ESMO GOOD SCIENCE BETTER MEDICINE BEST PRACTICE

ANNALS OF ONCOLOGY driving innovation in oncology

ORIGINAL ARTICLE

Predicted sensitivity to endocrine therapy for stage II-III hormone receptor-positive and HER2-negative (HR+/HER2−) breast cancer before chemo-endocrine therapy☆

L. Du[1], C. Yau[2], L. Brown-Swigart[3], R. Gould[1], G. Krings[3], G. L. Hirst[2], I. Bedrosian[4], R. M. Layman[5], J. M. Carter[6], M. Klein[7], S. Venters[2], S. Sonal[2], M. van der Noordaa[2], A. J. Chien[8], T. Haddad[9], C. Isaacs[10], L. Pusztai[11], K. Albain[12], R. Nanda[13], D. Tripathy[5], M. C. Liu[9], J. Boughey[14], R. Schwab[15], N. Hylton[16], A. DeMichele[17], J. Perlmutter[18], D. Yee[19], D. Berry[20], L. van't Veer[3], V. Valero[5], L. J. Esserman[2] & W. F. Symmans[1,21*]

[1]Department of Translational Molecular Pathology, The University of Texas MD Anderson Cancer Center, Houston; Departments of [2]Surgery and [3]Pathology, University of California, San Francisco; Departments of [4]Breast Surgery and [5]Breast Medical Oncology, The University of Texas MD Anderson Cancer Center, Houston; [6]Department of Pathology, Mayo Clinic, Rochester; [7]Department of Pathology, University of Minnesota, Minneapolis; [8]Department of Medicine, University of California, San Francisco; [9]Department of Medicine, Mayo Clinic, Rochester; [10]Department of Medicine, Georgetown University, Washington; [11]Department of Medicine, Yale University School of Medicine, New Haven; [12]Department of Medicine, Loyola University, Chicago; [13]Department of Medicine, University of Chicago, Chicago; [14]Department of Surgery, Mayo Clinic, Rochester; [15]Department of Medicine, University of California, San Diego; [16]Department of Radiology, University of California, San Francisco; [17]Department of Medicine, Perelman School of Medicine, University of Pennsylvania, San Philadelphia; [18]Gemini Group, Ann Arbor; [19]Department of Medicine, University of Minnesota, Minneapolis; Departments of [20]Biostatistics and [21]Pathology, The University of Texas MD Anderson Cancer Center, Houston, USA

Available online XXX

Background: We proposed that a test for sensitivity to the adjuvant endocrine therapy component of treatment for patients with stage II-III breast cancer (SET2,3) should measure transcription related to estrogen and progesterone receptors ($SET_{ER/PR}$ index) adjusted for a baseline prognostic index (BPI) combining clinical tumor and nodal stage with molecular subtype by RNA4 (*ESR1*, *PGR*, *ERBB2*, and *AURKA*).

Patients and methods: Patients with clinically high-risk, hormone receptor-positive (HR+), human epidermal growth factor receptor 2 (HER2)-negative (HR+/HER2−) breast cancer received neoadjuvant taxane−anthracycline chemotherapy, surgery with measurement of residual cancer burden (RCB), and then adjuvant endocrine therapy. SET2,3 was measured from pre-treatment tumor biopsies, evaluated first in an MD Anderson Cancer Center (MDACC) cohort ($n = 307$, 11 years' follow-up, U133A microarrays), cut point was determined, and then independent, blinded evaluation was carried out in the I-SPY2 trial ($n = 268$, high-risk MammaPrint result, 3.8 years' follow-up, Agilent-44K microarrays, NCI Clinical Trials ID: NCT01042379). Primary outcome measure was distant relapse-free survival. Multivariate Cox regression models tested prognostic independence of SET2,3 relative to RCB and other molecular prognostic signatures, and whether other prognostic signatures could substitute for $SET_{ER/PR}$ or RNA4 components of SET2,3.

Results: SET2,3 added independent prognostic information to RCB in the MDACC cohort: SET2,3 [hazard ratio (HR) 0.23, $P = 0.004$] and RCB (HR 1.77, $P < 0.001$); and the I-SPY2 trial: SET2,3 (HR 0.27, $P = 0.031$) and RCB (HR 1.68, $P = 0.008$). SET2,3 provided similar prognostic information irrespective of whether RCB-II or RCB-III after chemotherapy, and in both luminal subtypes. Conversely, RCB was most strongly prognostic in cancers with low SET2,3 status (MDACC $P < 0.001$, I-SPY2 $P < 0.001$). Other molecular signatures were not independently prognostic; they could effectively substitute for RNA4 subtype within the BPI component of SET2,3, but they could not effectively substitute for $SET_{ER/PR}$ index.

Conclusions: SET2,3 added independent prognostic information to chemotherapy response (RCB) and baseline prognostic score or subtype. Approximately 40% of patients with clinically high-risk HR+/HER2− disease had high SET2,3 and could be considered for clinical trials of neoadjuvant endocrine-based treatment.

Key words: prognosis, chemo-endocrine, response, prediction, hormonal, treatments

*Correspondence to: Prof. William Fraser Symmans, Departments of Pathology and Translational Molecular Pathology, The University of Texas MD Anderson Cancer Center, 2130 W Holcombe Blvd, Unit 2951, Houston, TX 77030, USA. Tel: +1-713-792-7962; Fax: +1-713-745-8221
E-mail: fsymmans@mdanderson.org (W. F. Symmans).

☆Note: This study was previously presented in part at the 2019 San Antonio Breast Cancer Symposium.

INTRODUCTION

Accurate estimation of residual risk of recurrence after adjuvant treatments depends theoretically on a calculus of sequential probabilities that micro-metastatic disease would be present, survive the combination and sequence of all systemic treatments administered, and progress to clinical recurrence during the period of follow-up. These probabilities can be estimated from the burden of cancer at presentation (stage) and its biological phenotype (molecular subtype), predictions of sensitivity or measures of response to each class of adjuvant treatment administered, and the expected pattern of risk over time.

Overall, patients who present with a clinical stage II-III breast cancer that expresses hormone receptors and does not overexpress human epidermal growth factor receptor 2 (HER2) (HR+/HER2−) have significant long-term residual risk. Disease-free survival (DFS) at 12-15 years following standard chemotherapy and then endocrine therapy is ∼70%.[1-3] Even molecularly low-risk stage II-III disease typically has 80%-85% DFS at 10 years, irrespective of the type of adjuvant treatment.[2,4] So, we considered how to predict sensitivity to adjuvant endocrine therapy within the context of disease burden, molecular prognosis, and sensitivity or response to other treatments administered.[5-7]

This work builds on the concepts learned from our research-based, 165-gene signature of estrogen receptor (ER)-related transcription that is not related to proliferation to predict sensitivity to endocrine therapy.[8] This was refined into the 28-gene $SET_{ER/PR}$ index of hormone receptor-related transcription.[9] However, predicted sensitivity to endocrine therapy for stage II-III disease should take baseline prognostic risk into account. In this article, we describe SET2,3 as a genomic algorithm that adjusts the measurement of endocrine-related transcriptional activity ($SET_{ER/PR}$) for baseline prognosis from clinical stage and molecular subtype based on four genes (Figure 1).[10,11] SET2,3 and its component signatures are measured accurately from routine pathology samples using a customized assay.[12] We have evaluated the prognostic contribution of SET2,3 and its components to other contemporary prognostic gene expression signatures measured from microarrays, and to residual cancer burden (RCB) as a prognostic surrogate for chemotherapy response.[6,7,11]

