*JNCI J Natl Cancer Inst* (2016) 108(9): djw118

doi: 10.1093/jnci/djw118
First published online April 29, 2016
Editorial

EDITORIAL

# Risk Stratification and Intrinsic Subtype Classification of Breast Cancer: A Multiparameter Test to Rule Them All?

Rachael Natrajan, Britta Weigelt

**Affiliations of authors:** Division of Breast Cancer, The Breast Cancer Now Toby Robins Research Centre, The Institute of Cancer Research, London, UK (RN); Department of Pathology, Memorial Sloan Kettering Cancer Center, New York, NY (BW).

**Correspondence to:** Britta Weigelt, PhD, Memorial Sloan Kettering Cancer Center, Department of Pathology, 1275 York Avenue, New York, NY 10065 (e-mail: weigeltb@mskcc.org).

Seminal studies in the early 2000s led to the development of a molecular-based classification of breast cancer as well as several multigene prognostic signatures for breast cancer patients (1,2). The 'intrinsic' gene subtypes (ie, luminal A, luminal B, human epidermal growth factor receptor 2 [HER2]–enriched and basal-like) (3-5) have constituted one of the biggest paradigm shifts in the way breast cancer is perceived from a research and clinical standpoint. These subtypes have not only become part of our lexicon but have also been shown to robustly identify prognostically relevant molecular subgroups of patients with breast cancer (1,4,6,7). Different gene sets and methods for the identification of the intrinsic subtypes have been published, and although they show equipoise in outcome across different patient populations, we and others have found that for individual patients, research versions of these different methods/gene lists only show a modest agreement and may assign the same patient to different intrinsic subtypes (8,9). It has been suggested, however, that technical limitations such as data normalization and platform annotation may have reduced the accuracy in subtype predictions and concordance (6,10,11), which would be improved if the comparisons were performed using validated diagnostic genomic assays (6).

In parallel with the class-discovery studies that resulted in the identification of the intrinsic subtypes, a plethora of microarray-based gene expression signatures used to predict the outcomes for individual patients with breast cancer have been developed (1,2). Some of these prognostic gene expression–based assays, namely Oncotype DX, EndoPredict, Prosigna (PAM50) and Breast Cancer Index, have recently been recommended by The American Society of Clinical Oncology to guide decisions on adjuvant systemic therapy for specific subsets of patients with early-stage breast cancer (12). Multigene prognostic assays have been shown to provide prognostic information that is complementary to that provided by tumor size, nodal status, and histologic grade in patients with estrogen receptor (ER)–positive disease, and some of these tests have been used also to define the subset of ER-positive breast cancer patients who are likely to benefit from chemotherapy (13). While there is minimal overlap between each of the signatures in terms of their gene composition, the discriminatory value of each test is comparable in patients with ER-positive disease (14,15). This stems from the fact that the prognostic power of these tests is derived from the quantification of proliferation- and ER-related genes (9,13,14,16–18). In addition to the gene expression signatures, immunohistochemistry-based assays such as IHC4, which assesses in a semiquantitative manner ER, progesterone receptor (PR), HER2, and the proliferation marker Ki67, have been developed and provide a risk score for recurrence in patients with ER-positive disease (19).

While retrospective bioinformatics comparisons demonstrated that these various gene expression signatures are prognostic (15,20–22), retrospective studies comparing commercial tests of these multiparameter assays revealed discrepancies not only in the number of patients assigned to a specific risk category but, more importantly, in the risk assessment of individual patients (23–25).

The study by Bartlett and colleagues in this issue of the Journal (26) addresses points that are germane to our understanding of the limitations of these assays in the context of precision medicine. It describes the first prospective direct comparison of commercially available multiparameter tests in a prospective randomized trial of patients with early breast

**Received:** March 16, 2016; **Accepted:** March 18, 2016

cancer—the Optimal Personalised Treatment of early breast cancer using Multiparameter Analysis (OPTIMA) (27). Samples from the feasibility study (OPTIMA-prelim) were tested using Oncotype DX recurrence score (RS), with a dichotomous cutoff greater than 25 (26). In addition, five multiparameter tests (MammaPrint/BluePrint, Prosigna [PAM50], MammaTyper, NexCourse Breast by Aqua [IHC4-AQUA], and IHC4 by conventional immunohistochemistry) were applied to all patients with available tissue (n = 302), and 236 patients had results from all tests. ER and HER2 status were centrally retested on all patients, and all multiparameter tests were carried out to the manufacturers' specifications within their own testing laboratories.

