OXFORD

JNCI J Natl Cancer Inst (2018) 110(2): djx137

doi: 10.1093/jnci/djx137
First published online August 4, 2017
Article

ARTICLE

Relevance of Spatial Heterogeneity of Immune Infiltration for Predicting Risk of Recurrence After Endocrine Therapy of ER+ Breast Cancer

Andreas Heindl, Ivana Sestak, Kalnisha Naidoo, Jack Cuzick, Mitchell Dowsett, Yinyin Yuan

Affiliations of authors: Centre for Evolution and Cancer, The Institute of Cancer Research, London, UK (AH, YY); Centre for Molecular Pathology, Royal Marsden Hospital, London, UK (AH, YY); Division of Molecular Pathology, The Institute of Cancer Research, London, UK (AH, YY); Centre for Cancer Prevention, Wolfson Institute of Preventive Medicine, Queen Mary University of London, London, UK (IS, JC); The Breast Cancer Now Toby Robins Research Centre, Institute of Cancer Research, London, UK (KN, MD); Cellular Pathology, Guy's and St Thomas' NHS Trust, Westminster Bridge Rd, London, UK (KN); Ralph Lauren Centre for Breast Cancer Research, Royal Marsden Hospital, London, UK (MD).

Correspondence to: Yinyin Yuan, PhD, Centre for Molecular Pathology and The Royal Marsden Hospital, 15 Cotswold Road, SM2 5NG, London (e-mail: yinyin.yuan@icr.ac.uk).

Abstract

Background: Despite increasing evidence supporting the clinical utility of immune infiltration in the estrogen receptor–negative (ER-) subtype, the prognostic value of immune infiltration for ER+ disease is not well defined.
Methods: Quantitative immune scores of cell abundance and spatial heterogeneity were computed using a fully automated hematoxylin and eosin–stained image analysis algorithm and spatial statistics for 1178 postmenopausal patients with ER+ breast cancer treated with five years' tamoxifen or anastrozole. The prognostic significance of immune scores was compared with Oncotype DX 21-gene recurrence score (RS), PAM50 risk of recurrence (ROR) score, IHC4, and clinical treatment score, available for 963 patients. Statistical tests were two-sided.
Results: Scores of immune cell abundance were not associated with recurrence-free survival. In contrast, high immune spatial scores indicating increased cell spatial clustering were associated with poor 10-year, early (0–5 years), and late (5–10 years) recurrence-free survival (Immune Hotspot: LR-$\chi^2 = 14.06$, $P < .001$, for 0–10 years; LR-$\chi^2 = 6.24$, $P = .01$, for 0–5 years; LR-$\chi^2 = 7.89$, $P = .005$, for 5–10 years). The prognostic value of spatial scores for late recurrence was similar to that of IHC4 and RS in both populations, but was not as strong as other tests in comparison for recurrence across 10 years.
Conclusions: These results provide a missing link between tumor immunity and disease outcome in ER+ disease by examining tumor spatial architecture. The association between spatial scores and late recurrence suggests a lasting memory of protumor immunity that may impact disease progression and evolution of endocrine treatment resistance, which may be exploited for therapeutic advances.

Estrogen receptor–positive (ER+) subtype accounts for about 80% of all breast cancers, the most common cancer in women. At diagnosis, the majority of ER+ patients have a good prognosis if treated with endocrine therapy. However, a subset of patients is at risk for disease recurrence and death, particularly after five years of adjuvant endocrine therapy. Differentiating these patients from low-risk patients who can safely avoid chemotherapy is a priority for breast cancer management (1). Currently available prognostic tests to predict risk in endocrine-treated patients include the widely used Oncotype DX 21-gene recurrence score (RS) (2), the PAM50 risk of recurrence (ROR) score (3), and the immunohistochemistry-based IHC4 test that is combined with the clinical treatment score (CTS) to integrate clinicopathological parameters (4). In particular, the amount of

Received: January 26, 2017; **Revised:** April 6, 2017; **Accepted:** June 5, 2017

prognostic information provided for long-term (0–10 years) and late (5–10) recurrence varies across these tests (5,6).

Immune infiltration is not explicitly accounted for in any of the above tests. Increasing evidence supports the role of tumor-infiltrating lymphocytes (TILs) in influencing disease progression and treatment response in breast cancer (7–10). Characterization of the nature of immune responses is key to understanding tumor immunity and empowering immunotherapy. However, the majority of reports focus on ER- and human epidermal growth factor receptor 2–positive (HER2+) breast cancers, where extensive immune infiltration is more common and immune scores were found to be highly predictive of survival and response to chemotherapy (8,11–15). In contrast, there is a lack of definitive data on the prognostic value of immune scores in ER+ breast cancer following endocrine treatment (7,16). A major reason for this is the absence of reproducible scoring methods to facilitate large-scale studies of ER+ breast cancer. This also limits the translation of immune scoring into clinical advances.

We have developed quantitative and reproducible approaches to score lymphocytic infiltration (LI) in breast cancer, based on fully automated image analysis of routinely generated hematoxylin and eosin (H&E)–stained histology sections. Quantitative immune scores of overall LI in tumors, as well as intratumor lymphocyte ratio (ITLR), were associated with good survival in ER- and ER-/HER2- breast cancer (17,18). In addition, our automated image analysis scheme enables the study of complex spatial patterns of TILs (19). The spatial interactions among TILs and cancer cells generate complex ecological dynamics that can ultimately impact tumor progression and response to treatment (8,20–22). In this study of 1178 postmenopausal breast cancer patients with ER+ disease, patients enrolled in the Arimidex or Tamoxifen Alone or Combined (ATAC) trial, our aims were to 1) establish the prognostic value of H&E-based, quantitative scores of TIL abundance as well as spatial heterogeneity for 10-year, early recurrence (0–5 years) and late recurrence (5–10 years) after endocrine therapy; 2) compare their prognostic value with established prognostic tests including RS, ROR, IHC4, and CTS; 3) evaluate a new histology-based model that combines IHC4 and immune scores as a cost-effective biomarker.

ARTICLE

Methods

Study Population

Classical clinicopathologic factors (age, nodal status, tumor size, grade, randomized treatment) were collected from patients with ER+ primary breast cancer in the ATAC trial who were randomly assigned to either anastrozole or tamoxifen (Supplementary Table 1, available online) (23). A total of 1178 eligible patients who did not receive chemotherapy and from whom H&E-stained slides from formalin-fixed, paraffin-embedded tissues were available were included (Figure 1). Of these, 963 patients were scored with prognostic scores including IHC4, RS, ROR46, and CTS (Table 1).

