Performance of Three-Biomarker Immunohistochemistry for Intrinsic Breast Cancer Subtyping in the AMBER Consortium

Emma H. Allott1,2, Stephanie M. Cohen1,3, Joseph Geradts4, Xuezheng Sun2, Thaer Khoury5, Wiam Bshara6, Gary R. Zirpoli6, C. Ryan Miller1,3,7, Helena Hwang6, Leigh B. Thorne7, Siobhan O’Connor7, Chi-Kit Tse8, Mary B. Bell1, Zhiyuan Hu1, Yan Li1, Erin L. Kirk2, Traci N. Bethea9, Charles M. Perou1, Julie R. Palmer9, Christine B. Ambrosone6, Andrew F. Olshan1,2, and Melissa A. Troester1,2

Abstract

Background: Classification of breast cancer into intrinsic subtypes has clinical and epidemiologic importance. To examine accuracy of IHC-based methods for identifying intrinsic subtypes, a three-biomarker IHC panel was compared with the clinical record and RNA-based intrinsic (PAM50) subtypes.

Methods: Automated scoring of estrogen receptor (ER), progesterone receptor (PR), and HER2 was performed on IHC-stained tissue microarrays comprising 1,920 cases from the African American Breast Cancer Epidemiology and Risk (AMBER) consortium. Multiple cores (1–6/case) were collapsed to classify cases, and automated scoring was compared with the clinical record and to RNA-based subtyping.

Results: Automated analysis of the three-biomarker IHC panel produced high agreement with the clinical record (93% for ER and HER2, and 88% for PR). Cases with low tumor cellularity and smaller core size had reduced agreement with the clinical record. IHC-based definitions had high agreement with the clinical record regardless of hormone receptor positivity threshold (1% vs. 10%), but a 10% threshold produced highest agreement with RNA-based intrinsic subtypes. Using a 10% threshold, IHC-based definitions identified the basal-like intrinsic subtype with high sensitivity (86%), although sensitivity was lower for luminal A, luminal B, and HER2-enriched subtypes (76%, 40%, and 37%, respectively).

Conclusion: Three-biomarker IHC-based subtyping has reasonable accuracy for distinguishing basal-like from nonbasal-like, although additional biomarkers are required for accurate classification of luminal A, luminal B, and HER2-enriched cancers.

Impact: Epidemiologic studies relying on three-biomarker IHC status for subtype classification should use caution when distinguishing luminal A from luminal B and when interpreting findings for HER2-enriched cancers. Cancer Epidemiol Biomarkers Prev; 25(3); 470–8. ©2015 AACR.

Introduction

Breast cancer is a heterogeneous disease, comprised of distinct tumor subtypes (1). Although some epidemiologic studies have used additional markers or assays to define etiologic subtypes (2–4), the majority relied on estrogen receptor (ER), progesterone receptor (PR), and HER2 to classify breast cancers as hormone receptor (HR) positive, HER2-positive, or triple negative (5–8). These studies have identified heterogeneity in breast cancer etiology, but some lack of agreement between studies may be attributable to discordant subtype classification. Few studies have systematically compared immunohistochemical, clinical record, and RNA-based intrinsic subtypes to estimate classification accuracy, particularly using tissue microarrays (TMA) (refs. 9, 10).

The advent of automated methods to analyze digital pathology data has begun to support high-throughput IHC-based breast cancer subtyping in large epidemiologic studies (11). Simultaneously, RNA-based methods have become more readily available for application in formalin-fixed paraffin-embedded (FFPE) tissues (12). In the current study, we carried out IHC staining for ER, PR, and HER2 using TMAs containing multiple cores per case for each of 1,920 cases in the African American Breast Cancer Epidemiology and Risk (AMBER) consortium. Herein, we describe (i) automated digital quantification of this three-biomarker IHC panel, (ii) methods for core-to-case collapsing, and (iii) IHC-based agreement with clinical scoring in the medical record. Finally, treating PAM50 as the reference standard for the
purposes of the current analysis, we evaluate (iv) agreement between three-biomarker IHC-based subtype classification and RNA-based PAM50 intrinsic subtyping. Although earlier studies employed a 10% threshold (13–15), current clinical guidelines recommend a 1% threshold for defining ER and PR positivity status (16), and we evaluate the impact of this revised threshold on agreement between IHC- and RNA-based intrinsic subtyping.