METHODS

Development of the baseline prognostic index

We interpreted prognostic risk votes from pT stage and pN stage in the published results of the Oxford overview, and refined this using subject-level clinical data from the control arms of the Breast Cancer International Research Group BCIRG-001 and BCIRG-005 adjuvant trials of chemotherapy in stage II-III HR+/HER2− breast cancer (permission obtained from Project Datasphere)[13-15] Clinical nodal (cN) and tumor (cT) stages were assigned integer risk votes from 0 to 3 corresponding to cN0, cN1, cN2, and cN3; and cT0-1, cT2, cT3, and cT4, respectively.[3]

RNA4 subtype classification (*ESR1*, *PGR*, *ERBB2*, *AURKA*) represents a convergence of a three-gene classifier by Haibe-Kains et al. (*ESR1*, *ERBB2*, *AURKA*) and the IHC4 classifier by Cusick et al. (genes *ESR1*, *PGR*, *ERBB2*, *MKI67*) because *AURKA* provides technically reproducible gene expression measurements, and *PGR* status helps to identify molecular subtypes.[11,16,17] We evaluated the public datasets of Affymetrix (Santa Clara, CA) microarrays from 1489 breast cancers, of which 990 were HR+/HER2−, and 399 of those were untreated node-negative with high *ESR1*, low *ERBB2*, and known outcomes for distant relapse-free survival (DRFS) at 5 years (Supplementary Table S1, available at https://doi.org/10.1016/j.annonc.2021.02.011). The individual transcripts (*ESR1*, *ERBB2*, *PGR*, and *AURKA*) had their log2 expression value (X) normalized to the mean of the 10 reference genes: X − mean + 2.[9] Three of four RNA4 genes (*ESR1*, *PGR*, *ERBB2*) have bimodal distribution of expression in breast cancers (Supplementary Figure S1A, available at https://doi.org/10.1016/j.annonc.2021.02.011).[18] Therefore, we defined positive *ESR1* and *PGR* expression status as expression exceeding a cut-point two standard deviations (2σ) below the mean gene expression value in the higher expression peak: *ESR1* = 8.93, *PGR* = 5.10. Similarly, we defined the cut point for *ERBB2* gene expression status (11.97) as 2σ above the mean value of gene expression in the lower expression peak. The expression level of *AURKA* was more prognostic when evaluated separately within *PGR* low and *PGR* high subsets. Therefore, we optimized a cut point for *AURKA* expression based on 5-year prognosis in untreated node-negative HR+/HER2− cancers as the 67th percentile in 103 cancers with low *PGR* expression and the

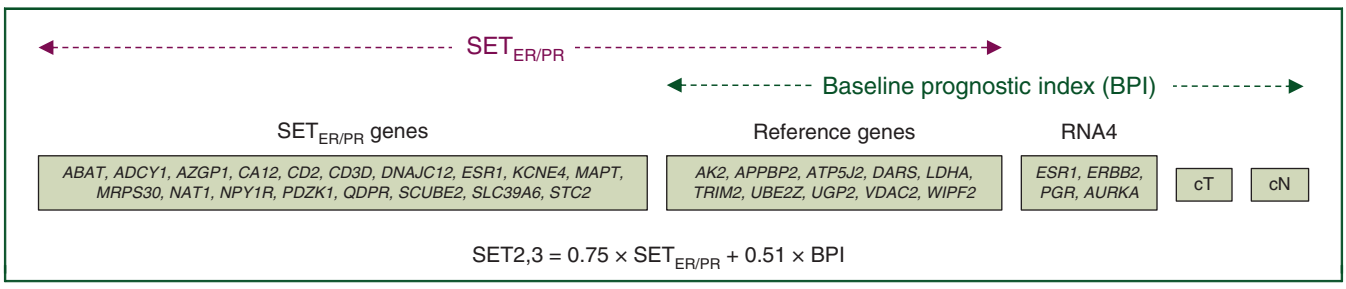

Figure 1. Description of SET2,3 algorithm.
Genes measured to calculate SET2,3 as a weighted sum of $SET_{ER/PR}$ index of hormone receptor-related transcription not associated with proliferation and the baseline prognostic index (BPI) that combines clinical T stage (cT), clinical nodal stage (cN), and molecular risk subtype from four genes (RNA4).

75th percentile in 296 cancers with high *PGR* expression (Supplementary Figure S2, available at https://doi.org/10.1016/j.annonc.2021.02.011).

The RNA4 classifier assigns low risk (0 votes) if *ESR1*-positive and *ERBB2*-low status, and *AURKA* expression value below the *PGR*-dependent cut point; and high risk (2 risk votes) if *ESR1*-negative, *ERBB2*-positive, or *AURKA* expression value above the *PGR*-dependent cut point. However, samples with *AURKA* expression level within 3% of the *PGR*-dependent cut point were classified as borderline status for RNA4 (1 risk vote). The baseline prognostic index (BPI) was defined as the total of risk votes (cT, cN, RNA4), subtracted from eight, and divided by two.

Clinical cohorts: chemo-endocrine therapy

Hormone receptor status was defined as positive if the percent of cancer cell nuclei staining for ER or progesterone receptor (PR) was at least 1% in the MD Anderson Cancer Center (MDACC) cohort or 5% in the I-SPY2 trial. HER2 status was defined as negative in both cohorts if not overexpressed (score $\leq$ 2) and not amplified (*ERBB2* $\leq$ 6 copies per nucleus and/or *ERBB2*/CEP17 ratio $<$ 2.0), and also in the I-SPY2 trial only, if *ERBB2* gene expression was not elevated (TargetPrint, Agendia, Irvine, CA). Clinical nodal status was determined before treatment from physical examination, ultrasound, and needle biopsy of suspicious nodes.

The MDACC cohort received neoadjuvant taxane–anthracycline chemotherapy, as previously reported.[7] The prospective I-SPY2 trial evaluates investigational treatments in HR+/HER2− breast cancers that are high risk according to the MammaPrint test (Agendia; using customized Agilent [Santa Clara, CA] 44K microarrays), combining the treatment with weekly paclitaxel in a neoadjuvant chemotherapy regimen of weekly paclitaxel followed by doxorubicin and cyclophosphamide (T/AC) every 2-3 weeks. This analysis includes 205 patients who received an investigational treatment and 63 controls who received chemotherapy (Table 1), representing the first six investigational treatments that were evaluated in HR+/HER2− cancers [neratinib, veliparib + carboplatin, trebananib (AMG-386), ganitumab (MK-2206), and ganetespib].[19-25]

In both the MDACC and I-SPY2 studies, response to neoadjuvant chemotherapy was evaluated using the RCB method.[5,6] Patients completed definitive local (surgical and radiation) treatments and were recommended standard adjuvant hormonal therapy of aromatase inhibitor or tamoxifen for at least 5 years.

