**Precision Medicine and Imaging**

# A Noninvasive Blood-based Combinatorial Proteomic Biomarker Assay to Detect Breast Cancer in Women over age 50 with BI-RADS 3, 4, or 5 Assessment

Meredith C. Henderson[1], Michael Silver[1], Quynh Tran[1], Elias E. Letsios[1], Rao Mulpuri[1], David E. Reese[1], Ana P. Lourenco[2], Joshua LaBaer[3], Karen S. Anderson[3], Josie Alpers[4], Carrie Costantini[5], Nitin Rohatgi[6], Haythem Ali[7], Karen Baker[8], Donald W. Northfelt[9], Karthik Ghosh[10], Stephen R. Grobmyer[11], Winnie Polen[12], and Judith K. Wolf[1]

## Abstract

**Purpose:** With improvements in breast cancer imaging, there has been a corresponding increase in false-positives and avoidable biopsies. There is a need to better differentiate when a breast biopsy is warranted and determine appropriate follow-up. This study describes the design and clinical performance of a combinatorial proteomic biomarker assay (CPBA), Videssa Breast, in women over age 50 years.

**Experimental Design:** A BI-RADS 3, 4, or 5 assessment was required for clinical trial enrollment. Serum was collected prior to breast biopsy and subjects were followed for 6–12 months and clinically relevant outcomes were recorded. Samples were split into training (70%) and validation (30%) cohorts with an approximate 1:4 case:control ratio in both arms.

**Results:** A CPBA that combines biomarker data with patient clinical data was developed using a training cohort (469 women, cancer incidence: 18.5%), resulting in 94% sensitivity and 97% negative predictive value (NPV). Independent validation of the final algorithm in 194 subjects (breast cancer incidence: 19.6%) demonstrated a sensitivity of 95% and a NPV of 97%. When combined with previously published data for women under age 50, Videssa Breast achieves a comprehensive 93% sensitivity and 98% NPV in a population of women ages 25–75. Had Videssa Breast results been incorporated into the clinical workflow, approximately 45% of biopsies might have been avoided.

**Conclusions:** Videssa Breast combines serum biomarkers with clinical patient characteristics to provide clinicians with additional information for patients with indeterminate breast imaging results, potentially reducing false-positive breast biopsies.

## Introduction

Breast cancer is the second leading cause of cancer-related deaths in U.S. women; with 246,600 cases diagnosed and 40,450 deaths in 2016 (1); however, if diagnosed early in a localized state, 5-year survival rates are > 98% (2). The gold standard in breast cancer diagnosis remains breast imaging followed by biopsy when warranted. Breast imaging results are scored by radiologists using the ACR BI-RADS classification (3); however, there exists a significant amount of inter-reader variability (4–6). In addition, breast imaging can be impeded by

breast tissue structures, for example, dense fibrous tissue or scar tissue from prior biopsy (7–10), thus false-positives remain a significant problem in the diagnosis of breast cancer (11, 12). Improvements in clinical sensitivity tend to correspond to decreases in clinical specificity (13–15). The standard-of-care (SOC), which is watch-and-wait or biopsy for BI-RADS 3 and 4 assessments, respectively (16), can result in potentially avoidable biopsies. The American Cancer Society recently recommended that age-at-first-mammogram be changed to 45 (from 40) for women of average risk and the U.S. Preventative Services Task Force recommends screening mammography every two years starting at age 50 (17, 18). Breast cancer detection could be greatly aided by the addition of a secondary, objective assessment.

Primarily used for recurrence monitoring and as prognostic indicators, serum biomarkers may have utility in breast cancer diagnostics (19, 20). Given the complexity of breast cancer and the heterogeneity of patients with breast cancer, no single biomarker has yet achieved the clinical sensitivity and specificity appropriate to serve as an adjunct to breast imaging; a combinatorial biomarker approach is likely the best approach for reliably detecting breast cancer (21). We have previously published a proof-of-concept article demonstrating the ability of a combinatorial serum biomarker panel, Videssa Breast, to accurately detect breast cancer (22). A later study demonstrated clinical validity when the results of Videssa Breast were paired with imaging results to better inform patient follow-up for women under the age of 50

[1]Provista Diagnostics, New York, New York. [2]Brown University and Rhode Island Hospital, Providence, Rhode Island. [3]Biodesign Institute, Arizona State University, Tempe, Arizona. [4]Avera Cancer Institute, Sioux Falls, South Dakota. [5]Scripps Cancer Clinic, San Diego, California. [6]Sutter Institute for Medical Research, Sacramento, California. [7]Henry Ford Hospital & Health Network, Detroit, Michigan. [8]St. Joseph's Hospital, Nashua, New Hampshire. [9]Mayo Clinic, Scottsdale, Arizona. [10]Mayo Clinic, Rochester, Minnesota. [11]Cleveland Clinic, Cleveland, Ohio. [12]Summit Medical Group, Berkeley Heights, New Jersey.

**Note:** Supplementary data for this article are available at Clinical Cancer Research Online (http://clincancerres.aacrjournals.org/).

**Corresponding Author:** Rao Mulpuri, Provista Diagnostics, Scottsdale, AZ 85255. Phone: 480-344-9207; Fax: 602-865-7574; E-mail: mulpurir@providtadx.com

**AACR**

## Translational Relevance

While improvements in imaging have increased breast cancer detection rates, false-positive rates have increased as well. Videssa Breast is a serum biomarker–based lab-developed test (LDT) that can be used in conjunction with breast imaging to determine appropriate follow-up, potentially sparing the time, costs, and stress associated with false-positive breast imaging/biopsies. Training and validation of Videssa Breast demonstrated high negative predictive value (NPV), which can offer patients and physicians a high degree of assurance that a negative Videssa Breast outcome indicates an absence of breast cancer.

with a BI-RADS 3 or 4 assessment (23). Because serum biomarkers can be affected by menopause status and the use of hormone replacement therapy (24, 25), we chose to design separate models for two populations (with age or FSH being the cutoff) to increase overall clinical accuracy. Although the biomarkers and the associated algorithm differ between the two populations, both are referred to as "Videssa Breast."