Materials and Methods
Study population and TMA construction
The AMBER consortium is comprised of observational studies of breast cancer in African American (AA) women (17). The current study includes 1,920 cases within the AMBER consortium for which paraffin-embedded tissue was available in TMAs. Cases were from the Carolina Breast Cancer Study (CBCS) phase III (n = 599), the Black Women's Health Study (BWHS; n = 199), and the Women's Circle of Health Study (WCHS; n = 309). In addition, we included 645 non-AA cases from CBCS and 168 non-AA cases from WCHS (Table 1) as these cases were present on the same TMAs and were relevant to our evaluation of subtype classification rates. Clinical ER, PR, and HER2 status were abstracted from medical records. Tumor grade was centrally assigned for CBCS cases by a single pathologist (J. Geradts) using the Nottingham breast cancer grading system (18). Each study was approved by Institutional review boards at participating hospitals and academic institutions.

Paraffin-embedded tumor blocks were requested from participating pathology laboratories for each case. Study pathologists (J. Geradts, H. Hwang, and T. Khoury) marked hematoxylin and eosin (H&E)-stained slides to indicate areas enriched for invasive breast cancer for coring, and TMA construction and sectioning were carried out at the Translational Pathology Lab, University of North Carolina at Chapel Hill (UNC) for CBCS and at the Roswell Park Cancer Institute for BWHS and WCHS. TMA blocks included between one and four tumor cores per case measuring 1.0 mm (BCBS), or between one and six tumor cores per case measuring 0.6 mm in diameter (BWHS and WCHS; Table 1). For CBCS, top and bottom sections from blocks containing a total of 4,783 cores were stained with H&E and examined by study pathologists (J. Geradts or L.B. Thorne) for presence of tumor cells, and 503 cores (11%) lacking sufficient tumor cellularity (<50 tumor nuclei per core) were excluded from analyses. For TMA blocks not manually evaluated for tumor cellularity (BWHS and WCHS), digital analysis of IHC was used to eliminate cores lacking sufficient tumor cellularity (<50 tumor nuclei per core) and, of a total of 2,060 cores, 290 (14%) were excluded due to low tumor cellularity.

IHC- and RNA-based assays
TMA blocks were cut into 5 μm serial sections and immunohistochemical staining protocols were optimized under pathologist’s supervision to achieve the best possible agreement with the clinical record. Final conditions were independently reviewed by multiple pathologists. Detailed immunohistochemical methods are described in Supplementary Materials and Methods.

Nanostring assays were used to measure the PAM50 gene signature on 535 cases from CBCS. Two cores from the same tumor block used for TMA construction were randomly sampled and pooled for analysis. The areas surrounding the holes left by the cores were examined by a study pathologist (J. Geradts or L.B. Thorne) to confirm the presence of tumor cells in the cores used for RNA extraction. RNA was isolated using the RNeasy FFPE Kit from Qiagen and nanostring analyses were performed in the Rapid Adoption Molecular laboratory at UNC. Nanostring probe sets included 200 genes of which 50 were the PAM50 genes (19) and five were housekeeping genes. Gene expression was median-centered and samples were standardized to zero mean and unit variance. The NanoStringNorm package in Bioconductor was used to eliminate samples that did not have sufficient nanostring data quality (n = 43) and the PAM50 predictor was performed as previously described (19) to categorize breast tumors into five intrinsic subtypes (luminal A, luminal B, HER2-enriched, basal-like, and normal-like). Tumors classified as normal-like may result from extensive normal epithelial or stromal content in the tumor (20), so we excluded normal-like tumors (n = 29) from our analysis. We also excluded 16 cases with equivocal HER2 IHC status, as we could not assess IHC-based subtype, giving rise to 447 cases with both IHC and intrinsic subtyping data.

Automated digital quantification of a three biomarker IHC-based panel
Automated digital image analysis of IHC staining (Fig. 1) was performed using a Genie classifier and the Nuclear v9 algorithm (for ER and PR) or Membrane v9 algorithm (for HER2; Aperio Technologies), and is described in more detail in Supplementary Materials and Methods.