A total of 1037 tumors were HER2-, 909 of which were scored with prognostic scores. A subset of 91 TransATAC samples were randomly selected and scored on H&E sections according to international recommendations (24) by a histopathologist (KN). Baseline demographics and clinical characteristics for all patients included in this analysis are provided in Supplementary Table 1 (available online). This study was approved by the South-East London Research Ethics Committee, and all patients included gave informed consent.

H&E Image Analysis and Validation

We curated a digital database of H&E histology slides for TransATAC and applied our histology image analysis pipeline (Supplementary Figure 1A, available online) (17). In brief, the image analysis pipeline exploits the nuclear morphological differences among cancer cells, lymphocytes, and stromal cells to differentiate them in H&E histological tissue sections. Cancer cell nuclei are generally large in size and demonstrate greater variability in appearance as compared with lymphocyte and stromal cell nuclei; lymphocyte nuclei are typically small, round, and homogeneously basophilic, and nuclei of stromal cells including fibroblasts and endothelial cells are more elongated. The pipeline consisted of four stages: 1) unsupervised segmentation of the nuclei; 2) supervised classification of individual cell nuclei into cancer, lymphocyte, other cell nuclei, and artefacts; 3) kernel smoothing to correct local sporadic errors; and 4) a hierarchical multiresolution model fitting to identify cancer cell clusters to further improve classification accuracy. The classifier was previously validated in molecular taxonomy of breast cancer international consortium (METABRIC) to have an overall accuracy of 90.1% and a high correlation between image analysis and pathological scores in the entire cohort (17).

To evaluate the accuracy of our image analysis pipeline for TransATAC, a test set of 627 cells randomly sampled from three images was annotated by a pathologist (DNR) blinded to image analysis results (Supplementary Figure 1A, available online). Accuracy rates for identifying the three cell types were 93.8% for cancer cells, 87.9% for lymphocytes, and 84.2% for stromal cells (Supplementary Figure 1B, available online). The balanced accuracy rates as the average for sensitivity and specificity for the three cell types were 0.864 for cancer, 0.839 for lymphocytes, and 0.876 for stromal cells (Supplementary Figure 1C, available online). On average, 217 101 cancer cells (SD = 178 677.5 cancer cells), 25 956 lymphocytes (SD = 35 365.21 lymphocytes), and 161 341 stromal cells (SD = 91 862.59 stromal cells) were identified in each TransATAC whole-section sample, with a total of 525 718 198 cells identified in the whole cohort.

Automated Scoring of Immune Cell Abundance and Spatial Heterogeneity

Immune cell abundance scores include Lymphocytic Infiltration (LI), which summarizes the fraction of lymphocytes in all cells in the tumor section; and Intratumor Lymphocyte Ratio (ITLR), Adjacent-to-Tumor Lymphocyte Ratio (ATLR), and Distal-To-Tumor Lymphocyte Ratio (DTLR), which are defined as the number of specific types of lymphocytes normalized by the number of cancer cells (17,18). Intratumor Lymphocytes (ITLs), Adjacent-to-Tumor Lymphocytes (ATLs), and Distal-To-Tumor Lymphocytes (DTLs) were identified using unsupervised clustering based on their spatial proximities to cancer cells, which was quantified using a kernel density method on the distribution of cancer cells.

Immune cell spatial scores quantify the amount of spatial clusters, or hotspots, formed by lymphocytes and/or cancer cells within the section (Figure 2) (13). Getis-Ord spatial analysis was carried out to identify tumor regions with statistically significant spatial clustering of immune cells, that is, immune hotspots. This means that the frequency of immune cells appearing at these locations is greater than expected by chance given the distribution of all cells in the entire tumor section and that, importantly, the difference between the actual and expected value is statistically significant. A P value was

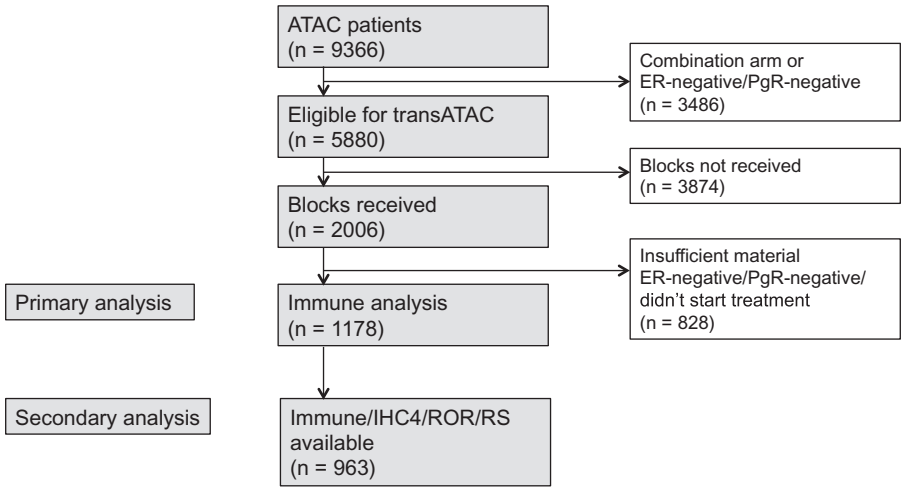

Figure 1. Consort diagram for the availability of samples for analysis from the Arimidex, Tamoxifen Alone or Combined (ATAC) trial. ER = estrogen receptor; IHC = immunohistochemistry; PgR = progesterone receptor; ROR = risk of recurrence score; RS = recurrence score.