Biomarker panel studies typically utilize samples drawn from subjects postdiagnosis. Despite the advantages, circulating biomarkers might be altered as a result of tissue damage caused by biopsy or other surgical intervention, thereby creating an aberrant biological signature that would not be useful prediagnosis. In addition, sample size is always a consideration in study design and the use of postdiagnosis serum samples permits the inclusion of a large number of cancer cases. However, this approach can result in selection bias, which often leads to a cancer incidence rate much higher than the true disease incidence, thereby drastically underestimating the proportion of false-positives. We approach this issue in a novel way by using prebiopsy samples for biomarker analysis.

In this study, we present the clinical validation of Videssa Breast, a lab-developed test (LDT), in women ages 50–75. Our primary goal was to design and validate a combinatorial proteomic biomarker assay (CPBA) that integrates patient-specific clinical data to produce a diagnostic score that distinguishes between benign conditions and breast cancer with high clinical sensitivity and a high negative predictive value (NPV). This study is unique in its use of serum samples drawn before breast biopsy, thus preserving the prediagnosis serum biomarker environment and providing realistic clinical performance metrics due to the proportion of disease and nondisease cases being representative of true disease incidence.

## Materials and Methods

### Study design and participants

The Provista-002 clinical trial (NCT02078570), sponsored by Provista Diagnostics, enrolled women ages 25–75, between April 2014 and July 2015, with enrollment capped at 1,000 subjects. Expected disease incidence was 15%, or 150 breast cancer cases. Inclusion and exclusion criteria are detailed in Supplementary Table S1. The study was approved by Institutional review board at each of the 12 U.S. clinical sites where subjects were enrolled (Supplementary Table S2). Written informed consent was obtained prior to enrollment and sample collection; the study

was designed and implemented in accordance with the Guidelines for Good Clinical Practice, with ethical principles detailed in the Declaration of Helsinki. All subjects (total $n = 1,021$) were categorized as either BI-RADS 3, 4, or 5 at the time of enrollment, as determined by mammography, ultrasound, MRI, tomography, or any combination of multiple modalities (Supplementary Figs. S1 and S2). All subjects had no personal history of breast cancer.

Of the 1,021 subjects enrolled, 30 were excluded for reasons shown in Supplementary Fig. S3. Of the remaining 991, a total of 663 subjects were assessed as being over the age of 50 or having serum FSH level > 20 mIU/mL (a biomarker for menopause; ref. 26) or both. The high FSH subjects were included to determine if better clinical performance could be attained by dividing subjects by FSH as opposed to age. Subjects under age 50 with FSH < 20 mIU/mL were not included in the current analysis because the algorithm developed previously (23) covers these subjects.

Clinicians were permitted to order follow-up imaging or surgical procedures as they deemed appropriate. While SOC was generally followed, 12% of BI-RADS 3 subjects did undergo biopsy or other procedure and 9% of BI-RADS 4 subjects did not undergo biopsy. All participants not diagnosed with breast cancer were followed for 6 months (additional 12-month follow up was available for $n = 506$ subjects), which included additional imaging and/or pathology results but did not include an additional blood draw. Blood samples were collected following BI-RADS assessment (within 28 days) and prior to biopsy (if ordered by the physician). Samples were excluded from analysis if consent was withdrawn during the study, if the sample had low volume remaining (< 2 mL), or if clinical or biomarker data were incomplete (Supplementary Fig. S3).

Videssa Breast results were not shared with clinicians to ensure that clinical decision making was unaffected. An overview of the study design is provided alongside the clinical management workflow, summarized in Supplementary Fig. S2.

### Sample collection and biomarker analysis

Following informed consent, blood was collected by the site using standard venipuncture and processed to isolate serum. All clinical sites utilized standard serum separating tubes and a standard serum collection protocol. Samples were batched and shipped by the sponsor site to Provista's laboratory in Scottsdale, AZ. Upon receipt by Provista, cryovials were accessioned and placed immediately into −80°C for storage.

Concentrations of 11 serum protein biomarkers (SPB) were determined using modified electro-chemiluminescent (ECL)-based ELISA Kits [Meso Scale Discovery (MSD)], as described previously (22, 23). Signal was detected using a Meso Sector S600 plate reader and sample concentration values were extrapolated by Discovery Workbench software (version 4.0).

Five biomarkers [two tumor-associated autoantibodies (TAAb) and three serum protein biomarkers (SPB)], along with FSH, were assessed using Abbott Architect i1000SR immunoassays, following manufacturer's specifications.

Serum was evaluated for the relative presence/absence of 34 TAAb (Supplementary Table S3) using an indirect ELISA, as described previously (22, 23). All recombinant proteins were purchased from Origene or Abnova. All samples were diluted and processed in duplicate with mean protein target values and median sample background values used for data analysis. Controls (serum positive for anti-GST or anti-myc/DDK antibodies) were included on each plate to monitor assay performance. Signal

was detected using a MSD Meso Sector S600 plate reader and Discovery Workbench 4.0 software.

All raw data from TAAb and SPB was transformed to reduce the influence of outliers and/or large values. Self-reported and physician-reported clinical information was collected for each subject. Clinical conditions and the criteria for categorization are detailed in Supplementary Table S4.