To validate automated analyses, study pathologists (J. Geradts, H. Hwang, and T. Khoury) carried out manual reviews of immunohistochemical staining within a training set of TMAs. For ER and PR, individual cores were classified as negative (<1% positive), borderline (≥1% and <10%), or positive (≥10%). For HER2, individual cores were classified as negative (0/1+), positive (3+), or equivocal (2+; ref. 21). Using a 10% threshold to define ER and PR status, agreement between automated and manual scoring was 89% for ER and 91% for PR (Supplementary Table S1). After excluding cores with equivocal HER2 (2+) status in either the TMA or clinical record (n = 63), HER2 status agreement between automated and manual scoring was 93%. As expected, agreement was slightly lower with additional categories [ER and PR status as negative (<1%), borderline (1%–10%), or positive (≥10%); HER2 status as negative (0/1+), equivocal (2+), or positive (3+); Supplementary Table S1].
Core-to-case collapsing

We tested two core-to-case collapsing methods to define a single ER, PR, and HER2 status for each case. The first assigned case-level status using a tumor cellularity-weighted approach. The weighted average of percent positivity was calculated by summing the product of percent positivity and core weight across all cores per case. Core weight was defined as the number of tumor nuclei in a given core divided by the total number of tumor nuclei across all cores for that case. Thresholds for ER, PR ($\geq 1\%$ and $\geq 10\%$), and HER2 ($\geq 10\%$) positivity were subsequently applied to define a dichotomous status for each of these three markers. Tumor cells staining positive at any intensity were considered positive for ER and PR, although only tumor cells staining at the $3+$ level were considered positive for HER2. Equivocal ($2+$) HER2 status was defined as $<10\%$ of tumor cells staining at the $3+$ level and $\geq 10\%$ of tumor cells staining at the $2+$ level, and negative HER2 status included all other cases. The $10\%$ threshold for HER2 was optimized to agree with manual scoring performed according to the then-current HER2 guidelines, indicating that a $30\%$ threshold should be used for HER2 (21). The second core-to-case collapsing method classified the case as ER, PR, or HER2 positive if any core was positive for ER, PR, or HER2, respectively. Core-level ER and PR positivity was defined considering both $\geq 1\%$ and $\geq 10\%$ thresholds, although HER2 positivity was defined as $\geq 10\%$ of cells staining at the $3+$ intensity level.

IHC-based agreement with clinical scoring in the medical record

Kappa statistics were used to examine case-level agreement between TMA and clinical ER, PR, and HER2 status in the medical record (22). Given that the majority of cases (76%) were diagnosed prior to the American Society of Clinical Oncology guidelines recommending the use of a $1\%$ threshold for classification of HR positivity (16), our primary analysis of case-level agreement between TMA and clinical ER, PR, and HER2 status was conducted using a $10\%$ threshold to define HR positivity status. Analyses were stratified by core-to-case collapsing method in addition to core diameter and median cellularity, to determine the impact of these factors on agreement with the clinical record. We conducted sensitivity analyses using a $1\%$ threshold for ER and PR, and excluding cases with only one core.

Agreement between IHC- and RNA-based subtyping

Using case-level ER, PR, and HER2 status, we defined IHC-based subtypes as follows: ER or PR positive, HER2 negative (luminal A); ER or PR positive, HER2 positive (luminal B); ER and PR negative, HER2 positive (HER2-enriched); ER and PR and HER2 negative (basal-like). We also defined IHC-based subtypes using a PR cut-off point of $20\%$ (23): ER positive or PR $>20\%$, HER2 negative (luminal A); ER or PR positive, HER2 positive or ER or PR $\leq 20\%$, HER2 negative (luminal B). In the absence of Ki67 data, tumor grade may distinguish luminal A and HER2-negative luminal B cancers (24), and we explored combined grade as follows: ER positive or PR $>20\%$, combined grade I or II, HER2 negative (luminal A); ER or PR positive, HER2 positive or PR or ER $\leq 20\%$ or combined grade III, HER2 negative (luminal B). Dichotomizing combined grade as I versus II/III reduced subtype classification accuracy, and mitotic grade offered no classification advantages over combined grade and therefore these results are not presented. Finally, we defined HER2-enriched cases as ER...
Results

IHC-based agreement with the clinical record: impact of core-to-case collapsing method

Using the weighted core-to-case collapsing method to define HR positivity status, agreement with the clinical record was 93% for ER and 88% for PR (using a ≥10% threshold; Table 2). These results were not impacted by a 1% positivity threshold (data not shown), nor was agreement substantially altered when using the any positive core-to-case collapsing method (Table 2). However, the any positive method with a 1% threshold resulted in reduced agreement with the clinical record (90% agreement for ER, 85% for PR). As such, we chose to proceed with the weighted method and 10% threshold to maximize agreement with clinical laboratory data for defining ER and PR positivity in our population.

