

REVIEW ARTICLE

Selecting postoperative adjuvant systemic therapy for early stage breast cancer: A critical assessment of commercially available gene expression assays

David M. Hyams MD¹ | Eric Schuur PhD² | Javier Angel Aristizabal MD³ |
Juan Enrique Bargallo Rocha MD⁴ | Cesar Cabello MD⁵ | Roberto Elizalde MD⁶ |
Laura García-Estévez MD⁷ | Henry L. Gomez MD⁸ | Artur Katz MD⁹ |
Aníbal Nuñez De Pierro MD¹⁰

¹ Desert Surgical Oncology, Eisenhower Medical Center, Rancho Mirage, California

² VMWA LLC, Palo Alto, California

³ Instituto Nacional de Cancerología, Bogota, Colombia

⁴ Instituto Nacional de Cancerología, Mexico City, Mexico

⁵ Universidade Estadual de Campinas, Campinas, Brazil

⁶ Hospital Dr. Ignacio Pirovano, Buenos Aires, Argentina

⁷ Centro Integral Oncológico Comprensivo Clara Campal, Madrid, Spain

⁸ Instituto Nacional de Enfermedades Neoplásicas, Lima, Peru

⁹ Hospital Sírio-Libanés, Sao Paulo, Brazil

¹⁰ Hospital Dr. Juan A. Fernandez, Buenos Aires, Argentina

Correspondence

David M. Hyams, MD, Desert Surgical Oncology, Eisenhower Medical Center, Suite K-207, 39000 Bob Hope Drive, Rancho Mirage, CA 92270. Email: dhyams.md@me.com

Risk stratification of patients with early stage breast cancer may support adjuvant chemotherapy decision-making. This review details the development and validation of six multi-gene classifiers, each of which claims to provide useful prognostic and possibly predictive information for early stage breast cancer patients. A careful assessment is presented of each test's analytical validity, clinical validity, and clinical utility, as well as the quality of evidence supporting its use.

KEYWORDS

adjuvant, biomarkers, breast neoplasms, chemotherapy, clinical decision-making, gene expression, genomics risk, prognosis, tumor

1 | INTRODUCTION

For 20 years, breast cancer mortality rates have steadily fallen.¹ Although adjuvant chemotherapy (CTx) has been a strong contributor to the overall effect, only a small proportion of

patients individually benefit from the treatments they receive.² Commercially available gene expression assays are promoted as a window into intrinsic tumor biology allowing a more accurate determination of prognosis and treatment benefit than traditional clinicopathologic classifiers.

Haddow and Polamaki's framework for the evidence-based evaluation of genomic assays identifies components essential for any new test.^{3,4} Analytic validity represents an assay's ability to accurately and reproducibly measure biomarkers of interest. Clinical validity requires the assay to provide accurate and reliable information regarding outcomes of relevance.⁵ Clinical utility reflects an assay's ability to favorably alter patient outcome.

[Correction added on 7 March 2017, after first online publication: In the first paragraph of the Oncotype DX section, 180 was replaced with 18.]

This is an open access article under the terms of the Creative Commons Attribution-NonCommercial License, which permits use, distribution and reproduction in any medium, provided the original work is properly cited and is not used for commercial purposes.

© 2017 The Authors. *Journal of Surgical Oncology* Published by Wiley Periodicals, Inc.

The Tumor Marker Utility Grading System (TMUGS) grades evidence on a progressive scale.⁶ The best evidence comes from completely prospective, controlled clinical trials (Level A). A lower level of evidence (LOE) are prospective-retrospective studies (Level B) with a prospective study design of archived tissue from a previous prospective clinical trial.⁷ Two concordant Level B trials can achieve Level I evidence. Prospective evaluation of previously collected patient registry data (level C), represents a lower LOE, but provides opportunity to confirm findings from higher LOE studies. Post hoc investigations, using convenience samples, in the absence of defined treatment, data acquisition, or follow-up, are prone to bias and represent the lowest LOE (level D). These should only be used for general confirmation of higher-level data or for hypothesis generation. TMUGS is summarized in Table 1.

Currently there are six commercial gene expression assays available worldwide for patients with early stage breast cancer (ESBC). This review identifies the relevant supporting data for these assays graded by TMUGS LOE.

2 | MATERIALS AND METHODS

2.1 | Article sources

A systematic search of the scientific literature was conducted in September 2016 that identified primary published studies describing the clinical validation of one of the six commercially available genomic classifiers for breast cancer (Table 2.) Databases searched included Medline and the Cochrane Library. In addition, relevant previously published review articles were consulted to be certain that appropriate reports were not overlooked.

2.2 | Search strategy

A keyword search strategy was implemented for the initial search of databases that included "BCI," "Breast Cancer Index," "HOXB13/IL17BR," "MGI," "Molecular Grade Index," "IHC4," "Oncotype," "21 gene," "Recurrence Score," "MammaPrint" "70 gene," "Blueprint," "TargetPrint," "PAM 50," "intrinsic subtype," "Endopredict," and "Prosigna."

2.3 | Study selection and data extraction

Only English-language publications were selected. Studies analyzed original data relevant to clinical validity or clinical utility and included a formal statistical test of pre-specified outcome endpoints. Epidemiological, exploratory, and purely descriptive studies were excluded. As this review focuses on decision-making in the immediate postoperative adjuvant setting, we have not included an evaluation of gene expression assays in the setting of neoadjuvant or extended adjuvant therapy, even when the LOE was high (e.g., Sgroi et al.⁸). Excluded also were publications not relating to female invasive breast cancer and studies without peer-reviewed publication. Final reports published in E-format were eligible. All eligible articles were fully reviewed by the investigators to determine that the study met the prespecified requirements. The PRISMA flow diagram for this review is presented in Fig. S1.

2.4 | Study analysis

A total of 34 articles were identified for inclusion in this analysis (Table S1). Data from each article selected was captured in a structured abstraction form. The framework from Simon and colleagues was used to evaluate the LOE of the identified studies.⁷ Each element of the

TABLE 1 Revised determination of levels of evidence, incorporating type of biomarker studies performed

Level	Category	Type of biomarker study performed	Validation studies
I	A	Prospective design to answer a biomarker question with predefined treatment, assessment, and specimen collection. Biomarker measurement and analysis is prospective with a contemporary statistical analysis plan (SAP).	None required
I	B	Prospective design to answer a biomarker question using a previously conducted prospective study with collection and archiving of tissue in the original study. Biomarker measurement and analysis is prospective with a contemporary SAP.	One or more with consistent results
II	B	Prospective design to answer a biomarker question using a previously conducted prospective study with collection and archiving of tissue in the original study. Biomarker measurement and analysis is prospective with a contemporary SAP.	None or inconsistent results
II	C	Prospective registry or case series designed to address a biomarker question, with prospective enrollment and tissue collection. Treatment and follow-up are standard-of-care. Biomarker measurement and analysis is prospective using a contemporary SAP.	Two or more with consistent results
III	C	Prospective registry or case series designed to address a biomarker question, with prospective enrollment and tissue collection. Treatment and follow-up are standard-of-care. Biomarker measurement and analysis is prospective using a contemporary SAP.	None, or one with consistent results, or inconsistent results
IV-V	D	Retrospective study design with convenience sample of archived specimens, without prospective patient recruitment. There is no defined treatment or follow-up. Biomarker measurement and outcome analysis does not use a prospective contemporary SAP.	Not applicable- Never satisfactory

Source: Simon RM, et al J Natl Cancer Inst. 2009;101:1446-1452.

TABLE 2 Assay platform summary

Characteristic	Breast cancer index	Endopredict	IHC4	MammaPrint	OncotypeDX breast cancer assay	Prosigna
Vendor	bioTheranostic	Myriad Genetics	Genoptix	Agendia BV	Genomic Health, Inc.	Nanostring, Inc.
Validated in	ER+, LN-, ESBC treated with TAM	ER+ HER2- ESBC treated with TAM from ABCSG-6 (tam arm only) and ABCSG-8	ER+ ESBC treated with TAM or anastrozole	Stage I, II; ≤5.0 cm LN-; ER+ or ER-; HER2: negative or positive; < 53 years of age	ER+ Node + or -, ESBC treated with TAM	ER+, node negative or positive, ESBC in post-menopausal women
Genes included	HOX B13/IL17BR ratio, Molecular Grade Index (5 genes) (Jenkowitz BCR 2011)	9 risk-related plus 3 normalization; Combined with clinical indicators for EPClin score (Filipits Clin Cancer Res 2011)	4 breast cancer related (Cuzick J Clin Oncol 2011)	70 (van de Vijver N Engl J Med 2002)	5 proliferation-related, 2 invasion-related, 4 ER-related, 5 Other, 5 reference (Paik N Engl J Med 2004)	50 PAM50-related, 8 normalization (Nielsen BMC Cancer 2014)
Assay processing	Central by vendor	Central by vendor	Local laboratories	Central by Vendor	Central by Vendor	Local Laboratories
Analytical platform	RT-PCR	RT-PCR	Immunohistochemistry	Microarray, central lab	RT-PCR	nCounter system from Nanostring Technologies, Inc.
Sample requirements	FFPE block or slides	FFPE block or slides	FFPE slides or tissue microarrays	FFPE block or slides	FFPE block, slides, cores	FFPE blocks or slides
Assay readout	Two risk categories	Continuous, two risk categories	Continuous risk score	Two risk categories	continuous risk score, three risk categories	Three risk categories plus continuous risk score
Referencing guidelines	ASCO (ER+ LN- ESBC, for prognosis); St. Gallen (ESBC for prognosis); ESMO (ER+ primary breast cancer for prognosis); St. Gallen (ESBC for prognosis); AGO (ER+ LN- ESBC for prediction)	ASCO (ER+ LN- ESBC, for prognosis); St. Gallen (ESBC for prognosis); ESMO (ER+ primary breast cancer for prognosis and prediction); AGO (ER+ LN- ESBC for prediction)	None	ESMO (primary breast cancer for prognosis and prediction), St. Gallen (ER+ ESBC, for prognosis and prediction), NCCN (ESBC for prognosis)	ESMO (ER+ primary breast cancer for prognosis and prediction), St. Gallen (ER+ ESBC for prognosis and prediction), NCCN (ER+ ESBC for prognosis and prediction), ASCO (ER+ LN- ESBC, for prognosis and prediction), NICE (ER+ LN- ESBC for prognosis and prediction), AGO (ER+ LN- ESBC for prognosis)	St. Gallen (ESBC for prognosis), NCCN (ER+ ESBC for prognosis), AGO (ER+ LN- ESBC for prognosis)
Commercial availability	Yes	Yes	Yes	Yes	Yes	Yes

reviewed articles was given a rating according to the best fit against the four study types of the Simon model (Table 1). If individual elements of a single study were not all rated as the same category, the lowest category of an individual element was used as the study's overall rating.