Table 1. Immune scores and prognostic scores*

Type and name	Definition	Reference
Immune abundance scores (based on H&E)		
Lymphocytic Infiltrate	Fraction of lymphocytes in all cells. Cells were identified using automated histology image analysis.	Yuan et al. 2012 (17)
Intratumor Lymphocyte Ratio	The number of intratumor lymphocytes normalized by the number of cancer cells. Intratumor lymphocytes were identified as the cluster of lymphocytes to be the closest to cancer cells based on their spatial proximity in unsupervised clustering.	Yuan et al. 2015 (18)
Adjacent-Tumor Lymphocyte Ratio	The number of adjacent-to-tumor lymphocytes normalized by the number of cancer cells. Adjacent-to-tumor lymphocytes were identified as the intermediate cluster of lymphocytes based on their spatial proximity to cancer cells in unsupervised clustering.	
Distal-Tumor Lymphocyte Ratio	The number of distal-tumor lymphocytes normalized by the number of cancer cells. Distal-tumor lymphocytes were identified as the cluster of lymphocytes to be the furthest away from cancer cells based on their spatial proximity in unsupervised clustering.	
Immune spatial scores (H&E)		
Cancer Hotspot	Fraction of tissue displaying spatial clustering of cancer cells. Tumor regions with spatial clustering of cancer cells were identified using Getis-Ord hotspot analysis, which assigned statistical significance of difference between observed local cancer cell density and global mean.	Nawaz et al. 2015 (13)
Immune Hotspot	Fraction of tissue displaying spatial clustering of immune cells. Tumor regions with spatial clustering of immune cells were identified in a similar way as cancer hotspots.	
Immune-Cancer Hotspot	Fraction of tissue displaying spatial clustering of immune and cancer cells simultaneously, in other words, tumor regions that are both immune hotspots and cancer hotspots.	
Prognostic scores		
IHC4 (IHC)	Immunohistochemistry-based score as a combination of ER, PgR, HER2, Ki67 expression, previously derived using TransATAC samples.	Cuzick et al. 2011 (4)
CTS (clinical)	Clinical treatment score that considers node status, size, grade, age, and treatment, previously derived using TransATAC samples.	Cuzick et al. 2011 (4)
RS (Molecular)	Oncotype DX 21-gene recurrence score based on RNA expression of 21 prespecified Oncotype DX genes.	Paik et al. 2004 (2)
ROR (Molecular)	PAM50 risk of recurrence score that combines molecular signatures with clinical information on tumor size.	Nielsen et al. 2010 (3)

*Unless specified otherwise, the scores were predefined for validation in external cohorts. H&E = hematoxylin & eosin; IHC = immunohistochemistry.

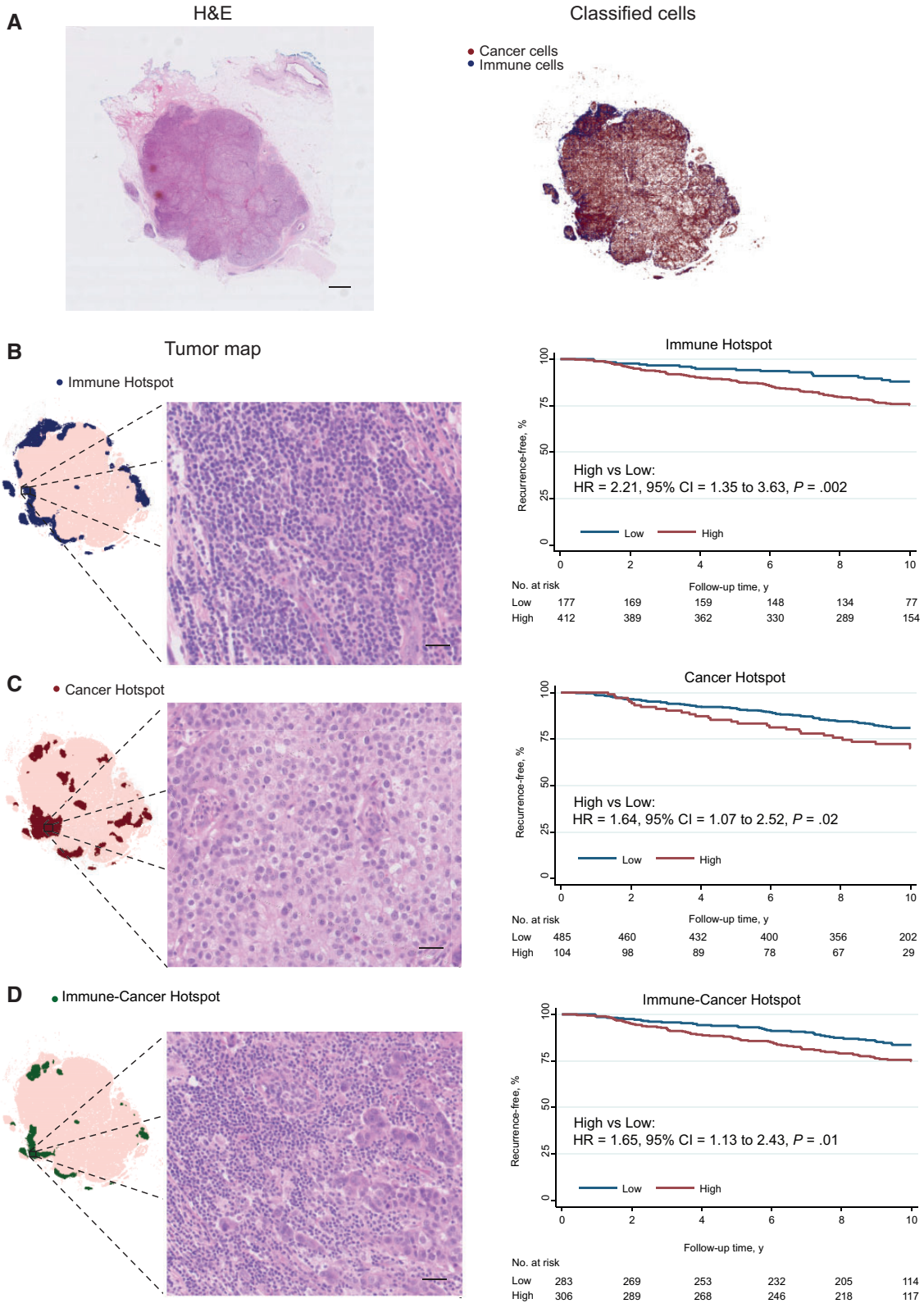

Figure 2. Illustration of the pipeline for identifying spatial hotspots with visual examples, and the Kaplan-Meier estimates for 10-year recurrence according to immune spatial scores in the validation set, split into two groups using cutoffs selected in the training set. **A)** An example of a TransATAC hematoxylin and eosin image and corresponding map of identified cancer and immune cells. **Scale bar** illustrates 2.5 mm. **B–D)** Visual examples of hotspots and Kaplan-Meier curves illustrating survival associations with immune spatial scores. **Scale bar** illustrates 35 μm. Kaplan-Meier curves were calculated and tested for equality using the log-rank test. The numbers of patients at risk in each group at various time points are given below each graph. All statistical tests were two-sided. HR = hazard ratio (95% confidence interval). CI = confidence interval; H&E = hematoxylin and eosin; HR = hazard ratio.