For HER2, agreement rates were similar for weighted and any positive methods (95% and 93%, respectively; Table 3). In addition, although both methods had high specificity (95% and 98% for any positive and weighted methods, respectively), the any positive method had higher sensitivity for identifying HER2-positive cases (82% vs. 75%, respectively). However, the any positive method produced a higher number of equivocal (2+) cases, relative to the weighted method (Table 3). A hybrid of both core-to-case collapsing methods, defining HER2 status based on being positive by either the weighted or any positive method, maximized sensitivity (78%) and minimized the number of equivocal cases, although maintaining similar levels of agreement with the clinical record (Table 3). Moreover, the number of equivocal cases using the hybrid method (n = 74; 4%) was similar to that of the clinical record (n = 50; 3%). As such, we chose to proceed with this hybrid method to define HER2 status.

For all three biomarkers, agreement between automated analysis of TMAs and the clinical record was similar to agreement rates reported between automated analysis of TMAs and manual review (Supplementary Table S1).

IHC-based agreement with the clinical record: impact of TMA characteristics

Core cellularity and diameter impacted agreement rates between TMAs and the clinical record. Using a 10% threshold to define HR positivity, cases with high tumor cellularity had higher agreement with the clinical record for ER and PR, as shown in Table 4. Similarly, cases with larger cores had higher ER agreement and PR agreement (Table 4). However, neither core cellularity nor diameter substantially impacted agreement with the clinical record for HER2, with similar agreement rates for tumors with high and low cellularity, and for 1.0 and 0.6 mm cores (Table 4). Using a 1% threshold to define ER and PR status did not alter agreement rates for cores with high cellularity and 1.0 mm diameter, but improved agreement for cores with low cellularity and 0.6 mm diameter (data not shown). Excluding cases with only one core (7% of cases) did not impact our findings (data not shown).

Agreement between IHC- and RNA-based PAM50 intrinsic subtyping

For the subset of CBCS cases with intrinsic subtyping data (n = 447), agreement between three biomarker IHC- and RNA-based subtypes is shown in Table 5. Given recent changes in clinical thresholds for ER and PR positivity (16), both 1% and 10% thresholds were considered to assess agreement. Although IHC-based definitions of luminal A (ER or PR positive and HER2 negative) and luminal B (ER or PR positive and HER2 positive)
resulted in high sensitivity for identifying luminal A tumors, specificity was low for both 1% and 10% thresholds. Conversely, luminal B specificity was high but sensitivity was low, regardless of threshold (Supplementary Table S2). Work by Prat and colleagues (25), reflected in St. Gallen guidelines (23), indicated that cases with negative HER2 status but low (<20%) PR positivity should be classified as luminal B, as an additional means of distinguishing luminal A from luminal B. Using this definition, the 10% ER threshold yielded greatest accuracy (Table 5). St. Gallen guidelines also suggest that Ki67 expression may distinguish luminal A and HER2-negative luminal B disease (23). Given that we lacked Ki67 data, we explored tumor grade as a surrogate (24). Overall, gains in sensitivity were offset by losses in specificity and vice versa, and accuracy was substantially unchanged (~70%–80%) regardless of grade (Supplementary Table S3).

HER2-enriched cases identified by RNA-based PAM50 analysis were difficult to accurately identify by IHC. Defining HER2-enriched cases as ER and PR negative, HER2 positive resulted in low sensitivity, although specificity was high (Supplementary Table S2). We found that defining HER2-enriched cases as ER negative and HER2 positive, regardless of PR status, resulted in slightly improved agreement with the intrinsic HER2-enriched subtype (Table 5). Using a 10% threshold, we found that 26% of HER2-enriched cases identified by RNA-based intrinsic subtyping were classified as luminal B by IHC, while the use of a 1% threshold gave rise to 43% of HER2-enriched cases identified by RNA-based intrinsic subtyping, which were classified as luminal B by IHC (Table 5). Finally, utilization of clinical HER2 status (which includes ISH data) to identify HER2-enriched cases did not improve on agreement rates observed using TMA HER2 status (data not shown).