3 | RESULTS

Key elements of each study and the assigned Simon LOE are summarized in Table 3.

3.1 | Breast cancer index

The Breast Cancer Index (BCI) is a multi-gene quantitative reverse transcriptase-polymerase chain reaction (RT-PCR) assay with two components, the HOXB13:IL17BR ratio (H/I) and the Molecular Grade Index (MGI).⁹ The H/I ratio was derived by evaluating differential gene expression in 60 tamoxifen-treated (TAM+) estrogen receptor-positive (ER+) postmenopausal ESBC patients.¹⁰ The MGI component of BCI was developed by identifying genes differentially expressed between tumors of low and high grade, focusing on those involved in invasive growth.^{9,11} The H/I ratio and the MGI have been the subject of independent prognostic validation studies. As the individual components are not used separately in commercial assays, their supportive studies are not reviewed here.^{9,12,13}

The BCI assay was trained in 314 samples of postmenopausal TAM+ ER+ lymph node-negative (LN-) ESBC patients from the Stockholm trial of adjuvant tamoxifen (TAM). This resulted in a continuous risk score from 0 to 10.¹⁴ The Stockholm trial, conducted between 1976 and 1990, randomized postmenopausal LN- patients to TAM for 2-5 years, versus no treatment.¹⁵ BCI was validated in 274 patients from the tamoxifen-untreated (TAM-) arm of the trial. BCI was significantly prognostic ($P < 0.001$) for distant recurrence (DR), independent of tumor size, grade, progesterone receptor (PR) status, and HER2/neu status.

A new BCI algorithm was subsequently developed from 283 postmenopausal subjects using the TAM- arm of the same Stockholm trial, and then validated in 317 TAM+ trial patients.¹⁶ Use of the same patient cohorts in these two separate studies, with different algorithms, diminishes the independence of the two studies and lessens their value for clinical validation.^{14,16,17} Patients were classified into low (64%), intermediate (20%), and high-risk (16%) groups using the new algorithm with 5-year distant recurrence free survival (DRFS) of 98%, 95.2%, and 87.8%. There was a significant ($P = 0.0063$) association of risk group with DRFS. External validation was performed in a multi-institutional convenience sample of 358 pre- (30%) and post-menopausal patients with ER+ LN- tumors.¹⁶ Treatment was heterogeneous in this group, with 32% having received CTx on an ad hoc basis. Univariate, and Cox multivariable analysis, demonstrated BCI as the only significant prognostic indicator among age, tumor size, tumor grade, PR-status, or CTx treatment in this sample.

Sgroi et al evaluated 665 primary tumor samples from postmenopausal ER+ LN- patients entered into the randomized ATAC trial of

TAM and anastrozole.⁸ The RNA analyzed had been prepared by Genomic Health for another study with the ATAC group using its proprietary technology. In the primary analysis, use of the cubic model of the BCI as a continuous variable was not associated with DR (interquartile HR 1.39 [95%CI 0.990-3.70]; $LR-\Delta\chi^2 = 3.70$; $P = 0.054$), even though the pre-specified categorical risk groups were associated with DR ($P < 0.0001$). In a secondary analysis, a revised linear model of BCI was utilized, demonstrating significant differences in absolute DR rate among the BCI risk groups ($P < 0.0001$). DR at 10 years for the low, intermediate, and high-risk groups was 4.8% (95%CI, 3.0-7.6), 18.3% (95%CI, 12.7-25.8), and 29% (95%CI, 21.1-39.1), respectively.

The cubic model of BCI was again tested in 292 patient samples from both arms of the NCIC MA.14 clinical trial of TAM and octreotide, versus TAM alone.¹⁸ Octreotide proved to be an inactive agent in the parent study. All patients were postmenopausal and 92% had ER+ disease. A total of 51% were LN- and 35% received adjuvant CTx. BCI risk groups had a significant univariate association with relapse free survival (RFS) (stratified log-rank $P = 0.005$). A stratified Cox stepwise multivariate model only demonstrated a significant hazard ratio for T-stage $\geq T2$ (HR, 2.22 95%CI 1.22-4.07; $P = 0.01$) and for higher continuous BCI ($P = 0.004$). The BCI low-risk (49.7%), intermediate-risk (23.6%), and high-risk (26.7%) groups had 10-year RFS rates of 87.5%, 83.9%, and 74.7%, respectively. BCI was prognostic for both LN- and lymph node-positive (LN+) patients.

There are no high LOE studies of BCI that predict the value of initial postoperative adjuvant ETx or CTx in ESBC patients.

3.2 | Endopredict

The Endopredict assay was developed using microarray gene expression data from 964 ER+, HER2/neu-negative (HER2-), LN- tumors from TAM+ patients.¹⁹ RT-PCR expression of eight cancer-related genes and three standardization genes were combined algorithmically to produce a continuous risk measurement. A cutpoint separated high-risk and low-risk groups, which was also combined with nodal status and tumor size to calculate an EPclin score.

Clinical validation using a prospective-retrospective design evaluated 1702 postmenopausal ER+, LN- or LN+ samples from ETx-only individuals in the ABCSG-6 (378 patients) or ABCSG-8 (1324 patients) trials.¹⁹ ABCSG-8 included subjects with low to moderate risk features, excluding grade three tumors. The study suggested that EP and EPclin scores were significant predictors of DR. The 10 year DR rates for patients with EP-low and EP-high were 8% and 22% in ABCSG-6 ($P < 0.001$) and 6% and 15% in ABCSG-8 ($P < 0.001$). More than 51% of the combined study population were EP-high despite the low to moderate clinical risk features of the ABCSG-8 patients. Nonetheless, multivariable Cox models demonstrated continuous EP score and nodal status to be independent predictors of DR.

When the same postmenopausal ABCSG-6 and ABCSG-8 populations were assessed in a study by Dubsy, EPclin, produced a greater difference between risk groups (HR 5.11, 95%CI 3.48-7.51; log rank $P < 0.001$) than National Comprehensive Cancer Network (NCCN) 2007 guidelines (HR 2.16, 95%CI 0.80-5.85; $P = 0.119$), German S3 2008 guidelines (HR 2.20, 95%

TABLE 3 Summary of discussed validation studies with assigned Simon Hayes level of evidence

Assay	Focus	Study	Setting	Pre-menopausal	Post-menopausal	ER+	ER-	LN+	LN-	Simon category
Blue print										
Prognostic										
		Krijgsman Breast Cancer Res Treat 2011	Various	✓		✓	✓	✓	✓	D
Breast Cancer Index										
Prognostic										
		Goetz Clin Cancer Res 2006	AET	✓		✓	✓	✓	✓	B
		Jerevall Br J Cancer 2011	AET	✓		✓	✓	✓	✓	B
		Sgroi Breast Cancer Res 2016	AET	✓		✓	✓	✓	✓	B
Prognostic and predictive										
		Sgroi Lancet Oncol 2013	AET	✓		✓	✓	✓	✓	B
		Zhang Clin Cancer Res 2013	AET	✓		✓	✓	✓	✓	B
Endopredict										
Prognostic										
		Filipits Clin Cancer Res 2011	AET	✓		✓	✓	✓	✓	B
		Dubsky Ann Oncol 2013	AET	✓		✓	✓	✓	✓	B
		Martin Breast Cancer Res 2014	AET/ACT	✓		✓	✓	✓	✓	B
		Buus J Natl Cancer Inst 2016	AET	✓		✓	✓	✓	✓	B
IHC4										
Prognostic										
		Cuzick J Clin Oncol 2011	AET	✓		✓	✓	✓	✓	B
		Park Oncology 2014	AET/ACT	✓		✓	✓	✓	✓	C
		Sgroi Lancet Oncol 2013	AET	✓		✓	✓	✓	✓	B
		Stephen Br J Cancer 2014	AET	✓		✓	✓	✓	✓	B
MammaPrint										
Prognostic										
		Bueno-de-Mesquita Lancet Oncol 2007	AET/ACT	✓		✓	✓	✓	✓	D
		Buyse J Natl Cancer Inst 2006	No Rx	✓		✓	✓	✓	✓	C
		Cardoso N Engl J Med 2016	AET/ACT	✓		✓	✓	✓	✓	A
		Mook Ann Oncol 2010	AET	✓		✓	✓	✓	✓	D
		Mook Breast Cancer Res Treat 2009	AET/ACT	✓		✓	✓	✓	✓	D
		van de Vijver N Engl J Med 2002	AET/ACT	✓		✓	✓	✓	✓	D
		Wittner Clin Cancer Res 2008	AET/ACT	✓		✓	✓	✓	✓	D

(Continues)

TABLE 3 (Continued)

Assay	Focus	Study	Setting	Pre-menopausal	Post-menopausal	ER+	ER-	LN+	LN-	Simon category
Oncotype	Predictive	Knauer Breast Cancer Res Treat 2010	ACT	✓	✓	✓	✓	✓	✓	D
		Prognostic								
PAM50	Predictive	Dowsett J Clin Oncol 2010	AET		✓	✓		✓	✓	B
		Paik N Engl J Med 2004	AET	✓	✓	✓		✓	✓	B
		Sparano N Engl J Med 2015	AET	✓	✓	✓		✓	✓	A
		Gluz J Clin Oncol 2016	AET	✓	✓	✓		✓	✓	A
PAM50	Prognostic and predictive	Albain Lancet Oncol 2010	AET/ACT		✓	✓				B
		Predictive								
PAM50	Prognostic	Paik J Clin Oncol 2006	AET/ACT	✓	✓	✓			✓	B
		Chia Clin Cancer Res 2012	AET	✓		✓	✓	✓	✓	B
		Dowsett J Clin Oncol 2013	AET		✓	✓		✓	✓	B
		Gnant Ann Oncol 2014	AET		✓	✓		✓	✓	B
		Gnant Ann Oncol 2015	AET		✓	✓		✓	✓	B
PAM50	Predictive	Cheang Clin Cancer Res 2012	ACT	✓		✓	✓	✓	✓	B
		Liu Breast Cancer Res Treat 2015	ACT	✓	✓	✓	✓	✓	✓	B

AET, adjuvant ETx; ACT, adjuvant CTx; ER+, estrogen receptor positive; ER-, estrogen receptor negative; LN+, lymph node positive; LN-, lymph node negative; LOE, level of evidence (Simon J Natl Cancer Inst 2009).