Table 1. Clinical and pathological characteristics of the patient cohorts with HR+/HER2− cancer

Characteristic	Category	MDACC *n* (%)	I-SPY2 *n* (%)
Neoadjuvant chemotherapy	T/AC[a]	307 (100)	63 (24)
	T + Exp Rx/AC		205 (76)
Age	>50	145 (47)	106 (40)
cT stage	0	3 (1)	0 (0)
	1	21 (7)	12 (4)
	2	160 (52)	163 (61)
	3	79 (26)	83 (31)
	4	44 (15)	10 (4)
cN status	Negative	106 (35)	129 (48)
	Positive	201 (66)	139 (52)
Grade	High	110 (36)	101 (55)
	NA	—	85
Histologic type	Ductal	247 (81)	226 (84)
	Lobular	24 (8)	13 (5)
	Ductal/lobular	29 (10)	20 (7)
	Other	7 (2)	3 (1)
	NA	—	6 (2)
Molecular prognostic test	Mamma Print	No testing done	Required high risk (100%)
Response	pCR	36 (12)	50 (19)
	RCB-I	25 (8)	36 (13)
	RCB-II	138 (45)	130 (49)
	RCB-III	62 (20)	52 (19)
	PD (RCB-III)	4 (1)	—
	NA	42 (14)	—
Progression	During NAC	0	0
Pathologists	Review	MDACC	Local sites
Local event	Yes	3	18
Distant event	Yes	104	38
Deceased	Yes	86	29
Follow-up for survivors	Median (years)	11.0	3.8

RCB classes used published cut points for RCB index score.
cN, clinical nodal; cT, clinical tumor; MDACC, MD Anderson Cancer Center; NAC, neoadjuvant chemotherapy; pCR, pathologic complete response; PD, progressive disease or inoperability during neoadjuvant chemotherapy has no RCB index score but is assigned RCB-III class; RCB, residual cancer burden; NA, not applicable or not available for assessment of RCB result.
[a] Patients with low-risk result from MammaPrint test were excluded from the I-SPY2 trial.

Gene expression assays and signatures

Patients in the MDACC and I-SPY2 clinical cohorts provided consent (Institutional Review Board approved) for gene expression profiling of pre-treatment research biopsies of tumor and comparisons of genomic signatures to treatment response and DRFS. Pre-treatment tumor biopsies (fresh frozen in RNA*later* [Qiagen, Valencia, CA] or optimal cutting temperature compound) underwent microarray-based gene expression profiling using Affymetrix U133A arrays in the MDACC cohort and Agilent 44K arrays in the I-SPY2 trial (customized for Agendia). Published R-scripts were applied to Affymetrix U133A data to estimate the 21-gene recurrence score (RS), 11-gene EndoPredict (EP), 70-gene MammaPrint (MP) signatures, and the PAM50 subtype using data from the U133A microarray platform.[11] We did not evaluate the H/I signature (*HOXB13*/*IL17BR*) because it is not reliably measured from microarray data.[8,26] Each transcript probe was calibrated *in silico* between Affymetrix U133A and Agilent 44K microarray platforms, including the cut points, before blinded validation analysis of data from the I-SPY2 trial (Supplementary Figure S3, available at https://doi.org/10.1016/j.annonc.2021.02.011). Corresponding MammaPrint scores and BluePrint subtype results for each I-SPY2 sample were provided by Agendia to the I-SPY2 statisticians (CY, DB).

Statistical methods

DRFS was defined from the date of diagnosis using published standardized criteria.[27] Survival probability within biomarker class was determined using the Kaplan–Meier

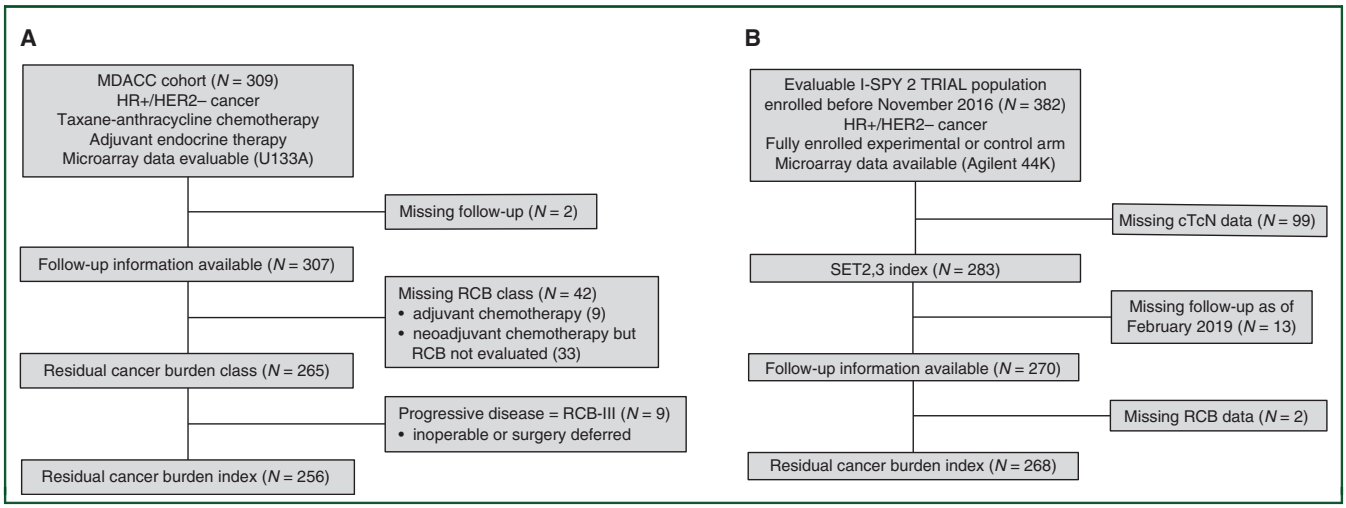

Figure 2. Consort diagrams of samples used to evaluate the clinical cohorts.
(A) The MD Anderson Cancer Center (MDACC) cohort and (B) The I-SPY2 trial. cN, clinical nodal stage; cT, Clinical T stage; RCB, residual cancer burden.

estimator, with 95% confidence intervals (CIs) estimated using the Greenwood formula with log—log transformation. Kaplan—Meier curves were truncated when the smallest subgroup had fewer than 10% of subjects remaining at risk.[28] Survival times were compared among RCB classes using the log-rank test. We used the R package 'survival' and 'survminer' for survival analysis. Hazard function for DRFS event was plotted for SET2,3 values in each group relative to the hazard for the lowest SET2,3 value in the group, and was plotted for RCB values relative to the hazard for all patients with pathologic complete response (pCR) (RCB = 0). Chi-square statistics for contingency table was used to compare the proportions of excellent response to neoadjuvant chemotherapy (pCR/RCB-I) between SET2,3 low and SET2,3 high subgroups in both MDACC and I-SPY2 cohorts.