Differences in biomarker averages were noted, as expected, in subjects with benign breast conditions when divided by FSH and by age (Supplementary Table S5). These results coincide with previous studies, indicating an algorithm that includes all ages would not achieve clinical significance.

### Model development and statistical analysis

Samples were categorized on the basis of cancer status [breast cancer/ductal carcinoma in situ (DCIS) or benign-confirmed or -presumed], BI-RADS (3/4/5), and breast density (dense vs. nondense vs. not recorded). Subjects were randomized to a training or blinded validation set (70% and 30%, respectively). An approximate 1:4 case:control ratio was used for both the training and validation sets. These numbers were selected to ensure adequate sample size according to the Clinical Laboratory Standards Institute for clinical validation of a LDT.

The primary objective was to determine the clinical performance of a CPBA, Videssa Breast, in differentiating benign conditions from breast cancer in a split training–validation cohort of women over age 50 or with elevated (>20 mIU/mL) FSH. High FSH samples were included because biological features of menopause (such as FSH) have been associated with changes in circulating biomarkers (24, 25). The inclusion of both sample sets boosted sample size and permitted the ability to determine which option resulted in better clinical performance, dividing samples by age or by FSH. The cut-off value for FSH was chosen following AUC analysis (Supplementary Fig. S4).

Previous studies suggested that models developed using SPB and/or TAAb markers were capable of predicting cancer with differing performance metrics (22, 23), thus models were created comprising only SPB, only TAAb, and SPB+TAAb to learn from each marker type and improve the overall cancer prediction. Models were created using R (version 3.0.3, 2014-03-06). Confidence intervals (CIs) were reported as two-sided binomial 95% CIs.

To eliminate any features with outlying values that could potentially skew analysis, feature selection was employed using a bootstrap elastic net where 200 bootstrapped samples were drawn from the training data. For each bootstrap, generalized boosted models were created. A selection frequency of 60% was selected as the cut off for acceptance as the number of features ($p$) is less than the number of subjects ($n$; model building algorithms allow for a high number of features when $p < n$). The 60% cutoff was selected as it provides enough distinction to eliminate outlying features or features that have no predictive attributes but will keep p large enough for robustness in model building algorithms.

We first carried out a training cohort analysis (469 women, cancer incidence 18.5%) consisting of the original set of biomarkers evaluated in this study (Supplementary Table S3). Logistic models were created with clinical factors in combination with predicted probabilities to improve clinical performance. Multiple clinical factors (age, family history, BI-RADS, smoking history, and breast density) were originally assessed; only age, family history, and BI-RADS were found to have a significant impact on

the biomarker-only models. To leverage information across the different models simultaneously, an algorithmic approach was employed to combine the predicted probabilities and generate a classification for each subject. Receiver operating characteristic (ROC) analysis was employed to evaluate model performance. Sensitivity and specificity were calculated at each unique combination of the three predicted probabilities (SPB model, TAAb model, and SPB+TAAb model).

The adjusted predicted probabilities from these logistic models were evaluated to determine optimal cut-off points for prediction (maximum sensitivity and specificity) and for biopsy rule-out (sensitivity > 90% and NPV > 95%). This "rule-out" approach aimed to achieve clinical relevance of the blood test by maximizing clinicians' confidence with fewer false negatives. The final CPBA, Videssa Breast, consists of 17 biomarkers (6 SPB and 11 TAAb; Supplementary Table S3). All analyses were conducted using SAS (version 9.4) and GraphPad Prism (version 6.03). AUC comparison $P$ values were calculated as described by Hanley and McNiel (27).

Videssa Breast was validated in an independent cohort ($n = $ 194), with clinical performance assessed as above. A blinded third-party data broker handled all validation data, keeping Provista blinded to the outcomes. Training model data (i.e., biomarker composition, clinical factors, and cut-off points) were locked until clinical outcome data for the validation set was received from the data broker.

## Results

### Study population

The Provista-002 study enrolled 1,021 women, ages 25–75, from 12 domestic sites (Supplementary Table S2). All subjects were assessed as BI-RADS 3, 4, or 5 at the time of enrollment. Blood samples were collected post-BI-RADS assessment and prior to biopsy to minimize any potential confounding biological factors (Supplementary Fig. S2). Of those enrolled, 663 women were assessed as over age 50 or biologically postmenopausal (as indicated by FSH > 20 mIU/mL). These samples were split into training and validation groups (70% and 30%, respectively).

Demographics and clinical characteristics of the training and validation subjects are detailed in Table 1. No statistically significant differences were noted between the cohorts.

### Videssa breast model development

All samples were analyzed for serum biomarkers as described in the Materials and Methods. Preliminary logit boost models were built to include SPB and TAAb biomarkers, resulting in a binary outcome. Clinical factors were assessed and added as a logistic multiplier to improve clinical performance. Of all the clinical parameters evaluated, only age, family breast cancer history, and BI-RADS were found to be significant. Logistic models were created with these clinical factors in combination with the output (predicted probabilities) from the combinatorial biomarker models. The resulting model score distributions differed greatly between each BI-RADS category (Fig. 1). To avoid bias introduced by BI-RADS being included in the model, two separate cut-off points were selected, one for BI-RADS 3 subjects and one for BI-RADS 4 and 5 subjects. Cut-off points were optimized for maximum sensitivity and NPV. Scores above the cutoff were designated "high-protein signature" and scores below the cutoff were designated as "low-protein signature." The final model