CI 1.16-4.19; $P = 0.014$), or St. Gallen 2011 guidelines (HR 2.78 95% CI 1.50-5.14; $P < 0.001$).²⁰ EPclin reassigned 58 (61%) of patients guideline-classified as high-risk to EPclin low-risk. However, the clinical guidelines of the era classified over 80% of the patients as high-risk to begin with. Ultimately, the authors suggested that the preponderance of truly low-risk patients in this cohort limited the statistical power of the study, proposing that the analyses for the remaining high-clinical-risk subgroup be regarded as exploratory. The authors suggested CTx for the high-risk group, but there was no evidence presented to support EPclin as a predictor of CTx benefit.

Martin studied the prognostic properties of EP and EPclin in an independent cohort of 555 patients with ER+ LN+ disease from the 1246 patient prospective GEICAM 9906 study treated with CTx and ETx.²¹ Both EP and EPclin-defined risk groups were significantly associated with distant metastasis free survival (DMFS) (log rank $P < 0.0001$). For EP, 25% (141) of evaluable patients were low-risk with estimated DMFS at 10 years of 93%. Among the 75% (414) of high-risk patients, 10-year DMFS was 70%. The absolute difference in risk between low-risk and high-risk EP groups was 23% (HR, 4.8 (95%CI 2.45-9.55), $P < 0.0001$). Using EPclin, only 13% were low-risk with 100% DMFS at 10 years. The remaining high-risk patients had a 10-year DMFS of 72%.

GEICAM 9906 included 300 (54%) premenopausal and 255 (46%) postmenopausal patients, with prognostic value evident in both groups (premenopausal, HR, 6.7; $P < 0.0001$; postmenopausal, HR, 3.3; $P = 0.0069$). EPclin was similarly prognostic for both premenopausal ($P = 0.0006$) and postmenopausal subsets as well. ($P = 0.0023$). As all patients in GEICAM 9906 received CTx, EP as a predictor of CTx benefit could not be separately assessed. To date, there are no LOE studies of EP or EPclin that identify a role for the assays in predicting CTx or ETx benefit.

The prognostic capability of EP and EPclin were compared to Oncotype DX using 928 RNA samples from the ATAC study, previously prepared by Genomic Health.²² In a multivariate analysis, the EP, EPclin, and the Oncotype DX assay were all prognostic for 10-year DR. EP provided more prognostic information than the Oncotype DX Recurrence Score (RS) in the 5-10 year interval, but the assays were equally prognostic from 0 to 5 years. Clinical factors in the EPclin score resulted in prognostic information that exceeded both the RS and EP. However, clinical factors were not generated by the assay and were not similarly incorporated into the RS data, making any direct comparison of EPclin and Oncotype DX problematic.

The genomic component of these assays sorts patients into low and high-risk groups of different sizes: EP classified more patients as high-risk (58.4%) compared to the Oncotype DX assay, which classified 38.3% with a "non-low" RS (RS ≥ 18). The hazard ratio for 10-year DRS between the low-risk and non-low risk patients was 2.98 (95%CI 1.94-4.58, $P < 0.001$) for EP compared to HR = 2.73 (95%CI 1.91-3.89, $P < 0.001$) for Oncotype DX). Although the HR's are similar and the confidence intervals overlap, a larger proportion of EP patients would be considered high-risk and might receive recommendation for CTx.

Martin, et al, more recently evaluated patients from the same GEICAM 9906 study and compared prognostic features of EP and

EPclin with a non-commercial research version of PAM50.²³ However, the comparison does not represent a real-world evaluation of the Prosigna assay, providing little value for practical decision-making.

3.3 | IHC4

IHC4 is an integrated immunohistochemistry (IHC) assay that utilizes four breast cancer protein biomarkers (ER, PR, HER2, and Ki-67) to assess risk in patients with ER+ ESBC.²⁴ Initial analysis was performed in a single laboratory using unique antibodies for IHC, careful H-scoring for ER, and specialized image analysis. Expression levels of these proteins are incorporated into the IHC4 algorithm. A clinical treatment score (CTS), derived from clinicopathologic features, may be added to obtain an overall risk score. The assay is performed with formalin-fixed paraffin-embedded (FFPE) tissue.

Development and validation was based on FFPE tissue samples from the ATAC trial.²⁴ The biomarkers were evaluated by likelihood ratio testing and shown to contribute prognostic information to time to DR (TTDR). In the ER+, LN- discovery cohort of ATAC, IHC4 coupled with CTS had greater prognostic value than CTS alone. (IHC4 LR- $\chi^2 = 29.3$, 95%CI 27.7-30.3).

IHC4 was also evaluated in a convenience sample of 786 ER+, post-menopausal, and pre-menopausal ESBC patients from Nottingham who were either untreated or treated with ETx alone.²⁴ Due to manual reading and the use of a different antibody, Ki-67 levels were nearly 2.5 times higher than in the ATAC cohort, requiring rescaling. After adjustment, the modified IHC4 score was prognostic for outcome when added to the clinical score (HR, 3.9; $P < 0.0001$).

Sgroi et al evaluated IHC4 in Cuzick's LN- ATAC population, the same population in which IHC4 was developed, to confirm the prognostic value of the immunohistochemical assay.^{8,24} In the 915 patients, the authors examined the likelihood ratio for both IHC4 and BCI; IHC4 appeared prognostic for recurrence to 5 years in multivariable analysis. Stephen and colleagues also showed that IHC4 was prognostic to 5 years using independent Cox analysis in samples from the Edinburgh BCI series and TEAM trial.²⁵ However, others have shown that a nomogram constructed from St. Gallen guidelines and Adjuvant Online (AOL), combined with IHC4 + CTS, provides prognostic information beyond that from IHC4 + CTS alone.²⁶⁻²⁸

IHC4 has not been validated to predict therapeutic benefit for ETx or CTx adjuvant regimens. Furthermore, Cuzick and colleagues have acknowledged that variations in consistency and qualitative output across distributed laboratories have limited the value of the assay.²⁴ Commercialization of the assay by Genoptix, may address some of these issues, but validation data for the commercial assay is currently unavailable.

3.4 | MammaPrint

MammaPrint is a microarray-based assay that uses the expression levels of 70 selected genes to classify tumors as "poor signature" or "good signature." Investigators used frozen tumor samples from 78 women with ESBC from archives at the Netherlands Cancer Institute

(NKI) to produce the RNA used to develop the prognostic signatures. Patients were less than 55 years old, and included 34 with distant metastases within 5 years. Early DR (<5 years) was chosen as a discriminator.²⁹

Clinical validation was performed in 295 consecutive NKI patients, younger than 53 years of age with ER+ (76.6%) and ER- (23.4%) early stage breast cancer.³⁰ There was significant heterogeneity among this population, with 49% LN+, 44% having received some form of adjuvant ETx and/or CTx. Nearly 21% (61/295) were in the original developmental group of patients, introducing potential bias. Overall, poor-signature patients had a higher rate of distant metastasis at 10 years (HR, 5.1, 95%CI (2.9-9.0); $P < 0.001$) than good-signature patients with DMFS of $50.6 \pm 4.5\%$, and $85.2 \pm 4.3\%$, respectively.

A subsequent validation study used 307 samples from ER+ (71%) and ER- (29%) subjects ≤ 60 years of age enrolled in the TRANSBIG registry who had not received adjuvant therapy.³¹ In this study, 85 of the 90 patients with ER- disease had a high-risk gene signature. The HR for overall survival (OS) by risk group was 2.79 (95%CI 1.60-4.87, $P < 0.001$). Also significant was TTDR (HR, 2.32, 95%CI 1.35-4.00, $P = 0.002$) and disease free survival (DFS) (HR, 1.50, 1.04-2.16 $P = 0.032$). In each case the 70-gene assay provided more robust prognostic information than clinicopathologic assessments, including AOL.

MammaPrint was also assessed in LN+ subjects from a multi-institution convenience sample that identified significant association between the gene signature and BCSS (log rank $P < 0.001$) or 10-year DMFS (log rank $P = 0.001$).³² In another convenience sample, Wittner demonstrated a negative predictive value of 100% in 100 LN- postmenopausal patients from the Massachusetts General Hospital. However, the corresponding positive predictive value was only 12%. Furthermore, there was no significant difference in TTDR between the gene signature groups.³³ Mook, used a sample of postmenopausal LN- patients from the NKI to show significant association between the gene signature and BCSS (log rank $P = 0.036$), but not DMFS (log rank $P = 0.07$).³⁴

MammaPrint is one of only two assays for which there is published Level I-A evidence supporting prognosis.³⁵ The MINDACT trial was a multicenter prospective randomized study that enrolled 6693 women with invasive ESBC, after screening 11 288. Of the 4595 screening failures, 26% were reported due to technical failure of the assay. Of enrolled patients, all underwent clinical risk assessment with a modified version of AOL, and had genomic risk assessment performed with MammaPrint. Low-clinical risk was defined as a 10-year BCSS in excess of 88% for ER+ patients, and of 92% for ER- patients. The difference between the ER+ and ER- thresholds was to account for a presumed absolute 4% 5-year benefit of adjuvant ETx.

