. *IEEE Trans Syst Man Cybern*. 1978;8:630–632.
26. Vincent L, Soille P. Watersheds in digital spaces: an efficient algorithm based on immersion simulations. *IEEE Trans Pattern Anal Mach Intell*. 1991;13:583–598.
27. Landini G. Colour deconvolution. Available at: http://www.dentistry.bham.ac.uk/landinig/software/cdeconv/cdeconv.html. Accessed December 4, 2011.
28. Ruifrok AC, Johnston DA. Quantification of histochemical staining by color deconvolution. *Anal Quant Cytol Histol*. 2001;23:291–299.
29. Landini G. How to correct background illumination in brightfield microscopy. Available at: http://imagejdocu.tudor.lu/doku.php?id = howto:working:how_to_correct_background_illumination_in_brightfield_microscopy. Accessed November 27, 2011.
30. Ihaka R, Gentleman R. R: a language for data analysis and graphics. *J Comput Graph Stat*. 1996;5:299–314.
31. Kendall MG, Smith BB. The problem of m rankings. *Ann Math Stat*. 1939;10:275–287.
32. Sing T, Sander O, Beerenwinkel N, et al. ROCR: visualizing classifier performance in R. *Bioinformatics*. 2005;21:3940–3941.
33. Mutterer J, Zinck E. Quick-and-clean article figures with FigureJ. *J Microsc*. 2013;252:89–91.
34. Cuzick J, Dowsett M, Pineda S, et al. Prognostic value of a combined estrogen receptor, progesterone receptor, Ki-67, and human epidermal growth factor receptor 2 immunohistochemical score and comparison with the genomic health recurrence score in early breast cancer. *J Clin Oncol*. 2011;29:4273–4278. Doi:10.1200/JCO.2010.31.2835.
35. Paik S, Shak S, Tang G, et al. A multigene assay to predict recurrence of tamoxifen-treated, node-negative breast cancer. *N Engl J Med*. 2004;351:2817–2826.
36. Sparano JA, Gray RJ, Makower DF, et al. Prospective validation of a 21-gene expression assay in breast cancer. *N Engl J Med*. 2015; 373:2005–2014.
37. Dabbs DJ, Klein ME, Mohsin SK, et al. High false-negative rate of HER2 quantitative reverse transcription polymerase chain reaction of the oncotype DX test: an independent quality assurance study. *J Clin Oncol*. 2011;29:4279–4285.
38. 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.
39. Acs G, Esposito NN, Kiluk J, et al. A mitotically active, cellular tumor stroma and/or inflammatory cells associated with tumor cells may contribute to intermediate or high Oncotype DX Recurrence Scores in low-grade invasive breast carcinomas. *Mod Pathol*. 2012; 25:556–566.
40. 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.