Association between SET2,3 and DRFS was evaluated in each cohort using multivariate Cox regression models based on the likelihood ratio test. One model was adjusted for RCB after neoadjuvant chemotherapy and its interaction with SET2,3, also including the treatment arm (experimental versus control) in the I-SPY 2 trial. Another model was adjusted for each gene expression signature (RS, MP, and EP) and PAM50 subtype in the MDACC cohort, and actual MP signature results from the I-SPY2 trial. A third model was adjusted for RCB and tested the ability of prognostic gene expression signatures to substitute for RNA4 subtype (part of BPI) or $SET_{ER/PR}$ index as components of SET2,3. All statistical analyses were carried out in R version 3.4.3 (https://cran.r-project.org) using and Bioconductor (www.bioconductor.org).

RESULTS

Characteristics of the clinical cohorts

The characteristics of the MDACC and I-SPY2 trial cohorts of HR+/HER2− cancers are summarized in Table 1 and Figure 2 (Consort diagrams). The main differences between cohorts were that 76% of I-SPY2 subjects received an experimental treatment with neoadjuvant chemotherapy, and all patients in I-SPY2 had a high-risk MammaPrint result. The MDACC cohort had a lower rate of clinical node positivity (52% versus 66%), lower rate of pCR or RCB-I (20% versus 32%), and longer duration of follow-up (11 versus 4 years). The I-SPY2 cohort did not include patients for whom RCB results were not recorded due to clinical progressive disease (PD), or not available due to incomplete pathologic review or primary surgical management. The cT and cN information was not available from 99 patients in I-SPY2 at the time of analysis.

Development of the composite SET2,3

The BPI and the $SET_{ER/PR}$ index have similar ranges, with higher values predicting a better outcome. SET2,3 was defined as the weighted sum of the BPI and the $SET_{ER/PR}$ index, using coefficients from their multivariate Cox regression model (from the MDACC cohort), as follows: $SET2,3 = 0.51 \times BPI + 0.75 \times SET_{ER/PR}$ (Figure 1). The rate of DRFS events at 10 years decreased linearly with increasing SET2,3 in the MDACC cohort, and from this we defined SET2,3 >1.77 as high SET2,3 status in the context of chemo-endocrine treatments (Supplementary Figure S1B, available at https://doi.org/10.1016/j.annonc.2021.02.011).

Prognostic performance of SET2,3 in the clinical cohorts

The rate of 3-year DRFS events decreased linearly with increasing SET2,3 in the test cohort from MDACC and the blinded validation cohort from the I-SPY2 trial (Supplementary Figure S4A and B, available at https://doi.org/10.1016/j.annonc.2021.02.011). A multivariate analysis of the MDACC cohort demonstrated that RCB, BPI, and $SET_{ER/PR}$ index all provided independently significant, long-term prognostic information (Table 2). $SET_{ER/PR}$ index and RCB were independently prognostic, but BPI was not independently prognostic in the I-SPY2 trial population with molecularly high-risk cancer (Table 2). SET2,3 and RCB did

not have significant interaction term, indicating an additive effect (Table 2). Treatment in the I-SPY2 trial (experimental versus control) was not prognostic [hazard ratio (HR) 1.02, 95% CI 0.50-2.07] when adjusted for SET2,3 (HR 0.43, 95% CI 0.28-0.65, $P < 0.001$) and RCB (HR 1.91, 95% CI 1.50-2.43, $P < 0.001$).

Prognostic performance of SET2,3 in patients with significant residual disease (RCB-II/III)

Moderate or extensive residual disease after neoadjuvant chemotherapy (RCB-II/III) was less likely if the pre-treatment tumor biopsy had low SET2,3 status: in the MDACC cohort (69% versus 86%, $P = 0.0005$) and the I-SPY2 trial (60% versus 82%, $P = 0.0007$). If patients had RCB-II or RCB-III after neoadjuvant chemotherapy, higher SET2,3 value in the pre-treatment biopsy was associated with lower relative risk of DRFS event, and this was observed in both clinical cohorts (Figure 3A and B).

Contribution of SET2,3 to molecular subtype and contemporary prognostic signatures

All three other prognostic gene expression signatures were prognostic in univariate analyses of the MDACC cohort, although MammaPrint test score was not prognostic in the I-SPY2 cohort that excluded the low-risk subset (Table 3). However, when SET2,3 was added to each signature in bivariate Cox models, SET2,3 was the only independently prognostic biomarker in every comparison (Table 3).

When evaluated specifically in patients with significant residual disease after neoadjuvant chemotherapy (i.e. RCB-II/III), SET2,3 added independent prognostic information to luminal molecular subtypes, whether described by PAM50 as luminal A versus luminal B in the MDACC cohort (Figure 3C), or using the pre-defined cut point to define high-risk (MP1) MammaPrint score, but not the ultra-high-risk (MP2) subset (Figure 3D). Corresponding results from the entire cohorts (including those with pCR or RCB-I) are shown in Supplementary Figure S5, available at https://doi.org/10.1016/j.annonc.2021.02.011.

The two cohorts of HR+/HER2− cancers had similar frequency of non-luminal subtype: 31% in MDACC and 28% in I-SPY2 (Supplementary Table S2, available at https://doi.org/10.1016/j.annonc.2021.02.011), with strong overlap of MP2 and basal-like subtypes (Supplementary Tables S3 and S4, available at https://doi.org/10.1016/j.annonc.2021.02.011). SET2,3 was low ($\leq$1.77) in almost all PAM50 basal-like (39/40), MammaPrint MP2 (69/72), or BluePrint basal-like (75/76) cancers. Indeed, SET2,3 was zero in 3/40 PAM50 basal-like and most MammaPrint MP2 (55/72) or BluePrint basal-like (61/76) cancers. Although luminal cancers had more RCB after chemotherapy than non-luminal cancers (Supplementary Table S2, available at https://doi.org/10.1016/j.annonc.2021.02.011), the frequency of high SET2,3 status was 81% in luminal A and 42%-51% in luminal B or MP1 cancers, with similar frequency observed across the RCB response classes (Supplementary Tables S3 and S4, available at https://doi.org/10.1016/j.annonc.2021.02.011).

Substitution of other prognostic signatures for components of SET2,3

We tested in the MDACC cohort whether other prognostic signatures could substitute for individual components of SET2,3. First, we recalculated BPI (cT, cN, prognostic signature), replacing RNA4 with each prognostic signature (PAM50, RS, EP, MP) scaled to the same range as RNA4. These recalculated BPIs retained independent prognostic significance, as did $SET_{ER/PR}$ index and RCB (Table 4). Next, we substituted $SET_{ER/PR}$ index with each prognostic signature (PAM50, RS, EP, MP) scaled to the same range as the $SET_{ER/PR}$ index. However, in those multivariate models, only BPI (cT, cN, RNA4) and RCB were independently prognostic (Table 4). So, other prognostic signatures could substitute for RNA4, but not for $SET_{ER/PR}$ index as a component of SET2,3.

Interpretation of RCB within subsets of high and low SET2,3

RCB was independently prognostic when compared to SET2,3 and its component signatures, or to other molecular prognostic signatures (Tables 2 and 3). Furthermore, SET2,3 contributed additive (rather than interactive) prognostic information to RCB (Table 2). Patients with more RCB after chemotherapy had a higher risk of DRFS event, relative to all the patients who had achieved pCR, irrespective of whether SET2,3 status was high or low (Figure 4A and B).