The MINDACT study utilized a complex design with multiple arms, randomizations, and study subpopulations. Analysis was made more challenging by risk assessment errors in 275 patients (4% of the total population). The clinical risk assessment was faulty in 103 cases, largely due to changes in local site assessment of tumor size, grade, or nodal status. However, 177 patients had a change in genomic assessment due to an assay quality control error. Additional complexity was introduced by 15% patient noncompliance with treatment assignment.

The primary endpoint of MINDACT was survival without distant metastasis at 5 years. This was assessed in a pre-defined "primary test population" derived from clinically high-risk genomic low-risk (cHgL) patients. The 644 patients in the primary test population, representing 9.6% of MINDACT, were cHgL and did not receive CTx. Excluded were 21 patients who had changes in risk assessment and 85 patients who did not comply with treatment recommendations. The rate of survival free of distant metastasis was 94.7% (95%CI, 92.5-96.2), exceeding the prospectively set non-inferiority criterion of 92% and suggesting prognostic value for the MammaPrint assay.

Several studies have suggested an association between MammaPrint risk signature and CTx benefit.^{36,37} In a retrospective meta-analysis study by Knauer using pooled data from previous non-trial clinical reports of adjuvant therapy, DDFS was 93% without CTx and 99% with CTx in good-signature patients (HR, 0.26, 95%CI 0.03-2.02; $P = 0.20$), and 76% versus 88%, in poor signature patients (HR, 0.35, 95%CI 0.17-0.71; $P < 0.01$).³⁶ The HR reduction was similar for the two groups, despite differences in significance, in this very underpowered study. The small sample sizes and the absence of randomized controlled CTx administration in any of the component studies limit any conclusion. Straver analyzed stage II and III patients with response to neoadjuvant chemotherapy assessed as a function of MammaPrint risk grouping. However, neoadjuvant studies cannot be directly extrapolated to predict degree of long-term benefit in the absence of late outcome information.³⁷

The prospective observational RASTER study, conducted in 16 community hospitals in the Netherlands, has been proposed to support MammaPrint use for therapy selection.³⁸ However, treatment decisions were made by the patient and treating physician, in a discretionary fashion, based on a combination of MammaPrint, prognostic indicators, treatment guidelines, and personal preferences. The heterogeneity of treatment and the absence of MammaPrint as the treatment discriminator make interpretation of the outcome data difficult.

MINDACT has provided the only high LOE data available to assess MammaPrint's role in therapeutic treatment benefit. CTx or no-CTx, was determined by randomization of 2745 patients in each of the discordant clinical and genomic risk assessment arms, on an "intent to treat" basis. The 5-year DMFS rate among the cHgL patients receiving CTx was 95.9% (95%CI, 94.0-97.2) and 94.4% (95%CI, 92.3-95.9) for those not receiving CTx. A 1.5% advantage in 5-year DMFS (with overlapping confidence intervals) was noted for those receiving CTx. In the per protocol population, the hazard rate reduction for CTx in the cHgL group ranged from 34 to 37% for DMFS, DFS, and OS. Notable was the statistically significant impact on DFS in this underpowered analysis (HR 0.64; 95%CI, 0.43-0.95, $P = 0.03$). For the clinical low-risk genomic high-risk (cLgH) discordant group, there was no apparent advantage to genomic assay-directed CTx, with 5-year survival free of distant metastasis 95% for both. In the per protocol population, there was no statistical advantage in DMFS, DFS, or OS for CTx administration. When reviewing both discordant groups, these data cannot exclude a small advantage for CTx in the cHgL group, but suggest no benefit for the use of chemotherapy in the cLgH group. As a result, MINDACT data provides no evidence of MammaPrint predictive CTx value in poor prognosis patients.

Two companion microarray-based assays are marketed with MammaPrint. TargetPrint offers quantitative assessment of ER, PR, and HER2 expression; Blueprint, when used with MammaPrint, classifies tumors into molecular subsets.^{27,30,39} Concordance between Blueprint/MammaPrint and the original intrinsic gene set from Perou et al is 92%.³⁹ However, analysis of the data suggests that these approaches characterize overlapping but distinct populations of luminal patients.

There are no studies with long-term outcome endpoints to assess the value of the MammaPrint/Blueprint combination. Whitworth has suggested a value in intrinsic subset reassignment with a resultant improvement in neoadjuvant pCR rates, evaluated in 403 patients from the NBRST prospective registry.⁴⁰ However, the study had no long-term outcome endpoints, relying only on the surrogate endpoint of pCR. Variations in the degree of pCR cannot be correlated to quantitative outcome differences across varied populations, particularly in hormone receptor positive (HR+) HER2- breast cancer.⁴¹

3.5 | Oncotype DX

Oncotype DX is an RT-PCR-based assay that measures expression levels of 21 genes, in RNA from FFPE tissue, to assess the risk of DR in ER+ ESBC. The assay was developed from 250 prognostic breast cancer genes, using a total of 447 patients from three independent studies.⁴² The 21 genes included 16 cancer-related genes, and five reference genes. A weighted algorithm produced a continuous RS from 0 to 100 which can be used to assign patients to low (0-17), intermediate (18-30), or high-risk (31) categories. Alternate risk cutpoints have been proposed.^{43,44}

Validation of Oncotype DX was performed in a prospective-retrospective study using 668 archived tissue samples from the TAM+ arm of the NSABP B-14 trial.⁴² The 10-year DR free rate (DRFR) was 93.2% in the low-risk group (51% of patients) and 69.5% in the high-risk group (27% of patients) ($P < 0.001$). Multivariate Cox Proportional Hazards Analysis demonstrated that likelihood of DR was associated with RS (HR, 3.21 per 50 units, 95%CI 2.23-4.61; $P < 0.001$), independent of age or tumor size. Continuous RS was a significant predictor of DR and OS ($P < 0.001$).

A second prospective-retrospective validation study of prognosis analyzed 1231 FFPE tissue samples from the ATAC trial of postmenopausal women with ER+, LN- and LN+ tumors who had received either 5 years of anastrozole or TAM.⁴⁵ Among 872 LN- patients there was a significant association of the RS with DR risk (HR per 50 units, 3.92, 95%CI 2.08-7.39; $P < 0.001$) in a multivariable model that adjusted for the effects of tumor size, local grade, age, and treatment. The assay had similar value in distinguishing TTDR for the LN+ population (HR, 3.47, 95%CI 1.64-7.38; $P = 0.002$). The 9-year DR risks were similar between LN- and 1-3 LN+ patients, when the RS was < 18 .

The prognostic value of Oncotype DX was further evaluated in LN+ postmenopausal patients using two arms of the SWOG 8814 study.⁴⁶ Of 367 patients with samples available, 148 received TAM+ only, while 219 received cyclophosphamide-doxorubicin-fluorouracil (CAF) CTx, followed by TAM+. Oncotype DX RS, evaluated in the TAM+ only group,

was prognostic for DFS. Cox regression analysis showed that the continuous RS was a significant predictor of DFS (HR per 50 units, 2.64, 95%CI 1.33-5.27; $P = 0.006$). The continuous RS was also prognostic for OS over a 10-year period, with HR 4.42 (95%CI 1.96-9.97; $P < 0.001$) adjusting for nodes.

The highest level of evidence for prognosis currently available comes from the prospective Cooperative Group trial, TAILORx, which was initiated in 2006 and which was specifically designed to evaluate Oncotype DX.⁴⁴ The study population of 10 253 eligible ER+, HER2-, LN- Oncotype DX tested women was divided into three predetermined risk groups with RS cutpoints at 11 and 25. Cutpoints lower than the prospectively defined cutpoints for the initial clinical validation studies were utilized to avoid patient undertreatment by accounting for the ends of the 95% confidence intervals for 10-year risk of recurrence rather than just the means. Patients in the RS < 11 group were assigned ETx alone, those in the mid-range group with RS 11-25 were randomized to CTx followed by ETx versus ETx alone, and those in the RS > 25 group were assigned CTx followed by ETx. Results from the mid-range group randomization are expected in 2017. However, data from the low-risk group of patients with RS < 11 have been published, providing TMUGS Level I-A clinical evidence for prognosis in this population.⁴⁴ Of the 1626 low-risk patients who were assigned ETx alone, the 5-year invasive DFS was 93.8% (95%CI, 92.4-94.9), the proportion free of DR was 99.3% (95%CI, 98.7-99.6), and the OS was 98.0% (95%CI, 97.1-98.6). For DFS, second primary breast cancers and deaths from other causes exceeded the 5-year risk of recurrence. Multivariate analysis that included age, tumor size, histologic grade, and surgery type showed no significant association with the rate of invasive DFS or freedom from DR. Only histologic grade showed any statistically significant association ($P = 0.02$) with freedom from any recurrence between intermediate and low grade (HR 8.07; (95%CI 1.06-61.45), and between high and low grade (HR 4.73; 95%CI 0.29-76.42).

Prospective data are also available from the West German Study Group Plan B trial.⁴⁷ This study, initiated in 2009, recruited 3198 patients. It was designed to assess alternative CTx regimens in patients with clinically high-risk LN- and LN+ ESBC. Oncotype DX testing was successfully performed in 2577 patients, representing 98% of the total HR+ population with tumor tissue samples. After an early protocol amendment, ETx treatment alone was recommended and accepted by 348 HR+, HER2- patients with pN0-1 disease and RS ≤ 11 . Although 31.4% were LN+ and 20% had grade three disease, the 3-year DFS rate was 98.4% (95%CI 97.0-99.8%). Among the intermediate-risk group (RS12-25), the 3-year DFS was 97.5% (95%CI, 95.9-99.0), and among the high-risk group (RS > 25) the 3-year DFS was 94.9% (95%CI, 91.4-98.4), with all patients in both groups assigned adjuvant CTx. Although only a 3-year DFS endpoint has been published, the authors suggested that the favorable outcome with ETx alone in the low-risk group precluded any potential benefit of CTx.