ARTICLE

computed for each spatial region, which can be used to determine statistical significance using a significance level of .05. Previously, tumors with a high amount of regions that are both immune hotspots and cancer hotspots (immune-cancer hotspots) were found to have good prognosis in ER- cancer (13). Here for comparison in ER+ breast cancer, Cancer Hotspot and Immune Hotspot scores that examine only one type of cell at a time were also included.

For validation of the intercorrelations among immune scores, H&E-stained, whole-section images of 743 ER+, treatment-naïve primary tumors from the METABRIC study (25) were analyzed using the same methods. All patient specimens were obtained with appropriate ethical approval from the relevant institutional review boards.

Statistical Analyses

Our primary objective was to assess whether immune scores had statistically significant prognostic information for predicting 10-year recurrence in postmenopausal women with breast cancer given either tamoxifen or anastrozole monotherapy but not chemotherapy. Secondary analyses included determining the prognostic ability of immune scores in predicting early (0–5 years) and late recurrences (5–10 years) in patients divided into subgroups by HER2 status, and the additional prognostic information provided by tests in multivariable comparisons including age (<65, ≥65 years), nodal status (0, 1–3, 4+), tumor size (≤1 cm, >1 to ≤2 cm, >2 to ≤3 cm, >3 cm), central read grade (poor, intermediate, well differentiated), and randomized treatment (anastrozole vstamoxifen). Hazard ratios are for a change in 1 SD in the overall data set to compare the effect size between different immune scores. The contribution of each of the variables was evaluated by the change in likelihood ratio χ^2 (LR-χ^2; 1 *df*, significance level $\chi^2 = 3.84$) in three ways: by univariate analyses, as an addition to a model containing only the clinical variables, and as a difference in LR-χ^2 when the variable was added to the IHC4 score. Sample splitting was used, in which the immune score was dichotomized by using half the data as the training set, and then the cutoff points for each score were evaluated in the remaining half of the data as the validation set. Because the use of optimizing cutoffs may lead to overestimation of prognostic power, dichotomized variables were only used for analysis presented in Figure 2, B–D, and continuous variables were used elsewhere. For measuring correlation among immune scores and with pathological TIL scores, Spearman's correlation was used. All statistical analyses were performed using STATA version 13.1 or R version 3.3.1. All statistical tests were two-sided, and a *P* value of less than .05 was considered statistically significant.

Results

Correlations Among Immune Scores

Four immune abundance scores (overall Lymphocytic Infiltration, Intratumor Lymphocyte Ratio, Adjacent-to-Tumor Lymphocyte Ratio, and Distal-To-Tumor Lymphocyte Ratio) and three spatial scores (Immune Hotspot, Cancer Hotspot, Immune-Cancer Hotspot) were calculated based on fully automated histology image analysis on whole-section slides (Figure 2, Table 1). There was a strong, negative correlation between ITLR and DTLR ($r = -0.888$) (Table 2), indicating that lymphocytes either infiltrate into close contact with cancer cells or largely stay in the stromal area. Correlations among spatial scores were also strong ($r = 0.502$–0.796), suggesting that spatial clustering of cancer cells and lymphocytes tends to co-occur in the same tumors. These data were further validated in the METABRIC cohort (n = 743) (Table 2) (25). We then compared the automated scores to a pathologist's TIL score (24) in a subset of 91 TransATAC samples. Overall, a weak correlation between the pathologist's score and all automated scores was found ($r < 0.260$), with the highest correlation observed between TIL scoring and DTLR ($r = 0.259$) (Supplementary Figure 2, available online).

Prognostic Value of Immune Scores

None of the immune abundance scores provided significant prognostic information for recurrence ($P > .1$). In contrast, high spatial scores were associated with poor recurrence-free survival across 10 years in the univariate analysis (Immune Hotspot: n = 1178, LR-$\chi^2 = 14.06$, $P < .001$) (Table 3). When dichotomized, immune spatial scores were also prognostic (Immune Hotspot training set: n = 589, $P = .01$, hazard ratio [HR] = 1.88, 95% confidence interval [CI] = 1.16 to 3.07; validation set: n = 589, $P = .002$, HR = 2.21, 95% CI = 1.35 to 3.63) (Figure 2; Supplementary Figure 3, available online). In addition, immune spatial scores were statistically significantly prognostic for early (0–5 years) and late recurrence (5–10 years; Immune Hotspot 0–5 years: LR-$\chi^2 = 6.24$, $P = .01$; 5–10 years: LR-$\chi^2 = 7.89$, $P = .005$) (Table 3). In the multivariable analysis adjusted for clinical variables as expressed by the CTS including node status, tumor size, grade, age, and treatment, spatial scores remained prognostic for all three time windows, except for Immune Hotspot for early recurrence and Immune-Cancer Hotspot for late recurrence (Table 3).

None of the patients in this study received trastuzumab if their tumors were HER2+, as is current practice. In the HER2- population (n = 1037), again spatial scores but not abundance scores were prognostic for all time windows (Table 3). In the multivariable analysis, spatial scores were prognostic for all time windows except for Immune Hotspot and Immune-Cancer Hotspot for early recurrence (Table 3). We henceforth focused on spatial scores only.

Comparison of Immune Spatial Scores With RS, IHC4, ROR, and CTS

The prognostic value of spatial scores for late recurrence (5–10 years) is similar to that of IHC4 and RS in both the overall population (Immune Hotspot: LR-$\chi^2 = 6.93$; IHC: LR-$\chi^2 = 6.75$; RS: LR-$\chi^2 = 6.79$) and the HER2- population (Immune Hotspot: LR-$\chi^2 = 9.80$; IHC4: LR-$\chi^2 = 10.87$; RS: LR-$\chi^2 = 7.78$) (Figure 3). None of these scores, however, added to IHC4 and RS for early recurrence (ΔLR-$\chi^2 \leq 3.84$) or was as prognostic as ROR and CTS in any time window (Figure 3). We then examined the additional prognostic value of spatial scores to IHC4 and RS for 0 to 10 years and late recurrence. Immune-Cancer Hotspot provided statistically significant prognostic value when added to IHC4 and RS for years 0 to 10 whereas Immune Hotspot also added prognostic value to IHC4 and RS for late recurrence (ΔLR-$\chi^2 > 3.84$) (Table 4). Cancer Hotspot, on the other hand, added prognostic information to IHC4 and RS in both time windows (Table 4). In the HER2- population, Immune Hotspot and Immune-Cancer Hotspot added statistically significant prognostic information to IHC4 and RS in both time windows (Table 4). Again, Cancer Hotspot added statistically significant information to IHC4 and RS in both time windows (Table 4).