A comparison of tests that classify samples into the intrinsic subtypes, namely Prosigna, BluePrint, and MammaTyper, found that although a similar proportion (around 60%) of patients were classified as having "luminal A" tumors, there was only modest agreement (Kappa scores = 0.39–0.55), with discordant results being recorded in 40.7% of patients (26). This is in agreement with previously published work demonstrating that differences in intrinsic subtype classification of breast cancers are strongly dependent upon the gene sets in the classification models (2,8,9).

A comparison of all five prognostic tests dichotomized as low/intermediate vs high risk revealed that the agreement was modest (Kappa scores = 0.33–0.60), with only 30.8% of patients being classified as low/intermediate risk by all tests and only 8.6% being consistently classified as high risk, even when just taking the extremes of the ranges into account, which would intuitively be less prone to discordances between assays (26). For 183 patients (60.6%), no consensus result across all five tests was found. Moreover, no test appeared to be superior in terms of consistency in agreement with another test.

The observed disagreement across all technically and clinically validated multiparameter tests, some supported by level 1B evidence (12,13), highlights the complexity of choice of test for patient risk stratification and questions the utility of these tests in individualizing therapy for breast cancer patients. Not only is there disagreement at the population level when classifying patients into high- and low-risk groups (19,23,26), but also the results for individual patients are often discordant. These discrepancies may stem from the fact that some assays are more accurate in predicting short vs late relapses and from differences in the approaches to define the cutoffs and weights for the proliferation- and ER-related gene sets (13,16,17). The study by Bartlett and colleagues has some important limitations, however. First, without long-term clinical outcome data, the test that would more accurately predict the outcome remains to be defined. In this group of ER-positive breast cancer patients studied, it may take years to obtain sufficient events and to answer this question. Second, not all tests supported by level I evidence were included in the comparisons. And third, the cutoffs employed for some of the assays included in this study are not necessarily the ones recommended by their respective commercial providers.

Although some multigene prognostic tests are supported by level I evidence, have been shown to be prognostic in material from numerous clinical trials, and are now recommended by guidelines provided by The American Society of Clinical Oncology and others (12,28), discrepant results are not uncommon when these assays are applied to the outcome prediction of individual patients. It is possible that different assays will be more appropriate in specific clinical contexts and types of prognostication desired. The study by Bartlett and colleagues reminds us of important points in the translation of biomarkers: that equipoise between tests when applied to the same population does not equate with concordance between the assays when applied to a given patient and that one size may not fit all when it comes to the use of multigene prognostic assays.

## Funding

RN is the recipient of a Breast Cancer Now Career Development Fellowship 2011MaySF01 and Breast Cancer Now programmatic funding.

## Note

The authors have no conflicts of interest to declare. The funders had no role in the writing of this editorial or the decision to submit it for publication.