Table 2. Multivariate Cox models for components of residual prognostic risk (DRFS)

Multivariate: components of residual risk	MDACC cohort ($N = 307$)		I-SPY2 trial ($N = 268$)	
	HR (95% CI)	*P*	HR (95% CI)	*P*
Baseline prognostic index (BPI)	0.55 (0.40-0.76)	<0.001	1.02 (0.64-1.63)	0.929
Predicted endocrine sensitivity ($SET_{ER/PR}$ index)	0.56 (0.41-0.75)	<0.001	0.41 (0.27-0.62)	<0.001
Residual cancer burden after NAC (RCB)	2.03 (1.61-2.54)	<0.001	2.02 (1.57-2.59)	<0.001
Multivariate: SET2,3 versus RCB	**HR (95% CI)**	***P***	**HR (95% CI)**	***P***
SET2,3	0.23 (0.09-0.62)	0.004	0.27 (0.08-0.89)	0.031
RCB	1.77 (1.25-2.50)	<0.001	1.68 (1.14-2.45)	0.008
Interaction term (SET2,3 × RCB)	1.19 (0.87-1.62)	0.257	1.17 (0.81-1.68)	0.401

CI, confidence interval; DRFS, distant relapse-free survival; HR, hazard ratio; MDACC, MD Anderson Cancer Center; NAC, neoadjuvant chemotherapy.

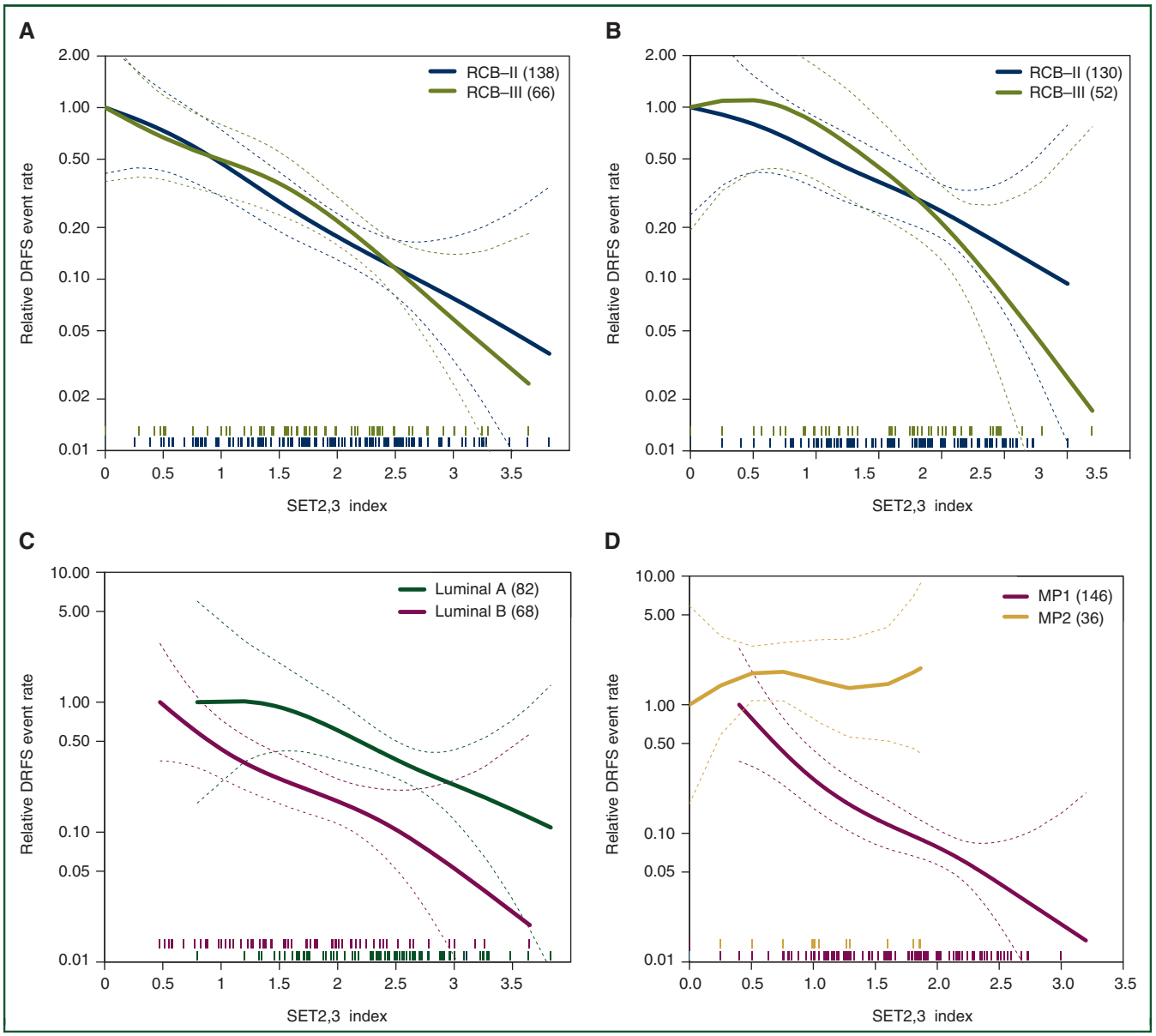

Figure 3. Prognostic performance of SET2,3 relative to extent of residual cancer burden (RCB) and molecular subtype.
The distant relapse-free survival (DRFS) hazard function for SET2,3 in the patients with significant residual disease after neoadjuvant chemotherapy (RCB-II/III). Risk of DRFS event for each SET2,3 value relative to lowest SET2,3 value in each group, according to RCB-II or RCB-III in (A) the MD Anderson Cancer Center (MDACC) cohort and (B) the I-SPY2 trial; and, in the patients with RCB-II/III, according to genomic subtype defined by: (C) PAM50 classifier as luminal A or luminal B in the MDACC cohort and (D) MammaPrint test as high-risk (MP1) or ultra-high-risk (MP2) subtype in the I-SPY2 trial. Note: Dashed lines represent 95% confidence interval bounds. RCB-III includes progressive disease. The MP1 and MP2 groups are equivalent to luminal B and basal subtypes, respectively.

Table 3. Performance of each prognostic signatures alone or in multivariate Cox model with SET2,3 (DRFS)

Prognostic signatures (continuous scores)	Univariate		Multivariate with SET2,3			
	Signature		Signature		SET2,3	
	HR (95% CI)	*P*	HR (95% CI)	*P*	HR (95% CI)	*P*
MDACC cohort study ($N = 307$)						
21-Gene recurrence score (RS)	1.01 (1.00-1.02)	0.030	0.99 (0.99-1.00)	0.242	0.49 (0.35-0.68)	<0.001
11-Gene EndoPredict (EP)	1.07 (1.02-1.14)	0.012	0.94 (0.86-1.01)	0.104	0.45 (0.32-0.64)	<0.001
70-Gene MammaPrint (MP)	10.27 (2.65-39.91)	0.001	1.34 (0.23-7.68)	0.741	0.57 (0.42-0.78)	<0.001
I-SPY2 trial ($N = 268$)						
MammaPrint test (MP)	0.54 (0.17-1.72)	0.299	2.82 (0.65-12.74)	0.178	0.39 (0.22-0.71)	0.002

CI, confidence interval; DRFS, distant relapse-free survival; HR, hazard ratio; MDACC, MD Anderson Cancer Center.