ARTICLE

Table 2. Correlations among immune scores and clinical variables in TransATAC and METABRIC*

Data set/variable	ITL	ATL	DTL	LI	Cancer Hotspot	Immune Hotspot	Immune-Cancer Hotspot	Age	Grade	Node	Size	Treatment
TransATAC												
ITL	–	–	–	–	–	–	–	–	–	–	–	–
ATL	–0.317	–	–	–	–	–	–	–	–	–	–	–
DTL	–0.888	–0.154	–	–	–	–	–	–	–	–	–	–
LI	–0.01	0.381	–0.174	–	–	–	–	–	–	–	–	–
Cancer Hotspot	–0.422	0.309	0.289	0.167	–	–	–	–	–	–	–	–
Immune Hotspot	–0.395	0.409	0.213	0.476	0.796	–	–	–	–	–	–	–
Immune-Cancer Hotspot	–0.408	0.203	0.327	0.341	0.502	0.75	–	–	–	–	–	–
Age	0.01	0.018	–0.019	–0.036	0.145	0.077	0.033	–	–	–	–	–
Grade	–0.168	0.126	0.114	0.144	0.268	0.287	0.243	0.128	–	–	–	–
Node	0.007	0.212	–0.018	0.213	0.096	0.096	0.107	0.188	0.099	–	–	–
Size	0.021	–0.012	–0.016	–0.051	0.231	0.194	0.155	0.247	0.171	0.322	–	–
Treatment	0.028	–0.03	–0.014	0.044	–0.023	0.021	0.048	–0.002	–0.014	0.007	0.024	–
METABRIC												
ITL	–	–	–	–	–	–	–	–	–	–	–	–
ATL	–0.276	–	–	–	–	–	–	–	–	–	–	–
DTL	–0.83	–0.306	–	–	–	–	–	–	–	–	–	–
LI	0.077	0.199	–0.191	–	–	–	–	–	–	–	–	–
Cancer Hotspot	–0.331	0.035	0.307	0.062	–	–	–	–	–	–	–	–
Immune Hotspot	–0.319	0.16	0.223	0.29	0.849	–	–	–	–	–	–	–
Immune-Cancer Hotspot	–0.253	0.051	0.221	0.152	0.796	0.824	–	–	–	–	–	–

*ATLR = Adjacent-to-Tumor Lymphocyte Ratio; DTL = Distal-To-Tumor Lymphocyte Ratio; ITLR = Intratumor Lymphocyte Ratio; LI = Lymphocytic Infiltration; METABRIC = Molecular Taxonomy of Breast Cancer International Consortium.

However, none of the spatial scores provided prognostic information beyond that of CTS or ROR in the HER2- or the overall population (ΔLR-$\chi^2 \leq 3.84$).

IHC4+Immune Spatial Score

To evaluate a new biomarker based entirely on histology slides, we sought to determine the prognostic value of a combined model of IHC4 and spatial scores for 0 to 10 years and late recurrence. The prognostic value of IHC4 and that of RS are similar in the two-time windows and populations (ΔLR-$\chi^2 \leq 3.84$). For predicting recurrence across 10 years, Cancer Hotspot combined with IHC4 achieved a better performance than RS alone in both cohorts (overall population: ΔLR-$\chi^2 = 9.23$; in HER2-: ΔLR-$\chi^2 = 6.54$) (Figure 3). For predicting late recurrence, prognostic value higher than RS was observed for IHC4 combined with any of the spatial scores in both populations, except for Immune-Cancer Hotspot in the overall population (Figure 3). However, none of the combined scores outperformed CTS and ROR. Finally, IHC4 combined with spatial scores added prognostic value to CTS, similar to that achieved by the RS but lower than ROR in the overall population (IHC4+ Immune Hotspot: ΔLR-$\chi^2 = 22.64$; IHC4+ Immune-Cancer Hotspot: ΔLR-$\chi^2 = 22.96$; IHC4+ Cancer Hotspot: ΔLR-$\chi^2 = 23.38$; RS: ΔLR-$\chi^2 = 23.53$; ROR: ΔLR-$\chi^2 = 29.18$).

Discussion

In this study, we aimed to establish the prognostic value of immune scores for recurrence in ER+ breast cancer patients treated with anastrozole or tamoxifen. While immune response and immunotherapy for ER- diseases have been under the spotlight, correlation of TILs with outcomes in ER+ disease is less clear (16), with many studies reporting the lack of statistically significant prognostic association (11,15,26,27). In line with these reports, we did not find prognostic value in immune abundance scores that only account for the amount of TILs in the entire histology section or in specific tumor regions including intratumor, adjacent tumor, and distal tumor, as scored by digital histology slide image analysis. In contrast, our immune scores based on the spatial heterogeneity of TILs were highly prognostic, particularly for late recurrence after five years of endocrine therapy. This suggests a lasting memory of tumor immunity on disease progression and evolution of treatment resistance in ER+ cancer. Such spatial heterogeneity may reflect spatial distribution patterns of different immune cell subsets. Intratumor heterogeneity of cancer cells may also in turn influence immune spatial distribution through cytokine secretion and neoantigen presentation. While the biological mechanisms remain to be investigated, our finding has clinically significant implication, which suggests that TILs have been previously overlooked in ER+ diseases because of the lack of in-depth analysis of TILs on the tissue spatial organization level. Our findings highlight the importance of examining not just cell abundance but also spatial patterns that can be indicative of immune functional phenotypes and disease prognosis.

In contrast to our observation in ER- tumors (13), high Immune-Cancer Hotspot score, indicating increased spatial clustering in immune and cancer cells, correlated with poor prognosis in ER+ breast cancer. However, this is consistent with our previous finding that immune gene signature was associated with poor response to endocrine therapy in a neoadjuvant setting (28,29). The difference may be due to immune composition and functionality in the two subtypes and mechanisms by which immune response contributes to hormonal therapy resistance (30,31). In a recent study of immune composition in 7270 breast cancers, a higher fraction of T-regulatory cells and M2 macrophages and a lower fraction of M1 macrophages were found in ER+ compared with ER- cancers (12). Therefore, compared with ER- subtype, the immune landscape of the ER+

Table 3. Comparison of prognostic value of immune scores, classical clinical variables, and prognostic scores tests in univariate and multivariable analyses adjusted for nodal status, grade, size, age, and treatment (all CTS components)