**Table 1.** Characteristics and demographics of subjects used to train and validate the Videssa Breast model

|  | Training set |  | Validation set |  | P |
|---|---|---|---|---|---|
| N | 469 |  | 194 |  | — |
| Age, median (range) | 58 (40–75) |  | 58 (40–75) |  | 0.94[e] |
| Race |  |  |  |  |  |
| Caucasian | 412 | 88% | 170 | 88% | 0.27[f] |
| Black/African American | 32 | 7% | 17 | 9% |  |
| Asian | 7 | 1.5% | 5 | 2% |  |
| American Indian/Alaska Native/ Hawaiian/Pacific Islander | 7 | 1.5% | 1 | 0.5% |  |
| Other[a] | 11 | 2% | 1 | 0.5% |  |
| BI-RADS Assessment |  |  |  |  |  |
| 3 | 155 | 33% | 59 | 30% | 0.35[f] |
| 4 | 300 | 64% | 125 | 65% |  |
| 5 | 14 | 3% | 10 | 5% |  |
| Biopsied subjects[b] | 326 |  | 133 |  | 0.37[f] |
| BI-RADS 3 | (22) |  | (5) |  |  |
| BI-RADS 4 | (290) |  | (118) |  |  |
| BI-RADS 5 | (14) |  | (10) |  |  |
| Benign breast conditions | 384 |  | 156 |  | 0.66[f] |
| Procedure-confirmed benign | (237) |  | (93) |  |  |
| Presumed benign[c] | (143) |  | (61) |  |  |
| Lobular carcinoma in situ[d] (LCIS) | (4) |  | (2) |  |  |
| Breast cancer (% Incidence) | 85 | 18.1% | 38 | 19.6% |  |
| Invasive carcinoma (BC) | (50) |  | (30) |  |  |
| DCIS | (35) |  | (8) |  |  |

NOTE: Includes women over age 50 and women with high FSH.
[a]Multicultural or not reported.
[b]Includes cyst aspiration and/or biopsy.
[c]Presumed all noncancer participants to be Benign.
[d]LCIS participants were categorized as noncancer (Benign).
[e]Statistical significance assessed by unpaired $t$ test.
[f]Statistical significance assessed by Fisher exact test or $\chi^2$ (based on group size).

resulted in an AUC of 0.82 in the training cohort (Fig. 2; Supplementary Fig. S5). This was significantly greater than the AUC of the biomarker algorithm alone, for which the outcome is binary (AUC = 0.65, $P < 0.001$). The use of separate cut-off points for BI-RADS 3 and BI-RADS 4 or 5 subjects resulted in an overall sensitivity of 94% and specificity of 47% (Table 2).
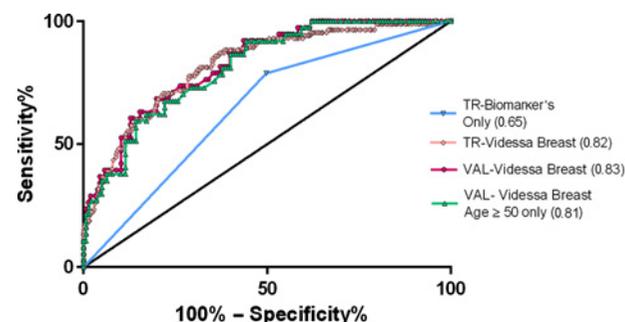
### Clinical validation

The locked, combined training model was tested on a blinded validation cohort ($n = 194$), resulting in an AUC of 0.83 (Fig. 2).
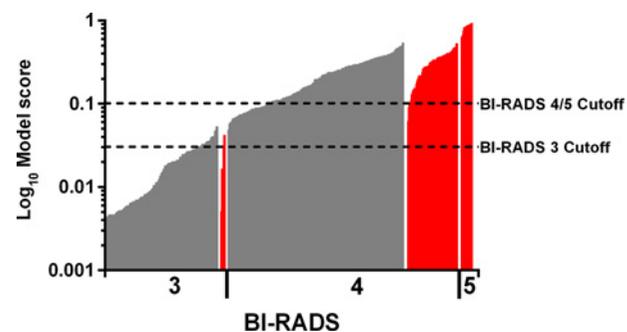


**Figure 1.**
Model score distributions by training subject BI-RADS. Subjects diagnosed with breast cancer are denoted with red bars. Because of differences in model score distributions between BI-RADS groups, two separate cut-off points were chosen (hatched lines). Any breast cancer samples below their corresponding cutoff are categorized as false-negative. Any non-breast cancer samples above their corresponding cutoff are categorized as false-positive.

Importantly, Videssa Breast resulted in only two breast cancer samples (out of 38) being scored as low-protein signature, resulting in a sensitivity of 95% and a NPV of 97% (Table 2). While high FSH subjects were initially included in this study, we acknowledge that subjects could easily be misclassified due to natural fluctuations in FSH. To better delineate an intended-use population, we note that clinical performance in the validation, age ≥ 50 years, was comparable to the performance within the entire validation cohort ($P = 0.87$). In addition, according to National Comprehensive Cancer Network (NCCN) clinical guidelines (28), subjects with a BI-RADS 5 assessment should always be recommended for biopsy. Omitting BI-RADS 5 subjects from the age



**Figure 2.**
Receiver operating characteristic (ROC) for the training and validation cohorts. Models were built using SPB and TAAb biomarkers only or biomarkers plus clinical factors (Videssa Breast). The validation cohort was assessed with all samples and in samples age ≥ 50 only ($n = 177$) and BIRADS 3 and 4 only ($n = 167$). TR, training; VAL, validation.

Henderson et al.