Table 4. Substitution analyses using other prognostic signatures for components of SET2,3: RNA4 (within BPI) and SET2,3 (DRFS)

Multivariate: components of residual risk	PAM50 subtype		21-Gene recurrence score		11-Gene EndoPredict		70-Gene MammaPrint	
	HR (95% CI)	*P*	HR (95% CI)	*P*	HR (95% CI)	*P*	HR (95% CI)	*P*
Signatures substituted in place of RNA4								
BPI: cT, cN, substitute signature	0.75 (0.64-0.86)	<0.001	0.75 (0.64-0.88)	<0.001	0.76 (0.65-0.90)	0.001	0.74 (0.63-0.87)	<0.001
$SET_{ER/PR}$ index	0.56 (0.41-0.75)	<0.001	0.54 (0.40-0.73)	<0.001	0.57 (0.41-0.78)	0.001	0.55 (0.40-0.74)	<0.001
RCB	1.95 (1.55-2.46)	<0.001	1.96 (1.56-2.47)	<0.001	1.98 (1.57-2.49)	<0.001	1.95 (1.55-2.45)	<0.001
Signatures substituted in place of $SET_{ER/PR}$ index								
BPI: cT, cN, RNA4	0.75 (0.64-0.88)	<0.001	0.72 (0.62-0.85)	<0.001	0.74 (0.64-0.87)	<0.001	0.73 (0.64-0.85)	<0.001
Substitute signature	1.16 (0.99-1.36)	0.069	1.09 (0.89-1.33)	0.419	1.14 (0.96-1.36)	0.139	1.22 (1.00-1.48)	0.051
RCB	1.82 (1.45-2.27)	<0.001	1.79 (1.43-2.25)	<0.001	1.84 (1.46-2.31)	<0.001	1.84 (1.47-2.30)	<0.001

BPI, baseline prognostic index; CI, confidence interval; cN, clinical nodal; cT, clinical tumor; DRFS, distant relapse-free survival; HR, hazard ratio; RCB, residual cancer burden.

RCB classes were strongly prognostic in the 55% of patients in the MDACC cohort whose cancer had low SET2,3 status (Figure 4C). On the other hand, RCB class was not significantly prognostic in the 45% of patients who had high SET2,3 status (predicted endocrine sensitive) (Figure 4D).

The results were similar in the I-SPY2 trial. RCB classes were strongly prognostic in 63% of patients whose cancer had low SET2,3 status (Figure 4E). Similarly, RCB class was not prognostic in the 37% of patients in I-SPY2 whose cancer had high SET2,3 status, although we did observe some DRFS events in the RCB-II group (Figure 4F).

DISCUSSION

Our findings underscore a concept that residual risk after adjuvant chemo-endocrine treatments is a function of stochastic risks from the original burden and biological nature of disease, chemosensitivity, endocrine sensitivity, and time. We have demonstrated their independently additive effects on risk using BPI from cT, cN, and RNA4 (or other contemporary signatures) for original burden and nature of disease, RCB for observed chemotherapy effect, and $SET_{ER/PR}$ index to measure hormone receptor-related transcriptional activity to predict endocrine sensitivity. The SET2,3 combines two of these three components, measuring endocrine-related transcription ($SET_{ER/PR}$) adjusted for BPI (Figure 1). We believe this is important for stage II-III disease because reliance on endocrine-based therapy should balance the baseline prognostic risk of the cancer with its predicted sensitivity to endocrine therapy. Indeed, the $SET_{ER/PR}$ index was independently prognostic in the I-SPY2 trial that preselected patients with high-risk cancer defined by the MammaPrint prognostic test (Tables 2 and 3). Furthermore, SET2,3 added prognostic information when residual disease was class RCB-II or RCB-III (Figure 3A and B) and whether luminal A or luminal B (or MP1) subtype in the patients with RCB-II/III (Figure 3C and D) and overall (Supplementary Figure S5, available at https://doi.org/10.1016/j.annonc.2021.02.011). Validation of longer-term prognostic performance will also be important.

Neoadjuvant endocrine-based therapy is an appealing way to de-escalate treatment for selected patients with clinical stage II-III HR+/HER2− breast cancer yet retain the ability to give chemotherapy in the adjuvant setting if there is insufficient response. Currently, there is no validated biomarker to identify suitable patients at the time of diagnosis, but SET2,3 is a promising biomarker for this purpose. Indeed, the $SET_{ER/PR}$ index within SET2,3 is a direct measure of endocrine transcriptional activity, and added unique information to contemporary prognostic signatures (Tables 3 and 4). Chemotherapy response (RCB classes) was less prognostic when SET2,3 was high (Figure 4D and F), and was statistically under-powered to observe a difference. However, we should not infer that standard endocrine therapy alone would be sufficient for every patient with high SET2,3, but we might consider these patients for clinical trials with neoadjuvant combining endocrine and other targeted treatments. Hence, we think it will be important to evaluate SET2,3 in samples from the relevant phase III adjuvant trials of contemporary endocrine-based regimens (augmented by combined targeted therapy) for stage II-III disease.[29,30]

Pathological response to chemotherapy (RCB) was strongly prognostic when SET2,3 was low (Figure 4C and E). Although 31% achieved pCR or RCB-I (with excellent prognosis), the residual prognostic risk of patients with RCB-II or RCB-III after neoadjuvant chemotherapy resembled that of hormone receptor-negative disease, despite their receiving adjuvant endocrine therapy.[6] These tumors include almost all basal-like cancers (BluePrint or PAM50) and many of the luminal cancers in each cohort. They are unlikely to respond to endocrine therapy alone, and patients' residual prognostic risk depended mostly on their tumor's sensitivity to chemotherapy.

Although our study demonstrates proof of concept, including blinded independent validation, there are relevant caveats. It is important to emphasize that the measurements of 21-gene RS Oncotype-DX, 11-gene EP, 70-gene MammaPrint, and PAM50 subtype from U133A microarrays represent approximations of these tests, so we avoided categorical assignments where possible, since cut points could be unreliable. However, the MammaPrint scores from the I-SPY2 trial were the clinical test results. On the other hand, measurements of SET2,3 in the I-SPY2 trial cohort were approximations from the adapted Agilent microarrays (Agendia), and we should assume less accurate categorical assignment (Figure 4E and F). Nevertheless, SET2,3 was independently additive to contemporary