Triple-negative status (ER, PR, and HER2 negative) using the three-biomarker IHC panel had relatively high accuracy for identifying the basal-like intrinsic subtype (Table 5). At the 10% threshold, we found that 10% of basal-like cases identified by RNA-based intrinsic subtyping were classified as luminal A or B by IHC (Table 5). However, rates of discordant classification between IHC- and RNA-based intrinsic subtyping were higher at the 1% threshold, with 27% of basal-like cases identified by RNA-based intrinsic subtyping classified as luminal A or B by IHC (Table 5). Including normal-like intrinsic cases (n = 26 with complete IHC data) in our comparisons between three biomarker IHC- and RNA-based subtypes did not alter our findings (data not shown).

**Discussion**

In this consortium comprising three large epidemiologic studies, we found that classification of ER, PR, and HER2 status using automated digital pathology resulted in high agreement with the clinical record for all three biomarkers. Agreement rates were highest when using a tumor cellularity-weighted method for core-to-case collapsing and in cases with high tumor cellularity. Furthermore, we report that this three-biomarker IHC-based panel conferred high sensitivity and specificity for identifying tumors classified as basal-like using RNA-based intrinsic subtyping. However, our findings highlight the need for further optimization of methods to identify HER2-enriched breast cancers and to discriminate between luminal A and B cancers, likely through inclusion of new biomarkers like Ki67 and/or other biomarkers uniquely present in luminal B or HER2-enriched tumors. As such, we recommend that epidemiologic studies relying on ER, PR, and HER2 IHC status for subtype classification

<table>
<thead>
<tr>
<th>Table 3. Case-level agreement between automated scoring of central tissue microarrays and clinical HER2 status: impact of core-to-case collapsing method</th>
</tr>
</thead>
<tbody>
<tr>
<td>Clinical HER2 status in the medical record</td>
</tr>
<tr>
<td>Central TMA HER2 status</td>
</tr>
<tr>
<td>Weighted core-to-case collapsing methoda</td>
</tr>
<tr>
<td>Positive</td>
</tr>
<tr>
<td>Total</td>
</tr>
<tr>
<td>Agreement</td>
</tr>
<tr>
<td>Kappa (95% CI)</td>
</tr>
<tr>
<td>Any positive core-to-case collapsing methodb</td>
</tr>
<tr>
<td>Negative</td>
</tr>
<tr>
<td>Positive</td>
</tr>
<tr>
<td>Total</td>
</tr>
<tr>
<td>Agreement</td>
</tr>
<tr>
<td>Kappa (95% CI)</td>
</tr>
<tr>
<td>Hybrid core-to-case collapsing methodc</td>
</tr>
<tr>
<td>Negative</td>
</tr>
<tr>
<td>Positive</td>
</tr>
<tr>
<td>Total</td>
</tr>
<tr>
<td>Agreement</td>
</tr>
<tr>
<td>Kappa (95% CI)</td>
</tr>
</tbody>
</table>

aWeighted method: Case level positivity is weighted by the tumor cellularity of each contributing core.
bAny positive method: Case is positive if classified as positive by either the weighted or any positive method.
cHybrid method: Case is positive if classified as positive by either the weighted or any positive method.

n = 106 (6%), n = 207 (12%), and n = 74 (4%) cases with equivocal (2+) central HER2 status. n = 50 (3%) cases with equivocal (2+) clinical HER2 status in the medical record.
use caution when distinguishing luminal A from luminal B cancers and when interpreting findings for HER2-enriched cancers.

The advent of digital pathology carries potential to advance molecular epidemiologic pathology research by providing a platform to obtain high-throughput and high-resolution quantitative data. Previous studies based on tumor-bank series have shown that interlaboratory agreement rates for ER and PR range from 80% to 90% (26–28). Our findings from this consortium, which drew from multiple clinical centers across the United States, showed that agreement rates were on the high end of this range for both biomarkers, thereby providing validation for digital IHC subtyping methods in epidemiologic research. Furthermore, our agreement rates for HER2 exceeded previous estimates from endocrine therapy (16). However, subtype distribution within these borderline cases, comprising approximately 6% of total breast cancer cases, is not well understood. One study indicated that approximately one quarter of ER borderline cases were luminal, although up to half are basal-like (32). Another study reported higher prevalence of luminals (44%) and lower prevalence of basal-likes (18%) among ER or PR borderline cases (33), suggesting that biologic characteristics of the borderline group may vary across study populations or according to technical factors including TMA characteristics.