**Table 2.** Clinical performance characteristics of training and validation cohort samples for models consisting of biomarkers only and biomarkers with clinical characteristics (Videssa Breast)

| | Training | | Validation | | |
| | Biomarkers only | Biomarkers w/Clinical | Videssa Breast | Videssa Breast (age ≥ 50) | Videssa Breast (age ≥ 50, BI-RADS 3 & 4 Only) |
|---|---|---|---|---|---|
| TN | 193 | 179 | 64 | 54 | 54 |
| FP | 191 | 205 | 92 | 86 | 85 |
| TP | 67 | 80 | 36 | 35 | 26 |
| FN | 18 | 5 | 2 | 2 | 2 |
| Sens | 79% (68%–87%) | 94% (86%–98%) | 95% (81%–99%) | 95% (80%–99%) | 93% (75%–99%) |
| Spec | 50% (45%–55%) | 47% (42%–52%) | 41% (33%–49%) | 39% (33%–47%) | 39% (31%–48%) |
| PPV | 26% (21%–32%) | 28% (23%–34%) | 28% (21%–37%) | 29% (21%–38%) | 23% (16%–33%) |
| NPV | 92% (87%–95%) | 97% (93%–99%) | 97% (89%–99%) | 96% (87%–99%) | 96% (87%–99%) |

≥ 50, validation cohort results does not impact clinical performance (93% sensitivity, 96% NPV). Importantly, the 96% NPV means a negative test value (low-protein signature) would mistakenly call a breast cancer subject as benign (false negative) in only 4% of cases. Therefore, we define the Videssa Breast intended-use population as women over age 50 with a BI-RADS 3, 4, or 5 on imaging.

### Analysis of clinical conditions

To determine whether Videssa Breast results could be influenced by the presence of clinical conditions or comorbidities [such as dense breast tissue, prior/other cancer diagnosis, hormone replacement therapy (HRT), endocrine conditions, or heart disease], *post hoc* analyses were conducted for all samples. Clinical conditions were categorized using the criteria described in Supplementary Table S4. We noted no significant association (ANOVA, $P = 0.08$) with any clinical conditions or comorbidities (Supplementary Table S6), indicating Videssa Breast performance is not directly influenced by these conditions. Importantly, the test performed equally well (ANOVA, $P = 0.08$) in women with dense and nondense breasts (Supplementary Table S6). Because dense breasts can impede certain imaging modalities, these results suggest the inclusion of Videssa Breast results can improve breast cancer detection in women with dense breast tissue, which is a major impediment to current screening modalities.

### Combined model performance—all ages

Previous studies reported on the clinical use of Videssa Breast in women ages 25–49 (23). These samples were assessed in the current (ages 50–75) model to determine whether the algorithm would be appropriate for all ages. Of the 17 total biomarkers in the Videssa Breast model described here, six are significantly different between women under age 50 and women ages ≥ 50 (Supplementary Table S5). Because of these differences, the model designed in this study is likely not appropriate for women under age 50. As shown in Supplementary Table S7, the AUC is lower for women ages 50+, although this difference is not statistically significant ($P = 0.19$). More importantly, sensitivity is dramatically lower in women under age 50 (53.8% compared with 94.9% in ages 50+), indicating the algorithm is not ideal for use as a biopsy rule-out in an all-ages population. This confirms our previous studies, which concluded that CPBA accuracy is improved when subjects are parsed into separate age groups (22, 29).

The current Videssa Breast model for women under age 50 has a sensitivity of 88% and a NPV of 99% (23). Development of this model had included subjects from a separate clinical trial that enrolled only women under age 50. When Videssa Breast data is combined into an all-ages population ($n = 1,145$), with subjects being parsed onto separate Videssa Breast models for under/over age 50, the combined clinical performance achieves a sensitivity of 93% and a NPV of 98% (Table 3). Thus, Videssa Breast is clinically significant in a comprehensive population of women ages 25–75 with suspicious breast imaging findings.

### Comparison of Videssa breast to imaging-based assessment on medical procedure rate

Of the combined, all ages subjects ($n = 1,145$), 722 subjects were assessed as BI-RADS 3 or 4 and had undergone one or more procedures (including, but not limited to, biopsy or cyst aspiration) to obtain a confirmed diagnosis. The total number of BI-RADS 3 or 4 subjects who underwent procedures was compared with the number of subjects (within the same population) scored as high-protein signature using Videssa Breast. The difference between these two values was used to determine the percent of subjects who could have been spared from biopsy, had Videssa Breast results been used in the clinical decision-making process (Table 4). Between the training and validation cohorts, a total of 71/397 subjects > 50 years (18%, or ~1 in 5) could have been spared biopsy. Similarly, a total of 254/325 subjects < 50 years (78%) could have been spared biopsy. By combining the performance of both <50 and ≥50 models, a total of 325/722 subjects (45%) could have been spared biopsy.

## Discussion

As breast imaging has improved in sensitivity, clinical specificity has decreased (13–15). The result has been a higher number of women being recommended for biopsy. While the detection and diagnosis of breast cancer are of high importance, breast biopsy is not without risk. Ideally, an adjunct biomarker test should be able to integrate imaging and clinical information to better identify and triage women who need breast biopsy. Furthermore, such a test should perform well in women with dense breasts, which is a limitation for some imaging modalities.

**Table 3.** Combined performance of Videssa Breast in women ages 25–75 years

| $n =$ | Age < 50<br>545 | Age ≥ 50<br>600 | Combined<br>1,145 |
|---|---|---|---|
| Sens | 88% (70%–96%) | 95% (89%–98%) | 93% (88%–97%) |
| Spec | 84% (80%–87%) | 43% (38%–47%) | 64% (61%–67%) |
| PPV | 25% (18%–35%) | 29% (24%–34%) | 28% (24%–32%) |
| NPV | 99% (98%–100%) | 97% (94%–99%) | 98% (97%–99%) |

NOTE: Clinical performance of Videssa Breast for women under age 50 (published previously; ref. 23) was combined with performance data for the current algorithm.