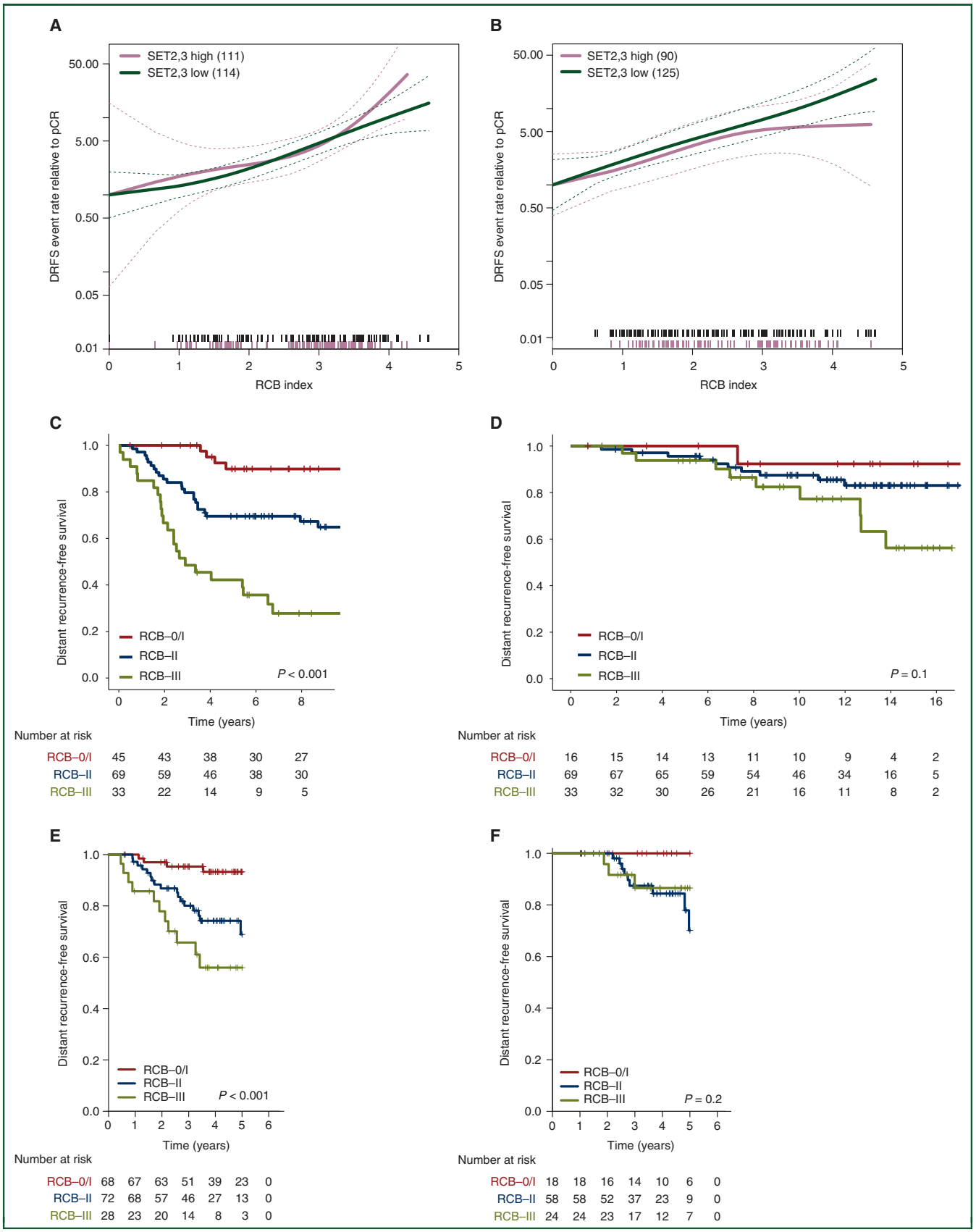

Figure 4. Prognosis of residual cancer burden (RCB) according to SET2,3 status.
RCB was evaluated within high and low SET2,3 classes, first as the hazard function for RCB values in the patients with residual disease (RD) by SET2,3 status (low SET2,3; high SET2,3), relative to all the patients with pathologic complete response (pCR) in (A) the MD Anderson Cancer Center (MDACC) cohort and (B) the I-SPY2 trial; and then as RCB classes (pCR or RCB-I, RCB-II, or RCB-III) in the MDACC cohort in cancers with low SET2,3 (C) or high SET2,3 (D), and in the I-SPY2 trial in cancers with low SET2,3 (E) or high SET2,3 (F). Note: Dashed lines represent 95% confidence interval bounds. In the MDACC cohort, there were 225 patients with RD and 36 with pCR (6 of 36 had high SET2,3). In the I-SPY2 trial, there were 215 patients with RD and 50 with pCR (8 of 50 had high SET2,3).

prognostic signatures in both cohorts. We also recognize that additional validation studies will be necessary to demonstrate generalizability and longer-term prognosis. Median follow-up was for shorter duration for the I-SPY2 trial (3.8 years), so we carried out sensitivity analysis of the MDACC cohort censored at 4 years of follow-up (Supplementary Tables S5-S7, available at https://doi.org/10.1016/j.annonc.2021.02.011), demonstrating consistent findings to the results at 11 years of median follow-up and to the I-SPY2 trial results (Tables 2-4). Also, investigational treatments were included with neoadjuvant chemotherapy in the I-SPY2 trial but did not alter the prognostic performance of SET2,3 or meaning of RCB (unpublished results). Finally, we note that SET2,3 is accurately measured from formalin-fixed paraffin-embedded (FFPE) tissues within and between laboratories.[12,31] So, it will be possible to evaluate SET2,3 in FFPE tumor samples from other clinical trial cohorts.

Overall, our study demonstrates that the SET2,3 index of sensitivity to endocrine therapy for stage II-III cancers adds meaningful and independent prognostic information to neoadjuvant chemotherapy response and molecular prognostic subtypes. These results were independently validated in the I-SPY2 clinical trial, notwithstanding the methodological differences. Our study also highlights the importance of neoadjuvant chemotherapy benefit when cancer has low SET2,3, and suggests a future opportunity to test within clinical trials whether patients with high SET2,3 benefit from a treatment plan that begins with neoadjuvant endocrine-based treatment (perhaps with additional targeted therapy), rather than chemotherapy. The potential utility of this assay deserves to be evaluated further.

ACKNOWLEDGEMENTS

This publication is partly based on research using information obtained from www.projectdatasphere.org, which is maintained by Project Data Sphere, LLC. Neither Project Data Sphere, LLC nor the owner(s) of any information from the web site have contributed to, approved, or are in any way responsible for the contents of this publication. The authors thank Dr Bastiaan van der Baan for supporting this analysis of Agendia test results from the I-SPY2 trial. Additional acknowledgements that relate to the overall I-SPY2 TRIAL have been included in the Supplementary Appendix, available at https://doi.org/10.1016/j.annonc.2021.02.011.

FUNDING

This work was supported in part by grant funding from the Breast Cancer Research Foundation (to WFS); National Institutes of Health [grant number P0513149] (to LE, LV, WFS); Cancer Prevention and Research Institute of Texas (CPRIT) [grant number RP180712] (to WFS, DT); and Safeway Foundation (I-SPY2). The I-SPY2 TRIAL is supported by Quantum Leap Healthcare Collaborative (2013 to present) and the Foundation for the National Institutes of Health (2010-2012) and by a grant [grant number 28XS197] from the National Cancer Institute Center for Biomedical Informatics and Information Technology.