The value of the Oncotype DX assay as a predictor of benefit from systemic chemotherapy was addressed in two separate prospective-retrospective studies, yielding Level 1-B validation evidence from both studies. Paik analyzed samples from ER+ LN- subjects in the NSABP B-20 trial that had randomized ER+ LN- patients to TAM, versus TAM

plus CTx.⁴⁸ Subjects in the high-risk group (RS \geq 31) had a significantly reduced DR when CTx was utilized (RR, 0.26; 95%CI, 0.13-0.53). In contrast, there was no significant benefit of CTx for the low (RS < 18) (RR, 1.31; 95%CI, 0.46-3.78) or intermediate-risk (RS18-30) (RR, 0.61; 95%CI, 0.24-1.59) groups. Cox models demonstrated significant interaction between the continuous RS and CTx ($P = 0.038$).

Albain analyzed the Oncotype DX assay in 367 (40%) available samples from the SWOG-8814 study randomized to sequential CAF and TAM+, versus TAM+ alone.^{46,49} Log rank test of CTx benefit, stratified by number of positive nodes, revealed a trend toward a DFS benefit for the entire population of analyzed patients (stratified log rank $P = 0.054$) consistent with the parent trial. However, when analyzed by the Oncotype DX test, DFS improvement for patients receiving CAF was confined to the high-risk group with RS \geq 31 (log rank $P = 0.033$; HR = 0.59, 95%CI 0.35-1.01). There was no apparent DFS benefit for the addition of CTx to patients with RS < 18 (log rank $P = 0.97$; HR = 1.02, 95%CI 0.54-1.93) or RS18-30 (log rank $P = 0.48$; HR = 0.72, 95%CI 0.39-1.31). OS benefit was similarly restricted to the high-risk group (stratified log rank $P = 0.027$), without improvement for the low (stratified log rank $P = 0.63$) or intermediate-risk (stratified log rank $P = 0.85$) groups.

3.6 | Prosigna

The Prosigna breast cancer test is a multiplex assay that measures the expression of 50 classification genes and eight housekeeping genes in order to assign a "risk of recurrence" (ROR) score to patients with ESBC. The assay, initially termed PAM50, was adapted and commercialized by Nanostring, Inc. for use in their nCounter analytic system, consisting of a single hybridization reaction with specific labeled probes.⁵⁰ Gene expression of 46 of the genes is used to compare a specimen to prototypical intrinsic biologic subsets originally described by Perou.⁵¹ A proprietary algorithm combines the gene expression data, a proliferation score, and tumor size to calculate an ROR score that corresponds to the 10-year risk of DR. Patients are assigned to low, intermediate, or high-risk groups, based on pre-determined cut-points. Unlike other tests, the cutpoints vary by clinicopathologic feature. All patients with ≥ 4 LN+ are considered high-risk. Patients with 1-3 LN+ and ROR ≤ 15 , or LN- patients with ROR ≤ 40 , are classified as low-risk. Satisfactory analytic validity has been reported when the test is conducted using standardized procedures in local molecular biology laboratories.⁵²

Dowsett evaluated Prosigna in 940 RNA samples previously extracted by Genomic Health from the ATAC trial.^{45,53} The prospective-retrospective study demonstrated that a continuous ROR score added prognostic value for DR beyond CTS for all patients in the LN- and LN+, ER+, postmenopausal population treated with ETx (LR- $\Delta\chi^2 = 33.9$, $P < 0.001$). In addition, Kaplan-Meier analysis showed ROR risk groups were prognostic for 10-year DR, varying by LN status.

Two prospective-retrospective reports of Prosigna utilized RNA and data from the previously described ABCSG-8 study.^{54,55} Each study analyzed the ABCSG-8 population of postmenopausal patients with grade 1-2 tumors either alone or with additional patients from ATAC. In their primary validation study, Gnant and colleagues analyzed 1478 patients from ABCSG-8 and developed a prognostic clinical

linear predictor (CLP) based on tumor grade, tumor size, and nodal status.⁵⁴ The ROR score provided a significant additional increase in prognostic information over the CLP (log -likelihood test: $\Delta LR\chi^2 = 53.49$; $P < 0.0001$). Subgroup analysis suggested that additional information was provided for all LN- and LN+ patients except those with HER2-positive disease. Pre-specified ROR risk groups were also prognostic for DRFS with a 10-year probability of 96.7% (95%CI 94.6-98.0) for the low-risk group, 91.3% (95%CI 88.1-93.8) for the intermediate-risk group, and 79.9% (95%CI 75.7-83.4) for the high-risk group.⁵⁴ Classification to luminal A versus luminal B subtype was prognostic for 10-year DMFS (HR, 2.85; 95%CI 2.04-4.00; $P < 0.0001$).

Evaluation of ROR in LN+ ER+ post-menopausal women was conducted in relevant patients from ABCSG-8 and ATAC. Analysis revealed prognostic value for DMFS in both the single LN+ subgroup (high- versus low-risk: HR, 3.56, 95%CI 1.62-7.80, $P = 0.0016$) and the 2-3 LN+ subgroup (high- versus low/intermediate-risk: HR, 3.023, 95%CI 1.462-6.249, $P = 0.0028$) using pre-defined cutpoints.⁵⁵ However, the LN+ number itself had influence on subgroup assignment beyond the actual gene expression though its impact on the ROR subgroup cutpoint. Intrinsic subtype analysis revealed improved DMFS for patients with Luminal A versus Luminal B tumors, in both N1 and N2-3 subgroups.

Chia further evaluated PAM50 in 398 premenopausal Stage I-III HR+ and hormone receptor-negative (HR-) patients who received adjuvant CTx as part of NCIC MA.12.⁵⁶ Unlike the commercial assay, RT-PCR rather than NanoString n-Counter technology was used. HR+ individuals represented 73% of patients, while 94% were either Stage I or II. Nearly 75% had positive nodes. In adjusted multivariate analysis PAM50 intrinsic subtyping was prognostic beyond clinicopathologic variables for both DFS ($P = 0.02$) and OS ($P = 0.02$). Continuous ROR, combined with a proliferation score and tumor size correlated with 10-year relapse in both LN- and LN+ patients using ROC analysis (DFS C-index = 57.6%, 95%CI 52.5-62.1%, and for OS C-index = 61.1%, 95%CI 55.8-66.3%). However, the study's primary focus was on intrinsic subtype classification and did not adequately pre-specify the specific ROR model that would be utilized. The study did suggest prognostic value for PAM50 in CTx treated premenopausal populations. However, as CTx was mandated in all patients, impact of CTx on outcome cannot be fully appreciated.

Cheang evaluated the prognostic properties of the quantitative RT-PCR version of PAM50 in 476 LN+ premenopausal patients from the NCIC MA.5 randomized controlled trial who received one of two CTx regimens.⁵⁷ Correlation between the assay, expressed as ROR-subtype, and the clinical outcome showed the ROR risk classifier was associated with a distinct 5-year RFS and OS prognostic advantage ($P < 0.0001$; log rank test). However, as all of the randomized arms of the original study received some form of CTx, any predictive value of ROR on selecting adjuvant PCTx, could not be assessed.

The only large prospective-retrospective study to evaluate PAM50 for adjuvant taxane therapy benefit prediction was published by Liu and associates, based on the NCIC CTG MA.21 clinical trial.⁵⁸ The original study enrolled 2104 patients, with 1105 available for

PAM50 analysis. Pre-menopausal and post-menopausal women up to 60 years of age with LN+, or high-risk LN- breast cancer were enrolled between 2000 and 2005. Patients were randomized to doxorubicin, cyclophosphamide, and paclitaxel (AC/T); dose-intense cyclophosphamide, epirubicin, and fluorouracil (CEF); or dose-dense, dose-intense epirubicin, cyclophosphamide, and paclitaxel (EC/T). None of the patients were randomized to ETx without CTx. CEF and EC/T proved superior to AC/T ($P = 0.01$). Although continuous ROR was associated with poorer RFS ($P = 0.03$), categorical ROR comparing the high ROR with the combined low and intermediate ROR risk groups was neither prognostic nor predictive. The lack of prognostic performance of the categorical ROR may have been related to the finding that 78.7% of patients were classified into the ROR high-risk group and only 3.4% were identified as ROR low-risk. The PAM50-based intrinsic subtypes also provided no predictive value for treatment selection or benefit. Only in the evaluation of RFS through multivariate analysis did intrinsic subtype and continuous ROR demonstrate prognostic, but not predictive, associations ($P = 0.03$ and $P = 0.002$, respectively).

4 | DISCUSSION

Molecular assays provide an opportunity for more accurate assessment of tumor biology than traditional clinicopathologic classifiers. The six commercially available gene expression assays reviewed in this report all provide prognostic information designed to estimate an individual's risk of tumor recurrence and/or mortality. Some manufacturers also suggest a predictive role for their tests in selecting among different systemic therapy options. However, each assay is based on different gene sets, has been developed in a unique population of varying heterogeneity, and was tested in studies of varying quality.

When assessing a genomic test, an appreciation of the analytic validity, clinical validity, and clinical utility are paramount. Peer-reviewed published evidence, rather than podium presentations or a marketing clearance, should be the prime consideration. Studies should be logical and easily interpreted, with unequivocal primary and secondary endpoints. Any assessment of chemotherapy benefit prediction should be conducted in a randomized clinical trial that compared patients treated with chemotherapy to patients who did not receive chemotherapy. Although differences in neoadjuvant pCR may suggest assay value, differences in the *rate* of pCR do not directly correlate with long-term outcome, particularly when assessing heterogeneous populations.^{59,60} Clinical endpoint data should capture short-term and relevant long-term cancer events.