			Univariate									Multivariable*								
			Immune Hotspot			Cancer Hotspot			Immune-Cancer Hotspot			Immune Hotspot			Cancer Hotspot			Immune-Cancer Hotspot		
Group	No.	No. of recurrences	HR (95% CI)	LR-χ^2	P†	HR (95% CI)	LR-χ^2	P†	HR (95% CI)	LR-χ^2	P†	HR (95% CI)	LR-χ^2	P†	HR (95% CI)	LR-χ^2	P†	HR (95% CI)	LR-χ^2	P†
All patients																				
0–5 y	1178	103	1.23 (1.06 to 1.43)	6.24	.01	1.31 (1.12 to 1.53)	9.83	.002	1.30 (1.14 to 1.49)	11.46	<.001	1.13 (0.97 to 1.32)	2.36	.12	1.20 (1.02 to 1.41)	4.56	.03	1.20 (1.05 to 1.38)	5.85	.02
5–10 y	1019	113	1.26 (1.09 to 1.47)	7.89	.005	1.30 (1.11 to 1.52)	9.34	.002	1.24 (1.05 to 1.47)	5.45	.01	1.16 (1.00 to 1.35)	3.61	.05	1.21 (1.03 to 1.42)	5.11	.02	1.14 (0.97 to 1.35)	2.27	.13
0–10 y	1178	216	1.25 (1.23 to 1.39)	14.06	<.001	1.30 (1.17 to 1.46)	19.16	<.001	1.28 (1.15 to 1.42)	16.72	<.0001	1.15 (1.03 to 1.28)	5.91	.01	1.20 (1.08 to 1.35)	9.59	.002	1.17 (1.06 to 1.30)	7.77	.005
HER2-negative patients																				
0–5 y	1037	78	1.23 (1.04 to 1.46)	4.63	.03	1.33 (1.11 to 1.59)	8.50	.003	1.26 (1.06 to 1.49)	5.56	.01	1.12 (0.94 to 1.34)	1.49	.22	1.21 (1.01 to 1.46)	3.78	.05	1.14 (0.96 to 1.36)	2.00	.16
5–10 y	909	102	1.33 (1.14 to 1.54)	10.9	.001	1.39 (1.18 to 1.62)	13.77	<.001	1.33 (1.12 to 1.57)	8.63	.003	1.23 (1.06 to 1.42)	6.31	.01	1.30 (1.11 to 1.53)	9.09	.002	1.22 (1.04 to 1.44)	4.95	.03
0–10 y	1037	180	1.28 (1.15 to 1.44)	15.10	<.001	1.36 (1.21 to 1.53)	22.17	<.001	1.29 (1.14 to 1.45)	13.98	<.001	1.18 (1.05 to 1.32)	7.01	.008	1.26 (1.12 to 1.42)	12.49	<.001	1.18 (1.05 to 1.32)	6.33	.01

*Adjusted for node, grade, tumor size, age, treatment. CI = confidence interval; CTS = clinical treatment score; HER2 = human epidermal growth factor receptor 2; HR = hazard ratio.
†Likelihood ratio χ^2, *P* value two-sided.

subtype is characterized by increased T-cell regulation and macrophage polarization toward protumorigenic M2, which is consistent with our results.

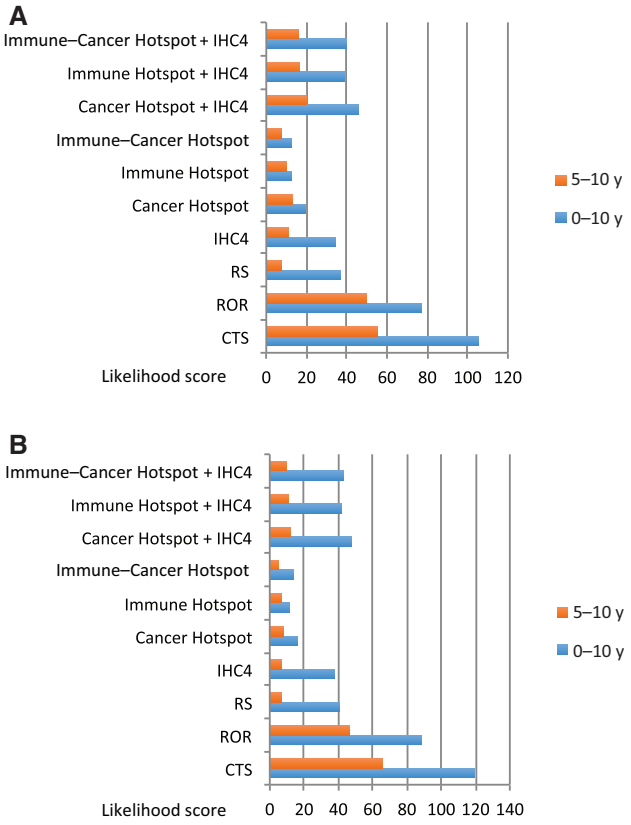

Figure 3. Barplots of likelihood scores for immune spatial and prognostic scores as well as combination of IHC4 and each immune spatial score (IHC4+I) for time windows of 0–10 and 5–10 years in **(A)** overall population and **(B)** human epidermal growth factor receptor–negative population. Kaplan-Meier curves were calculated and tested for equality using the log-rank test. The numbers of patients at risk in each group at various time points are given below each graph. All statistical tests were two-sided. CTS = clinical treatment score; ROR = risk of recurrence score; RS = recurrence score.

Strengths of this study include the large patient cohort with long-term follow-up systematically collected in a well-documented clinical trial, well-characterized samples that enable a direct comparison with established biomarkers, and fully automated and reproducible methods for immune scoring. These allowed us to evaluate quantitative immune scoring based entirely on H&E-stained tumor slides, which are readily generated as part of clinical routine. Comparison of the automated immune scores with a pathologist's score following recommendations for TIL evaluation in breast cancer (24) in a subset of samples showed a weak correlation overall. This is, however, unsurprising when one considers that the automated scores include regions of the tumor that are excluded on pathological evaluation. Indeed, the latter includes stromal TILs only, which are assessed as percentage surface area that is inflammatory as opposed to fibroblastic (25). In contrast, the automated scores count either the absolute number of lymphocytes in relation to cancer cells or the frequency of (co-)clustering of immune and cancer cells. Thus, the information gleaned from these methods potentially provides different biological information about the interaction between these cell types, and, as demonstrated herein, can provide valuable prognostic information. Developing a histology-based test as such thus has the advantages of cost-effectiveness and general applicability. The prognostic value of immune scores for late recurrence is similar to that of IHC4 and RS, which is the most widely used test for residual risk of recurrence following surgery and endocrine. Although they did not add prognostic value to CTS, in almost all occasions immune scores add prognostic information to the individual components that make up CTS, including node status, size, grade, age, and treatment. Because CTS was developed and optimized using TransATAC samples (4), the independent prognostic value of immune scores remains to be validated in further endocrine adjuvant therapy studies with homogeneous treatments such as POETIC (32). In addition, the strong prognostic value of CTS and ROR could be partly explained by the use of tumor size in their calculation, which is a highly prognostic factor. Furthermore, IHC4 combined with immune score is statistically significantly more prognostic than RS, particularly for late recurrence. The clinical utility of IHC4+ immune scores as a combined histology-based test is well worth exploring.