**Table 4.** Number and percent of subjects (excluding BI-RADS 5) in whom biopsy may have been spared had Videssa Breast results been incorporated in clinical decisions

| BI-RADS | Age < 50 | | | Age ≥ 50 | | | Combined | | |
|---|---|---|---|---|---|---|---|---|---|
| | 3 | 4 | ALL | 3 | 4 | ALL | 3 | 4 | ALL |
| n = PT w/Procedure(s) | 16 | 309 | 325 | 26 | 371 | 397 | 42 | 680 | 722 |
| Recommended by Videssa Breast | 2 | 69 | 71 | 4 | 322 | 326 | 6 | 391 | 397 |
| n = Spared | 14 | 240 | 254 | 22 | 49 | 71 | 36 | 289 | 325 |
| % Spared | 88% (62%–98%) | 78% (73%–82%) | 78% (73%–83%) | 85% (65%–96%) | 13% (10%–17%) | 18% (14%–22%) | 86% (72%–95%) | 43% (39%–46%) | 45% (41%–49%) |

NOTE: 95% confidence intervals are shown in parentheses (binomial exact calculation).
Abbreviations: TR, training; VAL, validation.

For the purposes of this study, both invasive breast cancer and DCIS were categorized as breast cancer. There is debate regarding how DCIS should be followed up, as it is rare for DCIS cases to become invasive (30). We chose to include DCIS in the breast cancer group in part to err on the side of caution, but also because women diagnosed with DCIS are at greater-than-average risk for developing subsequent invasive breast cancer (31).

We have previously described the ability of Videssa Breast to differentiate breast cancer from benign conditions in women under age 50 (23). The current study demonstrates consistent clinical performance in both the training and blinded validation cohorts, indicating a strong ability to identify breast cancer in women over age 50 assessed as BI-RADS 3, 4, or 5. Clinical performance did not differ significantly between clinical subpopulations or between women with dense and nondense breasts. It did, however, underperform in women under age 50 (Supplementary Table S7), further justifying separate algorithms for the two populations. The benefit of dividing subjects by FSH as opposed to age will be evaluated further in future model development studies.

Videssa Breast is the result of a multi-center, prospective study utilizing blinded split-validation. In designing this model, we sought to maximize clinical sensitivity and NPV, to minimize false negatives when used as justification for sparing or delaying a breast biopsy. The final model resulted in 95% sensitivity and 97% NPV in the blinded validation cohort. These results were combined with our previous study (women age < 50), resulting in 93% sensitivity and 98% NPV for an all-ages population of women with BI-RADS 3, 4, or 5. As such, a negative test result would be a false negative in only 2% of cases.

When assessing biopsy-sparing potential, it should be noted that BI-RADS 5 subjects have a >95% chance of malignancy. Therefore, these subjects should always be recommended for biopsy. In contrast, BI-RADS 3 subjects have a <2% chance of malignancy. Thus, an adjunct test should be able to determine that the majority of BI-RADS 3 subjects who had received biopsies/procedures could have been spared. Also, BI-RADS 4 subjects have a chance of malignancy between 2% and 95%, which is very broad. While the a, b, and c subdivisions provide better granularity regarding malignancy risk, not all centers utilize this system and reader variability exists. Thus, an adjunct test should be able to identify BI-RADS 4 subjects who could have been spared biopsy without causing undue concern that cancer might be overlooked.

Had Videssa Breast results been integrated into the clinical process prior to biopsy, approximately 325 of 722 subjects (45%) might have been able to forego biopsy. Videssa Breast utilizes both blood-based biomarkers and SOC clinical characteristics; this combination offers distinct advantages over SOC alone in decreasing the number of potentially avoidable breast biopsies. In addition, many subjects enrolled in this study underwent multiple imaging procedures prior to and during enrollment. In the same way that Videssa Breast could spare breast biopsy, the test might also be able to decrease the number of additional imaging procedures, resulting in medical cost savings and decreased time and anxiety experienced by patients. These benefits will be evaluated in future studies.

One limitation of this study is the lack of biopsy or other procedures in a large number (204/540, or 38%) of benign subjects. This could result in sampling bias, where breast cancer was missed due to lack of a biopsy and these subjects being

categorized as false-positive with Videssa Breast. Incorporation of Videssa Breast results in clinical decision making may help in this aspect, as a patient with a BI-RADS 4 assessment and a high-protein signature Videssa Breast result may be more likely to undergo a breast biopsy. Additional studies are needed to determine how many breast cancer cases might be missed by imaging but caught with the inclusion of Videssa Breast results.

Another limitation of this study is the total number of cancer cases available in the validation cohort. Our clinical studies were designed to study women ages 25–75. Clinical trial enrollment resulted in >150 breast cancer cases (consistent with the predicted breast cancer incidence of 15%). However, the unexpected need to develop separate models based on age resulted in a smaller number of breast cancer cases ($n = 123$) being available for the current study involving women of ages 50–75 years. Statistical power is further limited when considering only invasive breast cancer cases, as 38% of the breast cancer cases in the current study were DCIS. Overall, the number of breast cancer cases would be a limitation for a stand-alone test and/or to FDA approval or clearance of an *in vitro* diagnostic. However, the number of cancer cases studied, particularly when two age groups are combined ($n = 170$), are adequate to support the use of an LDT (developed by following the standards of CAP/CLIA) as an adjunct to the current SOC. Ongoing studies are being conducted to validate the assay in independent cohorts of different ethnic/demographic groups.