Analytic validation reflects the standardized performance and reproducibility of a specific assay technique. Ideally, an assay's methodology should be fixed prior to conducting definitive clinical validation studies so that the validation evidence adequately reflects performance of the commercial test. Dissonance between a validation test and the commercially marketed assay is illustrated by the BCI, which has undergone several iterations. The first two key BCI studies used opposite arms of the same Stockholm trial for development and

validation.^{14,16} This type of crosstalk between the studies weakens their independence and, therefore, their validation value. Furthermore, the assays developed from each strategy were, in effect, different tests. The third validation study in ATAC patients utilized two different approaches to scoring (BCI-cubic and BCI-linear), with the BCI-cubic (primary analysis) a non-significant prognostic test, and the BCI-linear (secondary analysis) highly prognostic. Although the ATAC BCI evaluation by Biotheranostics used commercial assay protocols, Genomic Health had performed the original ATAC RNA isolation with its own proprietary technology. As a result, BCI, as well as subsequent Prosigna and EP evaluations of patient cohorts from ATAC, incorporated a technique not reproduced in their ultimate commercial test.^{8,22,53} Similar variations between validation study procedure and the commercial assay are noted in the Prosigna test. The final Prosigna assay gene panel differed from the initial PAM50 assay by several gene panel deletions and additions, and the assay was performed using RT-PCR in some validation studies and using NanoString technology in others.^{56,61}

IHC4 analytic validity has also come under scrutiny. Because of the variability in IHC assays, particularly Ki-67, performed locally with varying reagents, and cutpoints, a fixed commercial version has been developed, marked by Genoptix as NexCourse Breast. However, there is no published validation to link the Genoptix NexCourse Breast commercial results with the previously published IHC4 validation data.

Although the genes selected and the algorithm utilized for the MammaPrint assay have remained unchanged throughout its development, there have been significant procedural changes, including a shift from fresh frozen/RNA-preserved tissue to formalin-fixed/paraffin-embedded tissue, and variations in the microarray plates utilized between studies and the commercial test. A study performed by Sapino and colleagues evaluated MammaPrint-FFPE to MammaPrint-fresh tissue in a series of 211 matched tissue samples.⁶² The two techniques demonstrated a concordance of categorical low-risk and high-risk classification of 91.5% (95%CI 86.9-94.5%); however, the clinical implications of having this level of discordance between platforms is unclear. A similar analysis was performed by Beumer and colleagues analyzing 552 MammaPrint-FFPE and MammaPrint-fresh sample pairs.⁶³ The results for the two types of tissue preservation had a Pearson correlation coefficient of 0.93 (95%CI 0.92-0.94) with the majority of the discordant samples in both studies within a 10% range of the cutpoints. This discordance, typically centering about the cutpoints, calls into question the strategy of a strict high-risk/low-risk reporting model promoted for this assay.

Of all the commercially available assays, only the Oncotype DX test made use of the commercial assay procedure for its entire suite of validation testing. The Oncotype DX assay was developed in FFPE tissue and all validation studies have used the same algorithm. Although clinical parameters may be used with the assay for clinical treatment decision-making, these parameters were not required for the assay's validation and are not incorporated into the RS calculation. Analytic validity should reflect both the quality controls related to pathology, histology, RNA extraction, and other pre-assay procedures and the precision capabilities of the test itself.

EPclin and Prosigna ROR Score add clinical parameters to the core genomic assay in order to produce a prognostic test with added precision. However, the primary purpose of genomic assays is to provide new information not otherwise available by an assessment of common clinical parameters. Clinical features should always be considered for ultimate real-world decision-making, but should be added to, rather than intrinsic to, the genomic assay. Combined scores may be misleading as noted in comparisons of EP and EPclin with the Oncotype DX test result.²² The Kaplan-Meier analysis showed a similar HR separation between low and non-low risk groups for both the EP and Oncotype DX tests for 10 years. The Oncotype DX assay characterized 48% more low-risk patients than EP alone, but only 5% more than EPclin. However, the one-sided use of recognized clinical prognostic features limits the value of such evaluations and provides little information about the quality of the core genomic component.

An important component of analytic validity is the ability of the assay to produce a reliable result. In the MINDACT trial, 26% of ineligible patients, or 10% of the total patients screened, suffered an analytic failure of the assay. Few of the remaining assays have published failure rates. However, the final analytic failure rate of the Oncotype DX test has been reported at less than 1% of 103 863 sequential specimens.⁶⁴ The MammaPrint assay has also suffered an important quality control (QC) lapse that was unrecognized for several months. This led to misreporting of results at both its European and American laboratory sites. There was a resulting over-reporting of risk and inappropriate chemotherapy recommendation for hundreds of patients. This quality control issue led to an FDA recall for the California laboratory, and complicated the conduct of the MINDACT assay in Europe.^{35,65} Although the error was reportedly due to a supplier change in a reagent used for RNA extraction, the episodes demonstrate the importance of internal ongoing QC checks for single tests that may have a dramatic impact on patient care.

Clinical validation requires clinical trials that are consistent with an assay's target population and that address relevant clinical questions. The study population must be sufficiently well defined to allow reasonable translation to common disease settings. Outcome statistics should not be confounded by widely varying treatments and group heterogeneity. Given the value of ETx in HR+ ESBC, appropriate prognostic validation studies in this population should include ETx as a base upon which chemotherapy decisions may be evaluated.^{66,67}

An example of confounding population heterogeneity is evident in the original MammaPrint validation study that included an assortment of different patient populations receiving an assortment of varying treatments.³⁰ Of the 295 patients in the study, 144 were node positive and 151 were node negative. ER+ was a characteristic in 69, with 226 ER-. Nearly 37% happened to receive adjuvant CTx, while 56% had been untreated. No patient treatments were randomized or protocol-directed. In the subsequent study of Buyse, none of the patients received any systemic adjuvant treatment.³¹ In the BCI validation study, using the MA.14 clinical trial, 35% of patients received adjuvant CTx, confounding the prognostic value of the assay. As a result, the impact of CTx on the prognostic value of BCI remains uncertain.¹⁸

The TMUGS process for assessing genomic assays, is designed to standardize validation by assigning levels of evidence to all supporting

published studies.^{6,7} The best support for any gene expression assay comes from a prospectively designed and prospectively conducted study. A single well-designed prospective study, conducted in a relevant population, with sufficient follow-up, can serve as the highest level of evidence. However, the cost of such studies may be prohibitive, and they may not always be feasible. The prospective-retrospective study provides an alternative for developing high level of evidence, when confirmed by two or more such studies with consistent results.

To date, only three completely prospective studies have been reported (TAILORx, PlanB and MINDACT). Peer-reviewed publications are available for the low-risk population of TAILORx (RS < 11), the German Plan B (RS ≤ 11) trial, and for several subsets of MINDACT.^{35,44,47} TAILORx and the Plan B studies are based on the Oncotype DX assay, whereas MINDACT is based on the 70-gene MammaPrint assay. All of the studies demonstrate excellent outcomes for CTx untreated patients in their pre-determined genomic low-risk populations. Prior prospective-retrospective studies have suggested that there is no benefit to CTx in the low-risk Oncotype DX group. However, MINDACT suggests that there may be a small CTx benefit (approximately 1.5% absolute) in the MammaPrint-determined genomic low-risk discordant group, with an associated improvement in DFS. In an editorial accompanying the MINDACT publication, Hudis suggested that this small benefit may be meaningful for some patients.⁶⁸ Beyond MINDACT, there are no TMUGS-defined high-LOE studies for MammaPrint. Although suggestive of prognostic value, the remaining studies are flawed by suboptimal design, population heterogeneity, as well as significant potential for treatment and detection bias. The remaining four genomic assays do have high-level evidence from at least one or more of prospective-retrospective studies suggesting their prognostic value. However, validation has been in varied patient populations without purely prospective trial data.

In the past year, a published prospective registry study of more than 38 000 Oncotype DX-evaluated ER+, LN-, HER2- patients in the United States produced results that paralleled those of prior prospective-retrospective studies.⁶⁹ In the SEER database, 21 023 low-risk LN- Oncotype DX patients (RS < 18) had a 5-year breast cancer specific mortality (BCSM) of 0.4% (95%CI, 0.3% -0.6%) with only 7% known to have received CTx.⁶⁹ Even among SEER patients with positive nodes (microscopic to three nodes), 5-year BCSM was only 1.0% (95%CI, 0.5-2.0) with 23% known to have received CTx. This registration study provides "real world" confirmation of the clinical utility of the Oncotype DX assay, and provides outcomes evidence that is remarkably consistent with results from earlier validation studies.

But, is the identification of very low-risk group by a well-validated genomic assay sufficient? Should all remaining patients above an assay's low-risk cutoff be expected to benefit from CTx? Prospective-retrospective data used to evaluate the Oncotype DX assay in the NSABP B-20 study showed no benefit for CTx in patients with low or intermediate RS results. However, there was a disproportionate benefit of CTx in patients with high-risk RS results (relative risk 0.26; 95%CI 0.13-0.53), and a statistically significant test for interaction

between chemotherapy treatment and the RS result. This relationship between the RS and CTx benefit was also found in the prospective-retrospective SWOG 8814 LN+ study with CTx/ETx compared to ETx alone.⁴⁶ Together these studies suggest a patient population (RS \geq 31) for whom chemotherapy may be remarkably active, providing recurrence risk reductions of up to 75%. As a quantum transition of CTx benefit at RS \geq 31 is unlikely, this data suggest that some small portion of the intermediate-risk population might be CTx responsive, perhaps at the highest RS levels within this group. However, the majority of intermediate group patients (RS18-30) do not appear to benefit from adjuvant CTx.

None of the remaining assays has TMUGS high-level evidence to convincingly predict CTx benefit in subgroups of tested subsets. MINDACT provided an opportunity to prospectively demonstrate the predictive value of MammaPrint. However, an analysis of the cHgL group showed a consistent risk reduction of 35% when chemotherapy was used in this group. This benefit was statistically significant for DFS, even in this underpowered subgroup. Conversely, the cLgH group provided no evidence of a benefit to chemotherapy, even when driven by the high-genomic risk. Thus, although 48% of the cHgL patients may potentially be spared chemotherapy, 47% of the cLgH patients might be recommended to receive unnecessary chemotherapy, without change in outcome measure. Although a very low-risk group for any genomic assay might not have sufficient risk or benefit to justify chemotherapy, predictive value does matter with increased risk, providing the Oncotype DX assay with the strongest clinical utility argument for deciding on CTx use in patients with ER+ and HER2- ESBC.