Although technical factors have been demonstrated to play a role (27, 30, 31), tumor biology and molecular factors may also drive discordance in biomarker classification. We and others (26, 28) have shown that PR agreement is often lower than that of ER, potentially due to more heterogeneous expression within a given tumor, or intratumoral heterogeneity. Intratumoral heterogeneity would be expected to pose a greater classification problem for cores representing a smaller sample of the tumor (as reflected by lower tumor cellularity and/or smaller core diameter); indeed, our findings suggest that agreement rates were lowest for PR. This explanation seems less likely to explain HER2 discordance, as neither core size nor cellularity was associated with agreement between TMAs and the clinical record. As such, intratumoral heterogeneity, in addition to technical factors including TMA characteristics, may contribute to the level of agreement between central and clinical biomarker classification.

Recent clinical guidelines have lowered the threshold used to define ER and PR positivity from 10% to 1%, given evidence that cases with borderline (\(\geq 1\% < 10\%\)) ER and PR positivity derive a benefit from endocrine therapy (16). However, subtype distribution within these borderline cases, comprising approximately 6% of total breast cancer cases, is not well understood. One study indicated that approximately one quarter of ER borderline cases were luminal, although up to half are basal-like (32). This is in general agreement with our findings that 25% of ER borderline cases were luminal, and 38% were basal-like. Conversely, another study reported higher prevalence of luminals (44%) and lower prevalence of basal-likes (18%) among ER or PR borderline cases (33), suggesting that biologic characteristics of the borderline group may vary across study populations or according to technical factors including higher-throughput and high-resolution quantitative data.

### Table 4. Case-level agreement between automated scoring of central TMAs and clinical estrogen receptor (ER), progesterone receptor (PR), and HER2 status: impact of core cellularity and diameter

<table>
<thead>
<tr>
<th>Cellularity</th>
<th>Clinical ER, PR, and HER2 status in the medical record*</th>
<th>HER2</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>ER Negative Positive Total</td>
<td>PR Negative Positive Total</td>
</tr>
<tr>
<td><strong>Cellularity = median</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Negative</td>
<td>271 34 305</td>
<td>336 32 368</td>
</tr>
<tr>
<td>Positive</td>
<td>6 509 515</td>
<td>31 376 407</td>
</tr>
<tr>
<td>Total</td>
<td>277 543 820</td>
<td>367 408 775</td>
</tr>
<tr>
<td>Agreement</td>
<td>95%</td>
<td>Agreement 92%</td>
</tr>
<tr>
<td>Kappa (95% CI)</td>
<td>0.89 (0.86–0.93)</td>
<td>Kappa (95% CI) 0.84 (0.80–0.88)</td>
</tr>
<tr>
<td><strong>Cellularity &lt; median</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Negative</td>
<td>156 66 222</td>
<td>229 76 305</td>
</tr>
<tr>
<td>Positive</td>
<td>9 586 595</td>
<td>43 427 470</td>
</tr>
<tr>
<td>Total</td>
<td>165 652 817</td>
<td>272 503 775</td>
</tr>
<tr>
<td>Agreement</td>
<td>91%</td>
<td>Agreement 85%</td>
</tr>
<tr>
<td>Kappa (95% CI)</td>
<td>0.75 (0.70–0.80)</td>
<td>Kappa (95% CI) 0.67 (0.62–0.73)</td>
</tr>
</tbody>
</table>

*Using a 10% threshold to define ER and PR positivity for ER and PR analysis and excluding HER2 2+ cases from HER2 analyses.
methods. We report that IHC-based surrogate definitions of PAM50-intrinsic subtypes were impacted by the threshold for defining ER and PR positivity, with a 1% threshold resulting in a greater proportion of basal-like cases classified as luminal A and B by IHC, and of HER2-enriched cases classified as luminal B by IHC, relative to the 10% threshold. Given the heterogeneity of borderline breast cancers, additional studies should refine classification for this group of tumors. Nonetheless, these findings demonstrate that changing clinical guidelines for HR positivity may adversely affect the accuracy of intrinsic subtype classification using a three biomarker IHC panel, potentially impacting the stability of epidemiologic findings over time.