DISCLOSURE

WFS is a co-inventor of a pending patent application for the sensitivity to endocrine therapy and co-founder with equity in Delphi Diagnostics that licensed the intellectual property. LP and WFS are co-inventors of an issued patent for the algorithm to calculate residual cancer burden that is freely available on the internet. LvV is a co-inventor of an issued patent for the MammaPrint test and holds equity in Agendia that licensed the intellectual property. The remaining authors have declared no conflicts of interest.

REFERENCES

1. Sparano JA, Zhao F, Martino S, et al. Long-term follow-up of the E1199 phase III trial evaluating the role of taxane and schedule in operable breast cancer. *J Clin Oncol*. 2015;33(21):2353-2360.
2. Liu MC, Pitcher BN, Mardis ER, et al. PAM50 gene signatures and breast cancer prognosis with adjuvant anthracycline- and taxane-based chemotherapy: correlative analysis of C9741 (Alliance). *NPJ Breast Cancer.* 2016;2:15023.
3. Pan H, Gray R, Braybrooke J, et al. 20-Year risks of breast-cancer recurrence after stopping endocrine therapy at 5 years. *N Engl J Med*. 2017;377(19):1836-1846.
4. 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(1):55-65.
5. Symmans WF, Peintinger F, Hatzis C, et al. Measurement of residual breast cancer burden to predict survival after neoadjuvant chemotherapy. *J Clin Oncol*. 2007;25(28):4414-4422.
6. Symmans WF, Wei C, Gould R, et al. Long-term prognostic risk after neoadjuvant chemotherapy associated with residual cancer burden and breast cancer subtype. *J Clin Oncol*. 2017;35(10):1049-1060.
7. Hatzis C, Pusztai L, Valero V, et al. A genomic predictor of response and survival following taxane-anthracycline chemotherapy for invasive breast cancer. *J Am Med Assoc*. 2011;305(18):1873-1881.
8. Symmans WF, Hatzis C, Sotiriou C, et al. Genomic index of sensitivity to endocrine therapy for breast cancer. *J Clin Oncol*. 2010;28(27):4111-4119.
9. Sinn BV, Fu C, Lau R, et al. SET-ER/PR—a robust 18-gene predictor for sensitivity to endocrine therapy for metastatic breast cancer. *NPJ Breast Cancer.* 2019;5:16.
10. 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.
11. 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.
12. Lau R, Du L, Chen E, et al. Technical validity of a customized assay of sensitivity to endocrine therapy using sections from fixed breast cancer tissue. *Clin Chem*. 2020;66(7):934-945.
13. Martin M, Pienkowski T, Mackey J, et al. Adjuvant docetaxel for node-positive breast cancer. *N Engl J Med*. 2005;352(22):2302-2313.
14. Eiermann W, Pienkowski T, Crown J, et al. Phase III study of doxorubicin/cyclophosphamide with concomitant versus sequential docetaxel as adjuvant treatment in patients with human epidermal growth factor receptor 2-normal, node-positive breast cancer: BCIRG-005 trial. *J Clin Oncol*. 2011;29(29):3877-3884.

15. Mackey JR, Pienkowski T, Crown J, et al. Long-term outcomes after adjuvant treatment of sequential versus combination docetaxel with doxorubicin and cyclophosphamide in node-positive breast cancer: BCIRG-005 randomized trial. *Ann Oncol*. 2016;27(6):1041-1047.
16. Sestak I, Buus R, Cuzick J, et al. Comparison of the performance of 6 prognostic signatures for estrogen receptor-positive breast cancer: a secondary analysis of a randomized clinical trial. *JAMA Oncol*. 2018;4(4):545-553.
17. Prat A, Cheang MC, Martin M, et al. Prognostic significance of progesterone receptor-positive tumor cells within immunohistochemically defined luminal A breast cancer. *J Clin Oncol*. 2013;31(2):203-209.
18. Karn T, Metzler D, Ruckhaberle E, et al. Data-driven derivation of cutoffs from a pool of 3,030 Affymetrix arrays to stratify distinct clinical types of breast cancer. *Breast Cancer Res Treat*. 2010;120(3):567-579.
19. Park JW, Liu MC, Yee D, et al. Adaptive randomization of neratinib in early breast cancer. *N Engl J Med*. 2016;375(1):11-22.
20. Rugo HS, Olopade OI, DeMichele A, et al. Adaptive randomization of veliparib-carboplatin treatment in breast cancer. *N Engl J Med*. 2016;375(1):23-34.
21. Albain KS, Leyland-Jones B, Symmans F, et al. Abstract P1-14-03: The evaluation of trebananib plus standard neoadjuvant therapy in high-risk breast cancer: results from the I-SPY 2 TRIAL. *Cancer Res*. 2016;76(suppl 4):P1-14-03.
22. Yee D, Paoloni M, van't veer L, et al. Abstract P6-11-04: The evaluation of ganitumab/metformin plus standard neoadjuvant therapy in high-risk breast cancer: results from the I-SPY 2 trial. *Cancer Res*. 2017;77(suppl 4):P6-11-04.
23. Chien AJ, Tripathy D, Albain KS, et al. MK-2206 and standard neoadjuvant chemotherapy improves response in patients with human epidermal growth factor receptor 2-positive and/or hormone receptor-negative breast cancers in the I-SPY 2 trial. *J Clin Oncol*. 2020;38(10):1059-1069.
24. Forero A, Yee D, Buxton MB, et al. Abstract P6-11-02: efficacy of Hsp90 inhibitor ganetespib plus standard neoadjuvant therapy in high-risk breast cancer: results from the I-SPY 2 trial. *Cancer Res*. 2017;77(suppl 4):P6-11-02.
25. Consortium IST, Yee D, DeMichele AM, et al. Association of event-free and distant recurrence-free survival with individual-level pathologic complete response in neoadjuvant treatment of stages 2 and 3 breast cancer: three-year follow-up analysis for the I-SPY2 adaptively randomized clinical trial. *JAMA Oncol*. 2020;6(9):1355-1362.
26. 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(11):1067-1076.
27. Hudis CA, Barlow WE, Costantino JP, et al. Proposal for standardized definitions for efficacy end points in adjuvant breast cancer trials: the STEEP system. *J Clin Oncol*. 2007;25(15):2127-2132.
28. Pocock SJ, Clayton TC, Altman DG. Survival plots of time-to-event outcomes in clinical trials: good practice and pitfalls. *Lancet*. 2002;359(9318):1686-1689.
29. Johnston SRD, Harbeck N, Hegg R, et al. Abemaciclib combined with endocrine therapy for the adjuvant treatment of HR+, HER2−, node-positive, high-risk, early breast cancer (monarchE). *J Clin Oncol*. 2020;38(34):3987-3998.
30. Mayer EL, Dueck AC, Martin M, et al. Palbociclib with adjuvant endocrine therapy in early breast cancer (PALLAS): interim analysis of a multicentre, open-label, randomised, phase 3 study. *Lancet Oncol*. 2021;22(2):212-222.
31. Lau R, Bossuyt V, Young B, et al. Reproducibility of the sensitivity to endocrine therapy (SET) assay for stage II/III breast cancer within and between pathology laboratories. *Lab Invest*. 2019;99:87 (abstr 191).