ARTICLE

Table 4. Additional prognostic value of immune spatial scores to IHC4 and RS in all patients and HER2- subgroup

Patient group	0–10 y			5–10 y		
	HR (95% CI)	ΔLR-χ^2	P*	HR (95% CI)	ΔLR-χ^2	P*
All patients		(n = 963)			(n = 824)	
Cancer Hotspot to IHC4	1.22 (1.08 to 1.37)	9.72	.002	1.24 (1.04 to 1.47)	5.57	.02
Cancer Hotspot to RS	1.25 (1.11 to 1.41)	12.46	<.001	1.27 (1.07 to 1.49)	6.97	.008
Immune Hotspot to IHC4	1.13 (1.01 to 1.27)	3.93	.05	1.19 (1.02 to 1.40)	4.10	.04
Immune Hotspot to RS	1.14 (1.02 to 1.28)	4.50	.03	1.21 (1.03 to 1.42)	4.71	.03
Immune-Cancer Hotspot to IHC4	1.17 (1.04 to 1.31)	5.87	.02	NS	NS	NS
Immune-Cancer Hotspot to RS	1.15 (1.02 to 1.29)	4.76	.03	NS	NS	NS
Her2 subgroup		(n = 848)			(n = 733)	
Cancer Hotspot to IHC4	1.26 (1.11 to 1.43)	11.73	<.001	1.32 (1.12 to 1.57)	9.31	.002
Cancer Hotspot to RS	1.29 (1.14 to 1.46)	14.28	<.001	1.37 (1.16 to 1.61)	11.82	<.001
Immune Hotspot to IHC4	1.14 (1.02 to 1.30)	4.56	.03	1.23 (1.05 to 1.44)	5.71	.02
Immune Hotspot to RS	1.15 (1.02 to 1.30)	4.84	.03	1.26 (1.08 to 1.47)	7.12	.008
Immune-Cancer Hotspot to IHC4	1.17 (1.04 to 1.33)	5.64	.02	1.24 (1.04 to 1.48)	4.95	.03
Immune-Cancer Hotspot to RS	1.13 (0.99 to 1.29)	3.22	.07	1.23 (1.03 to 1.47)	4.69	.03

*Likelihood ratio χ^2, P value two-sided. CI = confidence interval; HER2 = human epidermal growth factor receptor 2; HR = hazard ratio; RS = Oncotype DX 21-gene recurrence score; NS = not statistically significant.

Weaknesses of our study include the use of a single histology slide per tumor, such that intratumor heterogeneity cannot be fully addressed, and the lack of immune markers that may provide further insights into immune functions in ER+ disease and better predictors due to little residual tissue available. In addition, whether our methods can sufficiently address challenges arising from variability in factors that include fixation, staining, and acquisition in a clinical setting needs to be evaluated before implementation. Our results also only apply to women who are chemotherapy-free.

Furthermore, our study also provides relevant information for new treatment strategy in the high-risk ER+ population identified by immune scores. Our findings support different immunosuppressive mechanisms in the ER+ and ER- subtypes, and in light of these results call for the development of novel cancer therapeutics targeting the pathways that reverse these mechanisms specifically for ER+ disease. This may also help explain why anti-PD1 checkpoint inhibition, despite demonstrating activity as monotherapy in early-phase trials in ER+ breast cancer, had low response rates compared with triple-negative breast cancer and was highly variable among trials (33,34). Further, we speculate that immune scores may be useful as predictive biomarkers for immunotherapy, given the limited clinical utility of PDL1 expression in guiding patient selection (35).

In summary, enabled by fully automated image analysis of histology sections, our study provided an additional dimension to the analysis of immune functional phenotype in breast cancer. Spatial data provided by histology, once quantitatively analyzed, will aid the identification of clinically relevant features, potentially yielding predictions more powerful than measurements of cell abundance that ignore the spatial context.

Funding

This work was supported by the Royal Marsden National Institutes of Health Biomedical Research Centre grant A105. MD and YY acknowledge support by the Royal Marsden National Institutes of Health Biomedical Research Centre. MD acknowledges a Breast Cancer Now grant (CTR-Q4-Y1). JC and IS acknowledge a CRUK grant (C569/A16891). YY acknowledges support by CRUK (C45982/A21808), Breast Cancer Now (2015NovPR638) and the Wellcome Trust (105104/Z/14/Z).

Notes

The funder had no role in the design of the study; the collection, analysis, or interpretation of the data; the writing of the manuscript; or the decision to submit the manuscript for publication.

We thank Andrew Dodson and Daniel Nava Rodrigues for technical and pathological support. M. Dowsett has received commercial research grants and speakers bureau honoraria from AstraZeneca.

MD and YY designed the experiments. AH, IS, KN, and JC performed the analyses. AH, IS, and YY wrote the manuscript. All authors have approved the manuscript.