Although it has been proposed that additional biomarkers, such as Ki67, are required to accurately identify luminal A and luminal B cancers, these data are not commonly available in population-based studies. In the absence of Ki67, tumor grade and quantitative PR expression may be useful in distinguishing luminal A and luminal B cancers (23, 24). However, we found that maximal accuracy was approximately 80% for luminal A cases and approximately 75% for luminal B cases with the inclusion of these variables, suggesting that grade and PR thresholds are inadequate to accurately distinguish luminal subtypes. IHC is the most commonly used method for clinical classification of HER2-positive tumors, but even with inclusion of clinically ISH-positive cases in our HER2-enriched group, the sensitivity of the three-biomarker IHC panel was low for identifying HER2-enriched cases. It is well known that amplification or overexpression of an oncogenic pathway can occur via multiple molecular events, and a single marker cannot always accurately detect defects in a pathway. This molecular heterogeneity may contribute to discordance between the three biomarker IHC-based panel and the multigene RNA-based panel to identify HER2 pathway enrichment. Thus, both molecular and intratumoral heterogeneity may add to technical factors in producing discordance between central and clinical analyses of tumor subtype. These data underscore that population-based studies with three biomarker IHC-based data should avoid making etiologic inferences about luminal A, luminal B, and HER2-enriched intrinsic subtypes until better IHC-based subtyping methods can be identified for these subtypes. In contrast, high sensitivity and specificity of the three biomarker IHC panel for identifying basal-like breast cancers in our study suggests that additional biomarkers such as CK5/6 and EGFR may not be needed to classify basal-like cancers in epidemiologic studies. In sum, the current data suggest that a three-biomarker IHC panel is able to distinguish between basal-like and luminal, two key etiologic subtypes (34, 35), but that finer resolution of intrinsic subtyping in nonbasal–like cancers requires additional markers.

Our findings should be considered in the context of strengths and weaknesses of this study. First, although cores without invasive carcinoma were excluded based upon manual review of top and bottom H&E-stained sections by study pathologists, it is possible that benign epithelium or ductal carcinoma in situ was counted as invasive tumor in automated IHC quantification. However, high levels of agreement with the clinical record, exceeding that of previous studies, mitigate concerns about this potential source of bias. Second, although the current analysis offers insights into IHC-based surrogates for intrinsic subtyping within the context of a consortium of epidemiologic studies, it is important to note that this study represents an incomplete sample of AMBER, specifically representing those for whom we had TMA blocks. Because this sample does not represent the entire population of AMBER, it cannot be used to make inferences about the distribution of subtypes within the AMBER source population. Third, we had incomplete data on whether tumor blocks used clinically were the same blocks provided to our study. Thus, the biospecimens may have differed leading to a downward bias in the estimation of agreement between TMAs and the clinical record. Strengths of this study include validation of automated staining guided by multiple pathologists, availability of a validated RNA-based multigene assay for molecular classification of tumor subtype, and a large, diverse sample population representing African American and Caucasian women.

In conclusion, we report that automated digital analysis of IHC staining for ER, PR, and HER2 on TMA results in high agreement with the clinical record, and high sensitivity and specificity for identifying basal-like breast cancer cases classified by RNA-based intrinsic subtyping. However, we report reduced sensitivity and specificity for identifying luminal A, luminal B, and HER2-enriched cancers using IHC-based panels, indicating

Table 5. Agreement between three biomarker IHC-based and intrinsic subtypes: impact of using a 10% versus 1% threshold to define estrogen receptor (ER) and progesterone receptor (PR) positivity

<table>
<thead>
<tr>
<th>IHC subtype</th>
<th>Luminal A</th>
<th>Luminal B</th>
<th>HER2-enriched</th>
<th>Basal-like</th>
<th>Total</th>
</tr>
</thead>
<tbody>
<tr>
<td>ER positivity threshold</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1% ER positivity threshold</td>
<td>Luminal A</td>
<td>122 (76)</td>
<td>49 (57)</td>
<td>10 (14)</td>
<td>8 (6)</td>
</tr>
<tr>
<td>Luminal B</td>
<td>38 (24)</td>
<td>36 (42)</td>
<td>50 (43)</td>
<td>28 (20)</td>
<td>132</td>
</tr>
<tr>
<td>HER2-enriched</td>
<td>0 (0)</td>
<td>0 (0)</td>
<td>19 (27)</td>
<td>5 (4)</td>
<td>24</td>
</tr>
<tr>
<td>Basal-like</td>
<td>0 (0)</td>
<td>1 (1)</td>
<td>11 (16)</td>
<td>90 (69)</td>
<td>102</td>
</tr>
<tr>
<td>Total</td>
<td>160</td>
<td>86</td>
<td>70</td>
<td>131</td>
<td>447</td>
</tr>
<tr>
<td>10% ER positivity threshold</td>
<td>Luminal A</td>
<td>122 (76)</td>
<td>49 (57)</td>
<td>10 (14)</td>
<td>8 (6)</td>
</tr>
<tr>
<td>Luminal B</td>
<td>37 (23)</td>
<td>34 (40)</td>
<td>18 (26)</td>
<td>5 (4)</td>
<td>94</td>
</tr>
<tr>
<td>HER2-enriched</td>
<td>1 (1)</td>
<td>0 (0)</td>
<td>26 (37)</td>
<td>6 (5)</td>
<td>33</td>
</tr>
<tr>
<td>Basal-like</td>
<td>0 (0)</td>
<td>3 (5)</td>
<td>16 (23)</td>
<td>112 (64)</td>
<td>131</td>
</tr>
<tr>
<td>Total</td>
<td>160</td>
<td>86</td>
<td>70</td>
<td>131</td>
<td>447</td>
</tr>
</tbody>
</table>
that additional biomarkers are required for HFC-based classification of these intrinsic subtypes.