## References

1. Weigelt B, Baehner FL, Reis-Filho JS. The contribution of gene expression profiling to breast cancer classification, prognostication and prediction: a retrospective of the last decade. *J Pathol*. 2010;220(2):263–280.
2. Weigelt B, Pusztai L, Ashworth A, Reis-Filho JS. Challenges translating breast cancer gene signatures into the clinic. *Nat Rev Clin Oncol*. 2012;9(1):58–64.
3. Perou CM, Sorlie T, Eisen MB, et al. Molecular portraits of human breast tumours. *Nature*. 2000;406(6797):747–752.
4. Sorlie T, Perou CM, Tibshirani R, et al. Gene expression patterns of breast carcinomas distinguish tumor subclasses with clinical implications. *Proc Natl Acad Sci U S A*. 2001;98(19):10869–10874.
5. Parker JS, Mullins M, Cheang MC, et al. Supervised risk predictor of breast cancer based on intrinsic subtypes. *J Clin Oncol*. 2009;27(8):1160–1167.
6. Prat A, Ellis MJ, Perou CM. Practical implications of gene-expression-based assays for breast oncologists. *Nat Rev Clin Oncol*. 2012;9(1):48–57.
7. Prat A, Pineda E, Adamo B, et al. Clinical implications of the intrinsic molecular subtypes of breast cancer. *Breast*. 2015;24(Suppl 2):S26–S35.
8. Weigelt B, Mackay A, A'Hern R, et al. Breast cancer molecular profiling with single sample predictors: a retrospective analysis. *Lancet Oncol*. 2010;11(4):339–349.
9. Haibe-Kains B, Desmedt C, Loi S, et al. A three-gene model to robustly identify breast cancer molecular subtypes. *J Natl Cancer Inst*. 2012;104(4):311–325.
10. Dunning MJ, Curtis C, Barbosa-Morais NL, et al. The importance of platform annotation in interpreting microarray data. *Lancet Oncol*. 2010;11(8):717.
11. Perou CM, Parker JS, Prat A, Ellis MJ, Bernard PS. Clinical implementation of the intrinsic subtypes of breast cancer. *Lancet Oncol*. 2010;11(8):718–719.
12. Harris LN, Ismaila N, McShane LM, et al. Use of Biomarkers to Guide Decisions on Adjuvant Systemic Therapy for Women With Early-Stage Invasive Breast Cancer: American Society of Clinical Oncology Clinical Practice Guideline. *J Clin Oncol*. 2016; 34(10):1134–1150.
13. Reis-Filho JS, Pusztai L. Gene expression profiling in breast cancer: classification, prognostication, and prediction. *Lancet*. 2011;378(9805):1812–1823.
14. Zhao X, Rodland EA, Sorlie T, et al. Systematic assessment of prognostic gene signatures for breast cancer shows distinct influence of time and ER status. *BMC Cancer*. 2014;14:211.
15. Prat A, Parker JS, Fan C, et al. Concordance among gene expression-based predictors for ER-positive breast cancer treated with adjuvant tamoxifen. *Ann Oncol*. 2012;23(11):2866–2873.
16. Desmedt C, Haibe-Kains B, Wirapati P, et al. Biological processes associated with breast cancer clinical outcome depend on the molecular subtypes. *Clin Cancer Res*. 2008;14(16):5158–5165.
17. Wirapati P, Sotiriou C, Kunkel S, et al. Meta-analysis of gene expression profiles in breast cancer: toward a unified understanding of breast cancer subtyping and prognosis signatures. *Breast Cancer Res*. 2008;10(4):R65.
18. Gyorffy B, Hatzis C, Sanft T, et al. Multigene prognostic tests in breast cancer: past, present, future. *Breast Cancer Res*. 2015;17:11.
19. 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(32):4273–4278.
20. Fan C, Oh DS, Wessels L, et al. Concordance among gene-expression-based predictors for breast cancer. *N Engl J Med*. 2006;355(6):560–569.
21. Haibe-Kains B, Desmedt C, Piette F, et al. Comparison of prognostic gene expression signatures for breast cancer. *BMC Genomics*. 2008;9:394.
22. Iwamoto T, Lee JS, Bianchini G, et al. First generation prognostic gene signatures for breast cancer predict both survival and chemotherapy sensitivity and identify overlapping patient populations. *Breast Cancer Res Treat*. 2011;130(1):155–164.

23. 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(22):2783–2790.
24. Varga Z, Sinn P, Fritzsche F, et al. Comparison of EndoPredict and Oncotype DX test results in hormone receptor positive invasive breast cancer. *PLoS One*. 2013;8(3):e58483.
25. Kelly CM, Bernard PS, Krishnamurthy S, et al. Agreement in risk prediction between the 21-gene recurrence score assay (Oncotype DX(R)) and the PAM50 breast cancer intrinsic Classifier in early-stage estrogen receptor-positive breast cancer. *Oncologist*. 2012;17(4):492–498.
26. Bartlett JMS, Bayani J, Marshall A, et al. Comparing breast cancer multi-parameter tests in the UK OPTIMA prelim trial: All tests are equal – none are more equal than others. *J Natl Cancer Inst*. 2016;108(9): djw050.
27. Stein RC, Dunn JA, Bartlett JM, et al. OPTIMA prelim: a randomised feasibility study of personalised care in the treatment of women with early breast cancer. *Health Technol Assess*. 2016;20(10):1–202.
28. Senkus E, Kyriakides S, Ohno S, et al. Primary breast cancer: ESMO Clinical Practice Guidelines for diagnosis, treatment and follow-up. *Ann Oncol*. 2015; 26(Suppl 5):v8–v30.