References

1. Dowsett M, Goldhirsch A, Hayes DF, Senn HJ, Wood W, Viale G. International Web-based consultation on priorities for translational breast cancer research. *Breast Cancer Res*. 2007;9(6):R81.
2. 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(27):2817–2826.
3. Nielsen TO, Parker JS, Leung S, et al. A comparison of PAM50 intrinsic subtyping with immunohistochemistry and clinical prognostic factors in tamoxifen-treated estrogen receptor-positive breast cancer. *Clin Cancer Res*. 2010;16(21):5222–5232.
4. 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.
5. Sestak I, Dowsett M, Zabaglo L, et al. Factors predicting late recurrence for estrogen receptor-positive breast cancer. *J Natl Cancer Inst*. 2013;105(19):1504–1511.
6. Sestak I, Dowsett M, Ferree S, Baehner FL, Cuzick J. Retrospective analysis of molecular scores for the prediction of distant recurrence according to baseline risk factors. *Breast Cancer Res Treat*. 2016;159(1):71–78.
7. Savas P, Salgado R, Denkert C, et al. Clinical relevance of host immunity in breast cancer: From TILs to the clinic. *Nat Rev Clin Oncol*. 2016;13(4):228–241.
8. Denkert C, von Minckwitz G, Brase JC, et al. Tumor-infiltrating lymphocytes and response to neoadjuvant chemotherapy with or without carboplatin in human epidermal growth factor receptor 2–positive and triple-negative primary breast cancers. *J Clin Oncol*. 2015;33:983–991.
9. Denkert C, Loibl S, Noske A, et al. Tumor-associated lymphocytes as an independent predictor of response to neoadjuvant chemotherapy in breast cancer. *J Clin Oncol*. 2010;28:105–113.
10. Gu-Trantien C, Loi S, Garaud S, et al. CD4(+) follicular helper T cell infiltration predicts breast cancer survival. *J Clin Invest*. 2013;123:2873–2892.
11. Loi S, Sirtaine N, Piette F, et al. Prognostic and predictive value of tumor-infiltrating lymphocytes in a phase III randomized adjuvant breast cancer trial in node-positive breast cancer comparing the addition of docetaxel to doxorubicin with doxorubicin-based chemotherapy: BIG 02-98. *J Clin Oncol*. 2013;31:860–867.
12. Bense RD, Sotiriou C, Piccart-Gebhart MJ, et al. Relevance of tumor-infiltrating immune cell composition and functionality for disease outcome in breast cancer. *J Natl Cancer Inst*. 2017;109(1): djw192.
13. Nawaz S, Heindl A, Koelble K, Yuan Y. Beyond immune density: Critical role of spatial heterogeneity in estrogen receptor-negative breast cancer. *Mod Pathol*. 2015;28(12):1621.
14. Dieci MV, Criscitiello C, Goubar A, et al. Prognostic value of tumor-infiltrating lymphocytes on residual disease after primary chemotherapy for triple-negative breast cancer: A retrospective multicenter study. *Ann Oncol*. 2014;25:611–618.
15. Loi S, Michiels S, Salgado R, et al. Tumor infiltrating lymphocytes are prognostic in triple negative breast cancer and predictive for trastuzumab benefit in early breast cancer: Results from the FinHER trial. *Ann Oncol*. 2014;25:1544–1550.
16. Luen S, Virassamy B, Savas P, Salgado R, Loi S. The genomic landscape of breast cancer and its interaction with host immunity. *Breast*. 2016;29:241–250.
17. Yuan Y, Failmezger H, Rueda OM, et al. Quantitative image analysis of cellular heterogeneity in breast tumors complements genomic profiling. *Sci Translat Med*. 2012;4:157ra43.
18. Yuan Y. Modelling the spatial heterogeneity and molecular correlates of lymphocytic infiltration in triple-negative breast cancer. *J R Soc Interface*. 2015;12.
19. Yuan Y. Spatial heterogeneity in the tumor microenvironment. *Cold Spring Harb Perspect Med*. 2016;6(8).
20. Demaria S, Kawashima N, Yang AM, et al. Immune-mediated inhibition of metastases after treatment with local radiation and CTLA-4 blockade in a mouse model of breast cancer. *Clin Cancer Res*. 2005;11:728–734.
21. Fridman WH, Pages F, Sautes-Fridman C, Galon J. The immune contexture in human tumours: Impact on clinical outcome. *Nat Rev Cancer*. 2012;12:298–306.
22. Gentles AJ, Newman AM, Liu CL, et al. The prognostic landscape of genes and infiltrating immune cells across human cancers. *Nat Med*. 2015;21(8):938–945.
23. 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(11):1829–1834.
24. Salgado R, Denkert C, Demaria S, et al. The evaluation of tumor-infiltrating lymphocytes (TILs) in breast cancer: Recommendations by an International TILs Working Group 2014. *Ann Oncol*. 2015;26(2):259–271.
25. Curtis C, Shah SP, Chin SF, et al. The genomic and transcriptomic architecture of 2,000 breast tumours reveals novel subgroups. *Nature*. 2012;486:346–352.
26. Dieci MV, Criscitiello C, Goubar A, et al. Prognostic value of tumor-infiltrating lymphocytes on residual disease after primary chemotherapy for triple-negative breast cancer: A retrospective multicenter study. *Ann Oncol*. 2014;25:611–618.
27. Ali HR, Provenzano E, Dawson SJ, et al. Association between CD8+T-cell infiltration and breast cancer survival in 12 439 patients. *Ann Oncol*. 2014;25(8):1536–1543.

28. Dunbier AK, Ghazoui Z, Anderson H, et al. Molecular profiling of aromatase inhibitor-treated post-menopausal breast tumors identifies immune-related correlates of resistance. *Clin Cancer Res*. 2013;19(10):2775–2786.
29. Gao Q, Patani N, Dunbier AK, et al. Effect of aromatase inhibition on functional gene modules in estrogen receptor-positive breast cancer and their relationship with antiproliferative response. *Clin Cancer Res*. 2014;20(9):2485–2494.
30. Wei C, Cao Y, Yang X, et al. Elevated expression of TANK-binding kinase 1 enhances tamoxifen resistance in breast cancer. *Proc Natl Acad Sci U S A*. 2014; 111(5):E601–E610.
31. Jansen MP, Foekens JA, van Staveren IL, et al. Molecular classification of tamoxifen-resistant breast carcinomas by gene expression profiling. *J Clin Oncol*. 2005;23(4):732–740.
32. Dowsett M, Smith I, Robertson J, et al. Endocrine therapy, new biologicals, and new study designs for presurgical studies in breast cancer. *J Natl Cancer Inst Monographs*. 2011:120–3.
33. Dirix LY, Takacs I, Nikolinakos P, et al. Avelumab (MSB0010718C), an anti-PD-L1 antibody, in patients with locally advanced or metastatic breast cancer: A phase 1b JAVELIN solid tumor trial. *Cancer Res*. 2016;76(Suppl 4):S1–04.
34. Rugo HS, Delord JP, Im SA, et al. Preliminary efficacy and safety of pembrolizumab (MK-3475) in patients with PD-L1 positive, estrogen receptor-positive (ER+)/HER2-negative advanced breast cancer enrolled in KEYNOTE-028. *Cancer Res*. 2016;76(Suppl 4):S5–07.
35. Patel SP, Kurzrock R. PD-L1 expression as a predictive biomarker in cancer immunotherapy. *Mol Cancer Ther*. 2015;14(4):847–856.