Some have suggested that individual prognosis, and the correlation between recurrence risk and CTx responsiveness, are common across all the assays. However, the assays do not classify patients similarly, and outcomes cannot be extrapolated from one test to the next. Poulet and colleagues performed a prospective comparison between the commercial versions of the Oncotype DX and MammaPrint tests in 57 ER+ LN- patients with ESBC.⁷⁰ The disagreement was significant. Many of the MammaPrint high-risk patients were Oncotype DX low-risk, and had high ER expression. In this case, because ER expression correlates with ETx benefit, a poor-prognosis MammaPrint patient might do very well on ETx alone. Poor concordance was also evident ($r = 0.08$, 95%CI $-0.2-0.35$; Spearman correlation) between the Oncotype DX and Prosigna commercial assays in 52 patients studied by Alvarado and colleagues. They found a risk-classification concordance of only 54%.⁷¹

More recently, the OPTIMA Prelim Trial evaluated five assays, including four of the tests discussed in this review (Oncotype DX, MammaPrint, Prosigna, and NexCourse Breast-IHC4). Each assay was performed in its manufacturer's own commercial laboratory.⁷² Remarkably, there was significant disagreement between assays, with the overall level of agreement only "moderate" ($\kappa = 0.40-0.59$). The authors were perplexed that multiple assays, each with independent validation for prognosis, could have such disagreement. They suggested that assay concordance was no better than classic pathologic grading. However, OPTIMA Prelim did not measure long-term patient outcomes and could not provide independent validation

of each of the competing assays, nor could it identify the most accurate of the disparate tests. The poor concordance between all tests suggests that one well validated assay cannot reliably be replaced with others with less robust validation.

The challenge of medical decision-making is to provide the best care for today's patients while continually assessing new diagnostic and treatment paradigms for those in the future. Adjuvant therapy provides an opportunity to reduce the chance of DR and death, but carries its own risk of short and long-term toxicity. Modern therapies also generate significant costs to individuals and healthcare systems. Genomic classifiers that provide accurate prognostic information and reliable benefit prediction personalize the selection of treatment. They improve the likelihood of favorable long-term outcome while conserving healthcare resources. However, the choice should be made on the basis of a rigorous analysis of the ever-changing supporting data.

5 | CONCLUSIONS

Multiple marketed gene expression assays are now available to provide prognostic and predictive information in ESBC. A review of the available studies suggests that each of the six assays has some evidence supporting their use as prognostic tools in varying subpopulations. However only two assays, Oncotype DX and MammaPrint, have published evidence from large prospective clinical trials for prognosis relevant to many ESBC patients. Both of these assays identify a unique subpopulation with low-risk for which chemotherapy would add little benefit. To date, however, only Oncotype DX predicts the benefit of CTx in a broad population of HR+, pre- and postmenopausal, LN- and LN+ patients validated through adjuvant CTx trials that compared randomized CTx-treated with CTx-untreated populations in long-term follow-up. It is this combination of strong prognostic and reliable predictive information that provides an assay with maximum clinical utility.

ACKNOWLEDGMENT

No direct financial support has been provided for the conduct of this review.

AUTHORS' DISCLOSURES OF POTENTIAL CONFLICTS OF INTEREST

All authors completed a disclosure declaration. The following authors indicated a financial or other interest that is relevant to the subject matter under consideration in this article. Relationships marked with a "U" are those for which no compensation was received, those relationships marked with a "C" were compensated:

Employment or Leadership Position: None Consultant or Advisor Role: David M. Hyams, Biotheranostics (C), Genomic Health (C), Castlebiosciences (C), Exact Sciences (C); Eric Schuur, Genomic Health (C); Artur Katz, Bristol Myers Squibb (C), Merck Sharp Dome (C); Aníbal

Nuñez De Pierro, AstraZeneca (C), Roche (C) Stock Ownership: None
Honoraria: David M. Hyams, Genomic Health, Castle Biosciences,
Exact Sciences; Cesar Cabello, Genomic Health; Aníbal Nuñez De
Pierro, Astrazeneca, Novartis, Siemens Research Funding: None
Expert Testimony: None Other Remuneration: None

REFERENCES

- International Agency for Research on Cancer WHO. *Breast cancer: Estimated incidence, mortality and prevalence worldwide in 2012*. 2012 [cited 2016 January 9]; Available from: http://globocan.iarc.fr/Pages/fact_sheets_cancer.aspx.
- Clarke M, Collins R, Darby S, et al.; for the Early Breast Cancer Trialists Collaborative Group, effects of chemotherapy and hormonal therapy for early breast cancer on recurrence and 15-year survival: an overview of the randomised trials. *Lancet*. 2005;365:1687–1717.
- CDC Center for Surveillance Epidemiology and Laboratory Services (CSELS) Public Health Genomics. *Genomic testing: Acce model process for evaluating genetic tests*. 2010 [cited 2015 October 29]; Available from: <http://www.cdc.gov/genomics/gtesting/ACCE/index.htm>.
- Haddow JE, Palomaki GE; In: Khoury MJ, Little J, Burke W, editors. *Acce: A Model Process for Evaluating Data on Emerging Genetic Tests, In Human Genome Epidemiology: a Scientific Foundation for Using Genetic Formation to Improve Health and Prevent Disease*. New York: Oxford University Press; 2004:217–233.
- Hunter DJ, Khoury MJ, Drazen JM. Letting the genome out of the bottle – will we get our wish? *N Engl J Med*. 2008;358:105–107.
- Hayes DF, Bast RC, Desch CE, et al. Tumor marker utility grading system: a framework to evaluate clinical utility of tumor markers. *J Natl Cancer Inst*. 1996;88:1456–1466.
- Simon RM, Paik S, Hayes DF. Use of archived specimens in evaluation of prognostic and predictive biomarkers. *J Natl Cancer Inst*. 2009;101:1446–1452.
- Sgroi DC, Sestak I, Cuzick J, et al. Prediction of late distant recurrence in patients with oestrogen-receptor-positive breast cancer: a prospective comparison of the breast-cancer index (BCI) assay, 21-gene recurrence score, and IHC4 in the transatac study population. *Lancet Oncol*. 2013;14:1067–1076.
- Ma XJ, Salunga R, Dahiya S, et al. A five-gene molecular grade index and HOXB13: IL17BR are complementary prognostic factors in early stage breast cancer. *Clin Cancer Res*. 2008;14:2601–2608.
- Ma XJ, Wang Z, Ryan PD, et al. A two-gene expression ratio predicts clinical outcome in breast cancer patients treated with tamoxifen. *Cancer Cell*. 2004;5:607–616.
- Ma XJ, Salunga R, Tuggle JT, et al. Gene expression profiles of human breast cancer progression. *Proc Natl Acad Sci USA*. 2003;100:5974–5979.
- Goetz MP, Suman VJ, Ingle JN, et al. A two-gene expression ratio of homeobox 13 and interleukin-17b receptor for prediction of recurrence and survival in women receiving adjuvant tamoxifen. *Clin Cancer Res*. 2006;12:2080–2087.
- Ma X-J, Hilsenbeck SG, Wang W, et al. The HOXB13:IL17BR expression index is a prognostic factor in early-stage breast cancer. *J Clin Oncol*. 2006;24:4611–4619.
- Jerevall PL, Ma XJ, Li H, et al. Prognostic utility of HOXB13:IL17BR and molecular grade index in early-stage breast cancer patients from the stockholm trial. *Br J Cancer*. 2011;104:1762–1769.
- Rutqvist LE, Johansson H, and Stockholm Breast Cancer Study G, Long-term follow-up of the randomized stockholm trial on adjuvant tamoxifen among postmenopausal patients with early stage breast cancer. *Acta Oncol*. 2007;46:133–145.
- Zhang Y, Schnabel CA, Schroeder BE, et al. Breast cancer index identifies early-stage estrogen receptor-positive breast cancer patients at risk for early- and late-distant recurrence. *Clin Cancer Res*. 2013;19:4196–4205.
- Harris LN, Ismaila N, McShane LM, Andre F Reply to D. Sgroi et al, T. Sanft et al, M. S. Copur et al, and M. Goetz et al. *J Clin Oncol*. 2016;34:3946–3948.
- Sgroi DC, Chapman JA, Badovinac-Crnjevic T, et al. Assessment of the prognostic and predictive utility of the breast cancer index (BCI): an ancic ctg ma.14 study. *Breast Cancer Res*. 2016;18:1.
- Filipits M, Rudas M, Jakesz R, et al. A new molecular predictor of distant recurrence in er-positive, HER2-negative breast cancer adds independent information to conventional clinical risk factors. *Clin Cancer Res*. 2011;17:6012–6020.
- Dubsky P, Filipits M, Jakesz R, et al. Endopredict improves the prognostic classification derived from common clinical guidelines in er-positive, HER2-negative early breast cancer. *Ann Oncol*. 2013;24:640–647.
- Martin M, Brase JC, Calvo L, et al. Clinical validation of the endopredict test in node-positive, chemotherapy-treated ER+/HER2- breast cancer patients: results from the geicam 9906 trial. *Breast Cancer Res*. 2014;16:R38.
- Buus R, Sestak I, Kronenwett R, et al. Comparison of Endopredict and eclin with Oncotype DX recurrence score for prediction of risk of distant recurrence after endocrine therapy. *J Natl Cancer Inst*. 2016;108:djw149. doi: 10.1093/jnci/djw149
- Martin M, Brase JC, Ruiz A, et al. Prognostic ability of Endopredict compared to research-based versions of the PAM50 risk of recurrence (ROR) scores in node-positive, estrogen receptor-positive, and HER2-negative breast cancer. a GEICAM/9906 sub-study. *Breast Cancer Res Treat*. 2016;156:81–89.
- 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.
- Stephen J, Murray G, Cameron DA, et al. Time dependence of biomarkers: non-proportional effects of immunohistochemical panels predicting relapse risk in early breast cancer. *Br J Cancer*. 2014;111:2242–2247.
- Adjuvant Inc. *Adjuvant! Online*. 2013 [cited 2013 February 17]; Adjuvant! Online Home Page]. Available from: <https://www.adjuvantonline.com/>.
- Goldhirsch A, Wood WC, Coates AS, et al. Strategies for subtypes—dealing with the diversity of breast cancer: highlights of the st gallen international expert consensus on the primary therapy of early breast cancer 2011. *Ann Oncol*. 2011;22:1736–1747.
- Park YH, Im SA, Cho EY, et al. Validation and comparison of CS-IHC4 scores with a nomogram to predict recurrence in hormone receptor-positive breast cancers. *Oncology* 2014;86:279–288.
- van't Veer LJ, Dai H, van de Vijver MJ, et al. Gene expression profiling predicts clinical outcome of breast cancer. *Nature*. 2002;415:530–536.
- van de Vijver MJ, He YD, van't Veer LJ, et al. A gene-expression signature as a predictor of survival in breast cancer. *N Engl J Med*. 2002;347:1999–2009.
- Buyse M, Loi S, van't Veer L, et al. Validation and clinical utility of a 70-gene prognostic signature for women with node-negative breast cancer. *J Natl Cancer Inst*. 2006;98:1183–1192.
- Mook S, Schmidt MK, Viale G, et al. The 70-gene prognosis-signature predicts disease outcome in breast cancer patients with 1-3 positive lymph nodes in an independent validation study. *Breast Cancer Res Treat*. 2009;116:295–302.