Disclosure of Potential Conflicts of Interest

C.M. Perou is a board member at BioClassifier LLC; has ownership interest (including patents) in BioClassifier LLC; has ownership interest in PAM50 patent application; and is a consultant/advisory board member for BioClassifier LLC. No potential conflicts of interest were disclosed by the other authors.

Authors' Contributions


Acquisition of data (provided animals, acquired and managed patients, provided facilities, etc.): J. Geradiz, T. Khoury, W. Bshara, C.R. Miller, H. Hwang, L.B. Thorne, Y. Li, E.L. Kirk, T.N. Bethea, J.R. Palmer, C.B. Ambrosone, A.F. Olshan, M.A. Troester


Administrative, technical, or material support (i.e., reporting or organizing data, constructing databases): S.M. Cohen, X. Sun, G.R. Zirpoli, C.K. Tse, M.B. Bell, Y. Li, A.F. Olshan

Study supervision: X. Sun, M.B. Bell, C.B. Ambrosone, A.F. Olshan, M.A. Troester

Other (histologic slide review as pathologist): S. O’Connor

Grant Support

This work was supported by the NCI [P01CA151135-04 (AMBER Consortium); to J. Geradiz, T. Khoury, W. Bshara, G. Zirpoli, H. Hwang, C.M. Perou, J.R. Palmer, C.B. Ambrosone, A.F. Olshan, and M.A. Troester], P50-CA058223 (SPROE in breast cancer; to C.R. Miller, S. O’Connor, C.M. Perou, A.F. Olshan, and M.A. Troester), U01-CA179715 (to C.M. Perou and M.A. Troester), the University Cancer Research Fund, University of North Carolina at Chapel Hill (to E.H. Allott, J. Thorne, C.-K. Tse, M.B. Bell, E.L. Kirk, C.M. Perou, A.F. Olshan, and M.A. Troester), and the Breast Cancer Research Foundation (to C.B. Ambrosone and C.M. Perou). The Translational Pathology Laboratory (S.M. Cohen and C.R. Miller) is supported in part by grants from the NCI (1P01CA160866) and the University of North Carolina at Chapel Hill University Cancer Research Fund.

The costs of publication of this article were defrayed in part by the payment of page charges. This article must therefore be hereby marked in accordance with 18 U.S.C. Section 1734 solely to indicate this fact.

Received August 19, 2015; revised November 3, 2015; accepted November 9, 2015; published OnlineFirst December 28, 2015.

References


Performance of Three-Biomarker Immunohistochemistry for Intrinsic Breast Cancer Subtyping in the AMBER Consortium

Emma H. Allott, Stephanie M. Cohen, Joseph Geradts, et al.


Updated version  Access the most recent version of this article at: doi:10.1158/1055-9965.EPI-15-0874

Supplementary Material  Access the most recent supplemental material at: http://cebp.aacrjournals.org/content/suppl/2015/12/23/1055-9965.EPI-15-0874.DC1

Cited articles  This article cites 34 articles, 20 of which you can access for free at: http://cebp.aacrjournals.org/content/25/3/470.full.html#ref-list-1

E-mail alerts  Sign up to receive free email-alerts related to this article or journal.

Reprints and Subscriptions  To order reprints of this article or to subscribe to the journal, contact the AACR Publications Department at pubs@aacr.org.

Permissions  To request permission to re-use all or part of this article, contact the AACR Publications Department at permissions@aacr.org.