Another study limitation is the lack of serial sample collection. Many subjects were diagnosed not at enrollment, but on follow-up, which occurred 6–12 months later. Indeed, of the 6 total false-negative subjects (age $\geq 50$), three were diagnosed with breast cancer or DCIS on a follow-up visit. Absent the collection of serial samples, we cannot determine whether the Videssa Breast outcome for these subjects would have changed to a high-protein signature around the time of the follow-up visit (which would have categorized them as true positive instead of false-negative). Serial protein signature data are available for a separate clinical trial and will be assessed in future studies. We will also seek to evaluate how a subject's biomarker values change over time and the influence, if any, on Videssa Breast results. In addition, we have yet to determine whether any lead-time bias exists. With these limitations, it is recommended that physicians follow-up with high-protein signature patients per their routine practice.

## Conclusions

Videssa Breast is a noninvasive LDT assay that combines serum biomarkers and patient clinical characteristics. When used in conjunction with breast imaging results, Videssa Breast improved breast cancer detection compared with mammography alone in women over age 50 with a BI-RADS 3 or 4 assessment.

## Disclosure of Potential Conflicts of Interest

M.C. Henderson and E.E. Letsios hold ownership interest (including patents) in Provista Diagnostics. J. LaBaer holds ownership interest (including patents) in and is a consultant/advisory board member for Provista Diagnostics. K.S. Anderson is a consultant/advisory board member for Provista Diagnostics. No potential conflicts of interest were disclosed by the other authors.

## Authors' Contributions

**Conception and design:** M.C. Henderson, M. Silver, E.E. Letsios, R. Mulpuri, D.E. Reese, K.S. Anderson
**Development of methodology:** M.C. Henderson, M. Silver, R. Mulpuri, D.E. Reese, K.S. Anderson
**Acquisition of data (provided animals, acquired and managed patients, provided facilities, etc.):** M.C. Henderson, E.E. Letsios, R. Mulpuri, A.P. Lourenco, J. Alpers, C. Costantini, H. Ali, K. Baker, D.W. Northfelt, K. Ghosh, S.R. Grobmyer, W. Polen
**Analysis and interpretation of data (e.g., statistical analysis, biostatistics, computational analysis):** M.C. Henderson, M. Silver, E.E. Letsios, R. Mulpuri, D.E. Reese, A.P. Lourenco, J. LaBaer, K.S. Anderson, J. Alpers, D.W. Northfelt, J.K. Wolf
**Writing, review, and/or revision of the manuscript:** M.C. Henderson, M. Silver, Q. Tran, R. Mulpuri, D.E. Reese, J. LaBaer, K.S. Anderson, J. Alpers, C. Costantini, N. Rohatgi, D.W. Northfelt, S.R. Grobmyer, W. Polen, J.K. Wolf
**Administrative, technical, or material support (i.e., reporting or organizing data, constructing databases):** M. Silver, Q. Tran, E.E. Letsios, R. Mulpuri, N. Rohatgi
**Study supervision:** M.C. Henderson, Q. Tran, R. Mulpuri, A.P. Lourenco, N. Rohatgi, S.R. Grobmyer

## References

1. Howlader N, Noone A, Kraphco M, Miller D, Bishop K, Kosary CL, et al. SEER Cancer Statistics Review, 1975–2013 (Updated SEPT 2016). Bethesda, MD: National Cancer Institute; 2015. Available from: https://seer.cancer.gov/csr/1975_2013/.
2. American Cancer Society. Breast cancer facts and figures 2011–2012. Atlanta, GA: American Cancer Society; 2011.
3. D'Orsi C, Sickles E, Mendelson E, Morris E. ACR BI-RADS® Atlas, Breast Imaging Reporting and Data System. Reston, VA: American College of Radiology; 2013.
4. Michaels AY, Chung CSW, Frost EP, Birdwell RL, Giess CS. Interobserver variability in upgraded and non-upgraded BI-RADS 3 lesions. Clin Radiol 2017;72:694.e1–694.e6.
5. Lee AY, Wisner DJ, Aminololama-Shakeri S, Arasu VA, Feig SA, Hargreaves J, et al. Inter-reader variability in the use of BI-RADS descriptors for suspicious findings on diagnostic mammography: A Multi-institution Study of 10 Academic Radiologists. Acad Radiol 2017;24:60–6.
6. Gur D, Bandos AI, Cohen CS, Hakim CM, Hardesty LA, Ganott MA, et al. The "laboratory" effect: comparing radiologists' performance and variability during prospective clinical and laboratory mammography interpretations. Radiology 2008;249:47–53.
7. Boyd NF, Guo H, Martin LJ, Sun L, Stone J, Fishell E, et al. Mammographic density and the risk and detection of breast cancer. N Engl J Med 2007; 356:227–36.
8. van Breest Smallenburg V, Duijm LE, Voogd AC, Jansen FH, Louwman MW. Mammographic changes resulting from benign breast surgery impair breast cancer detection at screening mammography. Eur J Cancer 2012;48:2097–103.
9. Carney PA, Miglioretti DL, Yankaskas BC, Kerlikowske K, Rosenberg R, Rutter CM, et al. Individual and combined effects of age, breast density, and hormone replacement therapy use on the accuracy of screening mammography. Ann Intern Med 2003;138: 168–75.

10. Taplin SH, Abraham L, Geller BM, Yankaskas BC, Buist DS, Smith-Bindman R, et al. Effect of previous benign breast biopsy on the interpretive performance of subsequent screening mammography. J Natl Cancer Inst 2010;102:1040–51.

11. Pace LE, Keating NL. A systematic assessment of benefits and risks to guide breast cancer screening decisions. JAMA 2014;311:1327–35.