33. Wittner BS, Sgroi DC, Ryan PD, et al. Analysis of the Mammprint breast cancer assay in a predominantly postmenopausal cohort. *Clin Cancer Res*. 2008;14:2988–2993.
34. Mook S, Schmidt MK, Weigelt B, et al. The 70-gene prognosis signature predicts early metastasis in breast cancer patients between 55 and 70 years of age. *Ann Oncol*. 2010;21:717–722.
35. Cardoso F, van't Veer LJ, Bogaerts J, et al. 70-gene signature as an aid to treatment decisions in early-stage breast cancer. *N Engl J Med*. 2016;375:717–729.
36. Knauer M, Mook S, Rutgers EJ, et al. The predictive value of the 70-gene signature for adjuvant chemotherapy in early breast cancer. *Breast Cancer Res Treat*. 2010;120:655–661.
37. Straver ME, Glas AM, Hannemann J, et al. The 70-gene signature as a response predictor for neoadjuvant chemotherapy in breast cancer. *Breast Cancer Res Treat*. 2010;119:551–558.
38. Bueno-de-Mesquita JM, van Harten WH, Retel VP, et al. Use of 70-gene signature to predict prognosis of patients with node-negative breast cancer: a prospective community-based feasibility study (raster). *Lancet Oncol*. 2007;8:1079–1087.
39. Krijgsman O, Roepman P, Zwart W, et al. A diagnostic gene profile for molecular subtyping of breast cancer associated with treatment response. *Breast Cancer Res Treat*. 2012;133:37–47.
40. Whitworth P, Stork-Sloots L, de Snoo FA, et al. Chemosensitivity predicted by Blueprint 80-gene functional subtype and Mammprint in the prospective Neoadjuvant Breast Registry Symphony Trial (NBRST). *Ann Surg Oncol*. 2014;21:3261–3267.
41. Earl H, Provenzano E, Abraham J, et al. Neoadjuvant trials in early breast cancer: pathological response at surgery and correlation to longer term outcomes—what does it all mean? *BMC Med*. 2015;13:234.
42. 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.
43. Harbeck N, Sotlar K, Wuerstlein R, et al. Molecular and protein markers for clinical decision making in breast cancer: today and tomorrow. *Cancer Treat Rev*. 2014;40:434–444.
44. 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.
45. Dowsett M, Cuzick J, Wale C, et al. Prediction of risk of distant recurrence using the 21-gene recurrence score in node-negative and node-positive postmenopausal patients with breast cancer treated with anastrozole or tamoxifen: a transatac study. *J Clin Oncol*. 2010;28:1829–1834.
46. Albain KS, Barlow WE, Shak S, et al. Prognostic and predictive value of the 21-gene recurrence score assay in postmenopausal women with node-positive, oestrogen-receptor-positive breast cancer on chemotherapy: a retrospective analysis of a randomised trial. *Lancet Oncol*. 2010;11:55–65.
47. Gluz O, Nitz UA, Christgen M, et al. West German Study Group Phase III PlanB trial: first prospective outcome data for the 21-gene recurrence score assay and concordance of prognostic markers by central and local pathology assessment. *J Clin Oncol*. 2016;34:2341–2349.
48. Paik S, Tang G, Shak S, et al. Gene expression and benefit of chemotherapy in women with node-negative, estrogen receptor-positive breast cancer. *J Clin Oncol*. 2006;24:3726–3734.
49. Albain KS, Barlow WE, Ravdin PM, et al. Adjuvant chemotherapy and timing of tamoxifen in postmenopausal patients with endocrine-responsive, node-positive breast cancer: a Phase 3, open-label, randomised controlled trial. *Lancet* 2009;374:2055–2063.
50. Geiss GK, Bumgarner RE, Birditt B, et al. Direct multiplexed measurement of gene expression with color-coded probe pairs. *Nat Biotechnol*. 2008;26:317–325.
51. Perou CM, Sorlie T, Eisen MB, et al. Molecular portraits of human breast tumours. *Nature*. 2000;406:747–752.
52. Nielsen T, Wallden B, Schaper C, et al. Analytical validation of the PAM50-based prognostic breast cancer prognostic gene signature assay and ncounter analysis system using formalin-fixed paraffin-embedded breast tumor specimens. *BMC Cancer*. 2014;14:177.
53. Dowsett M, Sestak I, Lopez-Knowles E, et al. Comparison of PAM50 risk of recurrence score with Oncotype DX and IHC4 for predicting risk of distant recurrence after endocrine therapy. *J Clin Oncol*. 2013;31:2783–2790.
54. Gnant M, Filipits M, Greil R, et al. Predicting distant recurrence in receptor-positive breast cancer patients with limited clinicopathological risk: using the PAM50 risk of recurrence score in 1478 postmenopausal patients of the ABCSG-8 trial treated with adjuvant endocrine therapy alone. *Ann Oncol*. 2014;25:339–345.
55. Gnant M, Sestak I, Filipits M, et al. Identifying clinically relevant prognostic subgroups of postmenopausal women with node-positive hormone receptor-positive early-stage breast cancer treated with endocrine therapy: a combined analysis of ABCSG-8 and atac using the PAM50 risk of recurrence score and intrinsic subtype. *Ann Oncol*. 2015;26:1685–1691.
56. Chia SK, Bramwell VH, Tu D, et al. A 50-gene intrinsic subtype classifier for prognosis and prediction of benefit from adjuvant tamoxifen. *Clin Cancer Res*. 2012;18:4465–4472.
57. Cheang MC, Voduc KD, Tu D, et al. Responsiveness of intrinsic subtypes to adjuvant anthracycline substitution in the NCIC.CTG.MA5 randomized trial. *Clin Cancer Res*. 2012;18:2402–2412.
58. Liu S, Chapman JA, Burnell MJ, et al. Prognostic and predictive investigation of pam50 intrinsic subtypes in the NCIC.CTG.MA.21 phase iii chemotherapy trial. *Breast Cancer Res Treat*. 2015;149:439–448.
59. Cortazar P, Zhang L, Untch M, et al. Pathological complete response and long-term clinical benefit in breast cancer: the CTNOCBC pooled analysis. *Lancet* 2014;384:164–172.
60. De Mattos-Arruda L, Shen R, Reis-Filho JS, et al. Translating neoadjuvant therapy into survival benefits: one size does not fit all. *Nat Rev Clin Oncol*. 2016;13:566–579.
61. Filipits M, Nielsen TO, Rudas M, et al. The PAM50 risk-of-recurrence score predicts risk for late distant recurrence after endocrine therapy in postmenopausal women with endocrine-responsive early breast cancer. *Clin Cancer Res*. 2014;20:1298–1305.
62. Sapino A, Roepman P, Linn SC, et al. Mammprint molecular diagnostics on formalin-fixed, paraffin-embedded tissue. *J Mol Diagn*. 2014;16:190–197.
63. Beumer I, Witteveen A, Delahaye L, et al. Equivalence of Mammprint array types in clinical trials and diagnostics. *Breast Cancer Res Treat*. 2016;156:279–287.
64. Anderson JM, Shak S, Millward C, et al. Molecular characterization of breast cancer core biopsy specimens by gene expression analysis using standardized quantitative RT-PCR. *Cancer Res* 2009;69:Abstract nr 6021.
65. United States Food and Drug Administration. Class 2 device recall Agendia Mammprint. 2010 [cited 2016 September 27]; Available from: <http://www.accessdata.fda.gov/scripts/cdrh/cfdocs/cfRes/res.cfm?ID=89513>
66. Cuzick J. Aromatase inhibitors for breast cancer prevention. *J Clin Oncol*. 2005;23:1636–1643.

67. Early Breast Cancer Trialists Collaborative Group, Polychemotherapy for early breast cancer: an overview of the randomised trials. *Lancet*. 1998;352:930–942.
68. Hudis CA, Dickler M. Increasing precision in adjuvant therapy for breast cancer. *N Engl J Med*. 2016;375:790–791.
69. Petkov VI, Miller DP, Howlader N, et al. Breast-cancer-specific mortality in patients treated based on the 21-gene assay: a SEER population-based study. *Npj Breast Cancer* 2016;2:16017.
70. Poulet B, Jamshidian F, Butler S, et al. Risk classification of early stage breast cancer as assessed by MammPrint and Oncotype DX genomic assays. *Cancer Res* 2012;72:Abstract nr P6-07-03.
71. Alvarado MD, Prasad C, Rothney M, et al. A prospective comparison of the 21-gene recurrence score and the PAM50-based Prosigna in estrogen receptor-positive early-stage breast cancer. *Adv Ther*. 2015;32:1237–1247.
72. Bartlett JMS, Bayani J, Marshall A, et al. Comparing breast cancer multiparameter tests in the OPTIMA Prelim trial: no test is more equal

than the others. *J Natl Cancer Inst*. 2016;108:pii: djw050. doi: 10.1093/jnci/djw050

SUPPORTING INFORMATION

Additional Supporting Information may be found online in the supporting information tab for this article.

How to cite this article: Hyams DM, Schuur E, Angel Aristizabal J, et al. Selecting postoperative adjuvant systemic therapy for early stage breast cancer: A critical assessment of commercially available gene expression assays. *J Surg Oncol*. 2017;115:647–662. <https://doi.org/10.1002/jso.24561>