12. Bleyer A, Welch HG. Effect of three decades of screening mammography on breast-cancer incidence. N Engl J Med 2012;367:1998–2005.

13. Berg WA, Zhang Z, Lehrer D, Jong RA, Pisano ED, Barr RG, et al. Detection of breast cancer with addition of annual screening ultrasound or a single screening MRI to mammography in women with elevated breast cancer risk. JAMA 2012;307:1394–404.

14. Knuttel FM, Menezes GL, van den Bosch MA, Gilhuijs KG, Peters NH. Current clinical indications for magnetic resonance imaging of the breast. J Surg Oncol 2014;110:26–31.

15. Kim SA, Chang JM, Cho N, Yi A, Moon WK. Characterization of breast lesions: comparison of digital breast tomosynthesis and ultrasonography. Korean J Radiol 2015;16:229–38.

16. Breast Cancer Screening and Diagnosis, Version 2.2106. NCCN Clinical Practice Guidelines in Oncology. https://www.nccn.org/store/Login/Login.aspx. 2017.

17. Buist DS, Porter PL, Lehman C, Taplin SH, White E. Factors contributing to mammography failure in women aged 40-49 years. J Natl Cancer Inst 2004;96:1432–40.

18. Fedewa SA, de Moor JS, Ward EM, DeSantis CE, Goding Sauer A, Smith RA, et al. Mammography use and physician recommendation after the 2009 U.S. Preventive Services Task Force Breast Cancer Screening Recommendations. Am J Prev Med 2016;e123–31.

19. Harris L, Fritsche H, Mennel R, Norton L, Ravdin P, Taube S, et al. American Society of Clinical Oncology 2007 update of recommendations for the use of tumor markers in breast cancer. J Clin Oncol 2007;25:5287–312.

20. Füzéry AK, Levin J, Chan MM, Chan DW. Translation of proteomic biomarkers into FDA approved cancer diagnostics: issues and challenges. Clin Proteomics 2013;10:13.

21. Hollingsworth AB, Reese DE. Potential use of biomarkers to augment clinical decisions for the early detection of breast cancer. Oncol Hematol Rev. 2014;10:103–9.

22. Henderson MC, Hollingsworth AB, Gordon K, Silver M, Mulpuri R, Letsios E, et al. Integration of serum protein biomarker and tumor associated autoantibody expression data increases the ability of a blood-based proteomic assay to identify breast cancer. PLoS One 2016; 11:e0157692.

23. Lourenco AP, Benson KL, Henderson MC, Silver M, Letsios E, Tran Q, et al. A non-invasive blood-based combinatorial proteomic biomarker assay to detect breast cancer in women under the age of 50 years. Clin Breast Cancer 2017;17:516–25.

24. da Silva LH, Panazzolo DG, Marques MF, Souza MG, Paredes BD, Nogueira Neto JF, et al. Low-dose estradiol and endothelial and inflammatory biomarkers in menopausal overweight/obese women. Climacteric 2016;19:337–43.

25. Pradhan AD, Manson JE, Rossouw JE, Siscovick DS, Mouton CP, Rifai N, et al. Inflammatory biomarkers, hormone replacement therapy, and incident coronary heart disease: prospective analysis from the Women's Health Initiative observational study. JAMA 2002;288:980–7.

26. Reyes FI, Winter JS, Faiman C. Pituitary-ovarian relationships preceding the menopause. I. A cross-sectional study of serum follicle-stimulating hormone, luteinizing hormone, prolactin, estradiol, and progesterone levels. Am J Obstet Gynecol 1977;129:557–64.

27. Hanley JA, McNeil BJ. The meaning and use of the area under a receiver operating characteristic (ROC) curve. Radiology 1982;143:29–36.

28. Bevers TB, Anderson BO, Bonaccio E, Buys S, Daly MB, Dempsey PJ, et al. NCCN clinical practice guidelines in oncology: breast cancer screening and diagnosis. J Natl Compr Cancer Netw 2009;7:1060–96.

29. Weber D, Grimes R, Su P, Woods R, Baker P. Age-stratification's role in cytokine based assay development. Anal Methods 2010;2:653–6.

30. Morrow M, Katz SJ. Addressing overtreatment in DCIS: what should physicians do now? J Natl Cancer Inst 2015;107:djv290.

31. Narod SA, Iqbal J, Giannakeas V, Sopik V, Sun P. Breast cancer mortality after a diagnosis of ductal carcinoma *in situ*. JAMA Oncol 2015;1:888–96.

# Clinical Cancer Research

# A Noninvasive Blood-based Combinatorial Proteomic Biomarker Assay to Detect Breast Cancer in Women over age 50 with BI-RADS 3, 4, or 5 Assessment

Meredith C. Henderson, Michael Silver, Quynh Tran, et al.

| | |
|---|---|
| **Updated version** | Access the most recent version of this article at:<br>doi:10.1158/1078-0432.CCR-18-0843 |
| **Supplementary Material** | Access the most recent supplemental material at:<br>http://clincancerres.aacrjournals.org/content/suppl/2018/09/05/1078-0432.CCR-18-0843.DC1 |

| | |
|---|---|
| **Cited articles** | This article cites 27 articles, 1 of which you can access for free at:<br>http://clincancerres.aacrjournals.org/content/25/1/142.full#ref-list-1 |
| **Citing articles** | This article has been cited by 1 HighWire-hosted articles. Access the articles at:<br>http://clincancerres.aacrjournals.org/content/25/1/142.full#related-urls |

| | |
|---|---|
| **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, use this link<br>http://clincancerres.aacrjournals.org/content/25/1/142.<br>Click on "Request Permissions" which will take you to the Copyright Clearance Center's (CCC)<br>Rightslink site. |